



Estimating soil particle-size fractions and predicting soil texture from microwave remote sensing techniques with application in Ireland

Sandra Cristina Deodoro

Thesis submitted to National University of Ireland Maynooth in fulfilment of the requirements for the degree of

Doctor of Philosophy

Maynooth University – Ollscoil Mhá Nuad
Department of Geography

Submitted: October 2024

Head of Department
Dr Stephen McCarron

Research Supervisors
Prof Rowan Fealy, Prof Tim McCarthy,
Dr Réamonn Fealy and Dr Rafael de Andrade Moral

To my family, and in memory of my father Raimundo Deodoro.

Two roads diverged in a wood, and I –

I took the one less travelled by,

And that has made all the difference.

– Robert Frost

Once upon a time, everything seemed fixed and solid. Now, everything in the universe has begun to slide under our feet: mountains, continents, life, and even matter itself. To make future progress, science must peel away all the coverings of apparent stability in the world.

–Teilhard de Chardin

Declaration

I certify that this dissertation has not be submitted for a degree in this or any other university. The contents of the thesis are the original work of the author and all secondary materials and literature used have been adequately acknowledged and referenced.

Signed: *Sandra Cristina Deodoro*
Sandra Cristina Deodoro

Author note

This thesis has been prepared in accordance with the PhD regulations of Maynooth University and is subject to copyright.

The research presented in this thesis has been made possible through the financial support of the John and Pat Hume Doctoral Awards Programme – Maynooth University (2021-2025),, and the Department of Geography – Maynooth University.

The research articles and accompanying text presented in this thesis are the authors' original version(s) accepted by the respective publishers. The author retains the right to redistribute.

Acknowledgements

First and foremost, I want to express my sincere gratitude to my supervisor, Professor Rowan Fealy, as my PhD journey started with his trust in my research proposal four years ago. Then, the journey progressed with his constant support, advice, time, and shared knowledge – crucial for me to complete my doctorate. I want also to thank my co-supervisors, Professor Tim McCarthy, Dr Réamonn Fealy and Dr Rafael de Andrade Moral for their constant contributions to my research and PhD thesis.

I would also like to extend my thanks to my family for their unconditional support. Also to Alessandra and Estevão, from Hamilton Institute Maynooth University, for the friendship and support in R coding.

I am also grateful for the financial support provided by the Maynooth University through the Pat and Hume Doctoral Award 2021-2025.

Summary

Synthetic Aperture Radar is a low-cost alternative widely employed to estimate soil properties, especially over small map scales (i.e. large area). In contrast to deeper soil layers, topsoil is more consistent with the capability of a C-band SAR signal data (e.g., Sentinel-1) to reach the soil surface. However, few studies address the use of microwave-based sensors to estimate particle size fractions and soil texture classes (e.g., loam, sandy clay). As soil texture consists of the relative proportions of sand, silt, and clay, this soil property is compositional in nature (i.e., the sum of the components is equal to 100%), and such a constraint is not always considered in either explicitly spatial or non-spatial models. Moreover, retrieving information on soils from radar signals is a challenging task, particularly under vegetated soil conditions. This research seeks to address this challenge by employing H-alpha dual-pol decomposition over a study domain located in the Republic of Ireland, where soils are typically covered by grass/pasture. No study to date has employed this method for estimating sand, silt, and clay, or soil texture. Employing both a spatial and non-spatial framework and explicitly considering the compositional nature of soil texture, two different statistical modelling approaches – linear and tree-based – are used to derive soil estimates from Sentinel-1 data, in tandem with topographical and geophysical covariates. Five primary conclusions can be drawn from this work: (i) it is beneficial to treat soil texture as compositional data in the models employed; (ii) radar-based derivatives are not able to predict sand, silt or clay without the aid of covariates, since the models do not identify direct relationships between the backscattering coefficients (σ_{VV}^0 , σ_{VH}^0) and the soil particle size fractions; (iii) the non-spatial modelling approaches yield better estimates when fitted without the geophysical covariates, however, the geophysical covariates are useful in obtaining soil texture in the regression models with interactions terms; (iv) the H-alpha Dual-Pol Decomposition method provides, to a certain extent, soil information over low vegetation, and improved estimates of sand, silt and clay; and (v), the spatial models do not outperform the non-spatial models in estimating soil particle size fractions (PSF) in terms of numerical estimates, but are useful in capturing patterns and trends in the response variables. Hence, this research provides a methodological framework to inform the design of *in situ* soil surveys and a means to estimate soil properties over large spatial areas to generate new data products for use in hydrological, land surface, climate, and other model-based approaches that currently employ coarse global scale soil texture products. This research is also timely in light of the European and Irish policy initiatives around soils such as “A Soil Deal for Europe 2021-2030 (Mission Soil)” and “A Signpost for Soil Policy in Ireland 2021-2030”.

Contents

Summary	iii
Contents	i
List of tables.....	v
List of figures.....	vi
CHAPTER 1 INTRODUCTION	1
1.1 Overview.....	1
1.2 Soil properties: soil texture	4
1.3 Current approaches to measuring and mapping soil PSF and soil texture.....	7
1.3.1 Traditional methods	7
1.3.2 Soil property retrieval with SAR	8
1.4 Research motivation	16
1.4.1 Research rationale.....	16
1.5 Knowledge gaps.....	19
1.6 Research questions.....	20
1.7 Research aims and objectives	20
1.8 Thesis structure	21
CHAPTER 2 THEORY AND LITERATURE REVIEW	24
2.1 Introduction.....	24
2.2 Earth Observation platforms	24
2.2.1 Synthetic Aperture Radar.....	28
2.3.2 Dealing with vegetated soils in SAR images	37
2.3.3 The use of SAR for soil analysis in an Irish context.....	42
2.4 Modelling approaches for estimating soil particle size fractions.....	44
2.4.1 Soil particle size fractions as compositional data	44
2.4.2 Non-spatial models: Statistical machine learning algorithms.....	48
2.4.3 Spatial models: geostatistical and non-geostatistical methods	51
2.5 Summary.....	55
CHAPTER 3 STUDY LOCATION	57
3.1 Introduction.....	57

3.2 Geographical setting	58
3.3 Soil information for Ireland	62
3.3.1 LUCAS Survey	62
3.3.2 Irish soil mapping	65
3.4 Datasets	66
3.4.1 Soil information	67
3.4.2 Microwave remote sensing data.....	67
3.4.3 Environmental data	68
3.4.4 Geophysical datasets.....	69
CHAPTER 4 AN ASSESSMENT OF SENTINEL-1 SYNTHETIC APERTURE RADAR, GEOPHYSICAL AND TOPOGRAPHICAL COVARIATES FOR ESTIMATING TOPSOIL PARTICLE SIZE FRACTIONS	71
4.1 Context.....	71
4.1.1 Abstract.....	71
4.2 Introduction.....	72
4.3 Theoretical background	73
4.4 Data and Methods	75
4.4.1 Study Region	75
4.4.2 Data.....	76
4.4.3 Methods	79
4.5 Results.....	81
4.5.1 Statistical descriptive analysis	81
4.5.2 Soil particle size prediction.....	82
4.6 Discussion.....	88
4.6.1 General aspects of soil particle size estimation with Sentinel-1 C-band data.....	88
4.6.2 Y-compositional LRM – ILR transformation	88
4.6.3 Random Forest – CLR transformation applied.....	89
4.6.4 General aspects of the modelling approaches	90
4.6.5 Strengths and limitations of the work	91
4.7 Conclusion	92

CHAPTER 5 USING SURFACE SCATTERING MECHANISM FROM DUAL-POL SAR DATA TO PREDICT TOPSOIL PARTICLE SIZE FRACTIONS	94
5.1 Context.....	94
5.1.1 Abstract.....	95
5.2 Introduction.....	95
5.3 Theoretical background and related works	99
5.3.1 Radiometric calibration of SAR image.....	99
5.3.2 The H-alpha Dual Pol Decomposition.....	100
5.4 Data and methods.....	103
5.4.1 Study Region	103
5.4.2 Data.....	105
5.4.3 Methods	106
5.5 Results.....	110
5.5.1 Statistical models	115
5.6 Discussion.....	118
5.6.1 Strengths and limitations of the work	121
5.7 Conclusion	123
CHAPTER 6 EVALUATING THE USE OF EXPLICITLY SPATIAL MODELS AND NON-SPATIAL MODELS TO ESTIMATE SOIL PARTICLE SIZE FRACTIONS AND SOIL TEXTURE USING SENTINEL-1 AND COVARIATES	125
6.1 Context.....	125
6.2 Introduction.....	125
6.3 Background: Spatially explicit models and geostatistics.....	127
6.4 Data and Methods	129
6.4.1 Study Region	129
6.4.2 Data.....	130
6.4.3 Methods	132
6.4.4 Evaluation metrics for soil particle size predictions	135
6.5 Results.....	135
6.5.1 Spatial modelling.....	137
6.5.2 Non-spatial model.....	143

6.5.3 Soil texture classes – all models	143
6.6 Discussion.....	144
6.6.1 Strengths and limitations of the work	147
6.7 Conclusion	147
CHAPTER 7 DISCUSSION	149
7.1 Introduction.....	149
7.2 Results and findings from the methods employed	149
7.3 Limitations of the data, models, and methods employed in this study	155
7.3.1 Concerning field data from the soil legacy surveys	155
7.3.2 Concerning Sentinel-1 data.....	156
7.3.3 Concerning models	157
CHAPTER 8 CONCLUSION.....	159
8.1 Introduction.....	159
8.2 How this study has advanced the research for Geography and soil science	159
8.3 Potential areas and approaches for future research.....	160
8.4 Informing the design of future soil sampling surveys	161
8.5 Wider societal relevance of the research	161
8.6 Policy relevance.....	162
BIBLIOGRAPHY	164
Appendix A.....	181
Appendix B	213
Appendix C	220
Appendix D. Declarations of Co-authorship.....	237

List of tables

Table 1.1 Relationship between soil particle size fractions and other soil properties	5
Table 1.2 Values of available water content for different types of soil.....	7
Table 1.3 Summary of soil properties retrieval techniques from SAR data	10
Table 1.4 Glossary of the main terms related to remote sensing and geographical information system	22
Table 2.1 Approximate Microwave Measurement Depths for Bare Soil and Different Land Surface Covers	28
Table 2.2 Radar bands and applications	29
Table 2.3 Examples of polarisation types in radar systems	31
Table 2.4 Characteristics of major spaceborne SAR systems.....	33
Table 2.5 The dielectric constant for Earth materials	37
Table 2.6 Differences of common terminology between statistics and machine learning	48
Table 3.1 Summary of the procedures for deriving the topographic parameters from the DEM EEA-10 .	69
Table 4.1 Summary of data used in the modelling approaches	78
Table 4.2 Summary of methods and covariates employed in the two modelling approaches	80
Table 4.3 Summary of the evaluation metrics resulted from the modelling approaches validated with LOOCV	87
Table 5.1 Descriptive statistics for the dual-pol decomposition parameters retrieved for the soil samples (N=27).....	110
Table 5.2 The typical scattering mechanisms in an H- α feature space	111
Table 5.3 Model evaluation metrics on test data for models validated through LOOCV method. (N=27)	115
Table 6.1 Some related works regarding statistical and/or geostatistical models for soil properties estimation and prediction.....	128
Table 6.2 Distance band from neighbour	136
Table 6.3 Spatial autocorrelation: Global Moran Index	136
Table 6.4 Evaluation metrics resulted from the spatial modelling approaches validated with LOOCV ..	137
Table 6.5 Evaluation metrics resulted from the non-spatial modelling approaches validated with LOOCV	143

List of figures

Figure 1.1 An example of a basaltic soil sample and its respective textural class resulted from laboratory analysis. Soil ternary source: USDA/NCSS.....	5
Figure 1.2 An example of water behavior to permeability/percolation and infiltration into soils (topsoils and deep soils) as a function of soil texture and soil porosity. Soils with higher sand content, thereby lower porosity, exhibit a higher infiltration rate compared to soils with high clay content (higher porosity). Such hydrological conditions affects soil moisture, soil dielectric constant, and consequently, radar backscattering from soils. Source: Retrieved from Tobin <i>et al.</i> (2018).....	6
Figure 1.3 An example of soil particle size distribution analysed in laboratory, depicting the fractions settled in cylinder after sedimentation, based on Stokes' Law. According to Stokes Law, greater fractions (e.g. sand) fall faster in water than smaller ones. Source: USDA/ NSCS (n.d).....	8
Figure 1.4 Active and passive microwave remote sensors (including Sentinel 1) used for generation of the European Space Agency climate change initiative data sets, including soil properties. The vertical black solid lines in each box represent the start/end of missions; the arrows indicate the continuity of the mission. Source: Adapted from Babaeian <i>et al.</i> (2019).	8
Figure 1.5 Relationship between soil moisture, estimated from radar moisture maps, and soil particle size fractions (clay% and sand%). Source: Gorrab <i>et al.</i> (2015).	15
Figure 1.6 Idealised soil profile diagram with topsoil under pedological (eluviation) and geological contexts (parent material). Soil (or solum) consisting of A and B horizons. Source: Mahaney & Ramanathan (2003).	18
Figure 1.7 Idealised soil profile diagram with topsoil and organic layer in relation to the parent material, in terms of weathering results (regolith) and the original rock (bedrock). The combined O, A, E, and B horizons form the soil (or solum). Source: Hess & McKnight (2017).	18
Figure 2.1 The electromagnetic spectrum (a) with focus on microwaves region (b). Source: Canada Centre for Remote Sensing (CCRS). (n.d).....	25
Figure 2.2 Examples of Earth Observation platforms and ground sensors. Source: Adapted from Babaeian <i>et al.</i> (2019).	26
Figure 2.3 Typical SAR bands with their respective wave representation in the electromagnetic spectrum, depicting the microwave remote sensing. Source: NASA EarthData (n.d.)	29
Figure 2.4 An example of the interaction of different wavelengths with different land cover types and scattering objects. The intensity of backscattered energy changes with wavelength and the way radiation is scattered. Source: Schumann & Moller (2015).	30
Figure 2.5 Examples of dual polarisation. Source: ESA (2007).	32
Figure 2.6 An example of a theoretical scattering under two different soil-covering conditions (forest and grassland) and different water content on soil conditions. Source: Bourgeau-Chavez <i>et al.</i> (2009).	35

Figure 2.7 An example of scattering mechanisms under different land cover contexts: rough surface (bare soil), arboreal (tree) and urban (buildings) and relative scattering strength by polarisation. Rough surface scattering: $SVV > SHH > SHV$ or SVH ; Double bounce scattering: $SHH > SVV > SHV$ or SVH ; Volume scattering: main source of SHV and SVH . Source: NASA (n.d.).	39
Figure 2.8 Three canonical scattering mechanisms for vegetated soils. From the scattering scenario illustration, [T] refers to coherence matrix and is closer related to physical and geometrical properties of the scattering process, thus provides direct physical interpretation. Source: Jagdhuber <i>et al.</i> (2013). Module 2301: SAR Polarimetry.	40
Figure 2.9 An example of soil particle size fractions in the simplex SD (3-simplex space for two soil samples consisted of sand (58%), silt (29%) and clay (13%) classified as sandy loam (coarse loamy) (Point 1); and sand (40%), silt (29%) and clay (31%) classified as clay loam (fine loamy) (Point 2). The textural diagram is according to the USDA-NSCS system. Note that percentages of soil fractions sum 100 in each sample – a constraint of both soil property and the simplex (ternary diagram). Source: USDA-Soil texture calculator.	46
Figure 2.10 Machine Learning methods. Source: Dey (2016).	49
Figure 2.11 Example of a flowchart of soil property modelling based on machine learning application, developed by Araya <i>et al.</i> (2021) for soil moisture. Source: Araya <i>et al.</i> (2021).	54
Figure 3.1 Topographical map of Ireland with experimental areas of this study are displayed in black and red squares. Source of data: ESA-Copernicus Digital Elevation Model 10 m (DEM-EEA10).	59
Figure 3.2 Physiographic map of Ireland depicting the broad-scale physical landscape units. Source: Geological Survey Ireland (2018).	60
Figure 3.3 Mean monthly rainfall ranges for Ireland from 1991 to 2020. Source: Curley <i>et al.</i> (2023). The Irish Meteorological Service.	61
Figure 3.4 Topsoil physical properties for Europe (based on LUCAS topsoil data). Left: USDA soil textural classes derived from clay, silt and sand maps. Right: Zoom (unscaled) in the study area from the USDA soil texture classes map. Source: Adapted from Ballabio <i>et al.</i> (2016).	63
Figure 3.5 Clay, silt and sand content (%) in topsoil (0-20cm) modelled by Multivariate Additive Regression Splines. Topsoil physical properties for Europe (based on LUCAS topsoil data). Source: Ballabio <i>et al.</i> (2016). A yellow square was drawn on the figure to highlight the study area of this thesis.	64
Figure 3.6 Indicative soil texture map of agricultural soils in Ireland. Source: Creamer <i>et al.</i> (2016).	66
Figure 3.7 Approximate depths of investigation for typical Irish rock resistivities, for the four Tellus system frequencies, derived using the half-skin-depth estimator. Source: Ture <i>et al.</i> (2020).	70
Figure 4.1 Diagram for the radar equation that is used to determine the proportion of energy that is being returned from a target. Source: ESA-EO College (2021).	74
Figure 4.2 Map of the study area with the location of topsoil sampling sites available from different soil surveys and measurement programmes. Values in brackets indicate the number of samples for each soil dataset.	75

Figure 4.3 Topographic, morphometric and soil moisture for the study domain. Surface soil moisture data was obtained in the same date as that of the Sentinel 1 image (03/04/2021).	76
Figure 4.4 Flowchart of the methodology depicting the data used in the models, the processing steps performed for both SLC and GRD products from SAR Sentinel 1, and the modelling approaches (with geophysical and non-geophysical covariates).	80
Figure 4.5 Observed vs. Predicted plots from the Y-compositional linear regression model validated with LOOCV for Modelling approach 1 (SAR + Topography covariates), without 2-order interaction terms (Figures 4.5a, 4.5b, 4.5c) and with 2-order interaction terms (Figures 4.5d, 4.5e, 4.5f). The responses are measured in % content and the line in the plots refers to the 1:1 line. Corresponding soil texture classes are displayed in a soil ternary diagram. The soil ternary diagram in Figure 4.5g depicts the original soil texture classes (n=100% = 235) obtained directly from the measured data, while ternary diagrams in the 4.5h and 4.5i show classes resulting from predictions of soil particle size fractions without- and with 2 nd order interaction terms, respectively.	84
Figure 4.6 Observed vs. Predicted plots resulting from the Y-compositional linear regression model validated with LOOCV for Modelling approach 2 (SAR + Topography + Geophysical covariates) , without 2-order interaction terms (Figures 4.6a, 4.6b, 4.6c) and with 2-order interaction terms (Figures 4.6d, 4.6e, 4.6f). The responses are measured in % content and the black line in the plots refers to the 1:1 line. Corresponding soil texture classes are displayed in a soil ternary diagram. The soil ternary diagram in (4.6g) depicts the soil texture classes obtained directly from the measured data, while ternary diagrams in (4.6h) and (4.6i) show classes resulting from predictions of soil particle size fractions without- and with 2-order interaction terms, respectively.	85
Figure 4.7 Comparative plots of observed and predicted samples (Figures 4.7a, 4.7b, 4.7c) obtained from the Random Forest Modelling with CLR transformation applied to the response variables and validated with LOOCV(Approach 1: SAR + Topography covariates).The responses are measured in % content and the black line in the plots refers to the 1:1 line. The ternary diagrams on the bottom show the soil texture classes. The original (measured data) classes are depicted in Figure 4.7d and the classes resulting from predictions of soil particle size fractions are exhibited in 4.7e.....	86
Figure 4.8 Comparative plots of observed and predicted samples (Figures 4.8a, 4.8b, 4.8c) obtained from the Random Forest Modelling with CLR transformation applied to the response variables and validated with LOOCV (Approach 2: SAR + Topography + Geophysical covariates). The responses are measured in % content and the black line in the plots refers to the 1:1 line. The ternary diagrams on the bottom show the soil texture classes. The original (measured data) classes are depicted in Figure 4.8d, and the classes resulting from predictions of soil particle size fractions are exhibited in 4.8e.	87
Figure 5.1 Normalization areas for SAR backscatter (Small, 2011). β_0 refers to reflectivity in an area in slant range and measures brightness in SAR images; σ^0 is illumination corrected and measures the power returned to antenna from ground; γ^0 is plane perpendicular to the local look direction and is radiometrically corrected being suitable for volume scatterers.	100
Figure 5.2 The land use-land cover categories from an RGB composite applied to the Sentinel 1 imagery over the study area. Topsoil sampled sites in the study area (points) overlaid on RGB composite using VV	

and VH polarisations from Sentinel 1-C band. The RGB resulting image highlights some of the land use/land cover categories, such as man-made structures and urban areas (yellow) within which Dublin city is highlighted on the top-right of the figure, water bodies (blue), and bogs/peatlands/bare soil (brown), dense and higher vegetation are displayed in light cyan, and grassland, crops and low vegetation cover are shown in medium and dark shades of cyan.	104
Figure 5.3 Topsoil samples superimposed over topographical (morphometric) covariates images of the study area. The soil samples (n=27) were employed in the models. Elevation data refers to the DEM-EEA-10 (10 m spatial resolution) from which the other covariates were derivate.	104
Figure 5.4 Alpha and entropy maps of the entire scene (figures a, b) as retrieved from Sentinel 1 data using the dual H-alpha pol-decomposition. The corresponding H-Alpha plot for the entire scene (figure c). Square dots represent the soil samples selected (n=27). Blue: surface scattering; yellow: volume scattering; red: random and bragg scattering.	111
Figure 5.5 A subset of Alpha of the entire scene and a comparison to the RGB composite (R4G3B2) Sentinel 2 (optical imagery) showing a correspondence between the two images (a, b). The square box in the entire scene highlights the location of the subset. Blue colour in figure “a” (surface scattering) depicts bare soil and arable lands (with very low- or more-sparse vegetation), while yellow colours (volume scattering) correspond to vegetated areas as depicted in figure “b”. A spectral plot (bottom) was obtained from Sentinel 2 imagery for the selected soil sampling sites (14 points), which depicts spectral signatures of vegetation (green lines) and bare soil (red lines).	112
Figure 5.6 A detailed subset of Alpha (a) and Sentinel 2-RGB composite (b) with the corresponding H-alpha plane plot for the subset scene (c). The figure illustrates soils under different conditions of vegetation cover. Circles on the figures were only drawn to highlight low vegetation amongst soil.	113
Figure 5.7 H- α plane plots (right) of samples selected over bare soil and low vegetation (left) in the study area. Points (samples) located in Z3 and Z6 in the plots indicate surface scattering mechanisms as brag surface and random surface types, respectively, with low and medium entropy, respectively. In agricultural fields, brag surface may be associated with fallows while random surface could be associated with ploughed fields and fields with short grass (Anup Das, n.d: pp 27, 70, 77). The spectral plot on the bottom shows the spectral signature of a representative pixel within each polygon.	115
Figure 5.8 Scatterplots depicting the observed and predicted data resulted from the compositional linear model and different tree-based models with the H-alpha dual-pol decomposition applied (surface scattering partition) (a, c) and without surface scattering in the model (b, d). RF : Random Forest; GRF : Generalized Random Forests; LLF : Local Linear Forest; BART : Bayesian Additive Regression Tree; SBART : SoftBART.	117
Figure 5.9 Soil ternary (USDA) with the textural classes obtained from the statistical models with surface scattering data included in the models and log-ratio transformation applied to the response variables (CoDa approach). Soil ternary in figure a) refers to the classes obtained directly from the measured data. Soil textural classes inferred from the regression models are depicted in figure (b) with the compositional LRM and in figures (c-f) with the tree-model approaches.	118

Figure 6.1 Map of the study area with the location of topsoil sampling sites available from different soil surveys and measurement programmes.....	130
Figure 6.2 Neighbourhood analysis of response variables (measured data, content %) according to a) Distance Band and b) Delauney Triangulation (Voronoi tessellation) methods. A three-dimensional perspective of the data is depicted in “c”, wherein the locations of sample points are plotted on the x,y plane. The x,z plane and the y,z plane are scatterplots of the values projected onto such planes and represent the trend directions (blue and green colours). Above each sample point, the value is given by the height of a stick in the z-dimension.....	136
Figure 6.3 Scatterplots of observed and predicted values resulting from the CK approach with and without CLR transformation validated with LOOCV method. Model without interaction terms applied to the covariates.	138
Figure 6.4 Scatterplots of observed and predicted values resulting from the EBKR model with- and without CLR transformation validated with LOOCV method. Model without interaction terms applied to the covariates.	139
Figure 6.5 Surface predictions resulting from the EBKR approach with and without CLR transformation (1-3 and 4-5, respectively) and corresponding LOOCV validation plots.	141
Figure 6.6 Surface predictions resulting from the CK approach with- and without CLR transformation (1-3 and 4-5, respectively) and corresponding LOOCV validation plots.	142
Figure 6.7 Observed Soil texture classes displayed in the USDA ternary and the resulting classes from both non-spatial (b-e) and spatial models (f-i). a: Observed (measured) data; b: LRM model, without log-ratio transformation and without interaction terms; c: LRM with ILR transformation and interaction terms; d: LRM with ILR transformation, without interaction terms; e: RF model with CLR transformation; f: CK model with CLR transformation; g: CK without CLR transformation; h: EBKR model with CLR transformation; i: EBKR without CLR transformation.....	144

1.1 Overview

Soils play a vital role in the Earth-atmosphere system, given their connection with the atmosphere, biosphere, hydrosphere, and geosphere (Gerke *et al.*, 2022), and are important in terms of the ecosystem services they provide (i.e. provisioning, regulating, supporting, and cultural) (Baveye *et al.*, 2016). Such importance has increased the relevance of this natural resource in the socio-political agenda (Ballabio *et al.*, 2016). Soils also serve as a basis for human settlements and economic activities (i.e. physically supporting buildings and other man-made structures), as well as carrying information on past climate and cultural heritage in terms of paleoenvironmental reconstruction research and paleosol (or palaeosol) studies. Examples of which can be found in Sheldon & Tabor (2009), McClatchie *et al.* (2015) and Santos *et al.* (2021). Concerning the connections between soil functions and soil services (Baveye *et al.*, 2016), soils emerge as a crucial element in understanding current and future climate responses in light of their importance for carbon sequestration, climate regulation, (soil) biomass storage and flood-drought regulation (Lal *et al.*, 2021; Frelih-Larsen *et al.*, 2022; Certini & Scalenghe, 2023), as soil is an important component of both the carbon and water cycles. Hence, the role of soil in the regulation and mitigation of climate is also highlighted.

Soils, similar to other natural resources across the Earth (e.g. water and minerals), are experiencing increasing pressures, from agriculture, industry, forestry, and urbanisation, amongst other factors, resulting in the depletion of the resource and increased concerns around soil health and soil sealing. Increasingly, policies are focused on how to balance the demand pressures with appropriate conservation measures to maintain and increase soil health (Lal *et al.*, 2021; Frelih-Larsen *et al.*, 2022). More comprehensive knowledge and understanding of its intrinsic physical and chemical properties are ultimately required to underpin more sustainable soil management strategies (Hou *et al.*, 2020; Lal *et al.*, 2021; Frelih-Larsen *et al.*, 2022). This was reinforced when the Food and Agriculture Organization (FAO) of the United Nations declared 2015 as the International Year of Soil, a key focus of which aimed to increase awareness and understanding of the importance of soil for food security and provision of essential ecosystem services (e.g. nutrient cycling, carbon storage, water maintenance, filtering and buffering, biotic regulation, support of plant systems and human structures).

Consequently, there has been an increasing need for quantitative soil information for improved environmental monitoring and modelling (Minasny *et al.*, 2008; Mayr *et al.*, 2013; Gerke *et al.*, 2022) as well as for soil properties mapping. At the European level, there is also an increasing demand for soil data and information from policymakers to assess the state or health of soils (Tóth *et al.*, 2013; Panagos *et al.*, 2012), with over 60% of European soils considered unhealthy (European Commission, 2020).

Topsoil is the upper layer of soil or shallower surface soil, usually up to 20 cm. Because of this conditional position on topography (soilscapes), topsoil is susceptible to both interaction – as a driver within ecosystems – and pressure – as a resource for human activities. For example, topsoil is the primary interface between the land and atmosphere; it is the first portion of the soil to undergo intensive use, leaching, erosion (particularly bare soil), and to allow or avoid drainage. In terms of process and genesis, pedologically (i.e. colluvium and eluvium), topsoil may not have a direct relationship with the underlying parent material as this layer can be either naturally transported from one site to another via leaching (due to its surficial position on the landscape, which leads it to undergo transport and deposition across topography) or deliberately by anthropogenic activities employing topsoil application methods (e.g., ecosystem transfer technique for reconstructing valuable natural habitats in post-mining soils; site rehabilitation).

The soil's ability to filter water and absorb nutrients is primarily provided by topsoil. Thus, it plays a crucial role in agriculture, as plants are grown in the uppermost pedological layer. Another important feature of topsoil for environmental systems is related to climate change as most shorter term fluctuations in soil moisture content and total soil carbon stocks occur in this layer (Thomas *et al.*, 2020).

In terms of soil characteristics, they are generally classified on the basis of their properties, divided according to their physical, chemical and biological characteristics. Soil texture, which is a physical soil property, consists of a relative content of sand, silt and clay (i.e. soil particle size fractions – soil PSF or soil particle size distribution – soil PSD), thus, it is compositional in nature since the total sum of the three components (sand%, silt%, and clay%) is equal to 100. Hence, its analysis and statistical modelling requires specific methods to appropriately address this constraint. Particle sizes fractions range from less than 0.002 mm (e.g. clay) to greater than 2 mm (e.g. fine gravel) with distinct mineralogy – from primary silicate minerals and quartz (e. g. sand and most of silt sized particles) to secondary silicate minerals and other secondary minerals (most of clay fractions) (UNSW, 2007; Niang *et al.*, 2014).

Soil texture, due to the PSD, is one of the most important soil properties because it modifies and acts as a control on other properties (UNSW, 2007), such as permeability, fertility, chemistry, and moisture content, particularly the latter. For example, a high clay content may indicate high moisture content, high concentration of organic carbon (SOC) and low permeability (O'Geen, 2013; Gorrab *et al.*, 2015; Tobin *et al.* 2018; Byrne *et al.*, 2018; Coblinski *et al.* 2021). Soils with lower sand content – thus, higher fractions of silt and clay – will drain less easily and have higher average water contents; water content typically decreases with increasing sand percentage (Grote *et al.*, 2003). In engineering issues, such as road pavement damage, Benedetto and Tosti (2013) highlight the importance of developing methods for inferring reliable estimates of the moisture and clay content. Soil texture (as classes) and/or soil particle size (as fractions) are important for erosion, whether natural or anthropogenic causes. This is important for landscape change modelling or loss soil studies (Bezak *et al.*, 2024).

Many physical and chemical soil properties are strongly associated with clay fractions, while weathering-resistant primary minerals of the parent material are related to the sand and silt fractions (Niang *et al.*, 2014). Moreover, being a primary soil property, soil texture has a significant effect on the properties governing the retention and transport of water and chemicals in soils (Niang *et al.*, 2014), as well as on soil physical quality parameters (Reynolds *et al.*, 2002; Pittman & Hu, 2021). Another practical example is the utilisation of soil texture to estimate hydraulic parameters via pedo-transfer functions to avoid labour- and cost- intensive measurements of soil hydraulic parameters in the laboratory (Müller *et al.*, 2016).

The variability of a soil's texture has a significant influence on soil moisture distribution (vertically and horizontally in a soil layer), largely determined by its clay content. According to Gorraeb *et al.* (2015), the clay content is a relevant parameter since it has a strong influence on water storage or water retention capacity and availability for crop plants, field capacity and wilting point. The relationship between the size distribution of the voids in soil and the size distribution of the mineral particles it contains expresses the relationship between soil structure and texture.

Another soil property that links soil texture with soil moisture is the soil electrical conductivity, a soil quality indicator that is interesting for site-specific soil management (Molin & Faulin, 2013) and soil sensing. Due to the relationship between these, soil properties could be integrated into irrigation algorithms in smart agriculture and contribute to improving the accuracy of water prediction models (Gorraeb *et al.*, 2015). Microwave sensing applied to soils (proximal and remote) exploits the soil's electrical conductivity and topography conditions to measure soil properties as the radar backscatter signal is related to the soil dielectric property (moisture, and physical properties) and surface roughness (Ulaby *et al.*, 1978; Ulaby *et al.*, 1979; Dobson & Ulaby, 1981; Jagdhuber, 2012; Marzahn & Meyer, 2020). Building on this, soil information can be retrieved from a diverse range of Earth Observation (EO) platforms comprising satellite (spaceborne), aerial (manned or unmanned airborne platforms), and in situ to measure reflected or emitted electromagnetic radiation.

Examples of research employing remote sensing data applied to soils include the development of a global soil visible near-infrared (vis-NIR) spectral library (Rossel *et al.*, 2016) and the Global Soil Map project (Arrouays *et al.*, 2014). Ballabio *et al.* (2016) carried out a first approximation of the Global Soil Map products at a European scale, where they analysed to what extent the Land Use and Coverage Area Frame Survey (LUCAS) topsoil database could be used to map soil properties. These authors also modelled several soil properties (organic carbon, pH, sand, silt, clay, gravel, saturated water content, bulk density, to name a few), using remotely sensed data and hybrid approaches including regression kriging, and described the modelling of topsoil texture and related derived physical properties. Within this context, soil science has benefited from the development of digital soil mapping at local, regional and global scales, as well as from the availability of data from EO platforms. Examples of freely available data are those from satellite platforms such as Landsat, Sentinel, ALOS, Envisat, and Radarsat.

Notwithstanding the importance and potential of soil texture for understanding, analysing, estimating and predicting other soil properties, few studies address the use of microwave sensors – for example, Synthetic Aperture Radar (SAR) – and geophysical methods for soil texture retrieval (Gorrab *et al.*, 2015). The majority of such studies have focused on the utilisation of SAR for estimating soil moisture and carbon content and use soil texture as a secondary variable to retrieve these soil properties (Gorrab *et al.*, 2015; Zhou *et al.*, 2022). Furthermore, vegetation hinders radar signals from getting information from the underlying soil, especially with shorter wavelengths (higher frequencies). This situation has limited the use of SAR data over vegetated areas for studies focused on soils.

It is in this context that this research was undertaken. The work outlined here is concerned with the prediction and modelling of surface soil texture (topsoil) and seeks to evaluate the utility of SAR data (C-band) to estimate topsoil particle size fractions over a largely vegetated (grassland) study domain, located within the Republic of Ireland. Issues with using EO in places like Ireland, but also the potential for such use, and how to bridge the lack of studies employing SAR data for this aim are core topics in this study. The research here relies on legacy topsoil data and considers topsoil depth up to 10 cm. This depth was selected due to constraints related to the capabilities of SAR sensors to penetrate the soil (Babaeian *et al.*, 2019), as explained further in this work. Concerning the theoretical basis underpinning this research, in terms of understanding soil texture behaviour in the microwave spectrum, this work is based on pioneering studies that addressed the relationship between soil texture and soil moisture using remote sensing to analyse soil physical properties, which is found in Ulaby *et al.* (1978), Ulaby *et al.* (1979), Dobson *et al.* (1981; 1984), Hallikainen *et al.* (1985), and Rao *et al.* (1988). A detailed theoretical basis is provided in 1.3.2.1 and 1.3.2.2 sections.

This thesis is structured accordingly to present the main related concepts, research challenges/gaps, and the traditional methods to estimate soil PSF.

1.2 Soil properties: soil texture

Soil properties are divided according to their physical, chemical and biological characteristics (Figure 1.1). Soil physical properties include texture, moisture, structure, porosity, permeability, bulk density, and colour. Soil chemical properties (or soil chemistry) refers to colloids (organic matter, minerals), acidity, alkalinity, salinity, cation exchange capacity, anion exchange capacity, electrical conductivity, and pH, for example. Soil biological properties reflect the influence of the living organisms inhabiting soils, including bacteria, fungi, nematodes, earthworms, for example. As soil texture is the focus of this thesis, it is described here in detail.

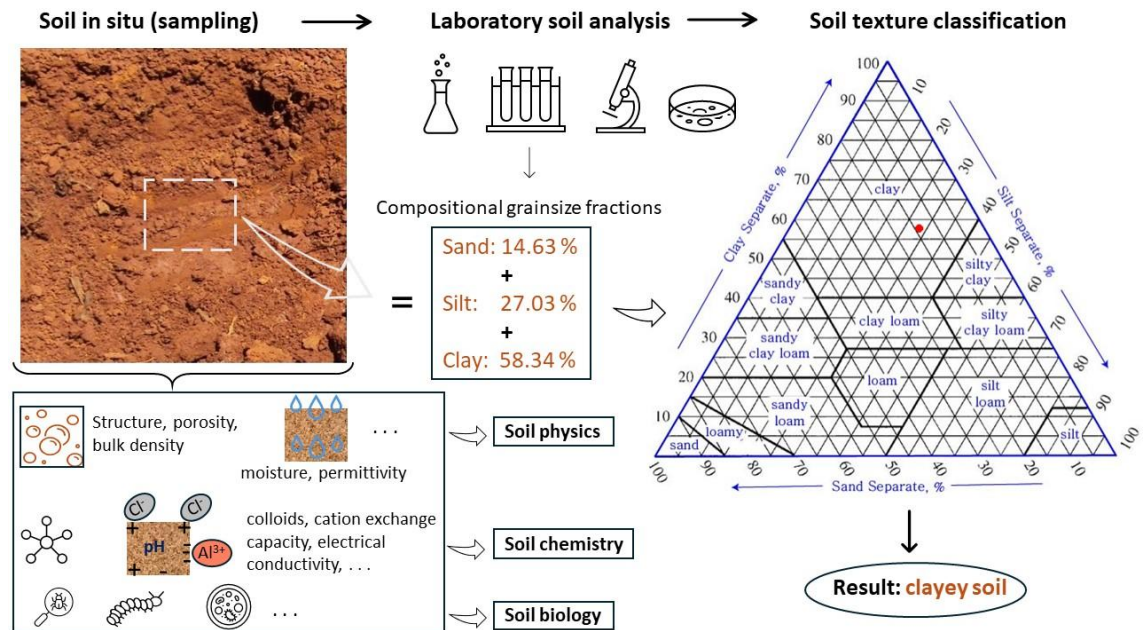


Figure 1.1 An example of a basaltic soil sample and its respective textural class resulted from laboratory analysis. Soil ternary source: USDA/NCSS.

Soil texture refers to the proportion of the soil related to the mineral component of soil expressed in terms of percentage of sand, silt, and clay – the particle size fractions (Figure 1.1). Particle size fractions control the properties and behaviour of soil (Table 1.1) due to its compositional nature, that is, they each impart their characteristics (UNSW, 2007). For example, considering soil moisture, soil texture strongly controls available water content (Müller *et al.*, 2016; Creamer *et al.*, 2016; Tobin *et al.*, 2018) into soils as water retention is mainly determined by clay content (Gorrab *et al.*, 2015). Thus, the smaller the size of a soil's particles, the greater its volumetric water capacity (Gorrab *et al.*, 2015), that is, clay soils have a greater water holding capacity than sandy soils (Creamer *et al.*, 2016). Soil texture has influence on a number of soil properties, including drainage, water holding capacity, aeration, susceptibility to erosion, cation exchange capacity, pH and buffering capacity (Creamer *et al.*, 2016) (Table 1.1).

Table 1.1 Relationship between soil particle size fractions and other soil properties

Property	Clay	Silt	Sand
Size range (mm) (AUS)	<0.002	0.002-0.02	0.02-2
Size range (mm) (USDA)	<0.002	0.002-0.05	0.05-2
Size range (mm) (WRB)	<0.002	0.002-0.063	0.063-2
Size range (µm) (AUS)	<2	2-20	>20
Size range (µm) (USDA)	<2	2-50	>50
Size range (µm) (WRB)	<2	2-63	>63
Observation	Electron microscope	Light microscope	Naked eye
Dominant minerals	Secondary	Primary; secondary	Primary
Attraction of particles	High	Intermediate	Low
Surface area	High-very high	Low-medium	Very low
Water holding capacity	High	Medium-high	Low
Aeration	Poor	Medium	Good
Compactability	High	Medium	Low
pH buffering capacity	High	Medium	Low
Element and nutrient retention	High	Medium	Low
Susceptibility to wind erosion	Low	Medium	Moderate (esp. fine sand)

Susceptibility to water erosion	Depends on degree of aggregation	High	Low (unless fine sand)
Consistency when wet	Sticky, malleable	Smooth	Loose, gritty
Consistency when dry	Hard clods	Powdery some clods	Very loose, gritty

Source: Adapted from UNSW (2007). Notes: AUS = Australia; USDA = United States Department of Agriculture; WRB = World Reference Base for Soil Resources. The soil classification/taxonomy system adopted by the Irish Soil Information System is that of USDA.

Soil texture can also affect the amount of pore space within the soil (O'Geen, 2013). Sand-sized soil particles exhibit large pores and a relatively small amount of total pore space (interstices). Conversely, clay-sized soil particles have small pores and more total pore space than soil made of sand-sized particles. In terms of permeability/percolation and infiltration rate, water behaviour (interstitial water content) is a function of soil texture and soil porosity (Figure 1.2). From this, sand and silt contribute less than clay to a soil's ability to retain soil water or nutrients (Table 1.1).

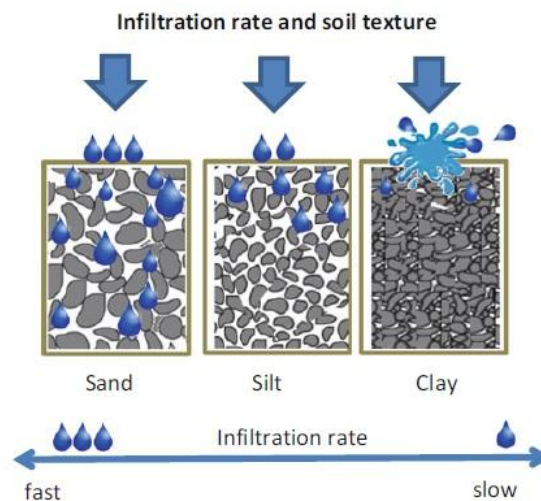


Figure 1.2 An example of water behavior to permeability/percolation and infiltration into soils (topsoils and deep soils) as a function of soil texture and soil porosity. Soils with higher sand content, thereby lower porosity, exhibit a higher infiltration rate compared to soils with high clay content (higher porosity). Such hydrological conditions affects soil moisture, soil dielectric constant, and consequently, radar backscattering from soils. Source: Retrieved from Tobin *et al.* (2018).

Soil texture (and mineralogy) is important for remote sensing (either optical or microwave types), since clay minerals (hematite, goethite, kaolinite, chlorite, and illite to name a few) affect soil reflectance (Coblinski *et al.*, 2021), dielectric properties (permittivity) (Hallikainen *et al.*, 1985; Das & Paul, 2015; Babaeian *et al.*, 2019) and electrical conductivity (Fletcher, 2022). Clay minerals influence the water dynamics in soil (Coblinski *et al.*, 2021) and both chemical and physical soil properties. Hence, soil particle size fractions play a crucial role in studies based on remote sensing techniques (either optical- or microwave electromagnetic spectrum-based) and geophysical methods for soil analysis.

Dielectric constant or relative permittivity is another parameter obtained from soil physical properties that is pivotal in microwave remote sensing. It refers to a measure of soil ability to store electrical energy (the amount of electric potential energy), in the form of induced polarisation, under the action of an electric field (Ilic, 2001). Some soil parameters are more

reliant on specific soil properties than others. For example, the available water content depends on the soil texture and structure (O'Geen, 2013). Table 1.2 shows an example of a range of values of water content for different types of soil in terms of the main textural classes.

Table 1.2 Values of available water content for different types of soil	
Soil texture	Available water content per soil depth (mm/m)
Sand	25 to 100
Loam	100 to 175
Clay	175 to 250

Source: FAO (1985).

Soil texture is sometimes confounded with soil structure. The former affects the latter but they are distinguished by the fact that soil structure can be damaged and changed by land management, whereas soil texture can rarely be changed (Emmet-Booth *et al.*, 2018). Soil structure refers to the aggregation of the individual particles of sand, silt, and clay assembled into clumps (Emmet-Booth *et al.*, 2018) or aggregates. Aggregation of soil particles can occur in different patterns, resulting in different soil structures that affects water circulation (FAO, n.d.). Soil structure is most usefully described in terms of grade (e.g., weak, moderate, strong), class (average size such as fine or thin, medium, coarse) and form of aggregates (granular/crumb, blocky/subangular, prismatic/columnar, and platy). Soil structural form has been referred to architecture of the soil (Emmet-Booth *et al.*, 2018) in terms of the physical arrangement of soil constituents.

1.3 Current approaches to measuring and mapping soil PSF and soil texture

1.3.1 Traditional methods

The most common dimensions (diametres) typically defined for soil grainsize or particle size fractions are those of USDA and WRB (Table 1.1). They are obtained from field sampling, the particle size is then determined by laboratory analysis and procedures (Figure 1.3). Once sand, silt, and clay are determined, they can be classified into soil textural classes using a soil triangle or soil ternary/diagram or simplex. Such graphic representation reflects the composition nature of soils as each side of the triangle refers to 100 units (i.e. %, g/Kg, etc.). Methods to analyse particle size include the pipette method (a reference method), sieving, hydrometer, laser beam, and field hand estimates or 'finger test' (less precise semi-quantitative method). Descriptions of each method are found in Ribeiro *et al.* (2020).

Regardless of method type, soil particle size fraction determination is a time-consuming and costly process particularly for larger areas, in a way that sampling design (including sub-samples), fieldwork and laboratory analysis are required. An example of sampling methodologies and laboratory analysis can be found in Jones *et al.* (2020).

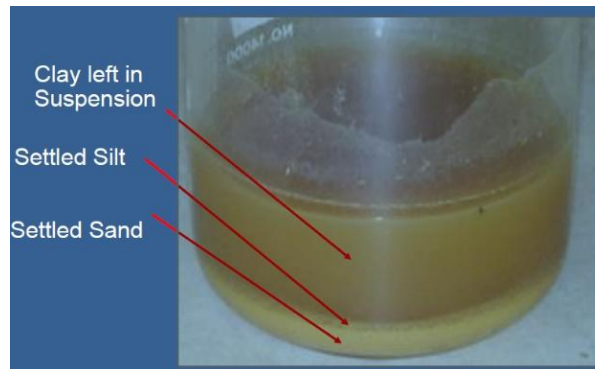


Figure 1.3 An example of soil particle size distribution analysed in laboratory, depicting the fractions settled in cylinder after sedimentation, based on Stokes' Law. According to Stokes Law, greater fractions (e.g. sand) fall faster in water than smaller ones. Source: USDA/ NSCS (n.d).

1.3.2 Soil property retrieval with SAR

Active and passive microwave remote sensors have been employed for soil properties estimation since the 1970s (Figure 1.4) employing both spatial and non-spatial approaches (Babeian *et al.*, 2019; Domenech *et al.*, 2020). Active sensors emit a pulse or beam towards an object and measure the received quantity of emitted energy; while passive sensors do not emit radiation but measure the backscatter or solar energy reflected by a target on the Earth's surface. With regards to active microwave remote sensing in soil science, SAR data has been used to map soil properties, including soil texture, moisture, organic matter content, and salinity (Beale *et al.*, 2019), since the radar backscatter – the received signal or response (echoes) from elements on the Earth's surface to a radar antenna – responds to soil characteristics, amongst other features on the Earth's surface.

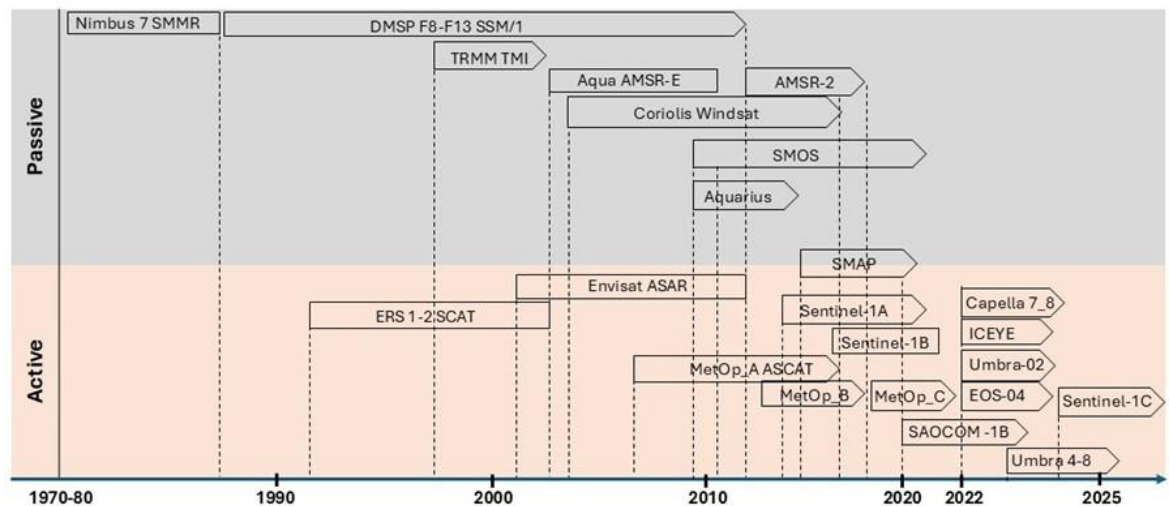


Figure 1.4 Active and passive microwave remote sensors (including Sentinel 1) used for generation of the European Space Agency climate change initiative data sets, including soil properties. The vertical black solid lines in each box represent the start/end of missions; the arrows indicate the continuity of the mission. Source: Adapted from Babaeian *et al.* (2019).

The SAR backscatter coefficient, measured as sigma nought (σ^0), is commonly used for retrieving or estimating soil properties. It is a normalised function of variables associated with the

viewing geometry and land surfaces (Beale *et al.*, 2019). More details on the backscatter coefficient is provided in Chapter 2.

The process of estimating soil properties from SAR data over vegetated soil is a challenging task (Wang *et al.*, 2023) due to scattering associated with vegetation, since either leaves (trees, crops) or grass may cover the entire soil surface. This results in attenuation of both the transmitted radar signal and the backscatter from the underlying soil (Beale *et al.*, 2019). Thus, vegetation is often the most significant factor to account for in SAR soil property estimation. On the other hand, some SAR system specifications, such as the incidence angle, polarisation, wavelength/frequency, may decrease the effects of soil roughness and vegetation on the radar signal (Ulaby *et al.*, 1978, 1979; Mattia *et al.*, 2006; Moran *et al.*, 2012). Barrett *et al.* (2009). Since incidence angle is the angle between the radar beam and the angle relative to nadir (normal) at a local surface (e.g. soil, leaf, water), it equals the angle of reflection and affects the backscattered signal from that surface. For example, concerning vegetation, at shallower incidence angles, most of the backscatter is expected to come from the plants (Moran *et al.* 2012; Morandeira *et al.*, 2021). Shallower incidence angles (> 40 degrees) maximise the response to crop conditions (Moran *et al.* 2012). Ulaby *et al.* (1979) showed that the effect of vegetation is minimised with incident angle lower than 20 degrees. Concerning soil roughness, Ulaby *et al.* (1979) observed that for bare fields, the effects of surface roughness are minimised with an incidence angle between 7 and 17 degrees. Wang *et al.* (2023) undertook a review of different approaches to retrieve surface soil moisture from SAR (Table 1.3), but it can also be extended to soil properties in general.

Despite the usefulness and potential advantages of SAR data for environmental analysis, particularly concerning soil and vegetation, Moran *et al.* (2012) point out some common constraints and drawbacks of satellite measurements of radar backscatter. These are (i) it requires a correction for inherent speckle (i.e. a grainy "salt and pepper" texture in an image) resulting in a lower operational resolution even if the image is acquired in fine resolution; (ii) cropped fields must be on the order of 1 km² to match the coarse spatial resolution of speckle corrected SAR data; (iii) the challenge of decomposing the signal associated with different incidence angles according to changes in crop/soil conditions; and (iv) total radar backscatter is a complex sum of the backscatter from vegetation and soil, making it complicated to determine if the signal is dominated by the crop or soil conditions (Moran *et al.*, 2012). Nonetheless, these are pre-processing procedures inherent to any SAR sensors.

Table 1.3 Summary of soil properties retrieval techniques from SAR data

Approach	Definition	Characteristics	Advantages	Disadvantages	References
1. Scattering Modelling • Empirical	Empirical scattering models are generally derived from specific data sets to capture the interaction of microwaves with natural surfaces through simple retrieval algorithms.	Regression fits between in situ measurements and backscattering coefficient (σ^0). The σ^0 is based on the amplitude information of the electromagnetic wave.	Simple and straightforward.	No physical basis behind the model. Usually only valid for the area under investigation.	Holah et al. (2005), Jiang et al. (2020), Wang et al (2023)
• Semi-empirical	The main structure of the model is based on theoretical models, and model coefficients are determined by a fit to experimental measurements. These type of models represent a compromise between the complexity of the theoretical models and simplicity of empirical models.	Based on theoretical models but have been extended according to empirical observations.	Offer a good compromise between the simplicity of empirical and complexity of theoretical models.	Each model has certain validity ranges. Generally only valid for bare soil surfaces (apart from Water Cloud Model).	Oh et al. (1992), Dubois et al. (1995), Wang et al (2023)
• Theoretical	Direct theoretical models simulate the backscattering coefficients for an area with known characteristics and specific roughness conditions. They comprise geometrical optics model, physical optics model, and the small perturbation model.	Simulates backscattering coefficient σ^0 as a function of volumetric soil moisture (m_v), root mean square (rms) of surface height and known radar configurations.	Not site-dependent.	Only accounts for single scattering terms. Many input parameters required making model implementation extremely complex.	Fung et al. (1992), Karam et al.(1992)
2. Change Detection • Image Differencing and Ratioing	<p>Image differencing involves the subtraction of backscatter intensity values between two different date images.</p> <p>Imaging ratioing divides the intensity values between the two dates, usually followed by a thresholding operation.</p>	Subtraction (& division) of intensity values from different dates.	Differences in σ^0 can be related directly to soil moisture.	Assumes surface roughness and vegetation remain time-invariant between dates.	Rignot & vanZyl (1993), Thoma et al. (2006), Thoma et al. (2008)
• Principal Component Analysis (PCA)	PCA or eigenvector analysis is a statistical technique to capture key spatial patterns in multi-dimensional datasets by transforming a number of correlated variables into a reduced number of uncorrelated variables.	Reduces the number of variables in a multi-dimensional dataset.	Enhances key patterns in the data. Change can be detected in the new “components”.	Assumes multitemporal data are highly correlated.	Lee et al. (1992), Knog et al. (2008)
• SAR Interferometry (InSAR): Interferometric coherence	InSAR exploits the phase (φ) information to calculate the interferometric coherence between two or more SAR scenes to provide complimentary information to that contained in the amplitude of the	Temporal (phase) decorrelation of σ^0 can be related to soil moisture changes.	Compliments the information contained in the amplitude of σ^0 .	Several different factors contribute to the phase decorrelation of σ^0 .	Borgeaud & Wegmueller (1996), Srivastava & Jayaraman (2001), Wig et al. (2022)

	<p>backscattering coefficient, through interferograms. The phase (ϕ) of a single SAR image is a measure of the two-way path length of the radar signal from transmitter to ground target back to receiver. Their phase difference ($\phi_1 - \phi_2$) can be used to generate topographic products such as Digital Elevation Models (DEMs).</p> <p>Interferometric coherence is a measure of the phase correlation between two co-registered complex SAR images, ranging from 0 (no coherence) to 1 (perfect coherence).</p>				
<p>• Differential SAR Interferometry (DInSAR)</p>	<p>DInSAR is a technique to separate topographic variations (i.e. DEM) and surface displacement to the path-length changes (InSAR phase), that is, two interferograms are subtracted from one another. This method extends the capability of InSAR for the measurement of small ground or surface deformations.</p>	<p>Topographic phase contribution is removed. Decrease in phase (surface displacement) is correlated to an increase in soil moisture.</p>	<p>Potential for estimating soil moisture with increased accuracies.</p>	<p>High coherence limits the application to sparsely or non-vegetated surfaces. A paucity of studies using this technique have been carried out to date.</p>	<p>Nolan et al. (2003) Zwieback et al. (2015)</p>
<p>3. SAR Polarimetry (PolSAR)</p> <p>• Polarimetric decomposition or target decomposition.</p>	<p>SAR polarisation means the modes in which radar signals can be transmitted and/or received (oscillation patterns), whose configuration (channels or bands) can be expressed, for example, as Vertical Transmit-Horizontal Receive Polarisation (VH) or Vertical Transmit-Vertical Receive Polarisation (VV). These channels can be used to determine structural information of the scatterer making PolSAR suitable for target recognition and detection applications.</p> <p>Pol-decomposition breaks down the backscattered signature into the various</p>	<p>Uses all polarisations to describe the complete complex scattering from within an imaged cell.</p>	<p>Advantage of using PolSAR over traditional SAR due to its ability to measure all the polarisation characteristics of a surface target simultaneously. Can separate the different types of scattering mechanisms within the imaged cell.</p>	<p>Limited by presence of vegetation (target decomposition theorems and Pol-InSAR interferometry can be used to compensate for these).</p>	<p>Hajnsek et al. (2003), Shi et al. (2005), Cui et al. (2017), Yang et al. (2019), Wang et al. (2023)</p>

	different scattering contributions coming from the imaged cell, yielding polarimetric parameters, such as coherence (γ), entropy (H), and alpha angle (α). The latter is the main parameter that can indicate surface scattering (e.g., soil) and volumetric scattering (e.g., vegetation).				
4. SAR Data Fusion • Active (SAR) and Passive (radiometer) microwave technologies	Data fusion in remote sensing combines multi-sensor datasets to create a better and accurate new image (output).	Integrates SAR and radiometer measurements to help discriminate between the multiple influences on σ^0 .	Vegetation (volume scattering term can be separated) and roughness effects can be minimised. Fine and coarse resolution data can be combined.	Issues in scaling and validation. Subpixel variability (passive data).	O'Neill et al. (1996), Bolten et al. (2003), Narayan et al. (2004), Gao et al. (2022)
• SAR and Optical		Same as above but with SAR and optical measurements.	Vegetation and roughness effects can be minimised.	Issues in scaling and validation.	Guo et al. (2019), Ai et al. (2021)

Source: Adapted from Barret *et al.* (2009) and Wang *et al* (2023).

1.3.2.1 Estimating soil texture

A theoretical explanation that underpins soil texture behaviour in the microwave spectrum refers to the relation between free water and bound water in soils (Dobson *et al.*, 1981), which is controlled by the distribution of soil particle size fractions. Water can move more freely around the soil particles when soil moisture increases, and it is important to highlight that free water has a dominant influence on the dielectric constant (Barret *et al.*, 2007). Thus, soil texture affects the backscattering coefficient, considering the fact that sandy soils contain a higher content of free water than clayey soils (Srivastava *et al.*, 2006; Das & Paul, 2015), whereas the latter contains more bound water. This was demonstrated by Jackson and Schmugge (1989) who found that water molecules are adsorbed onto the soil particles and effectively immobilize their dipoles, restricting bound water from interacting with the radar signal (Das & Paul, 2015).

Notwithstanding the importance of soil texture for understanding, analysing, estimating and predicting other soil properties, few studies address remote sensors for soil texture retrieval in vegetation-covered areas (Zhou *et al.*, 2022), notably grasslands. Concerning optical remote sensing, Zhou *et al.* (2022) highlight that multispectral data is useful to capture directly bare soil reflectance and, when representing vegetation properties, such data can indirectly predict soil properties. However, they point out that the applications of them in soil texture classification over vegetated soil are still limited. Typically, approaches to date have focused on the utilisation of SAR for estimating soil moisture and use soil texture as a secondary variable to model these soil properties as the water retention in soils are dependent on their particle size distribution (Gorab *et al.*, 2015).

Thermal remote sensing, a type of passive remote sensing, has been explored for estimating topsoil texture. Studies demonstrating such an approach are Rahimzadeh-Bajgiran *et al.* (2013), based on evaporative fraction and applied in the Canadian Prairies; Wang *et al.* (2015) and Müller *et al.* (2016) with the latter based on the concept of thermal inertia; and Sayão and Demattê (2018). However, Rahimzadeh-Bajgiran *et al.* (2013) point out that atmospheric conditions, such as clouds and water vapour, hinder the utilisation of remote sensing data within the optical and thermal bands with shallow measurement depth for surface and near-surface soil properties retrieval.

Rao *et al.* (1988), based upon both empirical and theoretical approaches, studied the effect of soil texture on microwave remote sensing through the analysis of three types of soil moisture measures – volumetric, gravimetric and field capacity – and the dielectric constant values of moist soils. The authors selected three test sites in India with different soil textures (clay, clay loam and sandy loam). The results of their analysis indicated a clear separation among the soil textures using gravimetric moisture, wherein discrimination increases as the soil moisture increases. On the other hand, for soil moisture expressed in terms of volumetric content, minimal textural dependence was observed on microwave reflectivity. Despite the objective to examine the best way to represent soil moisture while minimising the effect of texture on the microwave emission,

the study does provide an insight into the relationship between soil moisture and soil texture in terms of moisture measurement units, as most researchers normally focus on volumetric moisture. Such considerations play an important role in estimating soil texture, as soil particle size fractions influence the dielectric constant due to their interstitial water content (Das & Paul, 2015).

Hallikainen *et al.* (1985) evaluated the microwave dielectric behaviour of soil-water mixtures as a function of water content, temperature, and soil textural composition (clay, silt and sand) at frequencies between 1.4 GHz and 18 GHz., and found that soil texture had an effect on the dielectric behaviour over the entire frequency range, being most pronounced at frequencies below 5 GHz. According to Das and Paul (2015), the empirical curves demonstrated by Hallikainen *et al.* (1985) are the most common methods found in the literature to relate soil moisture and the dielectric constant in the absence of direct field measurement.

Zribi *et al.* (2012) proposed a method for the analysis and estimation of soil surface texture from TerraSAR-X over North Africa. Their study area included bare soils and olive groves, with the vegetation cover corresponding to approximately 5% of the soil surface area (Zribi *et al.*, 2012). The authors found a strong correlation between soil texture and a processed signal from a radar image acquired just after a rain event, and another image corresponding to dry soil conditions, acquired three weeks later.

1.3.2.2 Soil texture as a proxy for retrieving soil moisture

As satellite-coupled microwave sensors have been deployed since the end of the 1970s, there is a long history of research concerning the relationship between soil texture and soil moisture using remote sensing to analyse soil physical properties (e.g. Ulaby *et al.*, 1978; Ulaby *et al.*, 1979; Dobson *et al.*, 1981; Dobson *et al.*, 1984; Hallikainen *et al.*, 1985; Rao *et al.*, 1988). Soil texture and soil moisture are indicators of each other in models considering that the former affects the radar-backscattering coefficient through the trade-off between sand and clay fractions in terms of their interstitial water content, which is held by gravitational and electro-molecular forces, respectively. Bound water interact less with the radar signal (Das & Paul, 2015). Consequently, a higher proportion of sand results in an increase in the backscattering coefficient, as the amount of free water significantly contributes to the SAR backscatter (Das & Paul, 2015), while decreasing soil clay content increases the sensitivity of the radar signal, even in C-band SAR, such as Sentinel 1 (Das & Paul, 2015).

There is notable literature available concerning the retrieval of soil moisture from SAR using soil texture as an influencing variable. Works include Hallikainen *et al.* (1985), Rao *et al.* (1988), Niang *et al.* (2014), Pratola *et al.* (2015), Das & Paul (2015), Montzka *et al.* (2018), Beale *et al.* (2019), Bousbih *et al.* (2019), and Gu *et al.* (2021).

Gorrab *et al.* (2015) used a HH polarisation for retrieving both soil texture and soil moisture, at a 36° incidence angle. The imaging system used was TerraSAR-X in a flat area in Tunisia where vegetation cover is dominated by agriculture (cereals, olive trees, and market gardens).

Nevertheless, the test fields were targeted under bare soil. They demonstrated that it is possible to retrieve both soil texture and soil moisture using a single radar signal configuration (one incidence angle, one polarisation) (Gorrab *et al.*, 2015). In order to achieve this, they proposed a change detection method using the SAR data and ground auxiliary theta probe network measurements, for the soil moisture estimation, considering 3×3 pixel windows for the computation of effective soil radar signals. A linear relationship between the radar backscattering signals (in decibels) and soil moisture was modelled by Equation 1.1 (Gorrab *et al.*, 2015):

$$\sigma_{dB}^0 = S_0 mv + f(R) \quad \text{Equation. 1.1}$$

where S_0 is the radar signal's sensitivity to soil moisture (mv) and $f(R)$ is a function of the roughness R .

For the retrieval of clay and sand content (in %), they established two empirical relationships between the mean moisture values retrieved from the radar images and the two soil texture components over 36 test fields in soil (Figure 1.5).

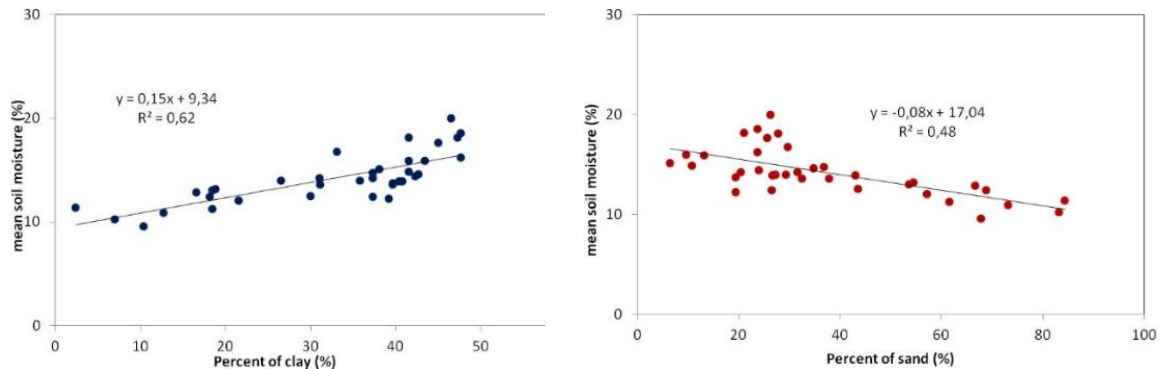


Figure 1.5 Relationship between soil moisture, estimated from radar moisture maps, and soil particle size fractions (clay% and sand %). Source: Gorrab *et al.* (2015).

In their methodological description for an algorithm based on the change detection approach for soil moisture retrieval, Gorrab *et al.* (2015) found a small improvement in the estimation of soil moisture when two TerraSAR images acquired under very dry conditions. amongst other five images acquired between late November 2013 and late January 2014, to account for temporal variations in surface roughness was introduced. Notwithstanding surface roughness is one of the factors that affect radar return. An accurate estimation of clay, with a root mean square error equal to 10.8%, was achieved after validation of the method in the second set of 34 fields. The authors proposed, and subsequently implemented, a simple algorithm to retrieve soil moisture from the TerraSAR-X radar sensor and field measurements, which allowed retrieval of soil texture, employing radar acquisitions with just one configuration (one incidence angle, one polarisation – HH). This approach is based on the approximately linear dependence of radar backscattering signals (in decibels) on soil moisture (Gorrab *et al.*, 2015). The authors also relied on the

relationship between the measured decrease in intensity of backscattered radar signals and clay present in the soil, as demonstrated by Zribi *et al.* (2012). Gorraeb *et al.* (2015) also identified the number of neighbouring pixels needed to minimize the effects of local terrain heterogeneities and speckle in the radar signals (i.e., 3x3). While these experiments were designed for bare soils, these authors concluded that it should also be possible to implement the proposed algorithm on surfaces with vegetation cover. They also concluded that the approach could also be applied to Sentinel-1 data.

1.4 Research motivation

The motivation for this research is twofold. In a more specific context, this work seeks to highlight one of the soil key attributes – particle size fractions and soil texture – for analysis and inference of other soil properties. Furthermore, the research develops a novel approach for estimating the proportion of sand, silt and clay (spatially and non-spatially), particularly at a regional scale, as a complement to the crucial but costly, labour and time-consuming, traditional methods of soil sampling and soil analysis.

The research addresses a critical gap in knowledge – the application of SAR data for soil textual retrieval over vegetation-covered areas. Currently, applications are mostly restricted to soil moisture retrievals over both bare and vegetated soils. Hence, this study is also motivated by contributions that can emerge for soil science and remote sensing knowledge.

In a broader context, the motivation relies on the consideration of soil as an important natural resource in a scenario of climate change and environmental crisis, when inadequate soil management can lead to erosion (loss and degradation), decreased soil health and fertility, and carbon dioxide (CO₂) release. Furthermore, soils are crucial for achieving the United Nations' Sustainable Development Goals (SDG) (Hou *et al.*, 2020) addressed directly and indirectly in the global policy framework of the SDG, such as goals number 2, 3, 6, 11, 12, 13 and 15, as well as one of the target pursued under European Union Mission programme through the “A Soil Deal for Europe” scheme.

1.4.1 Research rationale

The rationale for the research is addressed under three topics, according to the following:

- *Why soil particle size fractions / soil texture?*

Soil texture has a direct relationship with a number of soil properties (Creamer *et al.*, 2016), including permeability, fertility, moisture content, and organic carbon content, with the latter playing a central role in climate change (Lal *et al.*, 2021; Frelih-Larsen *et al.*, 2022). For example, there is an interplay between clay and soil organic matter (SOM) due to the ability of clay surfaces absorb and bind SOM and protect it from degradation (McDonald *et al.*, 2014). Soil texture can affect the ability of a soil to sequester carbon (C) following afforestation in a site (silt and clay have a greater capacity to store C and adsorb SOM) (Wellock, 2011). Most of the chemical and

physical properties of the soil (e.g., swelling – shrinking capacity, and cation exchange capacity) are governed by the presence of clay minerals in the soil (Kumari & Mohan (2021)).

Little attention has been given to soil texture in microwave remote, and even proximal, sensing application, notwithstanding the importance of this soil property in soil science. Various approaches have been proposed for soil moisture monitoring by using radar remote sensing systems (microwave-based and optical-based), whereas a paucity of studies have focused on the analysis of soil texture (Gorrab *et al.*, 2015). Typically, the latter is considered as an influencing and secondary variable when studying soil moisture or surface roughness from EO platforms (Gorrab *et al.*, 2015; Zhou *et al.*, 2022).

Gorrab *et al.* (2015) argue that soil texture has only a very limited influence on radar signals, and is difficult to directly retrieve one component of soil texture (e.g., clay) from such data (Gorrab *et al.*, 2015). On the other hand, empirical relationships based on a set of experimental data have been observed between soil texture and high-frequency X-band radar measurements (Gorrab *et al.*, 2015). The research here investigates to what extent there is an association between soil particle size fractions and radar backscattering for a given set of physical and meteorological conditions in Ireland.

It should be noted that soil mineralogy (i.e. analysis of constituents of soil grain size, such as vermiculite, illite, muscovite, goethite, hematite, silica, ilmenite, zircon, and monazite) is not explored in this research, despite its importance for remote sensing studies, particularly in optical-spectrum based sensors, in light of the Open Soil Spectral Library. Multi-spectral remote sensing is commonly employed for mapping surface mineralogy (Roberts *et al.*, 2019), and reflectance spectroscopy data (350–2500 nm) are successfully used in pedology as a predictor of soil colour and mineralogy (Poppiel *et al.* 2020). While dielectric models of the soils represent an important application in radar and radiometric remote sensing, the dependence of spectroscopic properties of the soil-bound water on soil mineralogy is yet not well-exploited (Mironov, 2004). Even if explored, such approaches using soil mineralogy would deserve another specific study.

▪ *Why topsoil?*

Topsoil, as the upper layer of soil, usually up to 20 cm depth, serves as support for ecosystem functioning and human activities. In most cases, topsoil is also related to the O horizon (organic horizon) if present and A horizon (surface horizon in mineral soils) (Figure 1.6 and Figure 1.7). Additionally, the top layer of soil is the first layer to undergo intensive use, leaching, and erosion (particularly bare soil), and to allow or avoid drainage, contamination, etc. The soil's ability to filter water and absorb nutrients is first provided by topsoil. Broadly, shallow soils are more susceptible to extreme environmental conditions (i.e. drought and waterlogging or flooding) (Patel *et al.*, 2021; Cybèle *et al.*, 2022). Furthermore, radar backscatter reflects the most surficial information from the terrain rather than underlying soil, particularly with C-band SAR such as

Sentinel 1. From this, it is feasible to use topsoil as a unit of analysis within the microwave region of the electromagnetic spectrum with short wavelength.

Terminology and relationship between segments of the Earth's Crust

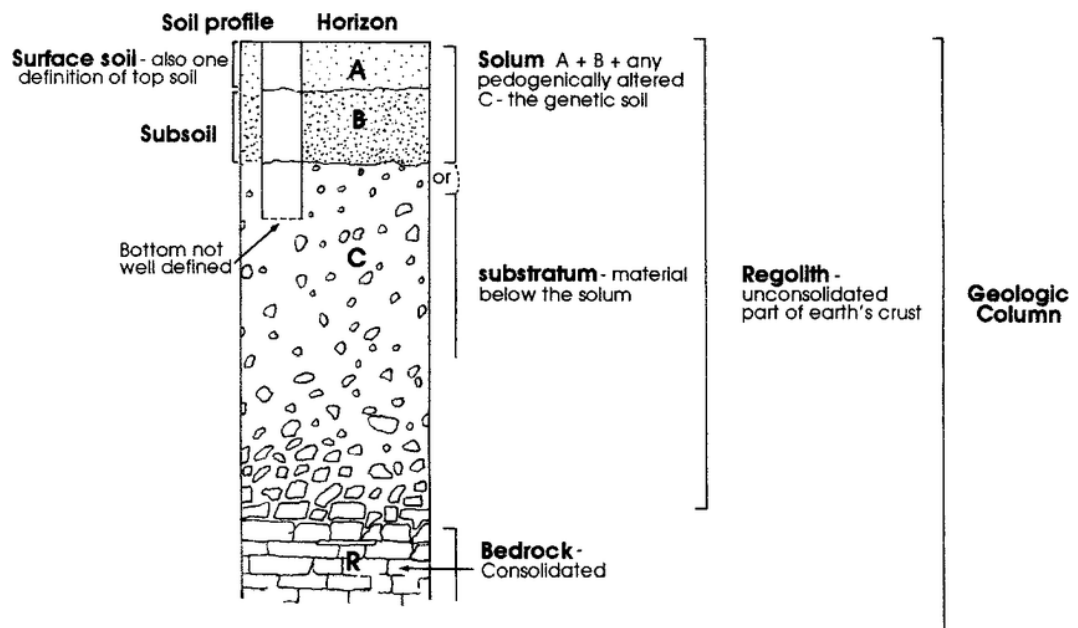


Figure 1.6 Idealised soil profile diagram with topsoil under pedological (eluviation) and geological contexts (parent material). Soil (or solum) consisting of A and B horizons. Source: Mahaney & Ramanathan (2003).

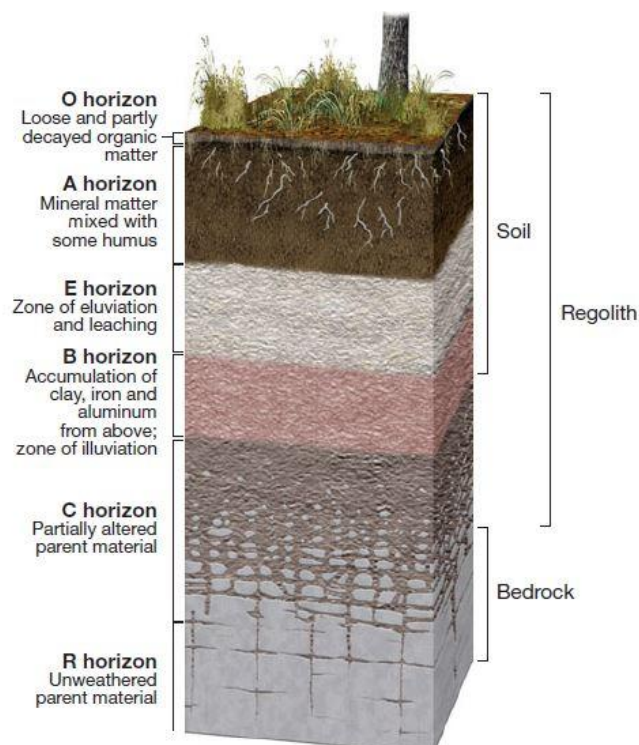


Figure 1.7 Idealised soil profile diagram with topsoil and organic layer in relation to the parent material, in terms of weathering results (regolith) and the original rock (bedrock). The combined O, A, E, and B horizons form the soil (or solum). Source: Hess & McKnight (2017).

- *Why SAR?*

Microwave remote sensing is an alternative to optical sensors under cloudy sky conditions, which is a common meteorological characteristic in Ireland. SAR sensor operates regardless weather conditions and luminosity level (i.e. it operates day and night). Furthermore, an active electromagnetic signal from a radar beam (e.g., from a C-band- and L-band SAR) can penetrate further into the soil (subsoil) than a pulse of light that bounces back (i.e. returns from the surface) to an optical sensor satellite (e.g. Sentinel 2).

On the other hand, the utilisation of SAR for the retrieval of soil parameters can be challenging, particularly where most natural surfaces contain some type of vegetation cover that influence the radar signal. This is the case for Ireland, where grassland is the dominant land-cover category (EPA, 2013), accounting for 54.33% of the country's territory (Eaton *et al.*, 2008).

So far, the investigation in other studies of soil parameters with polarimetric SAR remote sensing is constrained to bare soil areas and for retrieving soil moisture or soil roughness. SAR polarimetry exploits phase information and wave pattern oscillations from two orthogonal polarisations (i.e. H and V on each of transmit and receive) to determine the target response or scattering matrix. In addition, when dealing with soil properties retrieval under vegetation cover, it is important to separate individual scattering contributions from soil and vegetation, in order to extract only soil information (Jagd

huber, 2012). This research evaluates one method addressing this and relies on the work of Ulaby *et al.* (1978; 1979) and Dobson & Ulaby (1981), who investigated the microwave backscatter dependence on surface roughness, soil moisture, and soil texture, and Cloude & Pottier (1997) and Jagdhuber (2012), whose work focused on the separation of the individual scattering contributions from the soil and the vegetation.

The justification for the choice of a SAR approach in this thesis is based on the relatively greater ability of radar beam to penetrate the topsoil, and microwave remote sensors to deal with vegetation-covered soil due to wave phase information (polarimetry), compared to optical remote sensors. This work was conducted using a single sensor and satellite imagery at a specific date to be applied in a specified (single) study area. Notwithstanding the use of a time series of Sentinel 1 satellite images or a comparison of different SAR platforms (and wavelengths) and locations in Ireland are better approaches for results validation and demonstrating reproducibility, the decision on the datasets employed to carry out this research is mainly justified by a drier period/date criteria in order to ameliorate moisture effect due to rainfall in soil and vegetation. The specific rationale is highlighted accordingly in respective methodological chapters.

1.5 Knowledge gaps

The presence of vegetation over the soil surface reduces the backscattering sensitivity of soil. Concerning soil texture and particle size fractions, direct determination through SAR images is a difficult task as configuration of textural classes (i.e. particle size distribution) influences soil dielectric properties, arising from its water-holding capacity (Hallikainen *et al.*, 1985; Das & Paul,

2015). In studies that focus on retrieving soil parameters from airborne and spaceborne platforms, soil texture is usually considered as an influencing variable when studying soil moisture or surface roughness (Petropoulos *et al.*, 2015; Babaeian *et al.*, 2019). The majority of studies are restricted to bare soil areas or bare agricultural areas (Niang *et al.*, 2014; Gholizadeh *et al.*, 2018; Mondejar & Tongco, 2019) and to retrieval of soil moisture in covered soil areas. Within this context, this research also aims to evaluate a method that can deal with such an issue. The key knowledge gap is determining soil texture, as a primary variable, through microwave remote sensing, over vegetation covered-soil. This has particular relevance for countries like Ireland, where grassland is the dominant land-cover category (EPA, 2013).

1.6 Research questions

The key research questions underlying this thesis, which are linked to their corresponding objectives outlined in the next section, are as follows:

- How effectively can SAR data be employed to retrieve particle size fractions under vegetated topsoil? Is there a correlation between the backscattering coefficient and sand, silt, and clay fractions? (Objectives 1 to 4).
- To what extent can polarimetric decomposition predict soil particle size fractions? Can surface scattering ameliorate vegetation cover issues for soil particle size estimation? (Objective 5).
- What is the effect of treating soil particle size as compositional data into the models? (Objective 3).
- Can soil texture classes be indirectly obtained from the quantitatively predictive models? (Objectives 1 to 4).
- Should we employ spatial or non-spatial models to estimate soil grainsize fractions? What are the difference between them in terms of performance, accuracy and potential issues? (Objective 6).
- To what extent or under what environmental conditions can the method be reproduced in another study area? (Objectives 1, 2, 4, 5, and 6).

1.7 Research aims and objectives

The primary aim of this research is to identify, develop and evaluate a suitable methodological approach to estimate topsoil particle size fractions across scale using SAR data. A secondary aim is to test a method that can deal with the issue concerning radar signal interaction (or mixture) between soil-vegetation near the soil surface. In order to achieve these aims, a number of objectives are identified as follows:

Objective 1: Employ legacy data from soil surveys carried out in Ireland (point scale) to inform the selection of a suitable study area for training and testing the models (regional scale).

Objective 2: Utilise both linear-based and tree-based regression models to estimate sand, silt, and clay, in light of the variety and complexity of datasets.

Objective 3: Compare soil particle size fractions as compositional and non-compositional data and their statistical effects on modelling and estimation of soil texture.

Objective 4: Evaluate if topographical and geophysical covariates can improve the models.

Objective 5: Evaluate polarimetric decomposition adapted to dual-polarisation data such as Sentinel 1 as a method for dealing with soil under vegetation cover and the intrinsic microwave scattering effects of vegetation on soils.

Objective 6: Test two geostatistical methods (interpolators) for the spatial modelling approach.

1.8 Thesis structure

This thesis is divided into eight chapters. **Chapter 2** provides a combined theoretical background and review of literature underpinning the main concepts on this thesis's topic. The headings encompass topsoil and its layer depth considered in this research; the EO platforms for retrieving soil texture including the SAR technology; and soil properties modelling. This chapter provides a more general theoretical approach on the subjects addressed in this thesis whilst more details concerning theories are presented in the subsequent chapters according to their topic under analysis. Concerning the literature review, Chapter 2 outlines some related works on predicting soil texture and soil moisture from SAR, and how they deal with the effects of vegetation cover in soil backscattering in the models for those soil properties estimation. Chapter 2 concludes with a summary and review. It is worth noting that portions of chapter 2 are being included in a book, whose proposal was accepted for publication by Springer publisher. The related book chapter is in progress at the finalising stage of this thesis. **Chapter 3** presents the study area and highlights the reasons why Ireland was selected to evaluate the methods. It also provides a brief characterisation of soils and geographical setting in the context of the study area.

Chapter 4 focuses on estimating soil particle size fraction from a dual-polarised SAR. It outlines how soil texture was treated as compositional data within the model framework proposed. Chapter 4 was published in the European Journal of Soil Science (DOI: 10.1111/ejss.13414).

Chapter 5 presents the dual-polarimetric decomposition method employed to estimate topsoil particle size fractions under vegetation cover. It provides information on how the method was applied by bringing a statistical approach that underpins the idea behind this technique that was mainly developed for classification and segmentation purposes. Chapter 5 was published in the International Journal of Applied Earth Observation and Geoinformation (DOI: 10.1016/j.jag.103742).

Together, Chapters 4 and 5 address the challenges, gaps, objectives, and the methodology employed to achieve the aims/goals of the research outlined in this thesis. The proposed modelling

approach and methodology applied in this thesis are not without limitations, thus both these chapters also summarise the main points of the limitations of the study.

While Chapters 4 and 5 are rooted in a non-spatial modelling approach (i.e. without geographical coordinates), **Chapter 6** addresses two different spatially explicit models (e.g. built upon interpolation + semi-variogram and interpolation + regression) to estimate soil particle size fractions. That is, geographical location matters under this approach and all methods and analyses were carried out under geographical information systems (GIS). Chapter 6 provides a comparison of the results obtained from the non-spatial models addressed in Chapter 4 to those of spatially modelled, employing the same datasets, including environmental covariates, and considering the same study area. This was achieved using statistical evaluation metrics in a bid to answer one of the research questions previously mentioned in this thesis. In general, Chapter 6 is outlined through a theoretical background on spatially explicit models; a review of literature on the use of both statistical and geostatistical models for different soil properties estimates, but stressing soil particle size fractions and soil texture; a geostatistical exploratory analysis of the response variables, based on the approaches of spatial pattern, spatial dependence and neighbourhood. A description of the strengths and limitations of the work is also included in this Chapter. The three methodological sections of this study (Chapters 4-6) were designed to provide a continuous and coherent approach to addressing the research questions identified.

Chapter 7 provides a discussion of the results and findings by resuming the research questions and objectives presented in the introductory part and by informing to what extent they were achieved. It also discusses the limitations of the data, models, and methods employed in this study.

Finally, **Chapter 8** concludes this thesis by informing how this study has advanced the science for soil mapping and Geography. This final chapter also provides suggestions and directions for future potential works on this research topic and how might the approaches developed assist in the design of future soil sampling surveys – in Ireland or elsewhere. Chapter 8 also highlights the importance of the work beyond the research community, for example, by bringing in recommendations to soil legacy and soil policy, from which Ireland may also benefit.

In order to facilitate reading for those who are not familiar with some of the terminology in the field of remote sensing, Table 1.4 provides definitions of key terms used throughout this thesis.

Table 1.4 Glossary of the main terms related to remote sensing and geographical information system

Term	Definition	Acronym
Artificial intelligence	The ability of machines, computers or computer-controlled robot to perform and mimic tasks commonly associated with human mind.	AI
Backscattering	It is the received signal or response (echoes) from elements on the Earth's surface to a radar antenna, which serves as a measure of physical properties of targets (e.g., soil, vegetation, urban areas).	-
Data mining	Data mining is the process of deeply searching and analysing large raw datasets in order to identify patterns and extract useful information.	-
Digital mapping	Also known as digital cartography, is the process to create display data of locations and geographical phenomena (physical, social) as an image	-

	on the computer screen. Simply put, it is the art, science, and technology of making and using maps with a computer.	
Drone	Otherwise known as unmanned aerial vehicle (UAV) is a “flying robot that can be remotely controlled or fly autonomously using software-controlled” (Krishnan <i>et al.</i> , 2022)	UAV
Earth observation	It is a set of platforms (satellite and aerial) in remote sensing to measure reflected or emitted electromagnetic radiation in order to monitor land, marine and atmosphere.	EO
Electromagnetic spectrum	It refers to a ranking of distribution of electromagnetic radiation according to frequency or wavelength (radio waves, infrared radiation, visible light, ultraviolet radiation, X-rays, and gamma rays)	-
Geographical Information System	It is a computer system for capturing, storing, visualising, analysing, modelling and interpreting geographic data (spatial/geospatial).	GIS
Machine learning	It is a branch from artificial intelligence. It is a process to train a computer to learn from data in order to find patterns or extract information from data to make predictions and take decisions.	ML
Microwave remote sensing	A set of remote sensors that remotely (i.e. at a distance, without touching the target) operates using the microwave region of the electromagnetic spectrum. Example: radar.	-
Optical remote sensing	A set of remote sensors that operates using wavelengths of the electromagnetic spectrum from the visible to the near infrared up to thermal infrared.	-
Proximal sensing	Proximal sensing involves the use of sensors in close proximity to the targets (e.g., soil, plants) and is installed on platforms ranging from handheld, fixed installations, or robotics and tractor-embedded sensors.	-
Radar	Acronym: Radio detection and ranging. It is a sensing system that use radio waves to catch any reflected energy from targets.	Radar
Random forest	It is a popular statistical model and machine learning algorithm used for classification and regression problems. It was developed by Breiman (2001) and is based on collection of tree-structured classifiers and/or regressors.	-
Remote sensing	It is the science and technique of obtaining information about objects or areas on Earth’s surface from a distance, typically from aircraft or satellites.	-
Satellite	A type of EO platform designed to remotely monitor and survey the Earth.	-
Sensor	A device mounted on or embedded in EO platforms, which produces an output signal after sensing a target or physical phenomenon.	-
Spatial resolution	The physical dimension that represents a pixel of the image.	-
Tree-based model	Tree-based models are statistical models and machine learning algorithms that make predictions (classification and regression) by organizing data into a tree structure.	-

2.1 Introduction

This chapter brings in the concepts and theoretical principles, in tandem with a review of literature, that underpin the development of the proposed research. The theory approach addresses soil properties, electromagnetic spectrum, principles of SAR, modelling soil texture, statistical models and relevant literature on Irish soils.

Due to the nature of the relationship between soil moisture and soil texture and the methodological feasibility of the interplay between these variables, demonstrated in a number of studies, any relevant literature is also included. It comprises review articles that encompass recent advances and the state-of-the-art of remote sensing techniques on retrieving soil properties at various spatial and temporal scales. Some of the literature also highlighted how remote sensing techniques dealt with vegetation effects on the backscatter from the underlying soil.

The second section commences with a review on the state-of-art approaches employing remote sensing data and techniques to understand and retrieve soil properties in the microwave region of the electromagnetic spectrum. A discussion on SAR for retrieving soil properties and the main progress achieved and challenges faced by researchers is also presented. Insights about technical parameters of radars, such as polarisation and incidence angle, are identified and highlighted. A section on how soil moisture influences soil texture and vice-versa in radar signals is also included. The topic related to statistical modelling of soil particle size fractions encompasses both the statistical and geo-computation approaches to build spatial and non-spatial models in this research. The final subheading presents a summary and the knowledge gaps identified.

2.2 Earth Observation platforms

Earth observation (EO) is a subset of platforms to measure both reflected and emitted electromagnetic radiation, which comprises satellite, aerial (manned or unmanned) – remote sensors –, proximal sensors (e.g., Ground Penetrate Radar-GPR, and gantry or tripod-based), and ground sensors. Typically, there are three categories for measuring electromagnetic radiation from EO. They are visible, infrared and microwave. Hence, in terms of electromagnetic radiation, the EO can be divided into optical, thermal and microwave sensors. Frequency and wavelength determine the measurement depth of electromagnetic radiation (Figure 2.1 and Table 2.1).

The electromagnetic spectrum ranges from short to long wavelengths, wherein the former includes gamma and x-rays; the latter, microwaves and broadcast radio waves (Figure 2.1). The visible region comprises a small portion of the electromagnetic spectrum, whose wavelengths range from approximately 0.4 to 0.7 μm (400 to 700 nm). In the infrared region, the reflected infrared (IR) is of interest in soil analysis and is typically measured by optical remote sensing

platforms – those of making use of visible, near-infrared, and short-wave infrared sensors. In microwave spectrum, the shorter wavelengths have properties similar to the thermal infrared region while the longer wavelengths approach the wavelengths used for radio broadcasts (CCRS, nd). Electromagnetic waves undergo an atmospheric scattering and absorption effects, particularly in the visible spectral region. Radiation in shorter wavelengths is more readily scattered than longer wavelengths – associated with the interaction between the radiation wavelength and particles of different sizes in the atmosphere. The areas of the spectrum where atmospheric effects are limited or non-existent are known as atmospheric windows (Walsh, 2003; Jensen, 2009), that is, where radiation or emitted energy in particular wavelengths can pass through. These "windows" are used in remote sensing to obtain information concerning Earth's surface and objects on it.

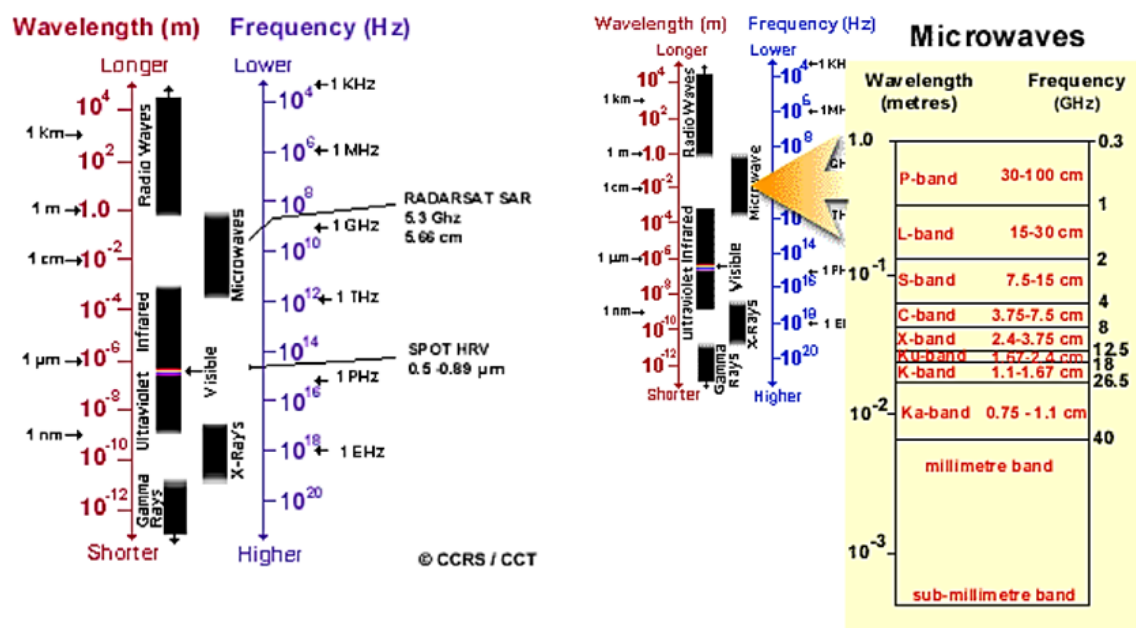


Figure 2.1 The electromagnetic spectrum (a) with focus on microwaves region (b). Source: Canada Centre for Remote Sensing (CCRS). (n.d).

The main advantages of satellite platforms for environmental monitoring are their capability of sensing differences in surface properties using different parts of the electromagnetic spectrum, and their ability for measuring different surface properties – for example, plant health, soil moisture, soil texture, surface roughness, surface temperature (O'Hara, 2019). On the other hand, such EO platforms have a number of limitations including, the wavelength-dependent attenuation of electromagnetic radiation in visible and infrared regions, potential for occlusion by clouds and cloud shadow and poor temporal resolution (O'Hara, 2019), particularly with regards to optical sensors. As a result, typical environmental conditions (e.g., clouds and illumination conditions) and technical characteristics of satellites (e.g., spatial, radiometric or spectral, and temporal resolutions), have to be taken into account regarding the choice of EO platforms for use in a

specific application or purpose. Figure 2.2 depicts a general overview of different types of remote sensing platforms with an example of a radar application to a soil property retrieval.

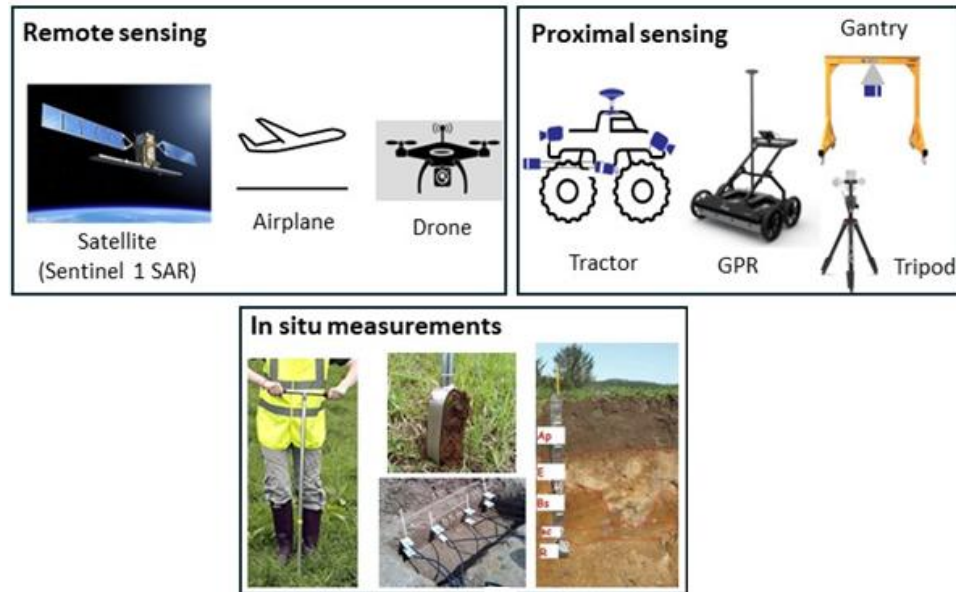


Figure 2.2 Examples of Earth Observation platforms and ground sensors. Source: Adapted from Babaeian *et al.* (2019).

Microwave or radar sensors are not limited by cloud cover or illumination conditions. Radar is an acronym that stands for radio detection and ranging whose wavelengths is longer than visible light (Figure 2.1). They are systems that actively send electromagnetic microwaves from a satellite or an aircraft and receive information from the Earth through a radar antenna that receives the backscattered echoes from the continent- or ocean surfaces.

While spaceborne SAR can be used to retrieve information on near-surface soil properties, surface features such as topography, roughness and wave interference affects radar signal return. The issue is that such interferences result in speckle that is inherent with active microwave systems. Thoma *et al.* (2008) point out that the methods designed to account for physical surface factors, image speckle and radar geometry influence on return signal can be grouped into empirical, semi-empirical and physically-based models. They highlight that regardless of retrieval method, imagery is typically filtered using speckle reduction filters and the use of relatively small (i.e., several pixels) moving windows. A limitation on this procedure is that the relationship between filter window size and retrieval accuracy is not determined in the evaluation of image filtering techniques but only in terms of image metrics (Thoma *et al.*, 2008). Thereby, they used field measurements and statistical analysis across multiple watersheds in order to determine backscatter, speckle removal and signal enhancement, as well as the appropriate scale for soil moisture retrieval, in three watersheds located in the United States of America (Arizona, Oklahoma and Georgia), using ERS-2 and RADARSAT-1 imagery (C-band SAR). They found that a 5×5 filter level (stepwise median filtering) was the optimum for estimating watershed averaged soil moisture and the optimum ground resolution for the study areas was 162 m (Arizona), 310 m (Oklahoma), and 1131 m (Georgia) (Thoma *et al.*, 2008).

Thoma *et al.* (2006) tested empirical, physical, semiempirical and image difference techniques (i.e., delta index) to estimate soil moisture employing Radarsat-1 and ERS-2, both C-band SAR. These can be useful for other retrieval purposes such as soil texture. Delta index implicitly accounts for topography, vegetation and roughness (Thoma *et al.*, 2006) as it relates the radar backscatter from dry soil and wet soil in a pixel for a given geographic location at a different time. It was calculated, per pixel and on a time-series basis, using wet and dry image pairs (Thoma *et al.*, 2006). Apart from bare soils, the experiments were carried out in sites with low-density grass and a mixture of forest, crops, and pastures.

Babaeian *et al.* (2019) described advances and the state-of-the-art of ground, proximal, and remote sensing techniques for retrieving soil moisture at various spatial and temporal scales. They highlighted (i) the straight relationship between wavelength and the depth of the soil layer measured by microwave signals (Babaeian *et al.*, 2019); and, (ii) the increase in attenuation of the microwave signal in vegetated surfaces, which affect the soil layer depth contributing to soil moisture retrieval (Babaeian *et al.*, 2019). The conclusions with respect to the remote sensing techniques are that (i) advances are needed for deriving physically based models for soil moisture retrieval from high spatial/temporal resolution optical/thermal remote sensing data; (ii) the combined use of active and passive microwave products may improve the resulting soil moisture retrieval accuracy and the spatial resolution. In general, such conclusions also apply to soil texture retrievals.

In another review encompassing studies covering the period from 1978 until 2015, Das and Paul (2015) described the fundamentals of active and passive microwave remote sensing technology and the models and algorithms to retrieve soil moisture, outlining a number of limitations that exist in soil moisture estimation algorithms and the major factors that affect radar configurations. Among their conclusions, the following are highlighted: (i) each of the methods and models has limitations of retrieval capacity or the target surface characteristics, in terms of their microwave bands (L, C and X) (Das & Paul, 2015); (ii) the contribution of other factors that influences the soil reflectance may not be effectively minimized (Das & Paul, 2015); (iii) at microwave frequencies, many natural surfaces do not fall into the validation regions of the theoretical models, and the available models fail to provide results in good agreement with experimental observations (Das & Paul, 2015).

A number of studies have proposed the implementation of multi-sensor (remote and proximal) and data fusion approaches for measurements of soil properties, instead of a single-sensor approach, even though such an approach is still at its early stages of development (Mouazen & Shi, 2021). Examples are found in Rahimzadeh-Bajgiran *et al.* (2013), Bousbih *et al.* (2019), Abdul Munnaaf *et al.* (2019) and Xu *et al.* (2020). Bousbih *et al.* (2019) combined Sentinel-1 (radar data) and Sentinel-2 (optical data) for estimation and mapping of soil clay using support vector machine and random forest, resulting in an accuracy of 63–65%. However, Mouazen and Shi (2021) highlighted that the multi-sensor data fusion approach should be applied

with caution, as the prediction accuracy of a soil property depends on interactions between data from different sensors, which might lead to deterioration of results. Using different sensors, different time-series, different in-situ measurement instruments, and same geographical location may not capture processes occurring in soils and sources of variabilities linked to soil dynamics in different scenarios of land management practices (Mouazen & Shi, 2021) due to the challenge task to meaningfully match- and obtain information contained in a pixel of different sizes and that in the field from different soil sampling units (e.g. soil profiles, boreholes, trenching). A main issue arise from successfully obtaining soil spectral- and soil backscattering dynamic feedback – from visible and microwave spectrum-based sensors, respectively – to account for multiple drying after rainfall (Mouazen & Shi, 2021). In this context, data fusion approaches depend on synergies between data coming from different sensors. (Mouazen & Shi, 2021).

Of relevance, Babaeian *et al.* (2019) provide a summary of the measurement depths of microwave bands (C-band, L-band, and P-band) for various land surface covers (Table 2.1. See also Figures 2.4 and 2.6), which shows the penetration capability of short- and long wavelengths SAR to reach the topsoil layer under different soil covering conditions – grass and forest). Babaeian *et al.* (2019) also explained that the measurement depth ranges for each microwave band were calculated based on one-tenth ($\lambda/10$), one fifth ($\lambda/5$), and one half ($\lambda/2$) of the wavelength ranges for bare soil, sparsely vegetated surfaces and fully vegetated surfaces, respectively. These authors also take heed of such depths being approximation values, as soil moisture content, soil texture, and surface roughness also affect the measurement depth of microwave signals. On the other hand, that information provides an idea about the microwave band type for a specific application.

Table 2.1 Approximate Microwave Measurement Depths for Bare Soil and Different Land Surface Covers

Land surface cover	Microwave bands			
	X-band (2.5 – 3.75 cm) (8 – 12 GHz)	C-band (3.75 – 7.5 cm) (4 – 8 GHz)	L-band (15 – 30 cm) (1 – 2 GHz)	P-band (30 – 100 cm) (0.3 – 1 GHz)
Bare soil	~1.25–1.87 cm	~1.87–3.75 cm	~7.5–15 cm	~15–50 cm
Agriculture and rangeland	~0.5–0.75 cm	~0.75–1.5 cm	~3–6 cm	~6–20 cm
Forest	~0.25–0.37 cm	~0.37–0.75 cm	~1.5–3 cm	~3–10 cm

Source: Babaeian *et al.* (2019).

2.2.1 Synthetic Aperture Radar

Synthetic aperture radar (SAR) is an instrument that is capable of retrieving bio- and geo-physical measures on the Earth's surface. These returned signals, whose intensity depends on target characteristics, are known as backscattering which serves as a focus of interest for studies and applications in remote sensing to detect physical properties, including soil properties.

The principle of Doppler beam sharpening or Doppler shift underpins SAR sensing (Van Zyl & Kim, 2010) - since Doppler assumes movement considering the object and observer, providing the apparent frequency change. It refers to the measurements of the frequency of the signal that is returned by the targets (multiple echoes) and indicates the direction from which those echoes

come from. Hence, targets can be separated in the along-track direction on the basis of their different Doppler frequencies (Van Zyl & Kim, 2010).

Radar waves (Figure 2.3) interact differently with soil, vegetation, water, ice and built surfaces, as their geometric and electrical characteristics are the primary factors influencing an object's return signal intensity (Liu *et al.*, 2019). SAR is relatively unaffected by illumination (i.e., it can image at night) or clouds. Such characteristics of SAR systems are due to the fact that the wavelengths of the spectral bands of these sensors are much longer than the average diameter of the particles and molecules found in the atmosphere (Sano *et al.*, 2020. Translated by Sandra Deodoro).

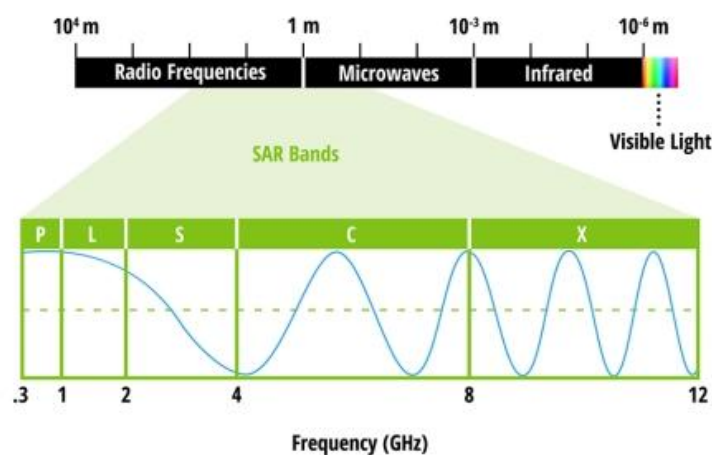


Figure 2.3 Typical SAR bands with their respective wave representation in the electromagnetic spectrum, depicting the microwave remote sensing. Source: NASA EarthData (n.d.)

The longer the wavelength (and the lower the frequency), the deeper the penetration into targets, particularly in forests (Table 2.2). Moreover, longer microwave wavelengths ($\lambda > 3.8$ cm) have some important advantages over visible region of the electromagnetic spectrum for agricultural applications (Moran *et al.*, 2012), owing to their ability of higher penetration within the vegetation volume, and the SAR signal can reach up to the underlying soil surface (Figure 2.4). For example, L-band sensors are minimally affected by vegetation (Beale *et al.*, 2019) depending on the geometry and leaves size, and foliar area density; and P-band (30cm – 100cm) can penetrate even tree cover to reach up to underlying soil volume (Srivastava *et al.*, 2006). Nevertheless, apart from wavelengths, the capability penetration of SAR signals also depends upon plant parameters such as plant moisture, plant height and plant volume (Srivastava *et al.*, 2006).

Table 2.2 Radar bands and applications

Band	Frequency (GHz)	Wavelength (cm)	Platform/Mission	Descriptions / Applications
P	0.3-1	100-30	Some products of UAVSAR	<ul style="list-style-type: none"> – Greatest penetration depth through vegetation and into the soil. – Ideal for soil moisture and biomass. – Difficult to operate from space due to ionospheric effects.

L	1-2	30-15	PALSAR, UAVSAR, AIRSAR, JERS-1, Seasat	<ul style="list-style-type: none"> – Provides vegetation penetration. – Applications included sea ice, geophysical monitoring, vegetation mapping (including tropical forest), soil moisture, and biomass. – Subject to ionospheric effects.
S	2-4	15-7.5	NISAR	<ul style="list-style-type: none"> – Little but increasing use for SAR-based Earth observation. – Agriculture monitoring.
C	4-8	7.5-3.8	ERS-1, ERS-2, RADARSAT-1, Sentinel-1	<ul style="list-style-type: none"> – Variety of applications, including sea ice, ocean winds, glaciers, and land-surface monitoring for motion risks; forest, water, and soil management.
X	8-12	3.8-2.4	DMSAR, ASTRIUM, SRTM	<ul style="list-style-type: none"> – Urban monitoring. – Ice and snow. – Vegetation cover (little penetration). – Fast coherence decay in vegetated areas. – Military reconnaissance.
Ku	12-18	2.4-1.7	- SNL sensors (Sandia National Laboratories-US) - SCLP (NASA)	<ul style="list-style-type: none"> – Rarely used for SAR (satellite altimetry). – Snow and Cold Land Processes
K	18-27	1.7-1.1	GRACE; NASA TDRS; GENEVO	<ul style="list-style-type: none"> – H₂O absorption (rarely used) – Vehicle navigation – Traffic supervision
Ka	27-40	1.1-0.8	KaSAR (ESA)	<ul style="list-style-type: none"> – Rarely used for SAR (airport surveillance) – Natural (land, water) and anthropogenic targets

Source: Adapted from ASF (n.d.), NASA (n.d.), and OSCAR (n.d.).

Imaging SAR sensors have demonstrated their potential for the effective measurement and monitoring of soil surface characteristics at a high spatial resolution (Gorrab *et al.*, 2015). Sentinel-1 is a C-band SAR, from the European Space Agency, that operates at a 5.4 GHz frequency, and an active sensor for near-surface soil moisture estimations (Babaeian *et al.* 2019).

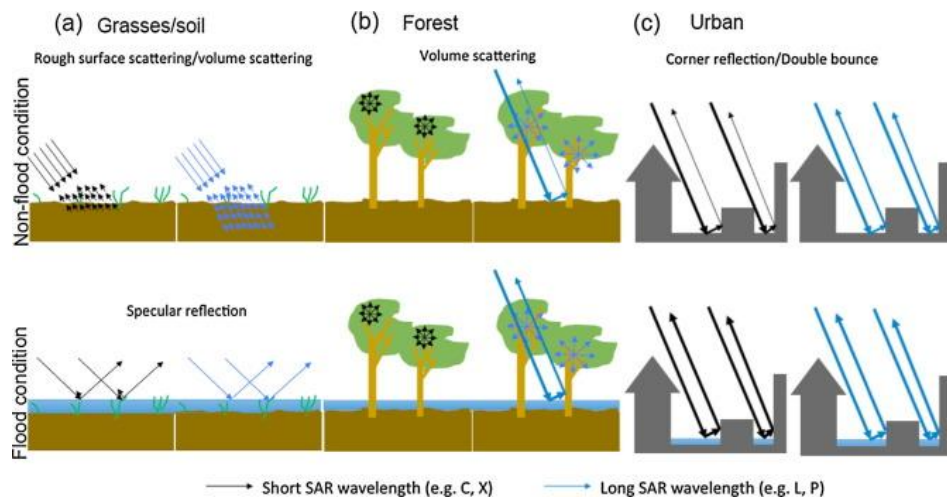


Figure 2.4 An example of the interaction of different wavelengths with different land cover types and scattering objects. The intensity of backscattered energy changes with wavelength and the way radiation is scattered. Source: Schumann & Moller (2015).

There is a crucial point in terms of the interaction between the geometry of the target and the wavelength (Liu *et al.*, 2019). Object geometry (e.g. trunk, large- and small leaf, stem, grass, soil aggregate) and orientation are related to the physical interaction between topographical and surface related effects. Thus, surface roughness, topography and the geometric appearance and

density of the vegetation are three main geometric parameters (Woodhouse, 2021). The size of the object influences the interaction between the object and the incoming microwave radiation – the wavelengths tend to interact more with objects that are similar in size to the wavelength. This is due to the Rayleigh's Criterion (Bole *et al.*, 2014; UIUC, 2010; Harvard, n.d.) that is usually applied to circular apertures, which is the case of SAR (i.e., cross-track direction and backscattering cross-section). As a result, longer wavelengths such as L or P bands, interact less with the smaller components (UIUC, 2010). For example, leaves and the small twigs have less impact on the microwaves that are passing through the canopy. Thus, in addition to the ability of microwave radiation to penetrate targets, there is also the ability to interact with them.

2.2.1.1 Radar polarisation

Radar polarisation refers to the mode in which radar signals are transmitted and/or received. In radar, waves are typically vertically or horizontally polarised. Depending on how the sensor is designed in terms of polarisation configuration, it can have a Vertical Transmit - Horizontal Receive Polarisation (VH) or Vertical Transmit - Vertical Receive Polarisation (VV), for example. Hence, radar polarisation refers to like-polarised waves (HH or VV) and cross-polarised waves (HV or VH), being all dual-polarisation (Figure 2.5). The former is also named parallel polarisation or co-polarisation and means that the scatterers reflect the wave in the same polarisation. Conversely, in cross-polarised, the signal may come back in a different plane as the transmitted and received polarisations are orthogonal to one another. Radar systems can have one, two or all four of these transmit/receive polarisation combinations (Table 2.3).

Table 2.3 Examples of polarisation types in radar systems

Type	Polarisation
Single polarised	HH or VV (or possibly HV or VH)
Dual polarised	HH and HV, VV and VH, or HH and VV
Alternating polarisation	HH and HV, alternating with VV and VH
Quad- or fully polarised	HH, VV, HV, and VH

Source: Canadian Space Agency (s.d).

Typically, the intensity of the backscattered signals in parallel polarisation is greater (usually strong) than in cross-polarisation (Sano *et al.*, 2020. Translated by Sandra Deodoro; Ibrahim *et al.*, 2023), which is usually weak. On the other hand, cross-polarisation is associated with multiple scattering and is useful when volume (multiple) scattering occurs. Thus, it can allow the separation of soil and vegetation, and forest and non-forest (Sowter, n.d.). Co-polarised backscatters are less sensitive to vegetation, easier to calibrate and less susceptible to system noise than cross-polarised signals, using incidence angles between 30° and 45° (Dubois *et al.*, 1995; Das and Paul, 2015).

A radar's polarisation type, to a certain extent, poses a penetration capacity in the targets on the Earth's surface, as the polarised signal is associated with geometry and depends on incidence angle (Beale *et al.*, 2019; Liu *et al.*, 2019). For example, parallel polarisation (or co-polarisation)

can penetrate vegetation more than cross-polarisation. In soil applications, polarisation implies the capability of an electromagnetic signal to reach the topsoil and backscatter, providing information about it.

In light of wave polarisation types, SAR polarimetry is an important concept in SAR applications. It refers to the polarisation state of the electromagnetic wave in an electromagnetic field including the magnitude and relative phase. SAR polarimetry also relies on the different oscillation patterns of the transmitted/received pulses. Thus, it is a useful technique for retrieving soil properties as polarimetry provides information for the physical interpretation of scattering processes and physical parameter retrieval. (Jagdhuber, 2012; Jagdhuber *et al.*, 2013; Niang *et al.*, 2014; Jagdhuber *et al.*, 2015). Due to this unique characteristic of SAR polarimetry, polarimetric decomposition is one of the main applications of SAR and a commonly established technique for separating elementary scattering processes in microwave remote sensing. Typically, such a method can be divided into *coherent* – only one dominant scatterer is present within the resolution cell, which seems quite unlikely for natural environments (Jagdhuber, 2012) – and *incoherent* types, in which more than one scatterer occurs in the resolution cell. Incoherent decompositions are based on second-order statistics including depolarization effects (Jagdhuber, 2012). More details on polarimetric decomposition methods are presented in later chapters, particularly Chapter 5.

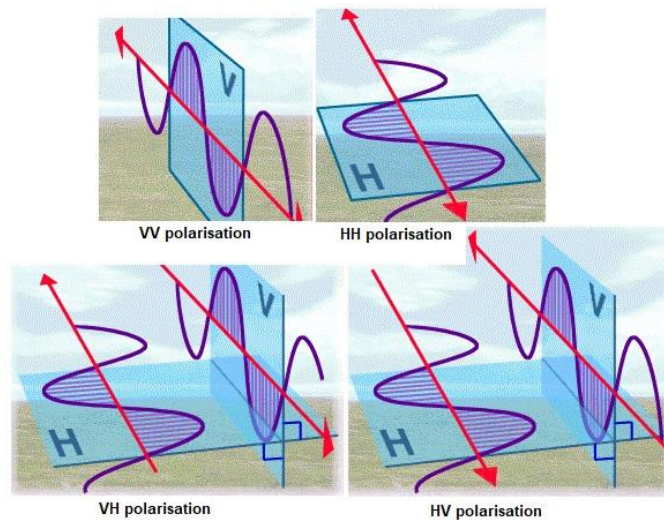


Figure 2.5 Examples of dual polarisation. Source: ESA (2007).

Niang *et al.* (2014) used SAR polarimetric data extracted from a RADARSAT-2 image as covariates for developing a digital map of soil surface texture in Montreal, Canada. The image's description is a quad polarisation mode or full polarimetric (HH, HV, VH, and VV), single look complex (SLC), fine spatial resolution (5.2 by 7.7 m, range by azimuth) and an incidence angle as of 23°. They applied the Lee refined filter (Lee and Pottier, 2009) to the radar data and extracted the polarimetric decomposition parameters (alpha, entropy, and anisotropy) employing different decomposition methods (e.g., Freeman and Durden, and Touzi decompositions). The authors

treated soil surface texture as compositional data applying the isometric log-ratio transformation to soil texture data.

Another important concept in radar remote sensing is the radar phase, which is usually applied to the oscillation of electromagnetic waves. Precise knowledge of phase properties in radar signal data is a key element in interferometric as well as in polarimetric SAR (ESA, n.d.). The phase of an image is exploited in polarimetric decomposition, interferograms, or digital elevation models (Sano *et al.*, 2020. Translated by Sandra Deodoro).

Babaeian *et al.* (2019) highlight the use of cross-polarisation in Sentinel-1 to correct for seasonal vegetation effects in the co-polarised backscatter measurements. In general, horizontally polarised (HH) is preferred due to the better penetration of these waves through vertically oriented vegetation (grasses, crops) to reach the soil (Brown *et al.*, 2003; Babaeian *et al.*, 2019). This is consistent with Dubois *et al.* (1995), Das and Paul (2015) and Sano *et al.* (2020) for soil moisture context, as outlined previously. Table 2.4 shows examples of spaceborne SAR systems with the respective radar bands and polarisation types.

Table 2.4 Characteristics of major spaceborne SAR systems

Platform	Sensor	Band	Polarisation	Highest Spat. Resolution(m)	Operational period
SEASAT	SAR	L	HH	25	1978
SIR-A	SAR	L	HH	40	1981
SIR-B	SAR	L	HH	25	1984
Almaz-1	SAR	S	HH	13	1991-1992
ERS-1	AMI	C	VV	30	199-2000
JERS-1	SAR	L	HH	18	1992-1998
SIR-C	SIR-C	L, C, X	VV, HH, HV	30	1994
X-SAR	X-SAR	L, C, X	VH, HH	30	1994-1996
ERS-2	AMI	C	VV	30	1995-
RADARSAT-1	SAR	C	HH	10	1995-
SRTM	C-SAR	C	VV, HH	30	2000
SRTM	X-SAR	X	HH	30	2000
ENVISAT	ASAR	C	VV, HH, HH / VV, HV/HH,VH/VV	30	2002-
ALOS	PALSAR	L	Quad-pol	10	2006-
TerraSAR-X	X-SAR	X	Quad-pol	1	2007-
RADARSAT, RADARSAT-2	SAR	C	Quad-pol	3	2012-2014
COSMO	SAR-2000	X	Quad-pol	1	2007-
TecSAR	SAR	X	HH, HV, VH, VV	1	2008
Kondor-5	SAR	S	HH, VV	1	2009
TanDEM-X	SAR	X	Quad-pol	1	2009-
RISAT	SAR	C	Quad-pol	3	2009-2017
HJ-1C	SAR	S	HH, VV	20	2009-2012
ARKON-2	SAR	X, L, S	-	1	2011
Sentinel-1	C-SAR	C	Dual-pol	5*, 10**	2011-
MapSAR	SAR	L	Quad-pol	3	2011-2019
KompSAT-5	SAR	X	HH, HV, VH, VV	20	2011-
SMAP	SAR	L	HH, HV, VV	3 km	2012-
DESDynI	SAR	L	Quad-pol	25	2015
SAOCOM-1	SAR	L	Quad-pol	7	2018-
Capella	SAR	X	Single-pol: HH, VV	0.5	2020-
Umbra 1, 2-3, 4-8	SAR	X	Single-pol: HH, VV	0.25	2021, 2022, 2023-
IceYE	SAR	X	VV	1	2023-

Source: Adapted from Barret *et al.* (2009). *SLC Wave-Mode product; **GRD product.

2.2.1.2 Backscatter

Radar measures the ratio between the power of the pulse transmitted and that of the echo received. This ratio is called the backscatter, that is, the signal that the object on the Earth's surface redirects directly back towards to radar antenna after the sensor emits the pulse. It also refers to a normalised measure called the backscattering coefficient, which is the expression for the returned signal. It is a normalised dimensionless number that can be defined either as per unit area on the ground or per unit area in slant range.

The backscattering coefficient has three main nomenclatures or notations depending on the type of measurement and on the plane reference area used to normalise the backscatter (Small, 2011). Beta nought (β^0) refers to the radar brightness (or reflectivity area in slant range) coefficient and does not require knowledge of the local incidence angle; Sigma nought (σ^0), typically expressed in decibel (dB), is the usual and conventional backscatter's notation for the measure between the incident electromagnetic power and the reflected electromagnetic wave power (distributed scatterer) in an area on the ground (Barret *et al.*, 2007; Zhang *et al.*, 2017; Periasamny, 2018; Nasirzadehdizaji *et al.* 2019; Mandal *et al.*, 2020; Qu *et al.*, 2020; Salma *et al.*, 2022). Sigma nought varies according to incidence angle, wavelength, and polarization, as well as with properties of the scattering surface itself. The intensity of σ^0 is a function of the physical and electrical properties of the target, along with the wavelength (λ), polarisation and incidence angle (θ) of the radar (Barret *et al.*, 2009). Gamma nought (γ^0) is the backscattering coefficient normalised by the cosine of the incidence angle, thus, it is recommend for areas with significant variation in topography (i.e., from flatten to mountainous terrain).

In order to determine the relationship between the radar signal and certain biophysical parameters, many backscattering models have been developed over the past four decades taking into account the various sensor configurations and surface parameters (Barret *et al.*, 2009). However, it is important to take into account the influencing factors on the backscattering coefficient. Major factors include radar configurations – polarisation, incidence angle and frequency or wavelength (bands) – and soil surface characteristics – roughness, vegetation cover, soil texture, soil moisture, and soil dielectric constant (Dobson *et al.*, 1981; Das & Paul, 2015). An increase in the roughness of the surface results in a decrease in the specular reflection and an increase in the surface scattering, which in turn increases the backscatter returned to the sensor (Ulaby *et al.*, 1982; Babaeian *et al.*, 2019). As plant geometry and plant moisture content affect the scattering mechanism, this process is more complex in the presence of vegetation cover. Based on field experiments, pioneering studies that demonstrated the dependence of the backscattering coefficient to soil texture and soil moisture are that of conducted by Ulaby *et al.* (1979), Wang (1980), and Dobson *et al.* (1981), particularly under vegetation-covered soil. Dobson *et al.* (1981) observed that water in a soil medium consists of two components, adsorbed and free (or absorbed), thus, the soil particle size may be considered as the ratio of the number of free water molecules to adsorbed molecules. (Dobson *et al.*, 1981). Ulaby *et al.* (1979) obtained a correlation of 0.92 between the scattering coefficient and soil moisture in a vegetation-covered soil (four crop types),

measured at 4.25 GHz, incidence angle of 10° , and HH polarisation, using a Microwave Active Spectrometer.

Saturation conditions also pose an effect on surface roughness backscatter, as soils with high moisture content, and supersaturated and flooded soils behave as specular surfaces (Figure 2.6), which yield lower backscattering at off-nadir angles (Dobson & Ulaby, 1981; Dobson *et al.*, 1985; Das & Paul, 2015). With a moisture content threshold of 35%, measured as volumetric moisture content, the radar signal becomes insensitive to the soil moisture (Gorab *et al.*, 2014; Das & Paul, 2015). The dependence of the radar signal on surface roughness in an agricultural area is mainly significant for low levels of roughness (Das & Paul, 2015).

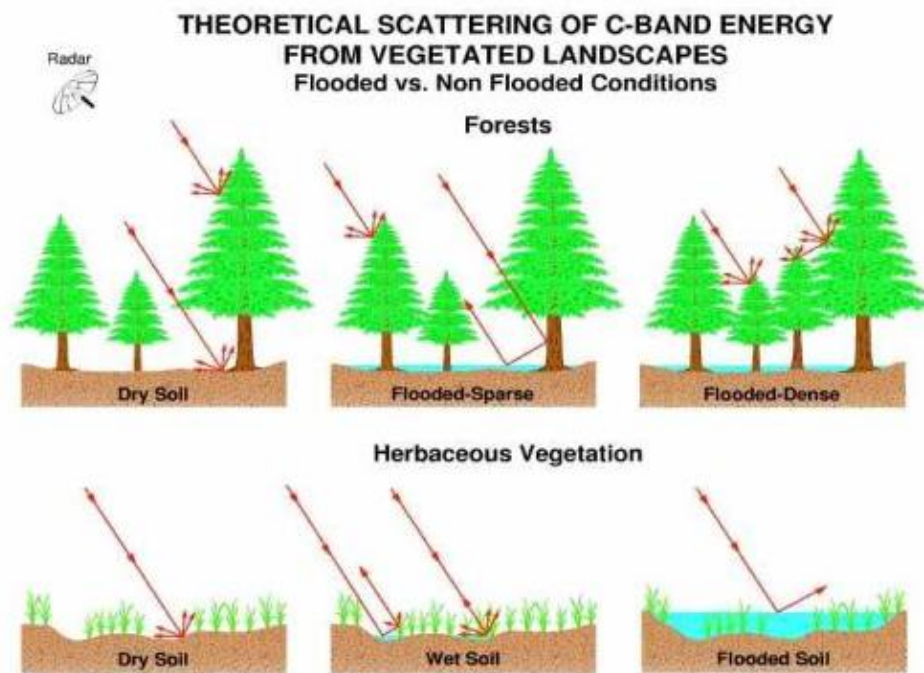


Figure 2.6 An example of a theoretical scattering under two different soil-covering conditions (forest and grassland) and different water content on soil conditions. Source: Bourgeau-Chavez *et al.* (2009).

Apart from radar polarisation, the utilisation of particular incidence angles improve the selective observation of vegetation (high incidence angles) or underlying soil (low incidence angles) (ESA, 2007:52), particularly in agriculture. According to Ulaby *et al.* (1979) steep angles tend to penetrate further into the soil. In their experiment, they observed that radars with frequencies below 6 GHz and angles lower than 20° exhibit minimal vegetation-attenuation effects on the σ° soil moisture relationship (Ulaby *et al.*, 1979).

In soil applications, direct determination of soil texture using SAR data, however, is a difficult task as soil textural configuration poses a change in soil dielectric properties regarding its water-holding capacity (Hallikainen *et al.*, 1985; Das & Paul, 2015). As previously mentioned in Chapter 1, soil texture affects the backscattering coefficient considering the fact that sandy soils contain a higher amount of free water than clay soils. In studies that focus on retrieving soil parameters from EO platforms such as satellites (including synthetic-aperture radar-SAR) and

airborne, soil texture is usually considered as an influencing variable when studying soil moisture or surface roughness (Petrooulos *et al.*, 2015; Babaeian *et al.*, 2019).

In remote sensing, the premise of purely deterministic targets is not always valid (Richards, 2009), likely due to the reflectance and backscattering contributions from neighbour targets. Thereby, incoherent target decomposition methods (based on reflection matrices), among models to decompose polarimetric synthetic aperture radar (SAR) images, are often used (Sano *et al.*, 2020. Translated by Sandra Deodoro). Examples of incoherent decompositions for backscattering analysis are the Yamaguchi model (Yamaguchi *et al.*, 2005), which consists of four components (surface scattering, double-bounce scattering, volume scattering and helix scattering), and Cloude-Pottier (Cloude & Pottier, 1997) with entropy, anisotropy and α angle components. Average entropy indicates that more than one scattering is contributing to backscattering, but does not estimate the number of mechanisms (Sano *et al.*, 2020. Translated by Sandra Deodoro). More details on polarimetric decomposition methods are presented and discussed later in this chapter, and an application in the study area is demonstrated in Chapter 5.

2.2.1.3 Electrical characteristics

The intensity of a radar's signal returning from the Earth's surface (i.e., backscattering) is a result of the interaction between the electrical characteristics of surfaces, terrain features (e.g., roughness, slope) and geometric characteristics of targets (Liu *et al.*, 2019; Beale *et al.*, 2019). The dielectric constant (ϵ), also called permittivity, is one measure of an object's electrical character, being a parameter that indicates the reflectivity and conductivity of various materials. It refers to a measure of the electrical potential energy of a substance under the influence of an electromagnetic field (Babaeian *et al.*, 2019). Most natural materials have, in the microwave spectrum, a dielectric constant in the range of 3 to 8 when dry, whereas water has a dielectric constant of approximately 80 (ESA, 2007). This indicates that soil or vegetation with moisture content will display greater reflectivity in SAR data.

Due to the clear distinction between the permittivity (dielectric constant) of water, solids, and air, this parameter provides a highly accurate determination of water content in the soil (Babaeian *et al.*, 2019). Thus, in contrast to optical (spectral surface soil reflectance measurement) and thermal (surface soil temperature measurement) methods, microwave remote sensing measures bulk dielectric soil properties that are governed by soil moisture content (Babaeian *et al.*, 2019).

Studies on the importance of soil properties, including soil texture and soil particle size fractions, on the dielectric constant date back to the 1980s, conducted by Schmugge (1980), Wang and Schmugge (1980), Dobson *et al.* (1985), and Jackson (1987) who developed dielectric mixture models for soil. Dobson *et al.* (1981) demonstrated that adsorbed water may have a much smaller permittivity than that of free water by as much as a factor of 10, due to the limited mobility of adsorbed water. In their study (a preliminary evaluation of the effect of soil texture on the radar response), they showed a weaker sensitivity to moisture content for "heavy soils" (high clay

content) than for “light soils” (Dobson *et al.*, 1981), thus observed that microwave scattering properties of a soil medium may, in turn, depend on its soil texture.

In soils, the dielectric constant is related to the bound water in soil surfaces which in turn is dependent on the total surface area of the soil available to the water molecules (Jackson, 1987). In the majority of soils, the dielectric constant is proportional to the number of water dipoles per unit volume (Dobson & Ulaby, 1986; Das & Paul, 2015). Apart from soil texture, it is also reliant on relative clay content (i.e. as a soil PSF) and clay mineralogy, such as montmorillonite (Dobson & Ulaby, 1981). Thus, soil particle size fractions are another key factor as they influence dielectric constant due to their interstitial water content (Hallikainen *et al.*, 1985; Das & Paul, 2015). For example, in the context of soil moisture, the dielectric constant in sands is greater than in clays at given volumetric moisture, and the energy status of water in clay soils is more “prone to hysteresis” than that present in sands (Dobson & Ulaby, 1981:60).

Using microwave remote sensing to estimate soil moisture, Srivastava *et al.* (2006) point out that the key factor underpinning this approach is the difference between the dielectric constant of water and that of dry soil at microwave frequencies. In wet soil, the dielectric constant value is ~ 20 (Barret *et al.*, 2007). An example of the influence of soil texture and soil moisture in the dielectric property (or permittivity) of soils is shown in Table 2.5.

Table 2.5 The dielectric constant for Earth materials	
Material	Dielectric Constant
Sand (dry)	3 – 6
Sand (saturated)	20 – 30
Silts	5 – 30
Shales	5 – 15
Clays	5 – 40
Humid soil	30
Cultivated soil	15
Rocky soil	7
Sandy soil (dry)	3
Sandy soil (saturated)	19
Clayey soil (dry)	2
Clayey soil (saturated)	15
Sandstone (saturated)	6
Limestone (dry)	7
Limestone (saturated)	4 – 8
Basalt (saturated)	8
Granite (dry)	5
Granite (saturated)	7

Hubbard *et al.* (1997)

2.3.2 Dealing with vegetated soils in SAR images

The backscatter signal over terrain covered by vegetation consists of two primary components – one from the underlying terrain and the other from the vegetation (tree canopy, sparse leaves, grass). As a result, estimates of a specific parameter for terrain (e.g., elevation and soil moisture) will generally include a vegetation bias (Kasilingam *et al.*, 2002).

On the other hand, some SAR system specifications, such as the incidence angle, may decrease the effects of soil roughness and vegetation on the radar signal. For example, for soil

moisture retrieval, Mattia *et al.* (2006) point out that steep incidence angles ($15 < \theta_i < 30$) are more useful for such an application. Moran *et al.* (2012) outline that some studies have reported the possibility of reducing or eliminating the sensitivity in radar look direction effects for linear cross-polarisations. In vegetation-covered soil, Ulaby *et al.* (1978) suggested keeping the microwave frequency as low as possible as the attenuation through the vegetation increases with frequency, and angles lower than 20° exhibit minimal vegetation-attenuation effects the σ° (Ulaby *et al.*, 1979).

Where there is potential to remove vegetation effects, the Water Cloud Model (WCM) is the tool most frequently used to estimate the backscatter coefficient of the land surface (σ_{soil}^0) to retrieve soil moisture (e.g. Attema & Ulaby, 1978; Graham & Harris, 2003; Beale *et al.*, 2019; Wang *et al.*, 2023). Other studies, such as Grote *et al.* (2010), highlight the use of non-invasive geophysical techniques including GPR as an alternative to both conventional point measurement techniques and remote sensing for soil parameters estimation when significant vegetation is present. However, GPR is not feasible for sensing large areas (e.g., regional geographical scale).

In general, the techniques used for removing and/or reducing vegetation effects on soil backscattering from radar can be grouped into two approaches, based on SAR polarimetry and SAR interferometry. The former concerns the amplitude and phase (φ) properties of the SAR signal, whereas the latter exploit phase information, precisely phase shift of waves (differential phase). The differential propagation of the electromagnetic waves into the soil causes the interferometric effects, that is, small scattering bodies in the soil or moisture gradients can produce phase and coherence variations (De Zan *et al.*, 2014).

Based on SAR polarimetry, polarimetric decomposition of vegetation and soil contributions is one of the methods commonly used to take into account the vegetation effects, that is, for separating the ground component from the vegetation component (Jagdhuber *et al.*, 2013, 2015; Babaeian *et al.* 2019). Typically, polarimetric decomposition or target decomposition is the process of extracting information about the scattering process and describing scattering properties by partitioning the total power or intensity into relative contributions of different idealised scatterers. As a target decomposition, the resulting scattering mechanisms can be surface, volume or double-bounce types (Figure 2.7). Rough surface scattering is most sensitive to VV polarisation (NASA, n.d.). Volume scattering to cross-polarized such as VH or HV (NASA, n.d.); and double bounce is most sensitive to an HH polarized signal (NASA, n.d.).

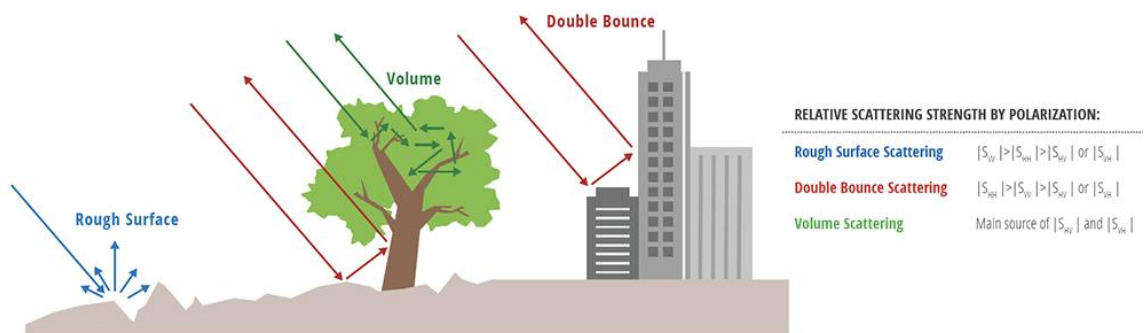


Figure 2.7 An example of scattering mechanisms under different land cover contexts: rough surface (bare soil), arboreal (tree) and urban (buildings) and relative scattering strength by polarisation. Rough surface scattering: $|S_{VV}| > |S_{HH}| > |S_{HV}|$ or $|S_{VH}|$; Double bounce scattering: $|S_{HH}| > |S_{VV}| > |S_{HV}|$ or $|S_{VH}|$; Volume scattering: main source of $|S_{HV}|$ and $|S_{VH}|$. Source: NASA (n.d.).

To apply a polarimetric decomposition using the polarimetric (non-interferometric) approach, SAR data has to be in a complex number output (i.e., the real in-phase component I and the imaginary quadrature component Q), which includes both magnitude and phase values required for performing polarimetric decomposition. Polarimetric decomposition is similar to the spectral mixing analysis or spectral unmixing methods employed in optical remote sensing. Such methods are used to decompose a reflectance or corrected radiance source spectrum into a set or a given endmember spectra or spectral signature, since not all pixel of a natural scene is pure, but a mixture reflecting the heterogenous nature of the surface (Somers *et al.*, 2011; Manolakis *et al.*, 2016; Halbgewachs *et al.*, 2022).

Jagdhuber (2012) investigated the main scattering contributions from the soil and the vegetation through polarimetric decomposition and inversion algorithms. The author also pointed out that L-band was an ideal frequency for soil parameter retrieval under agricultural vegetation cover due to the adequate penetration of the vegetation canopy combined with a sufficient signal-to-noise ratio (SNR) on bare fields. SNR is an important concept for any SAR application and technique as it refers to the ratio of signal power to noise power ($\text{SNR} = \text{received signal} / \text{system noise}$), typically expressed in decibels (dB) in remote sensing, and indicates whether a pixel contains more recorded signal or distortion or noise. The system noise is also called as ‘noise floor’, ‘noise equivalent’ or ‘noise equivalent sigma zero’ (NESZ). In Sentinel-1 data, the NESZ values are reliant upon the acquisition mode. Thus, maximising the SNR ratio is one of the goals in remote sensing since it means systematic noise in the backscatter signal. When $\text{SNR} < 1$, there is more noise than the amount of energy that is being received; conversely, for $\text{SNR} > 1$, the amount of energy received is larger than the system noise.

Jagdhuber *et al.* (2013), underline the success of polarimetric decomposition (proposed by Freeman and Durden in 1998) techniques used to filter the vegetation contribution and estimate soil moisture on the isolated ground components, enabling the interpretation and decomposition of different scattering contributions. These authors argue that radar polarimetry provides the basis to decompose the backscattered signal into different canonical scattering mechanisms (Figure 2.8), based on the assumption of separating scatterers by their polarimetric properties (i.e., measured polarimetric signatures). In general, the proposed approach combined polarimetric decomposition techniques with the concept of “multi-angularity” (Jagdhuber *et al.*, 2013). They used the full polarimetric E-SAR (a Germany Airborne Microwave Sensor in operation since 1989) data at L-band (thus, at lower frequency and longer wavelength) and applied a multi-angular polarimetric decomposition for the estimation of volumetric soil moisture under low agricultural vegetation (comprising bare and vegetated soils). Decomposition was obtained from the coherency matrix into three scattering components – a depolarising surface scattering

component, a deterministic dihedral scattering component and a generalized volume scattering component. Both the surface component and the dihedral component carry information about the ground (Jagdhuber *et al.*, 2013). After removing the vegetation component, soil moisture can be inverted (Jagdhuber *et al.*, 2013). Regarding incidence angles, they observed a higher sensitivity to the ground scattering components (surface, dihedral) at higher local incidence angles. As result, they obtained a root mean square error level of 6-8 vol. % by validation of the estimated soil moisture, against ground measured.

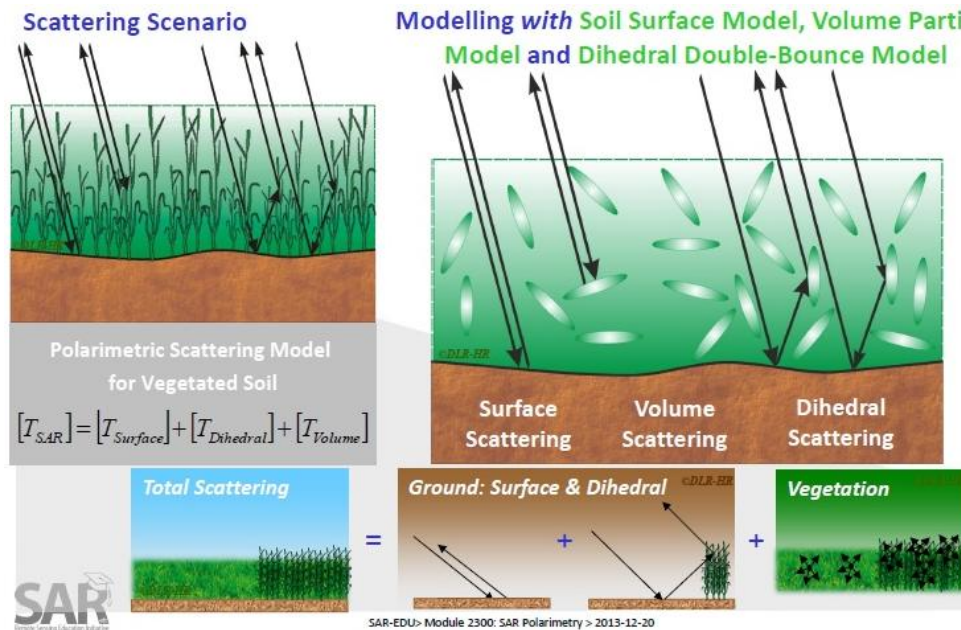


Figure 2.8 Three canonical scattering mechanisms for vegetated soils. From the scattering scenario illustration, [T] refers to coherence matrix and is closer related to physical and geometrical properties of the scattering process, thus provides direct physical interpretation. Source: Jagdhuber *et al.* (2013). Module 2301: SAR Polarimetry.

In a similar approach, Jagdhuber *et al.* (2015) tested the full polarimetric L-band E-SAR data with an iterative generalised hybrid decomposition method to retrieve volumetric soil moisture for vegetated land surfaces. They obtained root mean square errors (RMSE) ranging from 0.04 to 0.044. A drawback of the performance of model-based decompositions is their limitation to a certain range of local incidence angles ($\sim 20^\circ$ to $\sim 70^\circ$) (Jagdhuber *et al.*, 2013). Both studies (Jagdhuber *et al.*, 2013; 2015) employed airborne-based remote sensing platforms, extending application of pol-decomposition methods for datasets other than spaceborne platforms.

Barret *et al.* (2009) point out some strategies to deal with issues related to discriminating the effects of roughness and vegetation from the backscattered radar signal. They cited the use of multi-frequency, multi-incidence and/or multi-polarisation radar, and Pol-InSAR for accurate biophysical parameter retrievals.

Concerning SAR interferometry, the ground component resulting from surface scattering has different polarimetric signatures from the vegetation component (volume scattering), discriminated by exploiting interferometric phases of the ground and the vegetation components (Chen & Kasilingam, 2005). Interferometric methods (e.g., interferograms and coherence maps)

and SAR interferometry can estimate the vertical location of a scatterer within the pixel by exploiting the phase difference of two SAR images (Pichierri *et al.*, 2018), whereas amplitude-based techniques have a limitation concerning surface roughness that can influence backscatter (De Zan *et al.*, 2014). On the other hand, polarimetric techniques could be useful to separate the signal contributions from vegetated soils (De Zan *et al.*, 2014), that is, backscattering from soil and vegetation in an image cell (pixel).

Polarimetric SAR Interferometry (Pol-InSAR) has the potential to separate the scattering contributions from the canopy and the underlying soil by inverting a two-layer volume over ground scattering model (Pichierri *et al.*, 2018), and provides measurements to multiple polarisations. A scattering matrix $[S]$, or equivalently a scattering vector (commonly expressed in the Pauli basis) represents each single look complex (SLC) pixel in the image (Pichierri *et al.*, 2018). An example for separating mixing backscattering and for removing vegetation bias using Pol-InSAR is found in Kasilingam *et al.* (2002) to obtain digital elevation maps (DEM) over terrain. In this study, the authors developed a method based on Direction of Arrival (DoA) algorithms for identifying the largest peak of the interferometric spectrum that best discern the ground and vegetation components. They assumed the peak was associated with the ground component as it is the largest component over most natural surfaces, even for forested terrain with a significant vegetation component (Kasilingam *et al.*, 2002).

Another study case used the independent component analysis (ICA) for separating the vegetation and the ground components from Pol-InSAR measurements in the Glen Afric region, Scotland. ICA is a statistical and computational technique for separating sources and revealing features that underlie sets of random variables, measurements, or signals (Tannous & Kasilingam, 2009). In their study, Tannous and Kasilingam (2009) customised the FastICA algorithm developed by Hyvärinen and Oja (1997). Both authors applied an ICA algorithm in a physically based, three-component scattering mechanism, model proposed by Freeman and Durden (1998) – the depolarised canopy scattering component, the linearly polarised surface scattering component and the double bounce component. Tannous and Kasilingam (2009) assumed that (i) the individual scattering components are non-Gaussian and mutually independent (Tannous & Kasilingam, 2009); (ii) the measurements are linear mixtures of some unknown polarimetric scattering components (Tannous & Kasilingam, 2009); and (iii) the mixing process is unknown (Tannous & Kasilingam, 2009). Backscatter measurements were obtained at each pixel in HH, VV and HV polarisations in terms of a three-dimensional vector as follows (Tannous & Kasilingam, 2009) (Equation 2.1):

$$s = [S_{HH} \ S_{VV} \ S_{HV}]^T \quad \text{Equation 2.1}$$

where S_{HH} , S_{VV} and S_{HV} are the three polarimetric scattering components.

In the Freeman and Durden model, the surface component and the double bounce component, R_s and R_d , have similar signatures, and priory assumptions are that *If $Re\{S_{HH}^*S_{VV}\} > 0$* , then surface scattering is assumed to be dominant, otherwise double bounce is assumed to be the dominant component (Tannous & Kasilingam, 2009). Relied on the property of most of the polarimetric radar measurements between cross-polarised component and co-polarised components, Freeman and Durden (1998) estimated the canopy component in their model decomposition. However, in their study, Tannous and Kasilingam (2009) focused on the separation of scattering components by comprising only the two co-polarised components, S_{HH} and S_{VV} and no assumptions were made about surface scattering and double bounce scattering. In doing so, this procedure is consistent with Sano *et al* (2020), Das & Paul (2015) and Dubois *et al.* (1995), who argue that the intensity of the backscattered signals is greater in parallel polarisation (or co-polarisation) than that of in cross-polarisation. The findings are also consistent with Kasilingam *et al.* (2002) who found that the HV component from the ground is generally small. Tannous & Kasilingam (2009) tested the FastICA algorithm and validated it using both simulated and field data. They observed that (i) the success of the technique is dependent on how well separated are the polarimetric signatures of the two scattering components (Tannous & Kasilingam, 2009), (ii) noise and depolarisation limits the ability of the technique to accurately separate the two components (Tannous & Kasilingam, 2009), and (iii) the results were consistent with other radar metrics from the study area (Tannous & Kasilingam, 2009).

In their study, Chen and Kasilingam (2005) proposed a technique for separating independent components in a mixture of non-orthogonal sources, instead of in a standard orthogonal approach, by combining polarimetry with interferometry. According to these authors, although standard eigenvalue analysis of the polarimetric covariance matrix of the SAR measurement yielded orthogonal eigenvectors, the ground component and the vegetation component may not be necessarily orthogonal (Chen & Kasilingam, 2005). Based on this, they demonstrated that the non-orthogonal eigenvectors obtained by their proposed technique represent the ground and the vegetation scattering components. They observed that the ground component estimated in vegetated areas appeared to be a correction due to the vegetation bias, and it was strongly correlated to the interferometric coherence (Chen & Kasilingam, 2005). They concluded that their technique compensated for the vegetation bias, thus was able to separate the ground and vegetation scattering components.

2.3.3 The use of SAR for soil analysis in an Irish context

No studies addressing soil texture retrieval directly from SAR for Ireland were found up to the date of writing this thesis. The Web of Science (WoS), Scopus, Scielo and Google Scholar databases were searched for publications reporting applications of SAR to both soil PSF estimates and textural class retrieval. Only a few examples of applications for retrieving soil moisture

(taking soil texture into account) can be found in Barret *et al.* (2007), Pratola *et al.* (2014) and Pratola *et al.* (2015).

Barret *et al.* (2007) proposed a technique based on a multi-temporal analysis (in the spring-summer period) to derive relative changes in soil moisture in the near-surface (0-5cm) over a rural area in southern Ireland – on sloped terrain mainly vegetated with low pastures where the predominant soil type is silty loam. They employed an Advanced Synthetic Aperture Radar (ASAR) ENVISAT, C-band, at a 19° incidence angle and HH/HV polarisation. The authors found a correlation between the backscattering coefficient and volumetric soil moisture with R^2 values of 0.76 and 0.78, from April to May, respectively. Conversely, the sensitivity of the SAR signal from May to July displayed a correlation with R^2 values of 0.22 and 0.05. Barret *et al.* (2007) attributed this weak correlation to the advancing growth of vegetation (grass cover) that reached a height of 9.8cm in the July period, which might cause attenuation of the radar signal. Despite that, the authors confirmed the applicability of the incidence angles and polarisations used for Irish soils. The incidence angle chosen by Barret *et al.* (2007) agrees with Ulaby, *et al.* (1979), to whom steep angles tend to penetrate further into the soil. Radars with frequencies below 6 GHz and angles lower than 20° exhibit minimal vegetation-attenuation effects on the σ° soil moisture relationship. (Ulaby, *et al.*, 1979). The results and conclusions presented by Barret *et al.* (2007) highlighted the effect of vegetation on radar signals, as well as to what extent grass height can affect the backscattering coefficient.

In another approach for the same region in Ireland, Pratola *et al.* (2014) used the ASAR Wide Swath (WS) mode, at VV polarisation for retrieving of soil moisture, as well as for comparison and evaluation of the Global Soil Moisture Product – from the Global Climate Observing System – at Irish sites. The authors adopted a different approach as neither dense in situ station networks, nor hydrological models have been used, and they observed that the results obtained were consistent with those reported in other papers using different sensors and classical methods. According to them, this approach was also useful for understanding to what extent soil texture, terrain composition and altitude affect soil moisture retrieval in temperate grasslands. As result, they noted that during the winter, the distribution of soil moisture is rather homogenous over the ASAR pixels, whereas a larger spatial variability was observed in summer. They also noted that the coefficient of variation of sandy soil increases as the soil dries.

Pratola *et al.* (2015) highlight the understanding of the effect of seasonality on backscattering as heavy shower activity (frequent, but localised) characterizes summer precipitation patterns in Ireland. Regarding the vegetation cover effect on the C-band backscattering signal, according to Pratola *et al.* (2014), the influence of the vegetation was considered insignificant on the ASAR WS dataset since all of the study sites are cultivated with relatively short grass. Therefore, the Irish grassland effects could be generally neglected in the retrieval algorithms. They based this on Dobson *et al.* (1992), who carried out experiments with the European Remote Sensing satellite (ERS-1) under VV polarisation, and found that an average height of 40 cm of grass cover had

little influence on backscattering coefficients, attenuating the signal by less than 0.2 dB. From the different aspects of the effect of grass on radar data found in such works, it is important to consider the seasonality effect as noted in Barret *et al.* (2007).

In Ireland, where cloudy conditions are dominant throughout the year, it is important to consider the percentage of clouds in an image, which may hinder, or even make unfeasible, the use of optical EO data. In order to deal with such constraints, microwave-based satellites and aerial platforms (including drones) emerge as possible options. Notwithstanding the solutions brought by EO sensors to overcome illumination and cloud conditions, dealing with vegetation covered soil remains a challenge for soil's properties studies in places with large areas of grasslands, such as Ireland. Amelioration of such challenges is still limited in remote sensing (e.g. in preprocessing steps and retrieving/modelling soil information).

Studies for estimating soil particle sizes and soil texture in Ireland using microwave remote sensing are absent and there remains some gaps in the literature reviewed. Dealing with vegetation cover in SAR images in order to capture only soil information remains a challenge task.

2.4 Modelling approaches for estimating soil particle size fractions

2.4.1 Soil particle size fractions as compositional data

Compositional data (CoDa) (Aitchison, 1986) is defined as constrained or closed data with a constant sum constraint (or closure) equal to 1 or 100 (i.e., proportions or percentages), carrying relative information. This leads to the general principle of CoDa analysis. Due to this constraint, compositional data cannot be considered as univariate data or variables but rather as parts of the composition (Todorov, 2021). Thus, in soils, the sum of the components sand, silt and clay has to be guaranteed (e.g. sum up 100%) to obtain the respective soil textural class (e.g., loam, silty clay), which performs the total part of the composite.

Mathematically, compositional data is represented by points on a simplex wherein the only information is given by the ratios between components. This simplicial sample space S^D is also called the Aitchison geometry or the Aitchison simplex making up the structure of a real vector space. In the geometrical context, compositions are vectors whose components represent relative parts of a whole. Compositional data in three variables ($D=3$), which is the case of soil particle size fractions, are typically plotted in ternary plots or ternary diagrams. They are equilateral triangle whose barycentre graphically depicts the ratios of the three variables as positions.

All relevant information in compositional data is contained as ratios between components (parts) known as a compositional vector of D parts, $x = [x_1, \dots, x_D]$. A primary consequence of this condition is the fact that the use of standard statistical methods for the analysis of compositional data that is not consistent with specific geometrical properties and the principles of compositional data analysis (scale invariance, permutation invariance, and subcompositional coherence) leads to biased results (Filzmoser *et al.*, 2018) and spurious correlations. In other words, such a caveat refers to the raw components of CoDa being subjected to product moment correlation analysis

with the potential for unreliable interpretation based on standard methods of multivariate analysis designed for unconstrained multivariate data. This is the reason why standard multivariate statistical and geostatistical methods assume an unconstrained sample space and why they fail to give reasonable results.

In the late 19th century, Pearson warned the scientific community about such dangers by stating that interpretation of correlations between ratios whose numerators and denominators contain common parts must be done with caution (Pearson, 1897). This principle is applied in CoDa analysis on soil particle size fractions, for example, the total percentage p , is broken down into mutually exclusive parts with component percentages p_1, \dots, p_D and then transformed into a composition (Equation 2.2).

$$[x_1, \dots, x_D] = [p_1, \dots, p_D] / (p_1 + \dots + p_D) \quad \text{Equation. 2.2}$$

where $[x_1, \dots, x_D]$ is a D -part composition (e.g., soil texture), and p_1, \dots, p_D refers to the parts (e.g., soil particle size fractions).

In CoDa, the aim is to structure the simplex S^D in a Euclidean space of dimension $D-1$. In the Euclidean space, one component or dimension (variable) would be omitted for the traditional method, thus, such space is not appropriate for the analysis of soil particle size data (Zhang & Shi, 2019). Aitchison geometry is analogous to the Euclidean space in CoDa analysis and enables to express compositional data in interpretable real coordinates formed by log-ratios of pairs of compositional parts. With that geometry, standard statistical procedures can directly be applied, which means the development of the geometric and algebraic set up that is necessary for building up reliable probabilistic and statistical models for such data. Moreover, traditional statistical techniques (cluster, discriminant analysis, factor analysis, regression models, etc.) are being adapted to that geometrical structure (Aitchison geometry).

Since the original data are three-dimensional, soil PSF is commonly presented in a ternary diagram otherwise known as the textural diagram of soil, where soil textural classes are obtained – for example, loam, clayey, clay loam, sandy clay loam (Figure 2.9). On a triangular diagram, each side of the plot indicates the soil PSF and their composition (amount in %) and each corner represents a size particle. At the top, the soil would be clay; at the left corner, it would be sand, and at the right corner it would be silt. In order to find the soil textural class resulting from three PSF combinations, lines are projected from the percentages of each PSF and follow the respective directions indicated by the arrows. The intersection of these three lines corresponds to the class, which has a delimited region on the triangle..

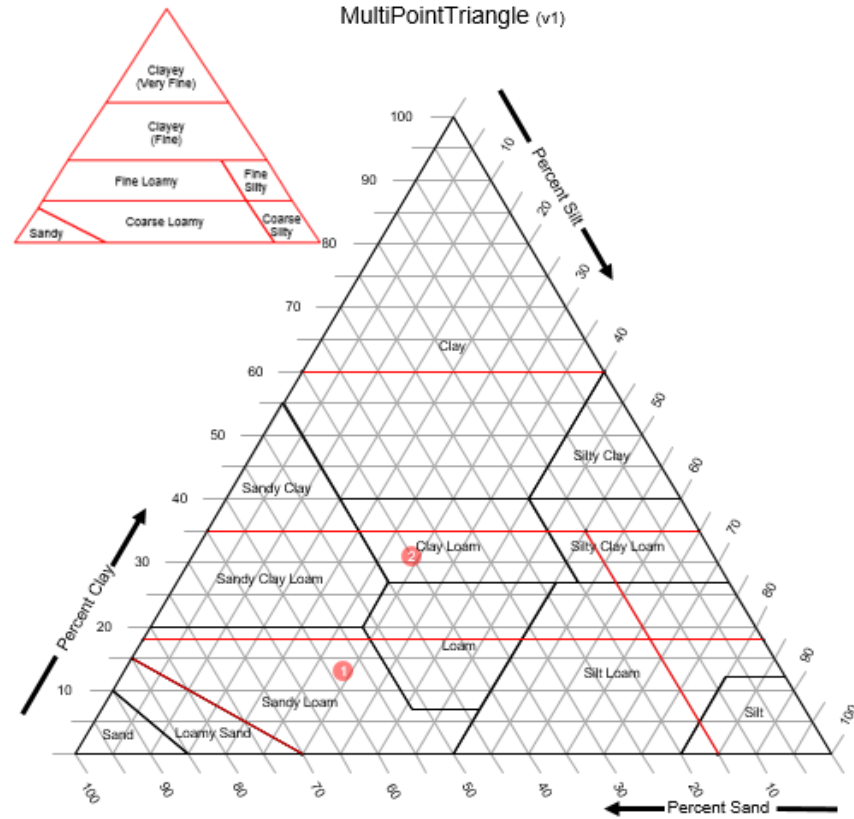


Figure 2.9 An example of soil particle size fractions in the simplex SD (3-simplex space for two soil samples consisted of sand (58%), silt (29%) and clay (13%) classified as sandy loam (coarse loamy) (Point 1); and sand (40%), silt (29%) and clay (31%) classified as clay loam (fine loamy) (Point 2). The textural diagram is according to the USDA-NSCS system. Note that percentages of soil fractions sum 100 in each sample – a constraint of both soil property and the simplex (ternary diagram). Source: USDA-Soil texture calculator.

Perturbation (\oplus) and Power (\odot) are the fundamental and the subsidiary operations in the simplex, respectively. They are used in statistics notations such as in linear models equation written specifically for CoDa – the simplicial regression. The former is analogous to translation in real coordinate space or Euclidian space (\mathbb{R}^D) and is used to find difference between compositions or change from one to another. The latter is the simplicial operation analogous to scalar multiplication in \mathbb{R}^D .

Given the compositional nature of soil texture, it requires the use of log-ratio approaches for adequate mathematical modelling. In this context, as CoDa analysis relies on log-ratio analysis, log-ratio transformations (Aitchison, 1982) are applied to the data. The most widely used approaches are the additive log-ratio (alr), centred log-ratio (clr) and isometric log-ratio (ilr) transformations. They are all linear transformation methods.

The additive log-ratio transformation (alr), or $S^D \Rightarrow \mathbb{R}^{D-1}$, is a transformation method wherein the ratios involve the division of each of the first $D-1$ components by the final component (Equation 2.3).

$$[y = \text{alr } x = [\text{alr}_1 x, \text{alr}_2 x, \dots, \text{alr}_{D-1} x] = [\ln(x_1/x_D), \ln(x_2/x_D), \dots, \ln(x_{D-1}/x_D)]] \quad \text{Equation 2.3}$$

The principal drawback of the alr transformation is that distances are not preserved, that is, a asymmetric transformation in the parts. The centred log-ratio transformation (clr), or $S^D \Rightarrow \mathbb{R}^D$, is symmetrical and the denominator part is given by the geometric mean of the components of the composition (Equation 2.4). In other words, it maps a D -part composition from the simplex isometrically, using the geometric mean as a common denominator (Todorov, 2021). The clr transformation is defined by:

$$z = \text{clr } x = [\text{clr}_1 x, \text{clr}_2 x, \dots ; \text{clr}_{D-1} x] = \left[\ln \frac{x_1}{g(x)}, \ln \frac{x_2}{g(x)}, \dots, \ln \frac{x_D}{g(x)} \right], \quad \text{Equation 2.4}$$

where z is the clr-transform of a composition or a data matrix of clr-transforms of compositions; x is a composition or a data matrix of compositions; D is the number of parts (components); $[\cdot]$ denotes a row vector; $g(x) = (x_1 \cdot x_2 \cdot \dots \cdot x_D)^{1/D}$ is the geometric mean of the components of the composition x in the napierian logarithm \ln .

The principal drawback of the clr transformation is the fact that the covariance matrix of a clr-transformed data set is singular and the Pearson correlation coefficient (Equation 2.5) is not informative.

$$\text{corr}(\text{clr}_k x; \text{clr}_j x), \quad \text{Equation 2.5}$$

where $k, j = 1, \dots, D$. This is due to the clr-transformed vector has D components but belongs to a $D-1$ dimensional subspace, thus, the resulting vector is also constrained because the sum of its parts equals to zero.

Notwithstanding the drawbacks, these two transformations of S^D to a \mathbb{R}^D allow using standard multivariate methods. Finally, the isometric log-ratio transformation (ilr), or $S^D \Rightarrow \mathbb{R}^{D-1}$, is a transformation method that preserves distance and is a representation of compositions by orthonormal coordinates. Thus, it is also known as the orthonormal log-ratio coordinates (olr-basis) (Martín-Fernández, 2019) in S^D (Equation 2.6).

$$\text{ilr } x = [\langle x, e_1 \rangle a, \dots, \langle x, e_{D-1} \rangle a], \quad \text{Equation 2.6}$$

where e_1, \dots, e_{D-1} are the olr-basis in S^D .

The ilr transformation hereinafter olr (Martín-Fernández, 2019) is a procedure for calculating the coordinates of a composition with respect to an orthonormal basis of S^D . The rationale of considering olr in our study relies in Martín-Fernández (2019) who coined this name in order to avoid confusion, since the clr transformation is also an isometric log-ratio transformation.

The alr transformation is historically recommended for models where the log-ratio normal distribution plays a role like parametric models. For example, in the classical linear discriminant analysis. On the other hand, alr is not recommended for models based on distances (or metric methods) such as cluster analysis as the Aitchison distance is not invariant under a change of

denominator in the alr (Martín-Fernández *et al.*, 1998). For such models, the clr transformation is normally suggested (Martín-Fernández *et al.*, 1998). It is not recommended, however, for parametric models because the components transformed sum up to zero, thus the clr-covariance matrix is singular (determinant = 0) causing issues in the models. Such covariance matrix does not allow application of most of the multivariate (Todorov, 2021). The olr transformation (ilr) has the advantages of both alr and clr methods, and not the drawbacks of both. It is generally agreed upon to use log-ratio orthonormal coordinates (Morais and Thomas-Agnan, 2021). The disadvantage of this transformation method lies on the fact that the interpretation of olr may be not easy or straightforward.

These transformations, particularly the olr method, play an important role in simplicial regression models. They transport the simplex space S^D , equipped with the Aitchison geometry, into the Euclidian space \mathbb{R}^{D-1} , thus eliminating the simplex constraints problem (Morais & Thomas-Agnan, 2021). A compositional data approach has been employed to estimate soil particle size fractions in both non-spatial modelling (Chappell *et al.* 2019; Loosvelt *et al.*, 2013) and explicitly spatial models (Odeh *et al.*, 2003; Niang *et al.*, 2014; Wang & Shi, 2017; Wang & Shi, 2018; Zhang & Shi, 2019; Shi & Zhang, 2023).

2.4.2 Non-spatial models: Statistical machine learning algorithms

Machine learning is well-known as a specific topic in artificial intelligence realm. As per its nomenclature, it trains a machine to learn from data or to “handle the data more efficiently” (Dey, 2016:1174) in order to interpret patterns (spatially explicit or not) or extract information from data to make predictions and take decisions. However, machine learning is also linked to statistics. For example, a simple machine learning model is the linear regression model, which can be used for generating predictions. Other common unsupervised and supervised statistical machine learning models are clustering, principal component analysis, logistic regression, and tree-based models. Hereafter such models are referred to as statistical machine learning (Hastie *et al.*, 2009; Sugiyama, 2016; James *et al.*, 2023), as statistical techniques and machine learning are combined together, particularly regarding to the tree-based models. It is worth noting that some classical statistical terms have their names changed in a machine learning context (Table 2.6), although definitions remain the same.

Table 2.6 Differences of common terminology between statistics and machine learning

Term definition	Statistics	Machine learning
Target	dependent variable / response	label
Variable	predictor variable / covariate	feature
Transformation	transformation	feature creation / engineering

Source: Adapted from Google Developers (2024).

The main advantage of machine learning algorithms is their capability to build models and provide researchers with accurate results, quickly and automatically, even on very large scale data

(SAS, n.d.; Dey, 2016). On the other hand, they are reliant upon large amounts of data, along with requiring significant computational processing and data storage resources.

The most widely adopted machine learning methods (or types of learning) are supervised learning, unsupervised learning and reinforcement learning (Figure 2.10). Regarding the former, algorithms are externally assisted and are trained as an input where the output is known. The input dataset is divided into training – which needs to be predicted or classified – and test dataset. The core tenet of this system is the recognition of patterns from the training dataset and application to the test dataset for prediction or classification (Dey, 2016). The learning algorithm receives a set of inputs along with the corresponding correct outputs, for example, soil samples collected in the field (in situ data). The algorithm learns by comparing its actual output with in situ data to find errors and modifies the model accordingly (SAS, n.d.). Through methods like classification, regression, prediction and gradient boosting, supervised learning uses patterns to predict the values of the label on additional unlabelled data. (SAS, n. d.). Thus, this method is commonly used to predict future events from historical data.

Unsupervised learning is used when data has no historical labels. The system does not know the "right answer", as the algorithm must figure out the patterns from the inputs (SAS, n.d.). In this approach, the algorithms learn new features from the data (Dey, 2016). The goal is to explore the data and find some structure within, thus, algorithms are also used to perform segmentation (image, texts) and identify outliers (Hastie *et al.*, 2009). Popular techniques include self-organizing maps, nearest-neighbour mapping, k-means clustering and singular value decomposition (SAS, n.d.). It is mainly used for clustering and feature reduction.

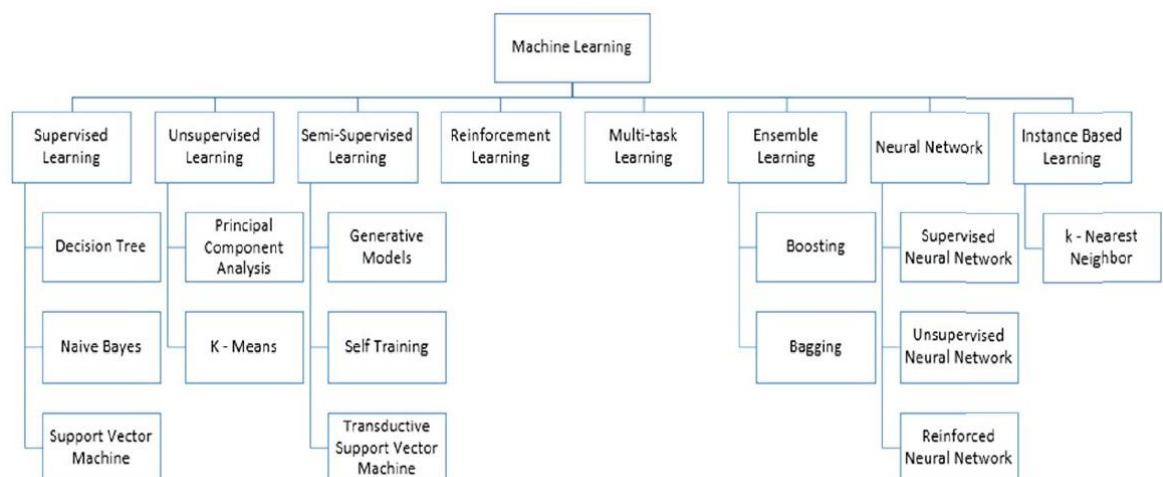


Figure 2.10 Machine Learning methods. Source: Dey (2016).

Machine learning has contributed to soil science, providing robust models for analysis and prediction of soil properties and soil parameters, given its ability to analyse different and complex soil datasets – remote sensor data, on-ground sensor data, and traditional point-field data (Padarian *et al.*, 2020; Motia & Reddy, 2021; Silvero *et al.*, 2023). For example, Padarian *et al.*

(2020) proposed a method to interpret a digital soil mapping model. They introduced and applied the Shapley additive explanations (SHAP) method to a convolutional neural network trained to predict soil organic carbon in Chile. According to them, the SHAP values are the contribution of a covariate to the final model predictions. As results, they found contribution of each covariate to the model predictions at a local and global levels, as well as a spatial interpretation of their contributions. They concluded that SHAP values should be included within digital soil mapping framework to address interpretability of more complex models. In their review, Motia and Reddy (2021) explored the use of machine learning methods for prediction and assessment of soil properties for improved decisions agricultural soil health management. Their analysis included algorithms, implementation tools, and performance metrics and concluded that the response for machine learning in the prediction and evaluation of soil properties for soil health management is promising for the sustainable growth of agriculture. Further, machine-learning methods are capable of tackling non-linear problems (Morellos *et al.*, 2016). For example, in a comparison of predictive performance between two linear multivariate and two machine-learning methods in soil parameters and properties prediction (e.g., total nitrogen, organic carbon and moisture content), Morellos *et al.* (2016) highlight that machine-learning methods performed better than a selection of multivariate regression methods.

The most common machine-learning algorithms for soil properties modelling are cubist, neural networks, decision trees, random forest, support vector machine, boosting, k-means clustering, regression splines, Bayesian networks and principal component analysis. For soil moisture retrieval from remote sensing, the artificial neural networks, support vector machines and the similar support vector regression, as well as tree-based models (e.g. random forest and boosted regression trees) are among the most commonly used algorithms (Araya *et al.*, 2021).

Random forests (Breiman, 2001) can be used for classification and regression purposes. The accuracy of a random forest depends on the strength of the individual tree and a measure of the dependence between them (Breiman, 2001). Growing an ensemble of trees and generating random vectors contribute to the accuracy, as the latter control the growth of each tree in the ensemble. As a random event, the method relies on the Strong Law of Large Numbers (an important theorem in probability theory) to cope with overfitting (Breiman, 2001), particularly for a large number of trees. Nevertheless, it is important to consider dropping features that do not contribute to the prediction process. In this context, hyperparameters play an important role, since tuning values are used to control the learning process, as well as to increase the predictive power of the model or to make estimation faster.

Random forest model/algorithm has been employed in soil sciences for varied purposes and under different climate context. For example, for obtaining a predictive model for the transfer factor soil-plant for radiocaesium in Germany (Urso *et al.*, 2023); to predict spatial distribution of soil pH, electrical conductivity, and soil organic carbon using several type of covariates in tandem with legacy soil data in India (Dharumarajan *et al.*, 2017); to predict soil properties, in

combination with environmental variables, in a saline semi-arid area in Russia (Suleymanov *et al.*, 2023); to generate digital mapping of soil classes in Southeast Brazil (Monteiro *et al.*, 2023), to predict soil textural classes, in Turkey (Kaya *et al.*, 2022), and to estimate particle size fractions in an area in Himalaya (Bashir *et al.*, 2024), and in Canada (Martinelli & Gasser, 2022).

There are many algorithmic approaches for a random selection of features using the training set in order to achieve classification accuracy. These may involve bagging, random split selection (at each node the split is selected at random from among the K best splits), a random set of weights, and Adaboost (adaptive reweighting of the training set). Using these validation techniques is important to improve model performance (Breiman, 2001).

An important issue, as in any modelling process, is to model the error. The generalisation error for forests converges to a limit, as the number of trees in the forest becomes large (Breiman, 2001). When using random forests for regression, the decrease in error from the individual trees in the forest depends on the correlation between residuals and the mean squared error of the individual trees (Breiman, 2001). Thus, a bound for the mean squared generalization error may be derived. For classification, the smaller the ratio expressed by the correlation between the strength of the individual classifiers in the forest divided by the square of the strength (c/s^2), the better the generalisation error (Breiman, 2001).

Understanding the mechanism of the random forest “black box” (i.e. models are created directly from data by an algorithm that make predictions based on complex internal processes) is essential (Breiman, 2001:6), that is, obtaining insights into the underlying relationship between inputs and outputs. However, it is difficult to decipher it, particularly in more complex models. The “black-box” characteristic of algorithms is largely discussed under both data science and critical data studies, as algorithms have direct relationship with data quality in terms of data completeness (Lauriault *et al.*, 2007), which may also affect the accuracy of the results or outputs. Examples of such discussions may be found in Kitchin, & Dodge (2011), Seaver (2013), Mahnke & Uprichard (2014), Dey (2016), Kitchin (2017), and Davies (2019). Further, due to their black box characteristic, in order to provide machine learning models with more reliability in their predictions, confidence intervals for the error can be obtained (Bayle *et al.*, 2020), particularly for random forest models as demonstrated by Marques F (2022).

2.4.3 Spatial models: geostatistical and non-geostatistical methods

Spatial modelling seeks to find and interpret spatial relationships, patterns, trends, and variations within a geographic area by leveraging the knowledge of geography and statistics (Singh & Sarma, 2023). This task is also facilitated by Geographical Information System (GIS) platforms by integrating in situ data and ancillary data such as thematic maps (e.g., soils, geology, topography, population) and satellite imagery (Singh & Sarma, 2023). In soil science geostatistics plays a crucial role in understanding the spatial variability of soil parameters (Singh & Sarma, 2023).

Janowicz *et al.* (2020) discuss spatial models in the context of Geographical Artificial Intelligence (GeoAI) and point out that the use of AI techniques in geography and other geosciences is not new. The authors cite works from the 80s and 90s, such as those of Couclelis (1986) and Smith (1984), who discussed the potential role of AI for geographic problem-solving, as well as Kitchin's book on Artificial Intelligence in Geography (Openshaw & Openshaw, 1997). GeoAI is also a subfield of spatial data science and its tasks include image classification, object detection, scene segmentation, simulation and interpolation, prediction, on-the-fly data integration, etc (Janowicz *et al.*, 2020).

According to Janowicz *et al.* (2020), spatially explicit models are those that fulfil at least one of the following requirements/tests and any of their combinations (Goodchild, 2001; Janowicz *et al.*, 2020):

- *Invariance test.* The results of spatially explicit models are not invariant under the relocation of the studied phenomena.
- *Representation test.* Spatially explicit models contain spatial representations of the studied phenomena in their implementations, which can be in the form of coordinates, spatial relations, and place names, to name a few.
- *Formulation test.* Spatially explicit models make use of spatial concepts in their formulations, for example, the notion of a neighbourhood.
- *Outcome test.* The spatial structures/forms of inputs and outcomes of the model differ.

Similarly, Singh and Thill (2021) discuss the properties of spatially explicit data that influence the performance of machine learning in a review paper. The properties are *spatial dependence, spatial heterogeneity, scale, specific representation of the data (polygon, line, point, regular grid, text), and measurement process*. These authors highlight that such properties are often ignored or inadequately handled in machine learning for spatial domains of application (Nikparvar & Thill, 2021). Based on their review, the authors recognised two broad practices in treating such properties in spatial domains, which are (i) the development of the properties of spatial data in the spatial observation matrix without amending the learning algorithm (Nikparvar & Thill, 2021); and (ii) spatial data properties are handled in the learning algorithm itself (Nikparvar & Thill, 2021). These authors argue that the latter practice offers the most promising prospects for the future of spatial machine learning.

Regarding the properties of spatially explicit data, Nikparvar-Thill (2021) highlight insights concerning scale due to the trade-off between sampling and training. The reliability of the learning process depends on the similarity of the spatial distribution of the samples and the test samples (Nikparvar and Thill, 2021). As data come from a specific geographic area, this fact poses a challenge in training a model for different geographic areas because the distribution of the test and training data sets is not similar (Nikparvar and Thill, 2021). The authors pointed out that “the sampling strategy for the training data set is essential to cover the heterogeneity of the phenomena of interest over the spatial frame of study” (Nikparvar and Thill, 2021:6).

Geostatistical (spatial-geographic nature) and non-geostatistical modelling methods are employed in soil science by using different types of data, from field sampling (in situ data) to both remote- and proximal sensors (on-field sensors), including unmanned platforms of Earth observation (drones). The geostatistical modelling approach is useful when the purpose is to create spatially explicit models of soil properties or soil quality parameters.

According to Singh and Sarma (2023), a complete geostatistical analysis process can be divided into (i) data exploration tools (e.g., histogram, normal Q-Q plot, voronoi maps, trend analysis, semi-variogram, general Q-Q plot and cross variance cloud); (ii) deterministic method for prediction or interpolation (e.g. inverse distance weighting-IDW, global polynomial interpolation-GPI, radial basis function-RBF, local polynomial interpolation-LPI); (iii) geostatistical method for prediction or interpolation (kriging, co-kriging, areal interpolation, empirical Bayesian kriging-EBK); and (iv) interpolation with barriers (kernel smoothing, diffusion kernel).

A deterministic method offers reproducibility and stability as it operates without any randomness (Singh & Sarma, 2023). This approach is valuable when and where repeatability, reliability, and precision are essential (Singh & Sarma, 2023); however, it does not provide uncertainty estimation. In turn, geostatistical methods aim to model spatially correlated data, thus, it considered the spatial dependence or autocorrelation in the data. Apart from creating spatial maps, geostatistical approaches seek to quantify uncertainty in predictions (Singh & Sarma, 2023).

There is an extensive body of literature concerning statistical and geostatistical models for soil property estimation and prediction, including texture, soil organic carbon, and moisture content in the digital soil-mapping realm (Akpa *et al.*, 2014; Gholizadeh *et al.*, 2018; Cisty *et al.*, 2019; Bousbih *et al.*, 2019; Araya *et al.*, 2021). Typically, they rely on Multiple Linear Regression, Regression Kriging, Ordinary Kriging, Geographically Weighted Regression, Random Forest, Support Vector Machine Regression, and Random Forests. Examples of studies that explicitly treat soil particle size fractions as compositional data in spatial models are found in Wang & Shi (2017), Wang & Shi (2018), and Shi & Zhang (2023).

Importantly, a specific method may perform in terms of accuracy and correctness, but its performance is dependent on the tuning parameters and nature of the data used. The optimum choice of method ultimately may represent a trade-off, depending on the characteristics of the study area and the aims of the research (Mei-Po Kwan, 2016; Janowicz *et al.*, 2020). A common question that arises before an experiment is carried out, or even remains afterward, refers to which method fits the data and how to measure model performance. Testing more than one method may shed light on this question or help to reduce bias (Mei-Po Kwan, 2016), by using error analysis. In analytical prediction, modelling the error or assessing the amount of algorithmic uncertainty involved (Mei-Po Kwan, 2016; Shi & Zhang, 2023) is as important as the results prediction, since the former controls the latter. For example, in their experiment with four types of machine

learning models, Araya *et al.* (2021) applied the mean absolute error (average deviation of predictions from the measured value), mean bias (average systematic bias measurement), and the coefficient of determination R^2 (a correspondence measurement between predicted and measured data) to assess the performance of models.

The ability to learn complex nonlinear mappings without any prior assumptions are often cited as benefits of using specific machine-learning techniques in remote sensing (Araya *et al.*, 2021), as they can integrate data from different sources. Conversely, there are some drawbacks of machine learning methods, namely (i) the need for a large volume of training data, requiring extensive ground truth data sets; (ii) they are not able to forecast situations that they have not been trained on (Araya *et al.*, 2021), as they are data-based methods, therefore interpretation of the interaction between predictors and a response is difficult. Nonetheless, such a disadvantage is also valid in any empirical/statistical model (extrapolation). Notwithstanding developing an application for soil moisture, Araya *et al.* (2021) research was underpinned by the general ideas behind each method employed in their study (artificial neural networks, support vector machine, the similar support vector regression, random forest, and boosted regression trees). Further, it sheds light on the shortcomings and benefits of the machine learning methods tested in their experiment, which may be useful for analytical modelling of other soil parameters such as soil texture.

Figure 2.11 depicts a workflow for soil property modelling based on a machine learning approach, developed for soil moisture content retrieval by Araya *et al.* (2021). The response variable and the covariates may differ in another study case, but it serves as a general model framework.

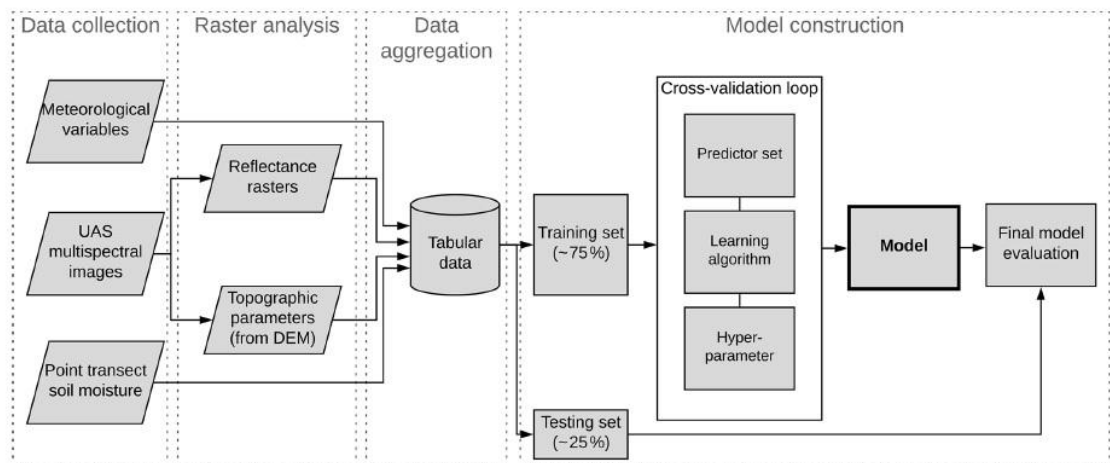


Figure 2.11 Example of a flowchart of soil property modelling based on machine learning application, developed by Araya *et al.* (2021) for soil moisture. Source: Araya *et al.* (2021).

A common approach is to randomly split the dataset at a proportion between 70 % - 90% and 30% - 10% for training and test sets, respectively (Shi & Zhang, 2023). Breiman (2001) withheld the remaining 10% for independent evaluation (test data) in their experiment with random forest models. However, when dealing with small datasets, cross-validation methods, such as leave-one-

out (LOOCV), are recommended (Shi & Zhang, 2023) instead of randomly splitting data. Due to iteration, LOOCV can be time-consuming if the number of observations is large (James *et al.*, 2023), but this method can be used with any predictive modelling approach (James *et al.*, 2023). Furthermore, data standardising is a recommended procedure for minimizing issues of scale (i.e., the units of the measurements) due to different types of variables, and for suitable and faster training (Araya *et al.*, 2021).

2.5 Summary

Two main gaps were identified from the literature review: the lack of studies for estimating soil particle sizes and soil texture in Ireland using microwave remote sensing; and an approach for dealing with vegetation cover in SAR images in order to capture soil information and to improve soil particle size estimates.

Concerning sensor-based data, the environmental conditions of the Earth's surface at study areas pose issues in obtaining information on soils across all types of EO platforms. The technical aspects of instruments (i.e., sensor configuration) were another factor highlighted in the literature searched, particularly regarding SAR. For example, their ability to retrieve soil information is influenced by microwave bands (wavelengths and frequency) and incidence angle. That is, the penetration capability into tree canopy, leaves and topsoil, together with the intrinsic radar signal sensitivity to and radar signal backscattering from surface conditions. Proximal sensors for measurements of soil properties, such as GPR, are limited to discrete measurements at particular locations, which do not represent the spatial distribution exhibited by highly variable soil properties (Petropoulos *et al.*, 2015). Remote sensing observations are only able to provide information about the upper few centimetres of the topsoil (Klotzsche *et al.*, 2018). Conversely, the advent of satellite-based remote sensing can provide spatially explicit maps of surface soil properties from space, with high or low resolution depending on the sensor type. GPR, by turn, may close the gap between large-scale remote sensing methods and small-scale point measurements (Klotzsche *et al.*, 2018), in addition to its potential to provide information about soil properties deeper in the soil rather than remote sensors.

Despite the extensive literature concerning soil properties retrieval from EO platforms (spaceborne, manned and unmanned airborne, proximal sensing), dealing with vegetation cover in soil applications, remains a challenge. Das & Paul (2015) also raise such an issue claiming that one of the most “interesting” problems to be resolved is how to invert the moisture, soil texture and roughness from remote sensing data. Soil texture, by turn, being a key soil property, has a significant effect on the other properties, as shown by scientific literature pointing out that soil texture controls the (geographical) behaviour of soil moisture. Nevertheless, it has received little attention in studies that address analytical retrieval and modelling from SAR data under vegetated soil. This may be related to the fact that translating the microwave signal into a soil characteristic is not straightforward (Han *et al.*, 2017) and separating individual scattering contributions from a

radar signal between soil and vegetation under vegetated fields in a pixel is a difficult task as the SAR backscatter comprises of scattering from soil, vegetation, and the interaction of these two elements (Dou et al., 2022). Thus, not all pixels of a satellite scene representing vegetation-covered soil contain only a single scattering mechanism (e.g., surface scattering from underlying soil). In this context, polarimetric decomposition emerges as a feasible method, as it distinguishes the scattering contributions of different targets (e.g. canopy and underlying soil) (Barret et al., 2009; Jagdhuber, 2012; Ji & Wu, 2015; Mascolo et al., 2021) and was originally employed for land use and land cover classification.

To conclude, amongst the available options of active microwave sensors and retrieval soil surface parameters methods, Barret *et al.* (2009) highlight that for a selection of a certain technique, it is important to take into account (i) careful consideration of the specific research purpose (Barret *et al.*, 2009); (ii) the major impact factors and accuracies required (Barret *et al.*, 2009); and (iii) prior knowledge of the study area conditions (Barret *et al.*, 2009).

3.1 Introduction

Ireland was selected for employing the methods outlined in this study due to its meteorological conditions and land use/land cover features. It is permanently cloudy, which presents considerable challenges for employing optical spectrum-based sensors, and has extensive areas of permanent vegetation, which provides a suitable study location to assess the potential for radar or microwave spectrum-based sensors.

Agriculture is the primary land use-land cover category in Ireland, accounting for 67% of the national land cover within which pasture is the main agricultural class (55%) according to the coordination of information on the environment – CORINE¹ land cover inventory of 2018. Grassland for cattle grazing and silage making is the dominant land cover and cultural backbone of the rural economy in Ireland (Eaton, 2008). It is worthwhile to distinguish two categories of grassland in Ireland, namely semi-natural grassland and improved grassland for agriculture. The former is not intensively managed and is protected under the European Union (EU) habitats directive; the latter is heavily managed to maximise production, and thus they are not protected under the EU habitats directive.

Thus, such land use/land cover context emerges as a challenge for employing EO data in areas like Ireland. Nonetheless, it assembles ideal conditions of land use management (croplands and grasslands) that allow remote sensors to capture information from soils, for example, leveraging the gaps between the leaves.

The study area of this research comprises two experimental areas (Figure 3.1). A larger area encompassing a central portion of Ireland, from west to east coast, was selected for testing the spatial and non-spatial modelling approaches to estimating soil PSF. This portion of Ireland was adopted in order to have as much heterogeneity as possible in terms of physiography, soil, and land cover characteristics, which is important for the replicability of the models. Moreover, this selected area is predominantly flattened, which is important for the SAR radiometric calibration performed as sigma nought (σ_0). The second experimental area comprises a small portion in the southeast of the larger area and was selected for comparing the dual-polarimetric decomposition results, especially the alpha parameter, with a Sentinel 2 image (optical remote sensing). This smaller area assembles key characteristics for this aim, such as being mostly topographically flat and exhibiting more patches of bare soil and very low vegetation (grasslands and croplands). Moreover, it was where the polarimetric decomposition method indicated more surface scattering mechanisms.

¹ An inventory of European land cover split into 44 different land cover classes.

3.2 Geographical setting

Parent material (i.e. geology based), topography, organisms and climate are the key soil-forming factors – and soil textural properties, such as clay, silt, and sand – through interaction among them. Thus, this section is structured according to such pivotal elements, except biological components, to interpret soils that generally characterise the Irish physical geography context.

In Ireland, parent materials consist of solid geology (bedrock), glacial till (unsorted), fluvio-glacial (sorted), and other (river alluvium, peat), but they can be broadly grouped as solid bedrock geology and bedrock derived glacial geology (O’Sullivan *et al.*, 2018). It is also important to highlight the influence of glaciation in Irish soils concerning drift geology, given the transportation of unconsolidated deposits or loose sediments by ice, water, wind or gravity.

Ireland is mostly comprised of lowlands (<150 m) (Figure 3.1), with flat gradient, and the midlands of Ireland are made up of a central plain of carboniferous limestone (O’Sullivan & Creamer, 2018). Altogether, the limestone lowlands account for approximately one-quarter of the land area (23.2%) (O’Sullivan & Creamer, 2018). Hill landscapes (i.e. upland areas between 150 and 500 m above sea level) is most concentrated in the southern half of the country, particularly in the uplands of Munster (O’Sullivan & Creamer, 2018). Mountain landscapes (i.e. above an elevation of 500 m above sea level) represent only a small portion of the Irish landscape – they are mostly found in the western seaboard counties, and Wicklow county in the east of Ireland (Wicklow Mountains) – whereas peatlands (e.g. raised bogs, blanket bogs and fens) account for nearly a quarter of the land area (Renou-Wilson, 2018). Noticeable drumlin landforms (i.e. shaped over glacial till) stretch from Strangford Lough to South Donegal, whose soil (drumlin soils) can vary from reasonably dry soils to very heavy clay wet soils (O’Sullivan *et al.*, 2018). This drumlin belt of the north midlands of Ireland that stretches across parts of Northern Ireland is one of the most extensive drumlin belts in Europe (O’Sullivan & Creamer, 2018). Moreover, altitude, aspect and slope determine the interaction between the site and solar radiation and precipitation, which lead south-facing slopes in Ireland receive more solar radiation and are consequently warmer and dryer with higher levels of evapotranspiration than slopes of a north-facing aspect (Walsh, 2012; O’Sullivan *et al.*, 2018).

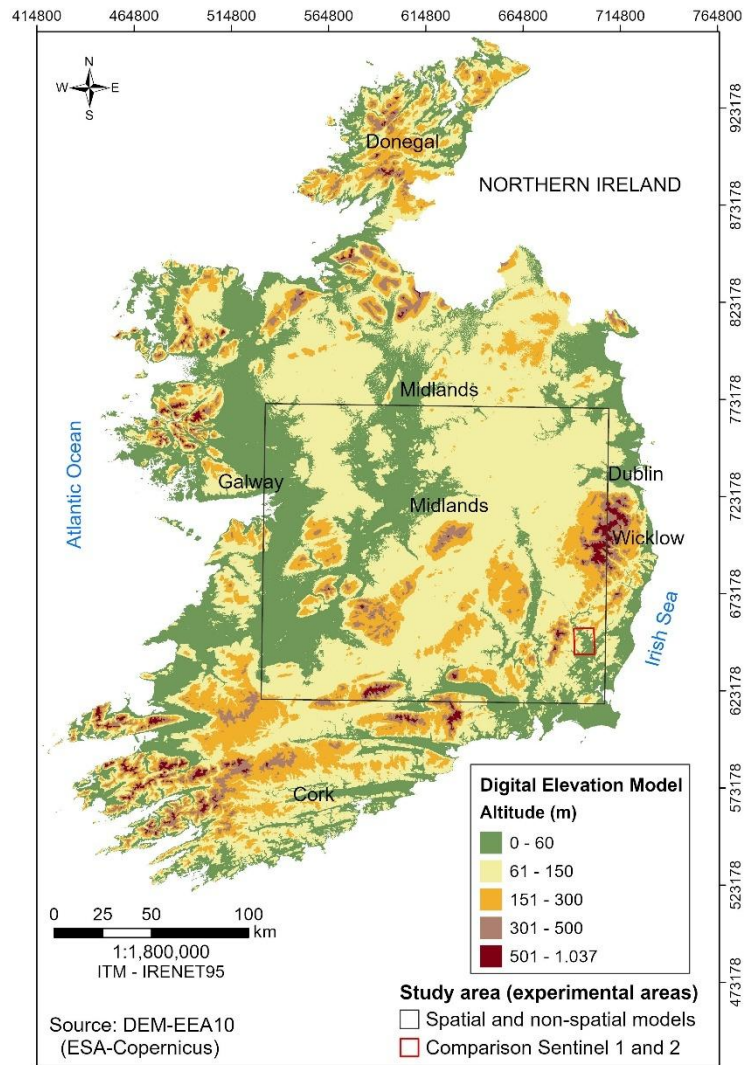


Figure 3.1 Topographical map of Ireland with experimental areas of this study are displayed in black and red squares. Source of data: ESA-Copernicus Digital Elevation Model 10 m (DEM-EEA10).

However, the defined threshold between hill and mountain adapted for the physiographic map of Ireland (Geological Survey Ireland-GSI) (Figure 3.2), considered hills as landforms extending above the surrounding landscape ranging between 100 and 300 m altitude, whereas a mountain as an elevation of at least 300 m within 7 km (Pellicer *et al.*, 2018). This threshold of 300 m is reinforced in Ireland as the minimum altitude “reached by mountains” (Pellicer *et al.*, 2018:4) in the country in most cases, and the classification of hills were “to some extent in agreement with classification of soils in Ireland from a soil characterization perspective” (Pellicer *et al.*, 2018:4), that is the Irish SIS (Creamer *et al.*, 2014). The physiographic units are based on Ireland’s bedrock geology, geomorphology, and Quaternary sediments (Pellicer *et al.*, 2018).

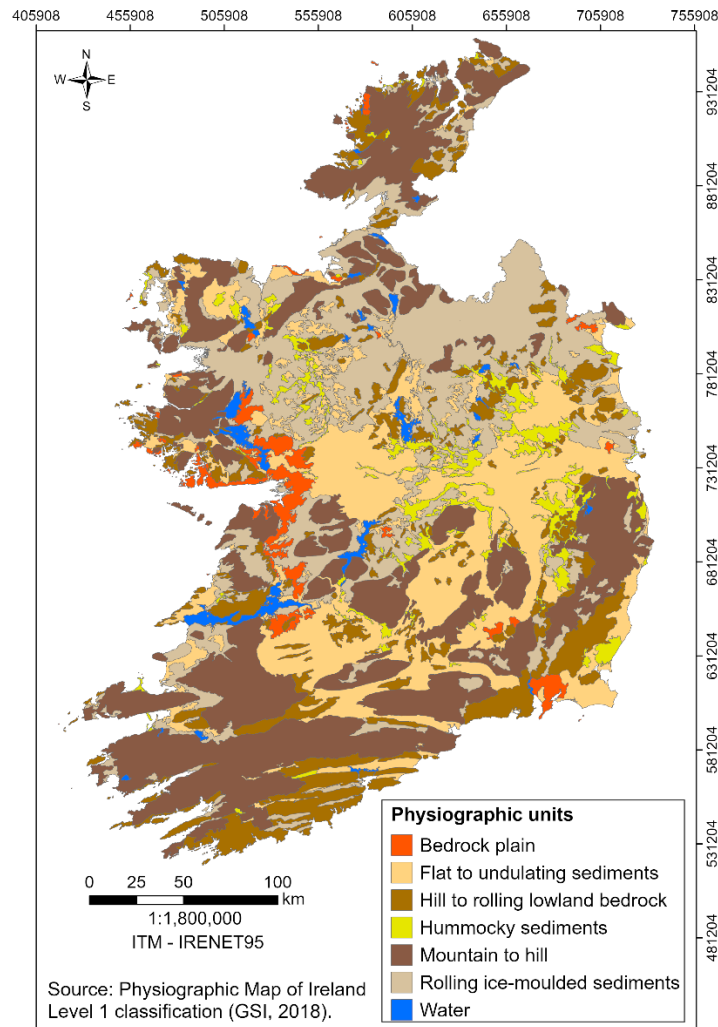


Figure 3.2 Physiographic map of Ireland depicting the broad-scale physical landscape units. Source: Geological Survey Ireland (2018).

The climate is characterised as mild maritime due to the influence of the Atlantic Ocean (Irish west coast) and the western fringes of the European continent with prevailing south-westerly winds (O’Sullivan *et al.*, 2018). In terms of dominant drivers and main processes in Irish soil formation, these authors point out the high rainfall rates in Ireland to be amongst the former, and leaching and gleying among the latter. They also highlight that excess soil moisture poses a constraint in Irish soils in terms of productivity and environmental perspectives.

The average annual rainfall in Ireland is approximately 1288 mm (Curley *et al.*, 2023) based on the period 1991-2020. The 30-year average annual distribution shows a typical west to east decrease (Curley *et al.*, 2023). The higher totals rainfall are found in western coastal areas that coincide with the rain-fed peats (Ombrotrophic Peats) as well with the highest mountains in the southwest, west and north-west, as rainfall increases significantly with height. Spring and summer are the driest seasons with 256 mm and 282 of rain, respectively (Curley *et al.*, 2023) while winter and autumn are wettest, with 380 mm and 369 mm, respectively (Curley *et al.*, 2023). The driest months are April and May (79 mm and 82 mm, respectively), whereas December is the wettest (142 mm) (Curley *et al.*, 2023) (Figure 3.3).

Soil temperatures in Ireland are moderated by the effect of the Gulf Stream (O’Sullivan *et al.*, 2018). Soil moisture related to precipitation represents, in general, the dominant climatic factor affecting soil formation in Ireland, as precipitation exceeds evaporation (O’Sullivan *et al.*, 2018). This characteristic, in turn, has a direct relationship with leaching along with both the topsoil and throughout the soil profile, whose process is dependent on the texture of the soil. Overall, leaching and gleying are the two main active processes driving soil development in Ireland due to the high rainfall rates occurring in the country (O’Sullivan *et al.*, 2018).

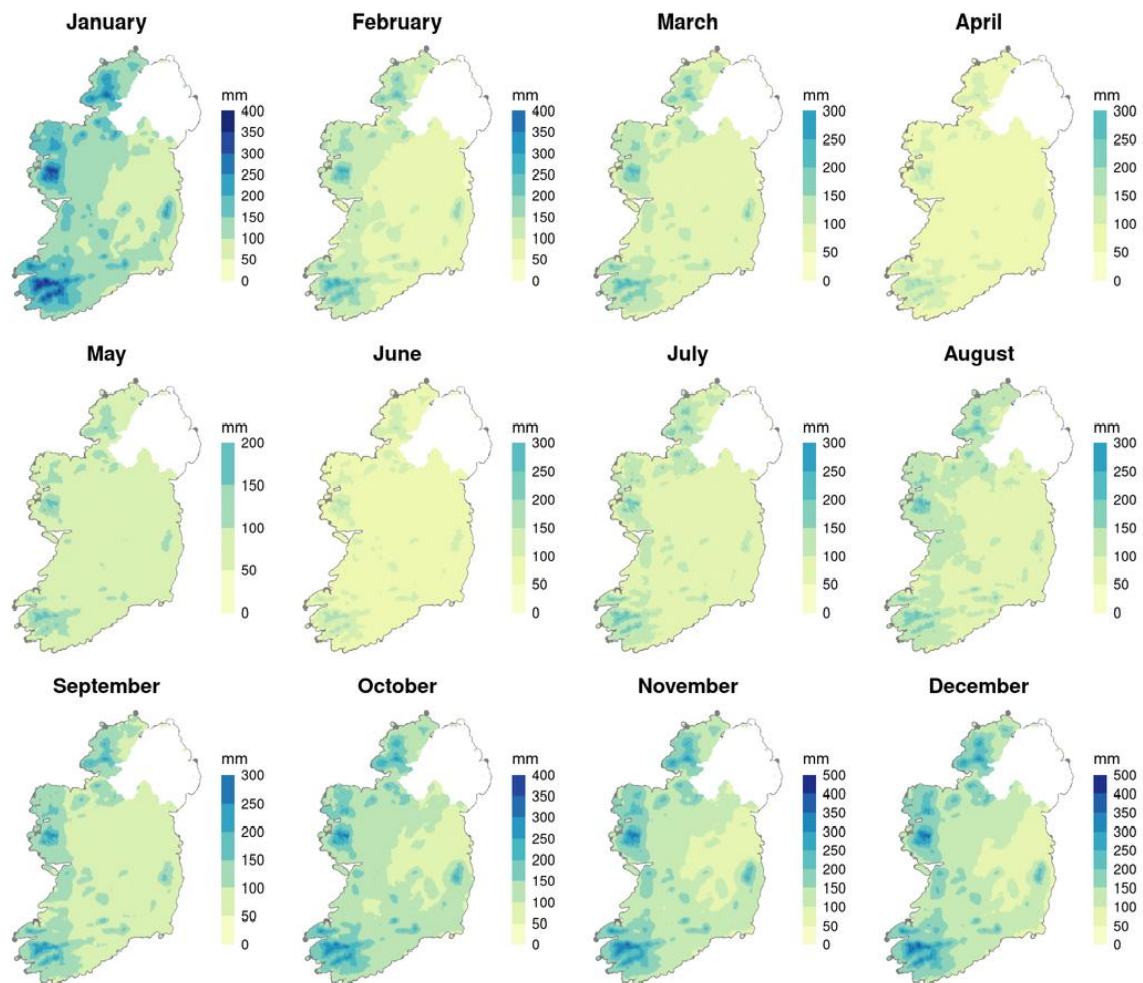


Figure 3.3 Mean monthly rainfall ranges for Ireland from 1991 to 2020. Source: Curley *et al.* (2023). The Irish Meteorological Service.

Soils in Ireland are relatively young (approximately 15,000 years) and their formation relates to the retreat of the last ice age of the Holocene epoch (O’Sullivan *et al.*, 2018). Irish soils formation mostly initiated from a glacial till which forms the majority of parent material in the country (O’Sullivan *et al.*, 2018). For example, clay-rich deposits from the Mid-landian glaciation, occurring in Wicklow and Wexford, on the east coast, can currently be associated with heavy, silty clay soils (O’Sullivan *et al.*, 2018). Thus, glaciation played a pivotal influence on Irish soils since the majority of them derived from glacial deposits (O’Sullivan & Creamer, 2018). Due to this fact, “mapping and describing Irish soils represents a complex task” (O’Sullivan &

Creamer, 2018:118), as the authors highlighted that nine Great Soils Groups in Ireland (out of 11) were found in one field in Co. Cork during one soil survey.

Amongst the key soil-forming factors, climate represents a major influence in soil formation, in particular temperature and precipitation. For example, the development of Podzol and Brown Podzolic soils are directly related to colder, wetter conditions that result in the development of lower horizons rich in Fe, Al and humus due to leaching (O’Sullivan *et al.*, 2018). Precipitation also plays an important role in the migration and accumulation of the soil particles size on the soil surface layer and across the soil profile. For example, Luvisols are associated with clay being elluviated or moved down the profile and redeposited in deeper horizons.

Limestone soils, which have a higher clay percentage, dominate the central lowlands of Ireland, and peat interspersed with these limestone soils is a prominent feature at this lowland (O’Sullivan & Creamer, 2018).

3.3 Soil information for Ireland

Limited soil information for Ireland has been generated using different digital soil mapping approaches – at continental and national scales. Currently, no directly derived soil textural map exists, rather, soil texture is derived from the soil modelling approaches, as presented below.

3.3.1 LUCAS Survey

At a continental scale, Ballabio *et al.* (2016) carried out a first approximation of the Global Soil Map products for Europe (Figure 3.4 and Figure 3.5), wherein they also developed the LUCAS topsoil database (soil samples collected at 20 cm depth). They predicted several soil properties (including sand, silt, clay) using remotely sensed data and spatial interpolation model. From the maps of each textural class (Figure 3.4) it can be noted that sand exhibits less occurrence than silt and clay in terms of content percentage, which resulted in a predominance of loam soil as depicted in USDA soil textural classes (Figure 3.4). Nonetheless, it is worth noting that such soil texture occurrence patterns in this instance are not accurate and have more to do primarily with the modelling approach and resolution of the datasets used for the predictive analysis. For example, the absence of peats and rock classes is noticeable.

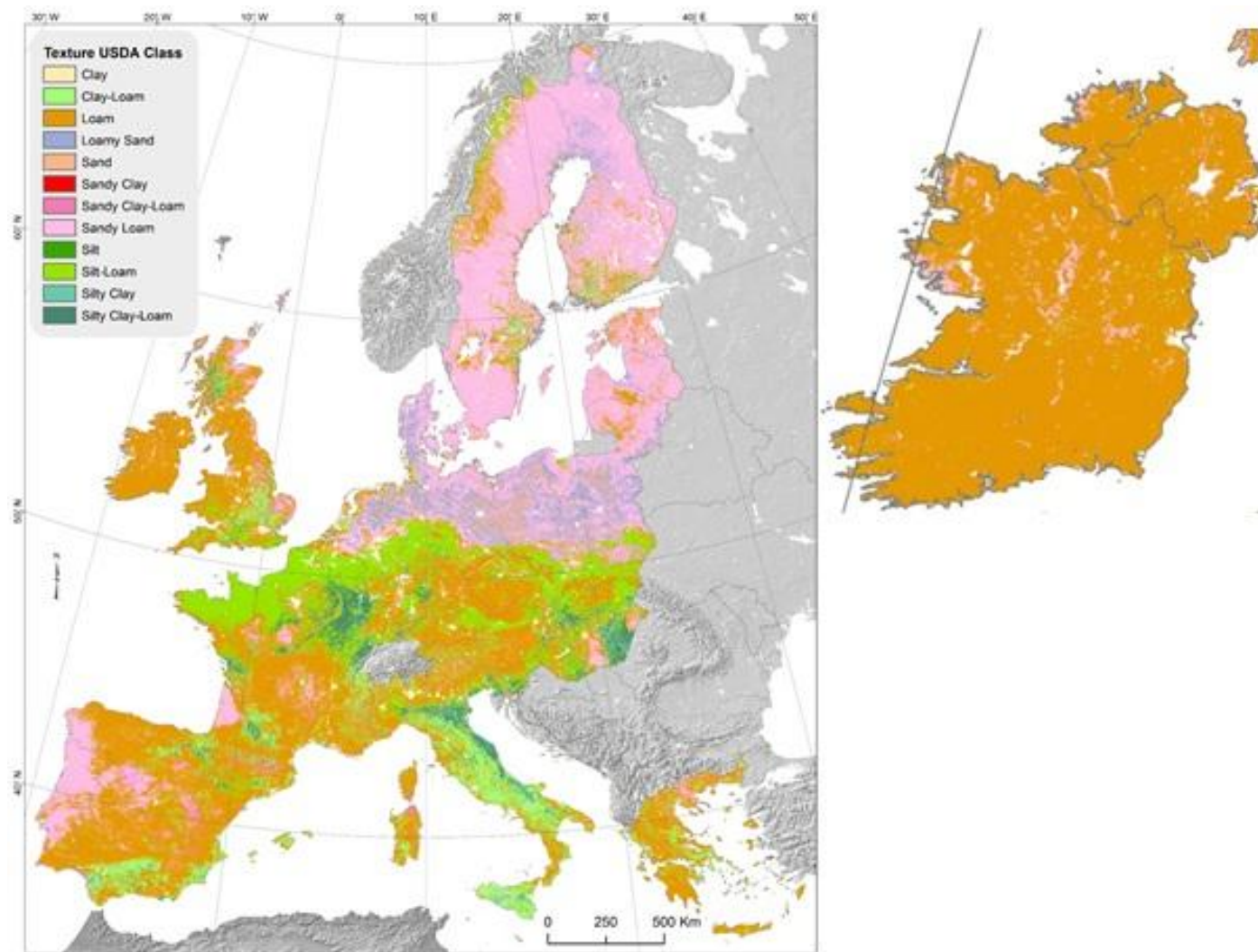


Figure 3.4 Topsoil physical properties for Europe (based on LUCAS topsoil data). Left: USDA soil textural classes derived from clay, silt and sand maps. Right: Zoom (unscaled) in the study area from the USDA soil texture classes map. Source: Adapted from Ballabio *et al.* (2016).

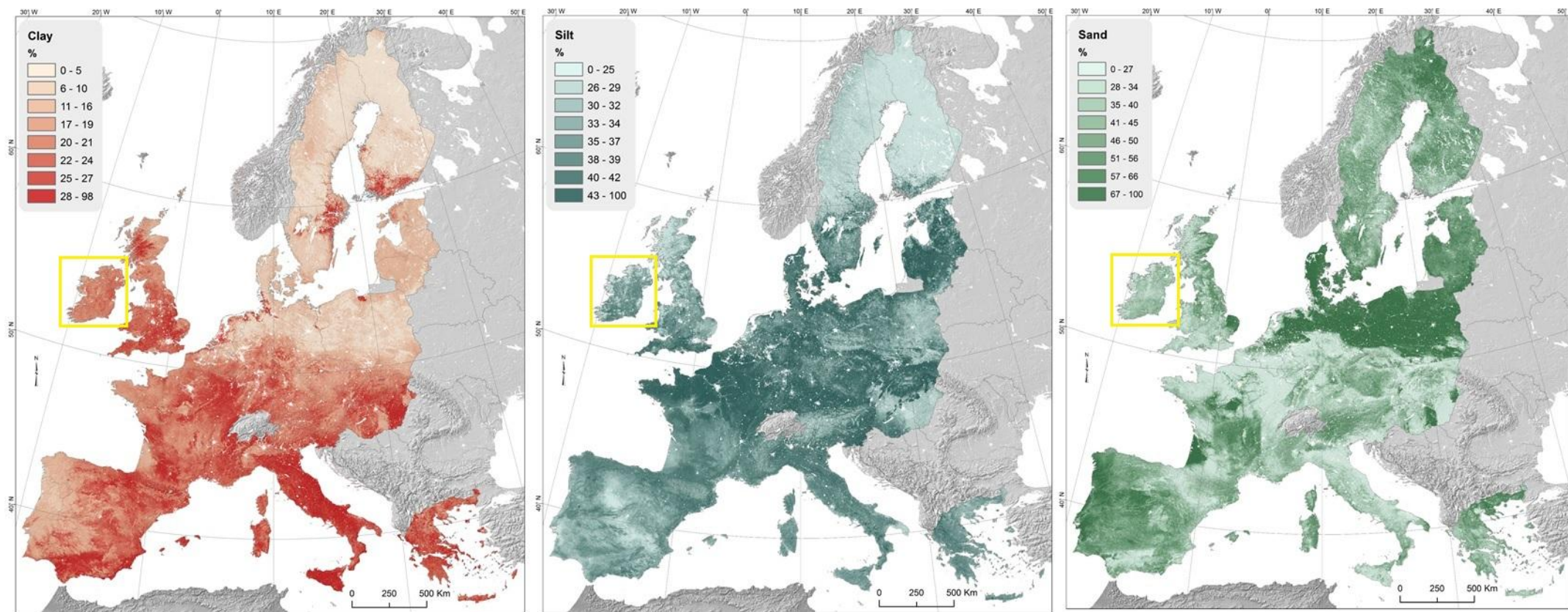


Figure 3.5 Clay, silt and sand content (%) in topsoil (0-20cm) modelled by Multivariate Additive Regression Splines. Topsoil physical properties for Europe (based on LUCAS topsoil data). Source: Ballabio *et al.* (2016). A yellow square was drawn on the figure to highlight the study area of this thesis.

3.3.2 Irish soil mapping

At a national scale, the Irish Soil Information System (SIS) developed a national soil association map for the country at a scale of 1:250,000, together with an associated digital soil information system. The project used existing data and maps from the previous National Soil Survey at different scales. The Irish Soil Information System was co-funded by Teagasc and the Environmental Protection Agency (EPA) as part of their Science, Technology and Research & Innovation for the Environment (STRIVE) programme.

A number of methods were employed, such as Random Forest, Bayesian belief networks, and neural networks in tandem with traditional soil survey applications, to estimate soil map units in counties where soil information was not available (Mayr *et al.*, 2013). Soil information such as soil classes, texture, and parent material (substrate) can be viewed for a specific area on a Web GIS platform on the Irish SIS website at <http://gis.teagasc.ie/soils/> (accessed 29 June 2024), which is freely available to access by the public.

A soil texture map was also derived at a national scale from the third edition soils map and associated database – the *Indicative Soil Texture Map of Agricultural Soils in Ireland* (Teagasc) at a scale of 1:250,000 (Figure 3.6). The inference of the main textural characteristics was supported by field observations taken all over the country, where soil texture was estimated by hand, and also where soil texture was measured in a laboratory for samples collected as part of the profile pit campaign. (Creamer *et al.*, 2016). According to the indicative soil texture map, fine loam is the predominant soil texture class in Ireland. Nevertheless, it is important to highlight that this map provides only an overview of the range of soil textural classes that occur in conjunction with the main geological landscapes found in Ireland, thus, it is not appropriate for detailed mapping purposes or field-level descriptions (Creamer *et al.*, 2016). The cut-off for the various texture categories is based upon clay content of the soil (%). In reality, soils will vary in their texture by 1–5% in most fields and by up to 20% in more extreme cases (Creamer *et al.*, 2016).

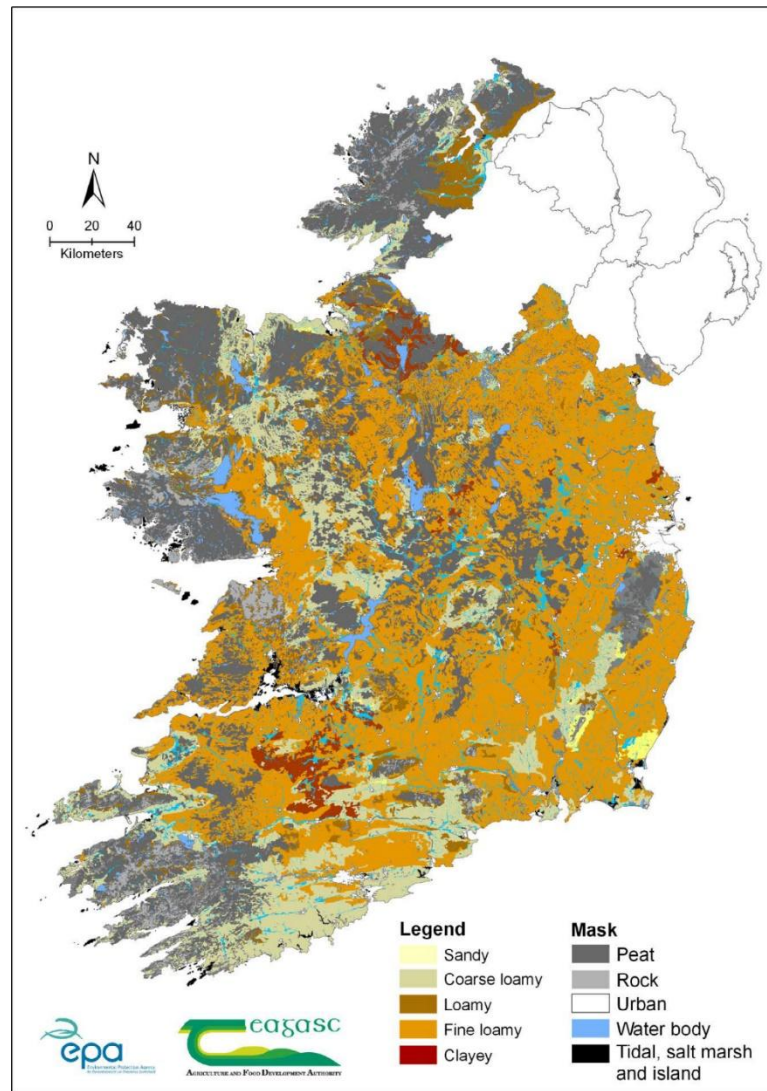


Figure 3.6 Indicative soil texture map of agricultural soils in Ireland. Source: Creamer *et al.* (2016).

In terms of national soils survey, Ireland has two complementary soil projects that encompass both topsoil and subsoil in soil properties analysis (including soil texture and soil particle size) and field observations. The Tellus Project survey, managed by the Geological Survey Ireland (GSI), aims to gather geochemical and geophysical data across Ireland to examine the chemical and physical properties of Irish soil, rocks and water. The Tellus programme involves two types of surveying: (i) airborne geophysical surveying using a low-flying aircraft, and (ii) ground-based geochemical surveying of soil, stream water and stream sediment. In the latter, researchers collect soil samples from both shallower (5-20 cm) and deeper (35 – 50 cm) depths. The second project, Terra Soil, is a collaborative research partnership between the Geological Survey of Ireland and Teagasc, which aims to produce information on nutrients, trace elements in soil along with soil texture information regionally. Terra Soil was launched in 2018 and analysed soil samples from both upper soil (5-30cm depth) and deeper soil (35-50cm depth) layers.

3.4 Datasets

Here, a brief overview of the data is outlined – with further detail provided in the respective methodological chapter. Specifically, this section addresses the data – soil, remote sensing, geophysical and environmental variables.

3.4.1 Soil information

Soil data, used to train and validate the models, consisted of 235 soil samples collected by different institutions and programs across Ireland covering both shallow and deep soil layers. The source datasets are as follows:

- Irish Soil Information System (SIS). Modal and Non-Modal profiles (Creamer *et al.*, 2014). Soil Survey carried out by the Irish Agriculture and Food Development Authority (Teagasc);
- LUCAS Topsoil 2009 Survey-v1 (Toth *et al.*, 2013; Orgiazzi *et al.* 2018) and LUCAS Topsoil 2015 Survey (Jones *et al.*, 2020), both soil campaigns were carried out by the European Soil Data Centre (ESDAC) under the programme named Land Use/Land Cover Area Frame Survey (LUCAS);
- Soil Carbon Project 2008. SoilC was a two-year project to estimate soil carbon stocks in representative Irish soils (by field sampling and laboratory soils analysis) and to model soil carbon cycling in Irish grassland. It was carried out by the University College Cork (UCC) and Teagasc. Details on the survey and results can be found in Kiely *et al.* (2009).
- WoSIS Soil Profile Database (June 2022). The World Soil Information Service (WoSIS) is a global soil database that aims to provide the user with standardised and ultimately harmonised soil profile data addressing 45 soil properties (physical and chemical).
- All the dataset encompasses soil particle size fractions as one of the soil properties analysed but only the topmost layer (0-15 cm in depth) was considered in this study as it relates to topsoil.

3.4.2 Microwave remote sensing data

The Sentinel 1 imagery (Level 1, acquisition date: 2021-04-03) was initially co-registered, multi-looked and speckle-filtered using a Boxcar method at a 7×7 window size, and radiometrically and geometrically calibrated. Calibration was applied to convert the raw images into backscattering coefficients (σ^0), for input in the modelling approach outlined in the methodological sections in Chapters 4-6. Terrain correction was applied by using the Copernicus Digital Elevation Model for Europe (ESA EEA-10) at 10 m spatial resolution. All datasets were reprojected to the Irish Transverse Mercator projection and IRENET-95 horizontal datum. These procedures were carried out using the Sentinel-1 Toolbox 9.0.0 version (SNAP-ESA).

Radar-based data consist of (i) the backscatter coefficients (backscatter intensity) provided by the VV and VH polarisations measured in sigma nought (σ^0) obtained for the in-situ soil sampling locations (i.e. soil legacy data, 235 soil samples), georeferenced; (ii) alpha parameter

resulting from the dual-pol decomposition, which varies between 0° and 90° and is used to determine the dominant scattering mechanism; (iii) the dual-pol radar vegetation index (RVI, Equation 3.1) (Nasirzadehdizaji *et al.*, 2019; Gururaj *et al.*, 2019) developed for crop monitoring using Sentinel-1; and (iv) surface soil moisture (SSM) which refers to the relative water content of the top few centimetres of the soil, measured by Sentinel-1 by the Global Land Service of Copernicus (raster 1 km, Europe, daily – version 1). Likewise the alpha parameter, RVI takes into account the vegetation effect on soil backscattering.

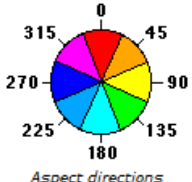
$$RVI = \frac{4\sigma_{VH}^0}{\sigma_{VV}^0 + \sigma_{VH}^0} \quad \text{Equation 3.1}$$

The rationale for focusing on Sentinel-1 datasets is threefold. Firstly, related to data availability and generalised technical reasons, such as their free-of-charge and prompt availability, their fine spatial resolution to be employed for larger areas and that matches the spatial resolution of the Digital Elevation Model for terrain correction – also available free of charge for Europe –, and their spectral resolution (i.e. C-band) that can pass through sparse and low vegetation to reach the top layer of soils. An X-band SAR has a lower ability to penetrate topsoil in comparison to a C-band SAR (Babaeian *et al.*, 2019). Conversely, an L-band SAR has a greater ability for this endeavour, however, the data are not freely downloaded and, when they are, their download is not straightforward. For example, users of SAOCOM products in the ASI Zone of Exclusivity are required to download the SAOCOM Registration Data Package and submit via email a registration request, including a brief description of the project in which the SAOCOM data are intended to be used. This process takes time to acquire data for immediate use. Secondly, there is a GIS-based platform, developed by ESA, dedicated to processing radar datasets, particularly Sentinel 1, including tools for performing dual-polarimetric decomposition. The open-code software Sentinel 1- Toolbox SNAP is freely available for download. Lastly, there is a solid and robust body of work exploring a wide range of applications from Sentinel 1 C-band – from geophysical (including soil properties) to biophysical characteristics of Earth’s surface.

3.4.3 Environmental data

Topographical data derived from the digital elevation model (DEM) ESA EEA-10 (10 m spatial resolution) was used because of the soil-landscape relationship and an earth-surface characteristic related to soil-surface roughness (micro- and macro-relief of a soil surface), which is a result of variations in soil grains, soil aggregates, soil clods and due to tillage. In this study, the topography covariates consist of altitude, slope, aspect, and curvature of the slope (concave surface or convex surface). They were obtained from the aforementioned DEM by using the Spatial Analyst tool – algorithms for surface spatial analysis – available in the ArcGIS® toolbox – 10.6 version (Table 3.1).

Table 3.1 Summary of the procedures for deriving the topographic parameters from the DEM EEA-10

Topographic parameter	Description
Altitude (elevation)	Surface elevation values were extracted directly from the cells (pixels) of the DEM (ESA EEA-10), at the sampling locations, by using the “extract values to points” algorithm, without additional procedure.
Slope	The slope tool identifies the steepness at each cell of a raster surface. The lower the slope value, the flatter the terrain; the higher the slope value, the steeper the terrain. The output slope raster can be calculated in degrees or percent (percent rise). The percent rise can be better understood when considered as the rise divided by the run, multiplied by 100. When the angle is 45 degrees, the rise is equal to the run, and the percent rise is 100 percent. As the slope angle approaches vertical (90 degrees), the percent rise begins to approach infinity. In this study, slope is measured in degrees.
Curvature	Calculates the curvature of a raster surface, optionally including profile and plan curvature. The primary output is the curvature of the surface on a cell-by-cell basis, as fitted through that cell and its eight surrounding neighbours. Curvature is the second derivative of the surface, or the slope-of-the-slope. Two optional output curvature types are possible: the profile curvature is in the direction of the maximum slope, and the plan curvature is perpendicular to the direction of the maximum slope. In this study, we used the primary output in which a positive curvature indicates the surface is upwardly convex, a negative curvature indicates the surface is upwardly concave, and a value of zero indicates the surface is flat.
Aspect	The Aspect tool identifies the direction the downhill slope faces. The values of each cell in the output raster indicate the compass direction the surface faces at that location. It is measured clockwise in degrees from 0 (due north) to 360 (again due north), coming full circle. Flat areas having no downslope direction are given a value of -1. Broadly, 0=North, 90=East, 180=South, 270=West. 

Source: Spatial Analyst ArcGIS® 10.6 tool help.

3.4.4 Geophysical datasets

The geophysical data were derived from airborne geophysical surveying and provided by the Geological Survey of Ireland (GSI), under the Tellus programme, a national programme to gather geochemical and geophysical data across the island of Ireland. The data used in this study relates to airborne surveys carried out during 2015-2021, which were merged and published in 2022. A selection of electromagnetic, radiometric (gamma-ray spectrometer: potassium concentrations (%), equivalent uranium (ppm) and equivalent thorium (ppm)) and magnetic data was obtained from the GSI. The former measures how electrical currents flow through the ground and how it changes due to different types of rock or soil. The radiometric and magnetic data provide information about soil and rocks.

Regarding the magnetic data (merged magnetic field anomaly in nT; 50 m x 50 m cell size), the First Vertical Derivative (1VD) dataset was selected, as it relates to the shallow magnetic responses, reflective of the top most soil layers (Ture *et al.*, 2020).

Electromagnetic data is commonly based on the mapping of apparent resistivity (ohm.m) or apparent conductivity; high resistivity (ρ) means low conductivity and vice-versa. All Tellus electromagnetic maps are published and presented in resistivity units. The electromagnetic resistivity at 25 kHz, reaching up to 50 m from the surface, in the 4F frequency composite image (from all four frequency images 0.9 kHz, 3 kHz, 12 kHz, and 25 kHz) was selected for this

research (Figure 3.7). The rationale for this selection relies on the fact that high frequency means low penetration, relevant for topsoil, and the composite image may reveal geological features not immediately apparent in each of the four individual resistivity maps (Ture *et al.*, 2020). The image cell size is 50 m by 50 m (an area on the ground of 2500 m²) containing the average resistivity of all the points located within that cell. When the soil contains clay, the apparent soil electrical resistivity is influenced by the clay content. The electrical resistivity of clay ranges from ~2 to ~100 ohm.m, whereas the electrical resistivity of sand is usually higher than 1000 ohm. (Besson *et al.*, 2004; Grandjean *et al.*, 2009).

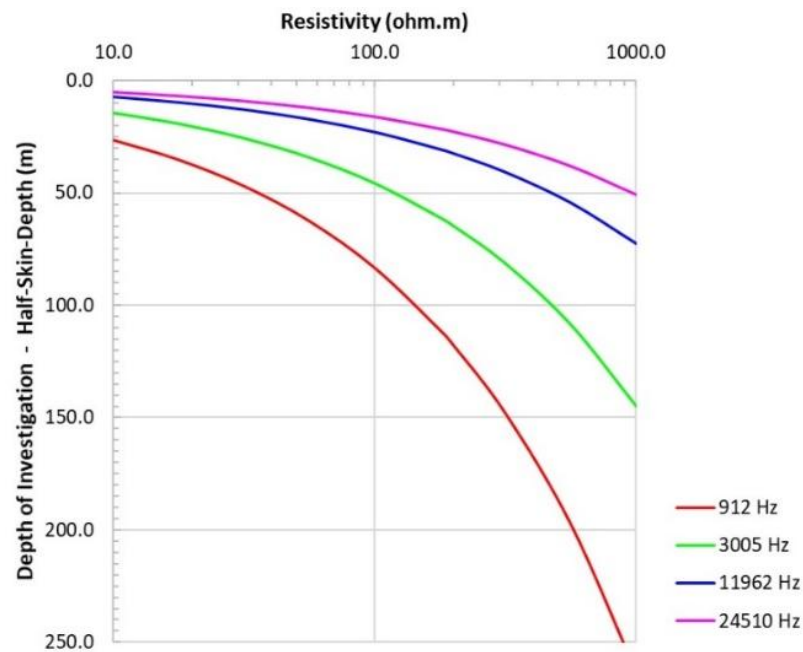


Figure 3.7 Approximate depths of investigation for typical Irish rock resistivities, for the four Tellus system frequencies, derived using the half-skin-depth estimator. Source: Ture *et al.* (2020).

CHAPTER 4 | AN ASSESSMENT OF SENTINEL-1 SYNTHETIC APERTURE RADAR, GEOPHYSICAL AND TOPOGRAPHICAL COVARIATES FOR ESTIMATING TOPSOIL PARTICLE SIZE FRACTIONS

This chapter is published as:

Deodoro, S.C., Moral, R. de A.; Fealy, R.M.; McCarthy, T., & Fealy, R. (2023). An assessment of Sentinel 1 SAR, geophysical and topographical covariates for estimating topsoil particle size fractions. *European Journal of Soil Science* volume 74, issue 5, September–October 2023, e13414, DOI: 10.1111/ejss.13414, <https://doi.org/10.1111/ejss.13414>

4.1 Context

This Chapter refers to the first part of this study's methodology (out of 3) and focuses on the core tenet of microwave remote sensing data applied to soils, as well as on the behaviour of sand, silt, and clay in the radar backscattering. It also serves as a basis for the second part of the methodology, which is a continued assessment of Sentinel-1 data for predicting topsoil particle size by employing a SAR polarimetric technique.

The data and methods outlined here seek to link two important concepts for soils and microwave remote sensing, as highlighted in the section "Theoretical background" (Section 4.3): topography and geophysics. Due to the lack of studies on this research subject, different aspects in terms of data and methods were considered for the statistical modelling approaches, such as the Sentinel-1 products available (SLC and GRD), the structure of the data (e.g., SAR-derived data, topography-based, geophysical-based), modelling approaches (e.g., linear-based and tree-based), and interactivity amongst covariates (e.g., models with- and without interactions).

Despite being motivated to answer two research questions – one related to the ability of SAR data employed to retrieve particle size fractions; and another related to the decision to consider soil particle size as compositional data in the models – findings from this Chapter also address the relationship between the backscattering coefficient and soil texture; obtaining soil texture classes indirectly from the predictive models; and the replication of the methods to another study area.

4.1.1 Abstract

Data derived from Synthetic Aperture Radar (SAR) are widely employed to predict soil properties, particularly soil moisture and soil carbon content. However, few studies address the use of microwave sensors for soil texture retrieval and those that do are typically constrained to bare soil conditions. Here, we test two statistical modelling approaches – linear (with and without interaction terms) and tree-based models, namely compositional linear regression model (LRM) and Random Forest (RF) – and both non-geophysical (e.g. surface soil moisture, topographic etc) and geophysical-based (electromagnetic, magnetic and radiometric) covariates to estimate soil

texture (sand %, silt % and clay %), using microwave remote sensing data (ESA Sentinel 1). The statistical models evaluated explicitly consider the compositional nature of soil texture and were evaluated with leave-one-out cross validation (LOOCV). Our findings indicate that both modelling approaches yielded better estimates when fitted without the geophysical covariates. Based on the Nash-Sutcliffe efficiency coefficient (NSE), LRM slightly outperformed RF, with NSE values for sand, silt, and clay of 0.94, 0.62, and 0.46, respectively; for RF, the NSE values were 0.93, 0.59, and 0.44. When interaction terms were included, RF was found to outperform LRM. The inclusion of interactions in the LRM resulted in a decrease in NSE value and an increase in the size of the residuals. Findings also indicate that the use of radar derived variables (e.g. VV, VH, RVI) alone were not able to predict soil particle size without the aid of other covariates. Our findings highlight the importance of explicitly considering the compositional nature of soil texture information in statistical analysis and regression modelling. As part of the continued assessment of microwave remote sensing data (e.g. ESA Sentinel-1) for predicting topsoil particle-size, we intend to test surface scattering information derived from the dual-polarimetric decomposition technique and integrate that predictor into the models in order to deal with the effects of vegetation cover on topsoil backscattering.

4.2 Introduction

Few studies address the use of microwave remote sensors for soil texture retrieval or for understanding, analysing and predicting these soil properties (Domenech *et al.*, 2020). Typically, research to date has focused on the utilisation of Synthetic Aperture Radar (SAR) for estimating soil moisture or carbon content and employ in-situ soil texture measurements to retrieve those soil properties (Pradipta *et al.*, 2022; Babaeian *et al.*, 2019; Beale *et al.*, 2019; Petropoulos *et al.*, 2015). Additionally, the investigation of soil texture with SAR remote sensing has largely been constrained to bare soil areas (Marzahn & Meyer, 2020; Mondejar & Tongco, 2019; Gholizadeh *et al.*, 2018; Niang *et al.*, 2014), which greatly limits its application. Similar to geophysical data based on electromagnetic signals, SAR data is reliant on dielectric properties (Marzahn & Meyer, 2020); namely, a measure of the electrical properties of a material that can impede a charge to move when subjected to an external electrical field. Consequently, both share a common difficulty, translating the measured response, or signal, into a meaningful soil characteristic (Pradipta *et al.*, 2022). An alternative approach is to understand the influence of the soil characteristic on the geophysical parameter (Pradipta *et al.*, 2022), provided that studied data refer to the same field conditions (Grandjean *et al.*, 2009). A pioneering study related to this subject can be found in Ulaby (1978; 1979).

Since soil texture comprises a relative percentage of sand, silt, and clay, the soil property is compositional in nature (i.e. the sum of the components D ($S^D=3$) is equal to 100%). In this context, we tested two different statistical modelling approaches – a linear model and tree-based model – to estimate the percentage of sand, silt and clay using microwave remote sensing data.

The modelling approach explicitly considers the compositional nature of soil texture and relies on the general principle of simplicial regression, that is, regression models fitted in S^D space. In addition, we also evaluated the use of both topographical and geophysical covariates.

In essence, we sought to address two key research questions: (i) to evaluate if radar-based methods, such as backscattering intensities from VH and VV polarisations, alpha, and a radar vegetation index (RVI) can be employed to estimate soil particle size; and, (ii) the value of explicitly considering the compositional nature of soil particle size [sand %, silt %, and clay %] in the modelling approaches, since their multivariate characteristic and sum of components are a constraint for obtaining soil texture (classes).

This paper is organised as follows. Section 4.3 presents a brief description of both the theory and relevant works. Section 4.4 outlines the methodological procedure and the datasets used to carry out the research, including a description of the selected study area. This section also outlines the treatment of soil particle size as compositional data within the modelling framework. Sections 4.5 and 4.6 present the results of the modelling approaches and discussion; finally, Section 4.7 concludes by outlining the relevance of findings.

4.3 Theoretical background

Soil texture affects the radar backscattering coefficient, as sandy soils contain a higher amount of free water than clay soils (Srivastava *et al.*, 2006; Das & Paul, 2015), whereas the latter contains more bound water. A theoretical explanation that underpins the behaviour of soil texture in the microwave spectrum refers to the relation between free water and bound water (Dobson *et al.*, 1981), both onto and into the soil, as the distribution of particle-size fractions controls the amount of free water in soils. This is consistent with Jackson and Schmugge (1989) who found that water molecules are absorbed onto the soil particles and effectively immobilize their dipoles, restricting the bound water's ability to interact with the radar signal (Das & Paul, 2015). Thus, the distribution of particle-size fractions controls the amount of free water in soils due to their interstice water content.

Consequently, soil porosity (i.e. open space between soil particles) and seasonality of soil moisture influence the soil's electrical resistivity and contribute to different soil texture characteristics (Grandjean *et al.*, 2009). For example, during summer, under dry conditions when the effect of porosity dominates, soil resistivity increases as soil moisture content decreases as water is a conductor of electricity. Therefore, higher moisture content in the soil lowers its resistivity. Conversely, when the porosity is low, the electrical resistivity can be used to discriminate soils with different textures (Grandjean *et al.*, 2009). During winter, under wet conditions when the soil pores (high porosity) are filled or almost filled with water, resistivity is highly sensitive to soil water (Grandjean *et al.*, 2009).

A core tenet for energy intensity measurements in microwave remote sensors is the radar range equation (Figure 4.1) which determines the proportion of energy that is returned from the

target. One of its parameters, the normalized radar cross section (σ^0), otherwise known as coefficient backscattering, is of significant importance as it is a measure of the polarisation intensity (i.e. VV, VH in Sentinel 1).

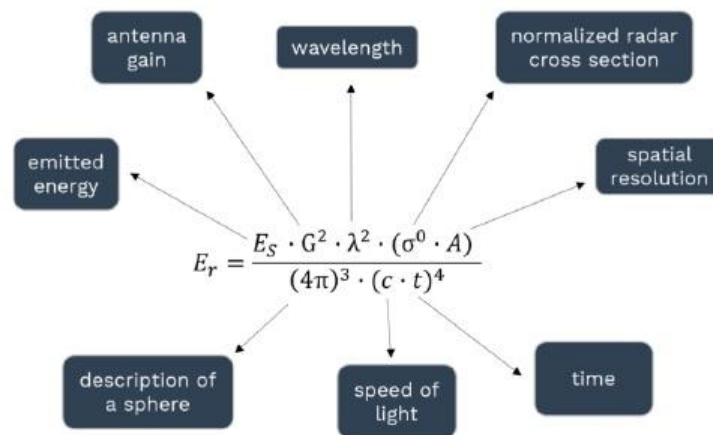


Figure 4.1 Diagram for the radar equation that is used to determine the proportion of energy that is being returned from a target. Source: ESA-EO College (2021).

A key challenge for utilising SAR for the retrieval of soil parameters is the presence of vegetation; most natural surfaces contain some type of vegetation cover that will impact on the signal. The SAR backscattering values reflect the physical properties of the earth's surface such as soil moisture content, surface roughness, topography, and dielectric constant (including soil and vegetation), which affect the measurement of microwave signals. Cross-polarisation is more sensitive to vegetation, as it captures vegetation response (backscattering) better than VV polarisation. Barret *et al.* (2007), in a study to derive relative changes in soil moisture in the near-surface (0-5cm) over an agricultural area in southern Ireland, observed that the growth of vegetation (grass) could cause attenuation of the radar signal (i.e. depolarisation effect due to volumetric scattering contribution). Pratola *et al.* (2014) in a similar study, also carried out in southern Ireland, found that in winter, the distribution of soil moisture is rather homogenous over the SAR pixels, whereas a higher spatial variability was observed in summer. They also noted that the coefficient of variation of sandy soil increases as the soil dries.

Tree-based models are alternatives to linear regression models due to their advantages of being relatively robust to overfitting and not requiring standardisation or normalisation of data, as they are insensitive to the ranges in the predictor values (Zhang & Shi, 2019). Decision trees are used in classification and regression problems whose predictor space is divided up by recursive partitioning. Random Forests (RF) (Breiman, 2001) is one example of tree-based models and it aims to merge less powerful learners to form a strong learner to minimize the residual sum of squares by tuning two main hyper parameters – the number of trees (ntree) and the number of features randomly sampled at each split (mtry). Successful examples of RF applications to estimate soil properties – without considering soil texture as compositional data – in remote sensing context are found in Mirzaeitalarposhti *et al.* (2022), Domenech *et al.* (2020),

Dotto *et al.*(2020), Cisty, Soldanova & Cyprich, (2019), Bousbih *et al.*(2019), and Ballabio *et al.* (2016).

Compositional data analysis (CoDa) is mostly performed in terms of log-transformations and relies on log-ratios between the parts or components of one sample. The theoretical foundations are found in Aitchison (1982). Three log-transformation methods are typically employed in CoDa analysis, namely additive log-ratio transformation (ALR), centred log-ratio transformation (CLR), and isometric log-ratio transformation (ILR). For the purposes of the work outlined here, we focused on the CLR and ILR methods, as they are symmetric transformations meaning that distances are preserved. The difference between these two methods is that, in the former, the identity (covariance) matrix is singular whereas, in the ILR transformation, the covariance matrix is non-singular. Compositional data approach has increasingly been employed to estimate soil particle size fractions in both non-spatial modelling (Chappell *et al.* 2019; Loosvelt *et al.*, 2013) and explicitly spatial models (Zhang & Shi, 2019; Wang & Shi, 2018; Wang & Shi, 2017; Odeh *et al.*, 2003).

Details about both CLR and ILR transformations as well as the simplicial linear regression model used in this study are found in Appendix A – Models.

4.4 Data and Methods

4.4.1 Study Region

The study area, representing approximately 24,989 km², encompasses a central swath of the Republic of Ireland (Figure 4.2). The land cover comprises croplands, grasslands, wetlands (peat bogs), and man-made structures, including urban areas. Agriculture, including pasture and arable land, is the primary land use-land cover category, based on the CORINE 2018 land cover inventory. (Appendix A-Figure A4.1).

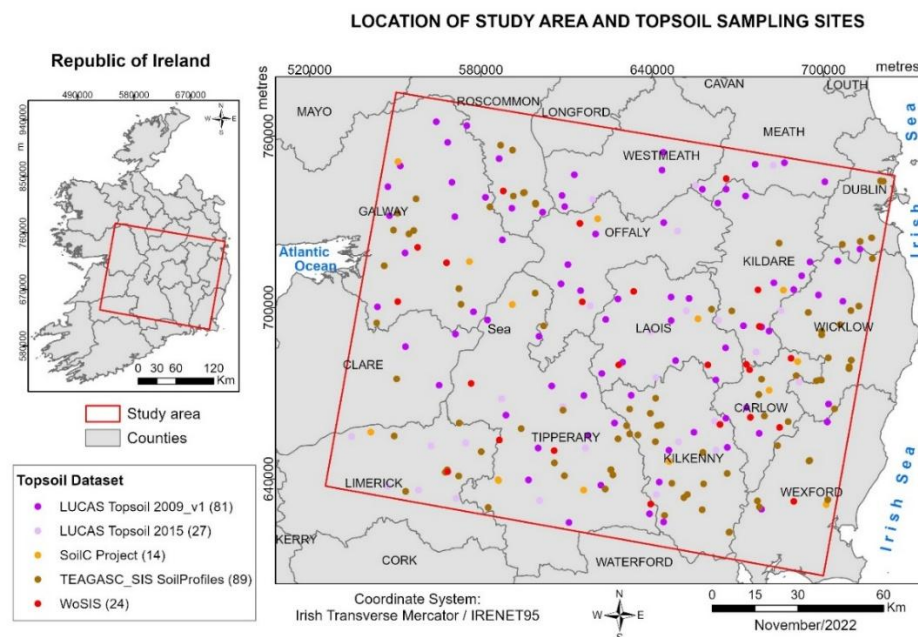


Figure 4.2 Map of the study area with the location of topsoil sampling sites available from different soil surveys and measurement programmes. Values in brackets indicate the number of samples for each soil dataset.

The study area is relatively heterogeneous concerning slope with variations in elevation across the region, ranging from 923 m on the uplands in the east of the domain to 1 m in the lowlands (i.e. flood plains in south-west) (Figure 4.3). Rainfall amounts are typically higher on the west coast reflecting the predominant maritime airflow off the North Atlantic Ocean interacting with orography along the west coast. Average annual rainfall in Ireland is approximately 1230 mm, but ranges from 750 mm to 1000 mm in the midlands/eastern areas, and between 1000 mm and 1400 mm along the west coast (Walsh, 2012). Spring and summer are the driest seasons with approximately 260 mm of rain (Walsh, 2012) while winter and autumn are the wettest, with approximately 350 mm of rainfall.

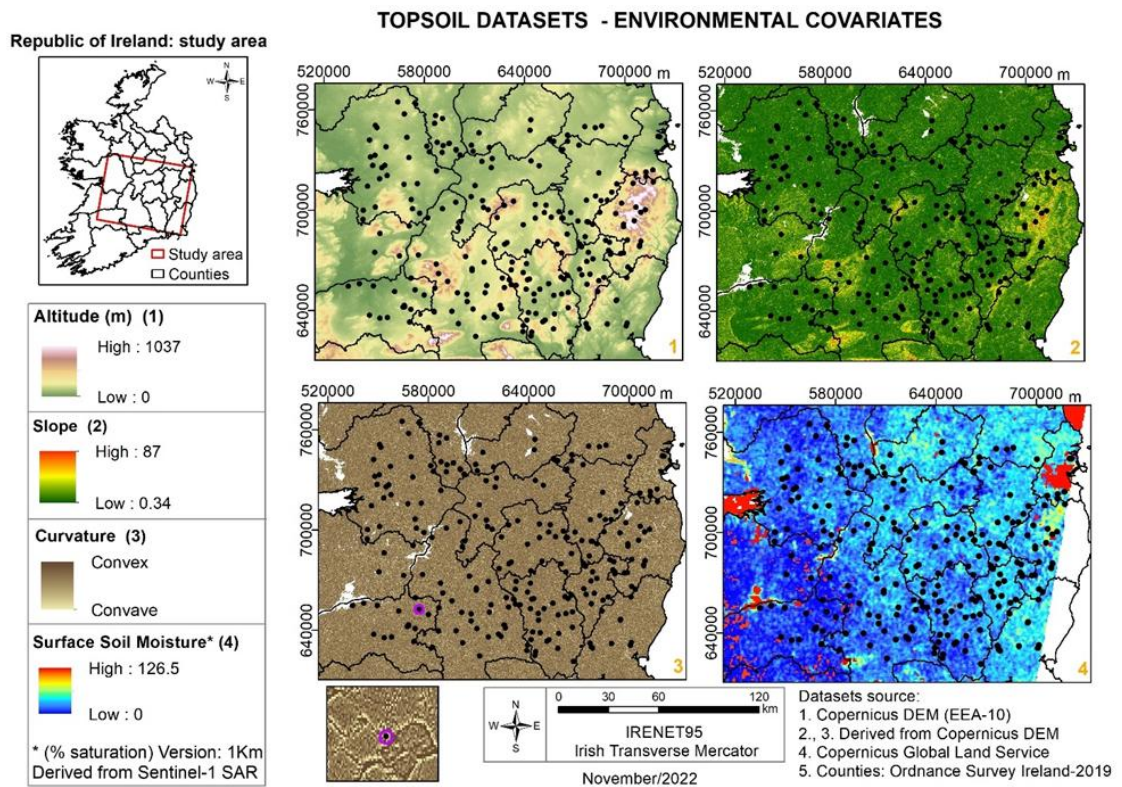


Figure 4.3 Topographic, morphometric and soil moisture for the study domain. Surface soil moisture data was obtained in the same date as that of the Sentinel 1 image (03/04/2021).

In terms of parent materials, they are broadly grouped into two main types, solid geology (and weathered materials) and drift geology (unconsolidated deposits or loose sediments) and account for the majority of parent materials across the Irish landscape (Creamer *et al.*, 2018). In an association between lithology and Quaternary sediments (GSI, 2017), the study area is composed of till derived from limestones, shales (Lower Paleozoic), and sandstones (Devonian); raised and blanket peats; till derived from granites (in the south-eastern portion); and alluvium, ordered by priority of area occurrence. The predominant soil texture class in Ireland is Fine Loamy (Creamer *et al.*, 2016).

4.4.2 Data

4.4.2.1 Soil Information

Soil textural information, used as the response variable to train and evaluate the models, consisted of 235 soil samples collected by a number of different institutions and programs across Ireland covering both shallow and deep soil layers. All the datasets encompasses soil particle-size fractions as one of the soil properties analysed but only the topmost layer (0-15 cm in depth) was considered in this study as it relates to topsoil. The source datasets are presented in Table 4.1 and details of each dataset are presented in the Appendix A-Datasets.

Soil sampling strategies included in legacy databases are typically derived based on a variety of methods, thus, it is important to highlight the procedures we adopted for using them in this study. In terms of the various depth intervals sampled in each survey, we considered the values for the 0-15 cm only. The rationale for this is due to the capability of a C-band SAR beam to reach the soil and capture information from it in the first 4 cm of the soil layer (Babaeian *et al.* 2019). This means that information beyond that depth of soil surface has no effect on the radar backscattering. The soil layer selected is within this ascribed depth. We did not employ a function-of-depth method to account for the differences in sampling depth to harmonize each soil profile from the field-described horizon (e.g. the equal area quadratic spline approach). This was justified for a number of reasons. We did not aim to model and/or predict the soil particle size fraction (PSF) by depth. Moreover, accounting for the error introduced by the spline procedure could lead to more uncertainty in the statistical models. We are explicitly considering particle size fractions as compositional data, and the spline procedure is applied independently to each particle size, thus, unity of the composition (e.g. sum=100) is not guaranteed in the spline-estimated values of sand, silt, and clay for each horizon (Saurette, 2022). Our approach is consistent with Read *et al.* (2018), who dealt with the same issues in relation to the soil datasets they employed to predict sand (%) and clay (%) from airborne geophysical data.

Regarding soil sampling, the methods differed between spade (e.g. LUCAS) and auger (e.g. LUCAS, SoilC-2008, SIS-Teagasc), whilst laboratory analysis methods differ from the pipette (SoilC2008, SIS-Teagasc) and laser diffraction (LUCAS). Despite slight differences in methods, they are consistent in considering topsoil only for the C-band SAR application in our study. Moreover, in a comparison of sampling with a spade and auger for topsoil monitoring for LUCAS, Fernández-Ugalde *et al.* (2020) found that the spade and gouge auger methods produced similar results for all soil properties, according to Lin's concordance correlation coefficient (LCCC ≥ 0.73). They concluded that, in general, the relation and average magnitude of the differences for clay, silt, and sand contents between the two sampling methods were satisfactory.

4.4.2.2 Radar Data

We used microwave remote sensing data obtained from the European Space Agency (ESA) for Sentinel-1 SAR (C-band), with Interferometric Wide (IW) swath mode acquisition, as radar-based predictors. The C-band SAR operates at a centre frequency of 5.405 GHz, which corresponds

to a wavelength of ~5.55 cm. Sentinel-1 data are available in dual polarisation (VV+VH). Both the Single Look Complex (SLC) and Ground Range Detected (GRD) (10 m x 10 m) products were obtained with the backscattering coefficients converted to the same units (σ^0). Sentinel-1 data (VV, VH) was acquired for the 3rd April 2021 (Appendix A- Figure A4.2) as the rainfall for the preceding month of March was below the long term average, associated with a high pressure that dominated the weather during March (Met Eireann, 2021); a number of meteorological stations within the study domain recorded their driest March since 2012, during 2021 (Met Eireann, 2021).

Radar-based data consist of (i) the backscatter coefficients (backscatter intensity) provided by the VV and VH polarisations measured in sigma nought (σ^0) obtained for the in-situ soil sampling locations, georeferenced; (ii) alpha parameter resulting from the dual-pol decomposition (Cloude & Pottier, 1997), which is used to determine the dominant scattering mechanism (e.g. surface); (iii) the dual-pol radar vegetation index (RVI) (Nasirzadehdizaji *et al.*, 2019; Gururaj *et al.*, 2019) developed for Sentinel-1; and (iv) surface soil moisture (SSM) which refers to the relative water content of the top few centimetres of the soil, measured by Sentinel-1 (ESA-Copernicus). The SSM product was also acquired on the same date (the 3rd April 2021). Likewise the alpha parameter, RVI takes into account the vegetation effect on soil backscattering.

4.4.2.3 Topographical and Geophysical data

A number of topographical and geophysical variables were also evaluated as covariates in the modelling approach. The topographical data consisted of altitude, slope, aspect and curvature of the slope (concave and convex surface). They were derived from the ESA EEA-10 (10 m spatial resolution) digital elevation model (DEM) using the surface Spatial Analyst tool in the ArcGIS® toolbox – 10.6 version (e.g. Aslam *et al.* (2021); Bogale (2021); Patton *et al.* (2018)). Details about the Spatial Analyst ArcGIS® applied in this work for deriving topographical data are presented in the Appendix A-Datasets.

The geophysical data refer to electromagnetic, magnetic and radiometric variables derived from airborne geophysical surveying undertaken by the Geological Survey of Ireland (GSI) as part of the Tellus programme, a national programme to gather geochemical and geophysical data across the island of Ireland. The former measures how electrical currents flow through the ground and how it changes due to different types of rock or soil. The radiometric and magnetic data provide information about soil and rocks. Details about the geophysical datasets are presented in the Appendix A-Datasets. Table 4.1 summarises the datasets used in this study.

Table 4.1 Summary of data used in the modelling approaches

Type of data	Original Dataset	Source	Covariate derived from the original dataset
Topsoil	Irish Soil Information System (SIS). Modal and Non-Modal profiles (Creamer <i>et al.</i> , 2014), which	https://data.gov.ie/dataset/irish-soil-information-system-national-soils-map	-

(response variables: sand, silt and clay)	includes the legacy survey <i>An Foras Talúntais-AFT</i> . Sampling depth varied according to survey purposes, but surficial horizons includes 0-5 cm and 5-10 cm.		
	LUCAS Topsoil 2009 Survey-v1 (Toth <i>et al.</i> , 2013; Orgiazzi <i>et al.</i> 2018). Sampling depth: 0-20 cm	https://esdac.jrc.ec.europa.eu/content/lucas-2009-topsoil-data	-
	LUCAS Topsoil 2015 Survey (Jones <i>et al.</i> , 2020). Sampling depth: 0-20 cm	https://esdac.jrc.ec.europa.eu/content/lucas2015-topsoil-data	-
	Soil Carbon Project 2008 (Kiely <i>et al.</i> , 2009). Sampling depth: 0-10 cm, 10-25 cm and 25-50 cm	https://www.ucc.ie/en/hydromet/oilcarbon/	-
	WoSIS Soil Profile Database (June 2022). Sampling depth: 0-5 cm, 5-15 cm, 15-30 cm, 30-60 cm, 60-100 cm, 100-200 cm (SoilGrids)	https://www.isric.org/explore/wosis	-
SAR (covariate)	Sentinel 1-C band (dual polarisation: VV, VH)	European Space Agency https://scihub.copernicus.eu/dhus/#/home	- Backscatter coefficients (VV, VH); - Alpha parameter obtained from a dual-pol decomposition (Cloude and Pottier (1997); - Dual-pol radar vegetation index – RVI. (Nasirzadehdizaji <i>et al.</i> , 2019; Gururaj <i>et al.</i> , 2019);
Soil moisture (covariate)	Surface soil moisture – SSM calculated from Sentinel 1 for the 3rd April 2021 (Copernicus Global Land Service)	European Space Agency https://land.copernicus.eu/global/products/ssm	-
Topography (covariate)	Digital Elevation Model-DEM EEA-10m (Copernicus European DEM)	European Space Agency https://spacedata.copernicus.eu/collections/copernicus-digital-elevation-model	- Altitude - Slope - Aspect - Curvature of the slope: concave and convex surfaces.
Geophysical: (covariate)	- Electromagnetic - Magnetic - Radiometric (potassium uranium, and thorium)	Tellus project (Geological Survey Ireland – GSI) https://www.gsi.ie/en-ie/data-and-maps/Pages/Geophysics.aspx	-

4.4.3 Methods

Two modelling approaches were evaluated in this study. The first focuses on the SAR radar-based data with the inclusion of topography covariates. The second approach employs both radar and geophysical-based data; topographical data are also considered in this approach. Two different statistical models were employed – linear- and tree- based models (Table 4.2 and Figure 4.4). Both approaches were evaluated with a leave-one-out cross validation (LOOCV), where the training set includes $n - 1$ observations and the evaluation set includes one observation. Here, we applied the ILR transformation to the response variables (sand, silt, and clay) for fitting the compositional LRM following the default partition built in the CoDaPack software (Comas-Cufí & Thió-Henestrosa, 2011). We applied the CLR transformation to the response variables to fit the

tree model-based regression (Random Forest model-RF). Details about both CLR and ILR transformations are presented in Appendix A – Models.

Table 4.2 Summary of methods and covariates employed in the two modelling approaches

Modelling Approach (Data)	Covariates	Statistical Model	Log-ratio transformation
Model 1: Non-geophysical (SAR + Topography)	VV, VH, alpha, RVI, SSM, altitude, slope, aspect, curvature	Y-compositional LRM	ILR
		Y-compositional RF	CLR
Model 2: Geophysical (SAR + Topography + geophysics)	VV, VH, RVI, SSM, altitude, slope, aspect, curvature, electromagnetic, magnetic, radiometric	Y-compositional LRM	ILR
		Y-compositional RF	CLR

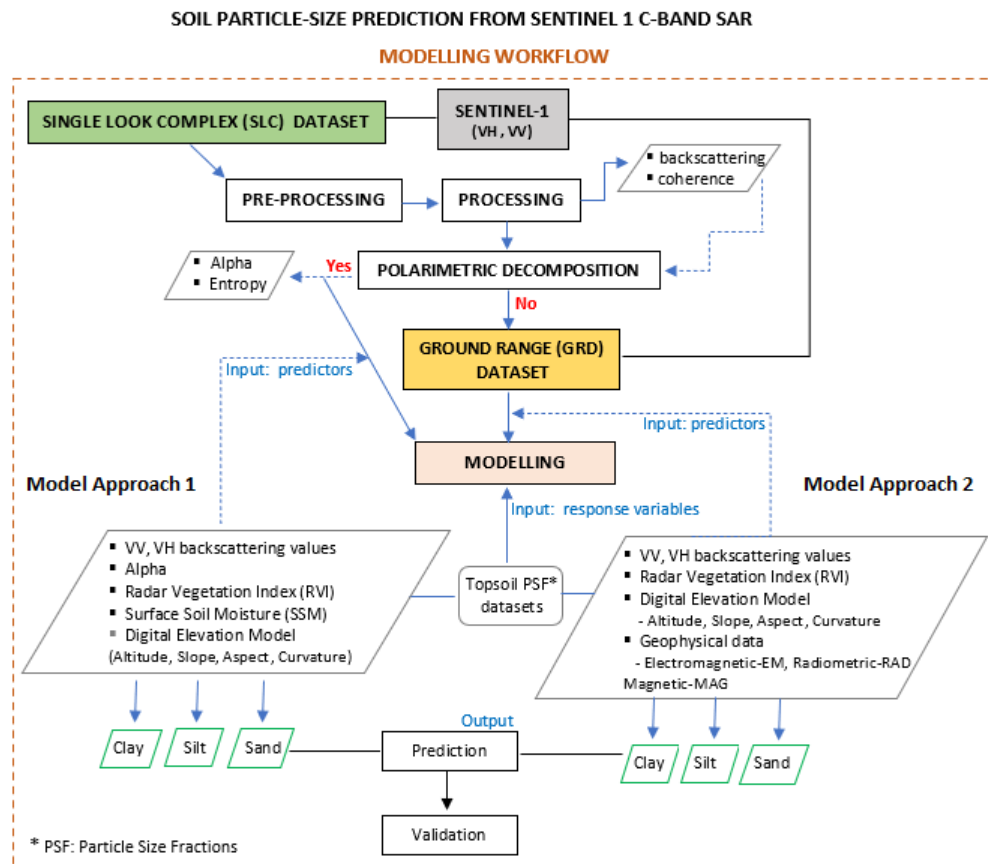


Figure 4.4 Flowchart of the methodology depicting the data used in the models, the processing steps performed for both SLC and GRD products from SAR Sentinel 1, and the modelling approaches (with geophysical and non-geophysical covariates).

We aimed to test both isometric log-ratio transformations (i.e. preserves distance in the transformation from the simplex space S^D to the Euclidian space R^D) in a linear model and in a tree-based model. A linear model is a least-squares problem (OLS) in the simplex space S^D (the natural sample space of compositional data), thus it can be formulated in terms of orthonormal log-ratio coordinates, following Morais and Thomas-Agnan (2021). We applied the CLR transformation to the response variables to fit the tree-based model since this transformation method produces a singular covariance matrix of the target compositional parts and RF does not rely on covariance matrices or other statistical assumptions. It is worth noting that the model was fitted using a multivariate approach, wherein the response variables were treated simultaneously.

After checking for the convergence of the tuning parameters for the RF model (Appendix A-Tables A4.1 and A4.2), we selected the default values available in the R package “Ranger” (Wright & Ziegler, 2017), which are $n_{tree} = 500$ and $m_{try} = 3$, since different values for n_{tree} and m_{try} did not significantly change the model’s performance. For the compositional LRM, we also relied on the default method in the CoDaPack software (Comas-Cufí & Thió-Henestrosa, 2011).

A key objective was to understand the nature of the relationship between the predictors and the response variables in order to investigate the effectiveness of the SAR-based covariates to predict soil texture. To achieve this, we also evaluated 2nd order interaction terms in the compositional LRM. This approach should also assist in understanding interactions between influential processes. It is also worth noting that the soil texture class was inferred from the compositional predictions of sand, silt, and clay content rather than directly using a categorical modelling approach. The resultant soil textural classes will also be outlined in the Results section.

4.4.3.1 Evaluation metrics for soil particle size predictions

The accuracy and performance of the original (untransformed) and transformation approaches were evaluated using three statistical indicators; the Root Mean Square Error (RMSE) (Equation 4.1), Mean Absolute Error (MAE) (Equation 4.2) and Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970) (Equation 4.3). The latter is a normalized statistical metric that determines the relative magnitude of the residual variance compared to the measured data variance (Nash and Sutcliffe, 1970). It indicates how well the plot of observed versus predicted (simulated) data fits the 1:1 line according to the following: $NSE = 1$ corresponds to a perfect match of the model to the observed data; $NSE = 0$ indicates that the model predictions are as accurate as the mean of the observed data; $-\infty < NSE < 0$ indicates that the observed mean is a better predictor than the model.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad \text{Equation 4.1}$$

where y_i is the actual value of the dependent variable, \hat{y}_i is the predicted value of the dependent variable, and n is the number of observations.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad \text{Equation 4.2}$$

where y_i , \hat{y}_i , and n represents the same descriptors that those of RMSE.

$$NSE = 1 - \left[\frac{\sum_{i=1}^n (Y_i^{obs} - Y_i^{sim})^2}{\sum_{i=1}^n (Y_i^{obs} - \overline{Y^{obs}})^2} \right] \quad \text{Equation 4.3}$$

where Y_i^{obs} is the i^{th} observation, Y_i^{sim} is the i^{th} simulated (predicted) value, $\overline{Y^{obs}}$ is the mean of observed data, and n is the total number of observations.

4.5 Results

4.5.1 Statistical descriptive analysis

Based on the initial descriptive analysis, the frequency distribution of the clay values displayed a highly skewed distribution in its original form (untransformed), with a skewness equal to -1.07. Following the CLR transformation, the symmetry improved (0.34) (see Appendix A- Figures A4.3 and A4.4). Conversely, for sand (referred to as sand.clr following the CLR transformation), this transformation method did not improve the skewness. Based on the selected soil textural information, the sand fraction exhibited greater variability as well as the minimum and maximum percentages, whereas clay showed lower variability and the lowest maximum percentages (see Appendix A- Figure A4.4). Moreover, the soil fraction contained a number of outliers, according to the Atypicality Index for CoDa (a measure of irregular data), under a level of confidence (threshold of atypicality) of 0.975 (default) (Appendix A- Figure A4.5). However, this may be due to differences in proportions between the components (i.e. sand %, silt %, and clay %), and those soil samples were considered in the modelling.

With regards to the predictor variables, the SAR VV polarisation provided the highest backscatter intensity values (0.35) whereas VH gave the lowest intensity values (0.07). Also, we did not find a significant correlation between the radar backscatter coefficients (σ_{VV}^0 , σ_{VH}^0) and the soil particle-size fractions for either the non-transformed or transformed response variables (Appendix A- Figures A4.6-A4.8). This result is not conclusive and it is limited as just relying on radar data from a single acquisition. Incorporating a time series might provide results with more insight into potential relationship. However, the use of a single acquisition, with a drier meteorological condition, was an approach to reduce the effect of high amount of moisture or water saturation on soils, since wet clay or heavy clay reflects radar waves. This is due to clay's high electrical conductivity, hindering radar signal penetration into soil layers and causing signal attenuation (i.e. wet clay absorbs radar waves). Hallikainen et al. (1985) and Das & Paul (2015) highlight the effect of soil textural configuration on soil dielectric properties, regarding water-holding capacity. Gorraeb et al. (2015) noticed a decrease in intensity of backscattered radar signals and with radar data acquired during rainfall months, particularly in the fraction of clay present in the soil.

4.5.2 Soil particle size prediction

4.5.2.1 Y-compositional LRM

▪ Modelling approach 1 : SAR + Topography (morphometry) data

The compositional model for approach 1 (Table 4.2 – Model 1), without interaction terms, yielded an overall R^2 value of 70.74 %. Curvature was identified as the most statistically significant predictor (p-value = $<2e-16$). In general, the distribution of the samples estimated by the model was similar to that of the original data in the ternary diagram (Appendix A- Figure A4.9). The compositional coefficients of the LRM indicated an effect of VH backscattering (σ_{VH}^0) on clay (referred to as inv.ilr. 3 after transformation) (Appendix A- Figure A4.10). Based on the NSE, the agreement between the observed and predicted samples in the test data were 0.94, 0.62,

and 0.46 for sand, silt and clay, respectively (Table 4.3 and Figures 4.5a-4.5c). Regarding the soil texture classes, the corresponding classes from the soil particle-size fractions predicted (test) are Sand (S), Loamy Sand (LS), SL (Sand Loam), Loam (L), Clay Loam (CL), Silty Clay Loam (SICL), Silty Clay (SIC) (Figure 4.5h).

The inclusion of 2nd-order interaction terms resulted in a disimprovement in the model; resulting in a lower NSE and increase in the RMSE and MAE values (Table 4.3 and Figures 4.5d-4.5f). The most statistically important variables and interactions were curvature, VH:slope, SSM:curvature, slope:curvature, SSM:Alpha, aspect:curvature, and, Altitude:Curvature. The derived soil texture classes are depicted in Figure 4.5i. Analysis of variance (ANOVA) for the models fitted with and without interaction indicated that the former is statistically significant (p-value = 9.929e-09).

These results are reflected in the soil texture diagram in terms of both number and typology of classes (Figures 4.5g, 4.5h and 4.5i). More clayey and loamy soil textural classes were obtained in the soil ternary when interaction terms were considered in the regression model, whereas the same is not observed in the model without the interaction terms. It is important to highlight that here, reference to clay is the textural class rather particle-size fraction.

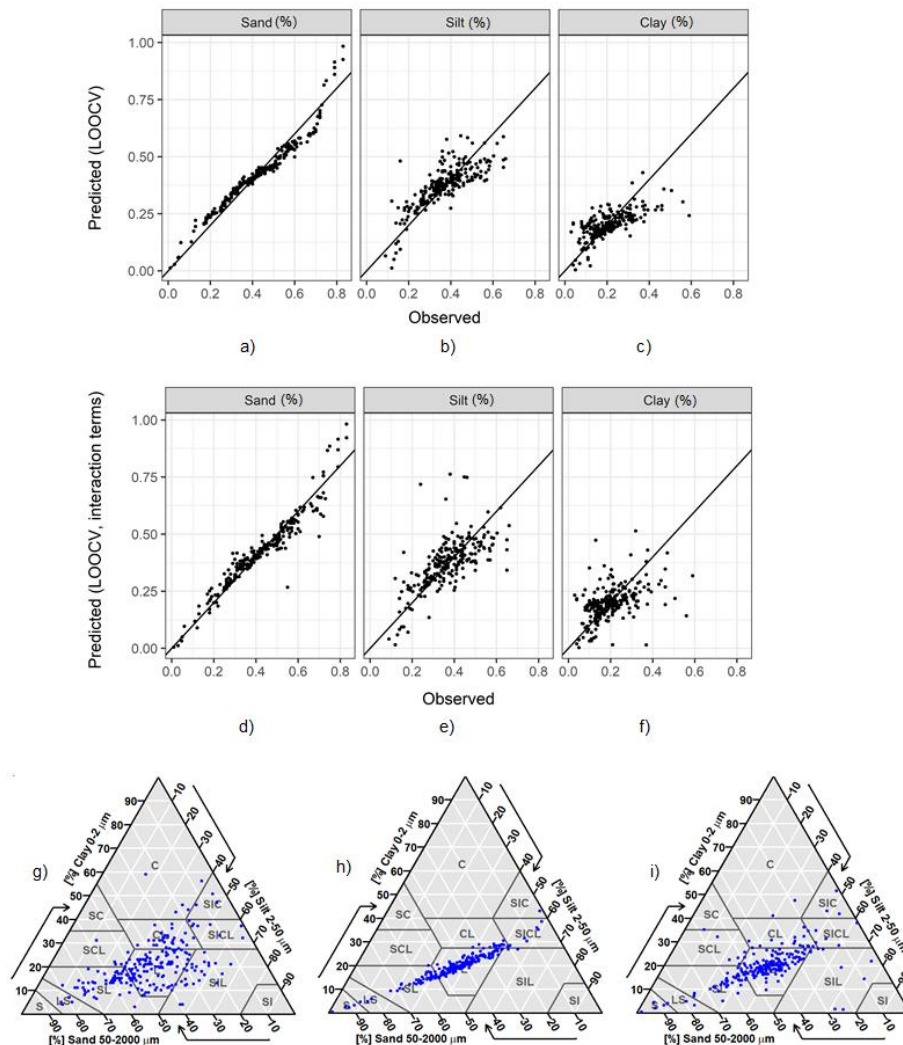


Figure 4.5 Observed vs. Predicted plots from the Y-compositional linear regression model validated with LOOCV for Modelling approach 1 (SAR + Topography covariates), without 2-order interaction terms (Figures 4.5a, 4.5b, 4.5c) and with 2-order interaction terms (Figures 4.5d, 4.5e, 4.5f). The responses are measured in % content and the line in the plots refers to the 1:1 line. Corresponding soil texture classes are displayed in a soil ternary diagram. The soil ternary diagram in Figure 4.5g depicts the original soil texture classes ($n=100\% = 235$) obtained directly from the measured data, while ternary diagrams in the 4.5h and 4.5i show classes resulting from predictions of soil particle size fractions without- and with 2nd order interaction terms, respectively.

It is worth highlighting that the sum of the soil particle-size fractions predicted (100% or 1) was guaranteed with the ILR transformation applied to the original data (See Appendix A- Table A4.3). Conversely, when performing a simple linear regression model, as a univariate approach and with no transformation applied to the data, the soil particle-size fractions predicted are not constrained to sum to 100% (Appendix A- Figure A.411). Moreover, some negative estimates were obtained from both a random proportional split (e.g. 75% and 25%) and LOOCV (Appendix A- Figures A4.11 and A4.12). These findings are likely a product of employing standard statistical methods for the analysis of compositional data, which can lead to biased results (Filzmoser *et al.*, 2018). Odeh *et al.* (2003) also noted that modelling individual components of composition was not meaningful.

▪ Modelling approach 2 : SAR + Topography (morphometry) + Geophysical datasets

The statistical descriptives indicated different symmetry and variability of the geophysical dataset (Appendix A- Figure A4.13). In Model 2, which employed the geophysical covariates and the ILR transformed compositional responses with LOOCV LRM, the overall R^2 obtained was 18.33 %. In terms of predicted and observed responses, the NSE was close to zero for sand, silt and clay (Table 4.3 and Figures 4.6a-4.6c). The radiometric covariates (thorium and uranium) and altitude were the most statistically significant predictors. An effect was observed for VV on the sand response (Appendix A- Figure A4.14). The corresponding classes from the soil particle-size fractions predicted (test) are Loamy Sand (LS), SL (Sand Loam), Loam (L), Clay Loam (CL) and Silt Loam (SIL) (Figure 4.6h). Note that LS and SIL were obtained with only one sample.

When 2nd order interaction terms were considered in the model, the agreement between the observed (actual) and the predicted data decreased (Figures 4.6d-4.6f) as indicated by the evaluation metrics (Table 4.3). Additionally, more than five soil texture classes were obtained (Figure 4.6i). Analysis of variance (ANOVA) for the models fitted with and without interactions indicated that the former is statistically significant ($p\text{-value} = 0.04$).

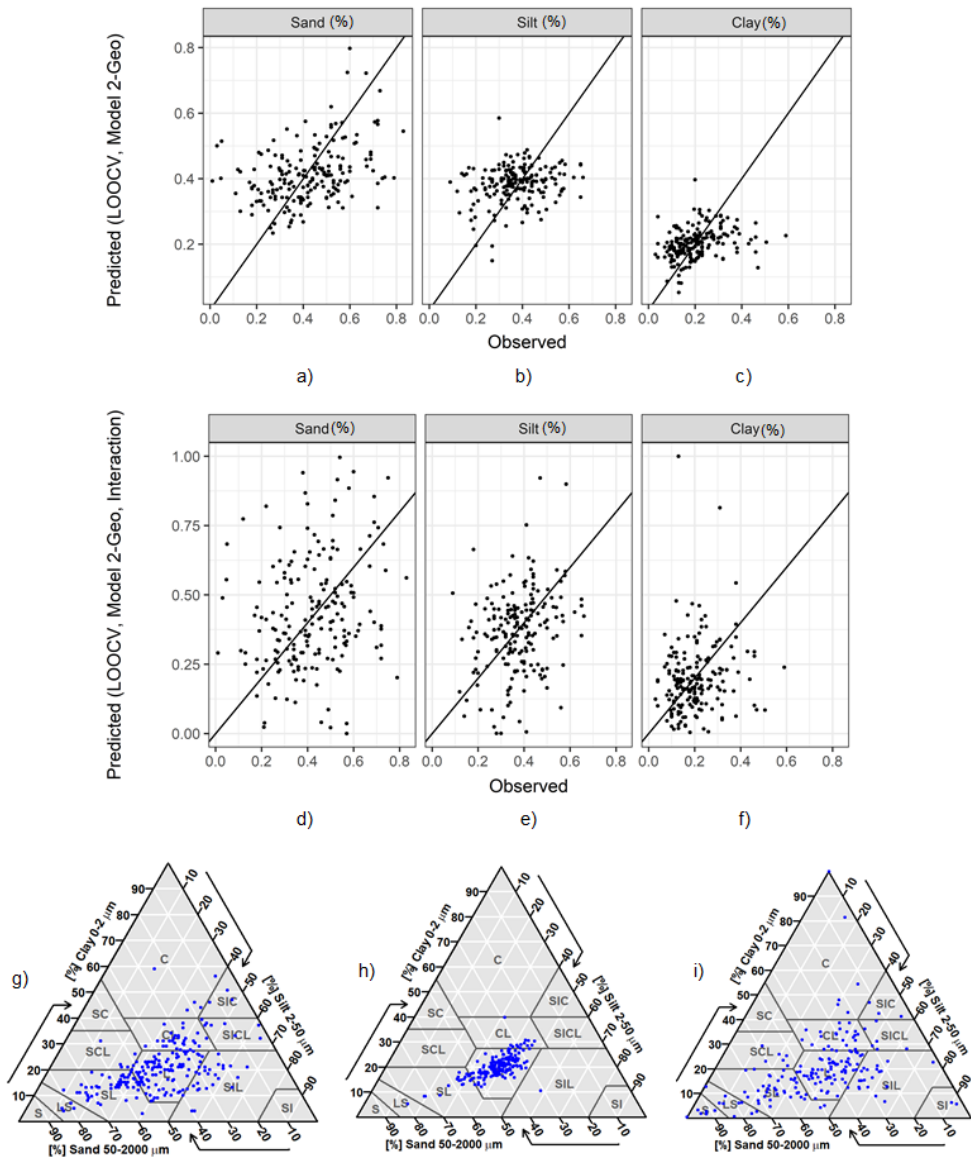


Figure 4.6 Observed vs. Predicted plots resulting from the Y-compositional linear regression model validated with LOOCV for Modelling approach 2 (SAR + Topography + Geophysical covariates), without 2-order interaction terms (Figures 4.6a, 4.6b, 4.6c) and with 2-order interaction terms (Figures 4.6d, 4.6e, 4.6f). The responses are measured in % content and the black line in the plots refers to the 1:1 line. Corresponding soil texture classes are displayed in a soil ternary diagram. The soil texture diagram in (4.6g) depicts the soil texture classes obtained directly from the measured data, while ternary diagrams in (4.6h) and (4.6i) show classes resulting from predictions of soil particle size fractions without- and with 2-order interaction terms, respectively.

4.5.2.2 Random Forest – CLR transformation applied

▪ Modelling approach 1 : SAR + Topography (morphometry) datasets

For Model approach 1, employing RF and CLR transformed responses, the NSE values obtained for sand, silt and clay were 0.93, 0.59 and 0.44, respectively (Table 4.3 and Figures 4.7a-4.7c). Similar to the LRM model, the most important variable was curvature (concavity and convexity), according to impurity method, which is a measure of the variance of the responses for regression in Random Forest (Ranger package) (Appendix A- Figure A4.15). Regarding soil texture classes, the corresponding classes from the soil particle-size fractions predicted are Sandy Loam (SL), Loam (L), Clay Loam (CL), and Silty Clay Loam (SICL) (Figure 4.7e).

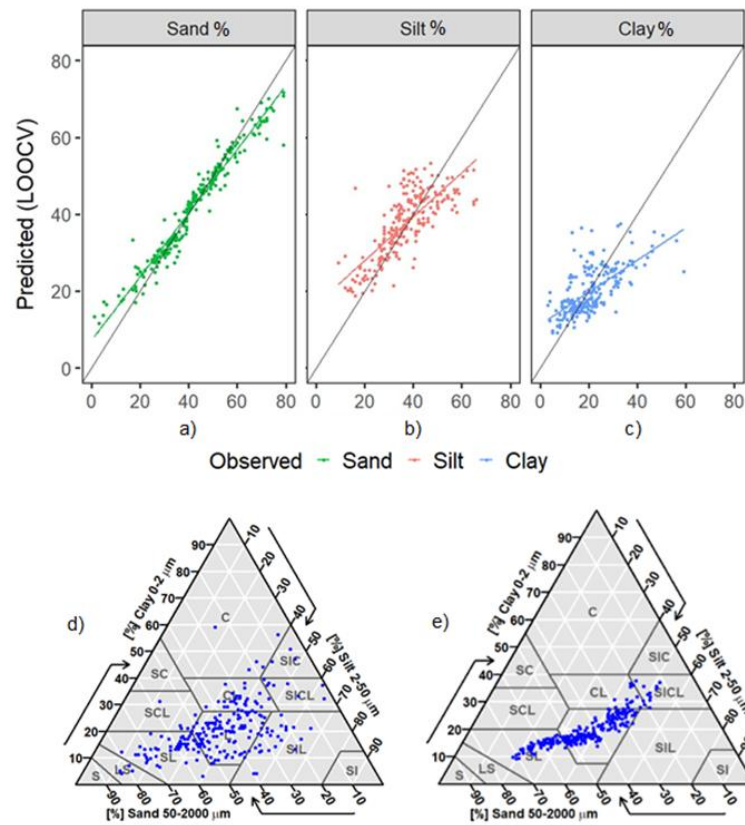


Figure 4.7 Comparative plots of observed and predicted samples (Figures 4.7a, 4.7b, 4.7c) obtained from the Random Forest Modelling with CLR transformation applied to the response variables and validated with LOOCV (Approach 1: SAR + Topography covariates). The responses are measured in % content and the black line in the plots refers to the 1:1 line. The ternary diagrams on the bottom show the soil texture classes. The original (measured data) classes are depicted in Figure 4.7d and the classes resulting from predictions of soil particle size fractions are exhibited in 4.7e.

Consistent with the LRM approach, the sum of the soil particle-size fractions predicted (100% or 1) was guaranteed with the CLR transformation applied to the original data (See Appendix A- Table A4.4).

▪ Modelling approach 2 : SAR + Topography (morphometry) + Geophysical datasets

The inclusion of the geophysical covariates resulted in a model with a NSE of close to zero (Table 4.3) indicating that the model simulations have the same explanatory power as the mean of the observations (Figures 4.8a-4.8c). Loam was the main soil textural class (Figure 4.8e) (e.g. LS, SL, L, SIL, CL). This is likely explained by the higher proportion of sand and silt estimated by the model in the test data following the observed data in which the proportion of sand and silt are also higher than clay. The RF method showed a low magnitude of importance with altitude, uranium, electromagnetic, RVI and slope being the most important predictors (Appendix A- Figure A4.16).

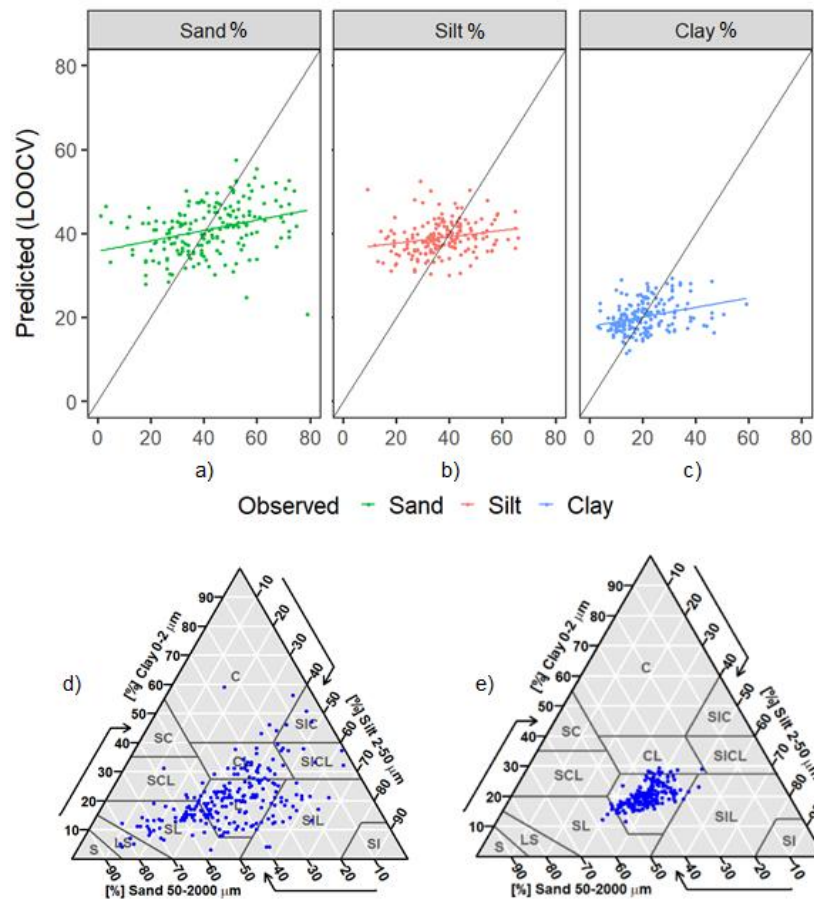


Figure 4.8 Comparative plots of observed and predicted samples (Figures 4.8a, 4.8b, 4.8c) obtained from the Random Forest Modelling with CLR transformation applied to the response variables and validated with LOOCV (Approach 2: SAR + Topography + Geophysical covariates). The responses are measured in % content and the black line in the plots refers to the 1:1 line. The ternary diagrams on the bottom show the soil texture classes. The original (measured data) classes are depicted in Figure 4.8d, and the classes resulting from predictions of soil particle size fractions are exhibited in 4.8e.

Table 4.3 Summary of the evaluation metrics resulted from the modelling approaches validated with LOOCV

MODELLING APPROACH (validation method : LOOCV)	SPSF*	MODEL EVALUATION METRICS		
		RMSE	MAE	NSE
Model 1 : Y-compositional LRM (ilr transf.) without 2-order interaction terms	Sand	4.12	3.37	0.94
	Silt	7.08	5.44	0.62
	Clay	7.00	5.02	0.46
Model 1 : Y-compositional LRM (ilr transf.) with 2-order interaction terms	Sand	4.88	3.35	0.91
	Silt	9.08	6.45	0.37
	Clay	8.93	6.16	0.13
Model 1 : Y-compositional RF (clr transformation)	Sand	4.28	2.91	0.93
	Silt	5.55	5.52	0.59
	Clay	7.13	5.27	0.44
Model 2 : Y-compositional LRM (ilr transf.) without 2-order interaction terms	Sand	15.21	12.04	0.11
	Silt	11.48	9.16	- 0.03
	Clay	9.37	7.00	0.08
Model 2 : Y-compositional LRM (ilr transf.) with 2-order interaction terms	Sand	23.05	17.86	-1.05
	Silt	16.33	12.83	-1.09
	Clay	14.84	10.61	-1.30
Model 2 : Y-compositional RF (clr transformation)	Sand	15.26	11.71	0.10

	Silt	11.26	8.62	0.01
	Clay	9.33	7.03	0.09
Non-CoDa approach (without log-ratio transformation) Simple LR (univariate responses)	Sand	7.10	5.35	0.81
	Silt	7.92	6.10	0.52
	Clay	7.36	5.33	0.41

Model 1: non-geophysical data; Model 2: with geophysical data; * SPSF: soil particle size fraction

4.6 Discussion

4.6.1 General aspects of soil particle size estimation with Sentinel-1 C-band data

We obtained different goodness of fit metrics depending on the soil fraction. This denotes the complexity of soil texture for statistical modelling due to the size magnitude of the soil fractions and the mineral constitution, as well as the reliance of the signal intensity from the SAR beam on how clayey or sandy the soil is. Typically, cross polarisation (VH) measurements are significantly lower than the co-polarisation (VV) (Ulaby *et al.*, 1978; Ulaby *et al.*, 1979; Sano *et al.*, 2020), and this was observed over our study area wherein the VV intensity values (in σ^0 units) ranged from 0.01 to 0.31, whereas for VH polarisation the coefficient backscatter varied between 0.02 and 0.07. It is worth noting that the Sentinel-1 dataset was acquired during a dry period (03/04/2021). This is important in terms of the amount of free water (e.g. in sand fractions) that can interact with the incident microwave, affecting the SAR backscatter (Das & Paul, 2015), since soil texture in the microwave spectrum is a function of the relation between free water and bound water (Jackson & Schmugge, 1989; Dobson *et al.*, 1981). The best prediction was obtained for the sand fraction, compared to clay, reported by the evaluation metrics.

We did not find a significant correlation between the radar backscatter coefficients and the soil particle-size fractions. Notwithstanding the log-transformations slightly improved the correlation coefficients, but the values remained low. Han *et al.* (2017) also found a weak correlation in a study investigating the association between radar backscatter (UAVSAR) and in situ soil property measurements on an approximately 3 km long section of earthen levees along the lower Mississippi river (riverside and landside). On the landside, the correlation coefficients found with HV polarisation were -0.16, 0.16 for clay and sand, respectively; and 0.16, -0.25 with VV polarisation. Such results would seem to confirm the fact that translating the microwave signal into a soil characteristic is not straightforward. While UAVSAR is an aerial Earth Observation (EO) platform, Sentinel 1 is an orbital-based EO platform (about 700 km altitude) and thus more distant from the targets. This means a loss of energy equal to the power of 4 (Wolff, 2007) according to the radar range equation (Figure 4.1).

4.6.2 Y-compositional LRM – ILR transformation

The Y-compositional LRM applied to the **Model Approach 1 (SAR + Topography covariates)** yielded significant predictions for sand (NSE = 0.94), silt (NSE = 0.62) and clay (NSE = 0.46) with lowest RMSE and MAE. It is worth noting that the model was fitted using a

multivariate approach, wherein the response variables were treated simultaneously. Whilst interaction terms decreased the goodness-of-fit metrics (e.g. NSE values equal to 0.91, 0.37, 0.13, for sand, silt, and clay, respectively) and increased both the RMSE and MAE, the soil texture classes obtained from the particle-size fractions estimations appeared to be more consistent with the texture classes derived from in situ data.

Notwithstanding the low values of the metrics obtained from **Model approach 2 (SAR + Topography + Geophysical covariates)** without interaction applied to the model, it raises some points regarding the relationship between the response variable and the predictors, as well as amongst the predictors. Results suggest that the geophysical-based variables reduced the prediction importance of both the radar-based and topography-based covariates by returning poor models. This was also observed with a more generalised- and elastic modelling in which any variable with a coefficient of zero was dropped from the model – See Appendix A: Models. Similar to Model approach 1 (LRM without geophysical data), interaction terms decreased the goodness-of-fit metrics and yielded more clayey classes. Furthermore, the inclusion of interaction terms resulted in soil textural classes that were more consistent with the textural classes derived from the observed data. This finding is likely due to the effect of soil texture on radiometric signals. Read *et al* (2018) found that gamma radiometric thorium (Th) and potassium (K) both had strong relationships with clay (%) and sand (%). Cattle *et al.* (2003) found that topsoils with strong clay content were distinguishable from sandy soils using radiometric Th and K. However, our findings require further exploration of geophysical data and soil PSF, since such covariates contributed to obtaining more consistent classes (soil texture) rather than estimates (soil PSF %) in the regression models with interactions.

4.6.3 Random Forest – CLR transformation applied

The Random Forest algorithm with the CLR transformation applied to the response variables also yielded good results in predicting soil particle-size fractions predictions with SAR-derived and topography-based covariates (**Model approach 1**). In general, the evaluation of the observed and predicted data is similar to that of the compositional LRM with ILR transformation (olr-coordinates) and without 2nd order interaction terms. When interactions between covariates was specified in the linear model, the RF model outperformed the compositional LRM, exhibiting lower RMSE and MAE and higher NSE.

Model approach 2 (**SAR+Topography+Geophysical data**) fitted with the Random Forest method and CLR transformation applied to the compositional responses resulted in poor predictions, based on the evaluation metrics. Despite the poor predictions, the mostly Loam class resulting from the soil fractions estimation appears to follow the pattern observed in the study area since fine loamy is the predominant soil texture class found in Ireland. Notwithstanding the low performance of the models and low importance magnitude of the geophysical covariates, the effect of the geophysical-based variables on the other covariates is clearly noticed in the Appendix

A- Figures A4.15 and A4.16, wherein the measure of variable importance was enhanced (e.g. altitude, slope, RVI, SSM) while others are weakened (e.g. curvature).

4.6.4 General aspects of the modelling approaches

In terms of soil dynamics, the results suggest that the models captured processes rather than genesis, with topography parameters, especially curvature, being the most important predictor. This variable may have reflected variations in topography in radar signals since the concavity, convexity, and flatness of the terrain represent the ability of a radar beam to bounce back to the sensor. Curvature may also have captured surface roughness conditions. In the context of the soilscape, concave surfaces are areas of accumulation and deposition of materials (e.g., water, silt, sand, clay), while convex surfaces tend to disperse materials downslope. To a certain extent, such an observation is consistent with the topsoil position on the landscape. As a first soil layer, it is more prone to undergo surface interactions than subsurface interactions. In relation to the topography covariates being identified as important predictors and machine learning models (e.g. RF) not performing well in some situations, these general findings are consistent with Mirzaeitalarposhti, *et al* (2022) and Schönbrodt-Stitt *et al* (2021) who also applied Sentinel-1 data and topography covariates to estimate soil particle-size fractions and soil moisture.

Notwithstanding their low importance, the VV and VH covariates (σ_{VV}^0 , σ_{VH}^0) were found to be related to sand, silt, and clay in the compositional LRM. Whilst we observed an effect of VH on the clay fraction in Model approach 1 – an increase in VV resulted in a decrease in the clay estimate; an increase in VH led to an increase in the clay estimate. We also found an effect of VV on sand, in Model approach 2 – an increase in VV led to an increase on sand.

Regarding this case where the compositional linear model slightly outperformed the tree-based model, this could be due to the small sample size used to fit the model. These results are not related to the parameter tuning of the RF model, since we tested different values for *ntree* and *mtry* and the model's performance did not change significantly (Appendix A-Tables A4.1 and A4.2). Moreover, the evaluation strategy did not play significant role in this case, since data splitting and leave-one-out cross-validation produced similar results. On the other hand, when the linear model were fitted with interaction terms, the RF approach outperformed the LRM. This is likely due to the fact that tree-based models consider variables sequentially, that is, they consider interactions without specifying them. Additionally, RF models do not rely on formal model assumptions such as linearity, normality, (multi)collinearity and homoscedasticity.

Results obtained with both the linear-based and tree-based models, indicated that geophysical data are not good predictors for estimating PSF in topsoil. This was also observed with a more generalised- and elastic modelling under different penalty coefficients – See Appendix A: Models. However, the use of geophysical covariates resulted in better estimates of the soil textural classes. Electromagnetic, magnetic and radiometric signals may be more related to deeper soil layers, due to their relationship with the underlying rock. Moreover, topsoil may be influenced by topography

conditions such as slope and curvature in terms of surficial materials deposits (geomorphological process). Our study region, as an extensive landscape, is relatively heterogeneous concerning these geomorphological variables. Interpretation of radiometric data where parent materials of transported sediments have different origins is also more challenging, which suggests a natural limit on soil textural interpretation of soil radiometric data (Read *et al.* 2018; Wilford *et al.*, 1997). Our findings highlight the need for further investigation of radiometric data and soil properties in our study region.

4.6.5 Strengths and limitations of the work

Our findings denoted how and to what extent interaction terms between variables work (explicitly specified) in a linear model and (underlying) in a tree-based model to estimate soil PSF in a large and physically heterogeneous area. Furthermore, they provide a potential basis and methodological approach to obtain either estimates of soil PSF (as continuous results) or soil textural classes (as categorical) based on two different statistical modelling and three different types of data.

Our work also demonstrated the potential relevance of the statistical modelling approaches for soil particle-size fractions mapping and soil textural mapping. Moreover, since our work is grounded in CoDa analysis, it provided some important information to Digital Soil Mapping as such techniques may not guarantee that the sum of the predicted clay, silt, and sand contents would be equal to 100% (Azizi *et al.*, 2023; Taghizadeh-Mehrjardi *et al.*, 2020).

Another highlight regarding the potential relevance of this work to soil properties mapping builds upon the fact that the models indicate soil particle-size estimation as more processes-based rather than genesis-based, as topography-based variables – especially curvature – were the most important covariate and geophysical-based data did not improve the models for topsoil.

We addressed different aspects of the prediction models in terms of the Sentinel-1 data products (e.g. SLC and GRD), the structure of the data (e.g. SAR-derived data, topography-based, geophysical-based), modelling approaches (e.g. linear-based and tree-based), and interactivity amongst covariates (e.g. models with- and without interactions). Such aspects brought robustness to the study.

Nevertheless, our work is not without limitations, which can be summarised as follows. Firstly, the lack of the dielectric constant of the soil, which could be useful to determining the relationship between soil particle-size fractions and backscattering coefficients, as well as to investigate whether only radar-based variables (σ_{VV}^0 and σ_{VH}^0) could be used as predictors. The intensity of the backscattering coefficient σ^0 (a normalised measure of the radar return) is a function of the physical and electrical properties of the target, along with the wavelength (λ), polarisation and incidence angle (θ) of the radar (Barret *et al.*, 2009). We expected that the geophysical electromagnetic data could fill this lack with electrical resistivity information, however, according to the results, the geophysical datasets did not improve the model.

Secondly, it could be challenging to select radar images that match the soil vegetated-covering conditions for all legacy topsoil datasets to address only bare soil samples in the modelling approaches. On the other hand, controlling different land cover conditions in this case could be more of a challenge-based issue than a limitation issue (Appendix A- Tables A4.5 and A4.6). There is a paucity of national soil information requiring the use of multiple different data sources. While not ideal, due to the difference in methods, we adopted a pragmatic approach to using the available data for application on a small geographical scale (large areas).

4.7 Conclusion

In this work, we tested two different statistical modelling approaches – linear models and tree-based models – to predict percentages of sand, silt and clay from Irish topsoil using microwave remote sensing data topographical and geophysical covariates, on different vegetation cover conditions, taking into account the compositional nature of soil texture.

Based on the results from the prediction models, it can be concluded that it was worthwhile treating soil PSF as compositional data in the models. We also found that radar-based variables were not able to predict soil particle size without the aid of covariates, since the models evaluated here did not identify their importance as covariates and no linear or direct relationships were found between the backscattering coefficients (σ_{VV}^0 , σ_{VH}^0) and the soil particle size fractions.

Based on the analysis outlined here, we highlight the following strategies to predict soil PSF from Sentinel 1-SAR:

- The linear model (compositional Y-LRM), without 2nd order interaction terms, was found to be the optimum approach; where it is necessary to consider interactions between variables, the RF approach should be employed.
- To obtain the soil textural classes from the estimates of PSF (sand %, silt %, clay %) – instead of predicting the classes using a classification method – the compositional Y-LRM with interaction terms applied to the covariates is useful as the classes were more consistent with the ternary diagram of the measured data than those obtained without interaction. This finding may be useful for subsequent use in models, which require soil textural class and not soil textural properties
- LOOCV is a better validation method over randomly splitting data for dealing with small sampling size (e.g. n=235).
- Employing geophysical data with SAR data did not improve the model estimates of soil particle fractions; the inclusion of geophysical covariates resulted in poor model estimates. However, the use of geophysical data was found to result in soil textural classes that better matched the textural classes derived from the measured data.

As part of the continued assessment of microwave remote sensing data (Sentinel-1) for predicting topsoil particle size fractions, we intend to test surface scattering information derived from the dual-polarimetric decomposition technique to deal with the vegetation cover and

integrate that predictor into the models. The goal is to separate different types of scattering mechanisms (i.e. ground or surface and volume or vegetation contributions) using the H- α dual-pol decomposition. This method yields surface scattering that will be used for predicting sand, silt and clay contents (%) since target decomposition theorems and Pol-SAR can be used to compensate for the presence of vegetation cover (Barret *et al.*, 2009). This procedure will also follow Jagdhuber (2012:4).

CHAPTER 5 | USING SURFACE SCATTERING MECHANISM FROM DUAL-POL SAR DATA TO PREDICT TOPSOIL PARTICLE SIZE FRACTIONS

This chapter is published as:

Deodoro, S.C., Moral, R. de A.; Fealy, R.M.; McCarthy, T., & Fealy, R. (2024). Using the surface scattering mechanism from dual-pol SAR data to estimate topsoil particle size fractions. *International Journal of Applied Earth Observation and Geoinformation*, 128, e10374, <https://doi.org/10.1016/j.jag.2024.103742>

5.1 Context

This Chapter is the second part of the methodology of this study following up on the results of the first methodology section. It draws upon a continued assessment of Sentinel-1 data for predicting topsoil particle size by employing a SAR polarimetric technique to deal with the effects of vegetation cover on topsoil backscattering and integrating the results as a predictor into the models. It sought to answer two key research questions: (i) *To what extent can polarimetric decomposition predict soil particle size fractions?* and, (ii) *Can surface scattering alleviate vegetation cover issues for soil particle size estimation?*

The rationale for using dual-pol decomposition for estimating soil PSF is based on the principle of separating individual scattering contributions between soil and vegetation under vegetated fields to extracting the soil properties (Jagdhuber, 2012). This aim relies on polarimetry to exploit the phase difference of wavelength from a returned radar beam (i.e. echo or backscattering), and on the aim of target decomposition to extracting information about the scattering properties of targets. This is achieved by partitioning the total power or intensity into relative contributions of different idealised scatterers (e.g. surface, volume, and double bounce), whose information is given by the resulting alpha, isotropy, and anisotropy parameters. The assumption behind this method is that there is always a dominant scattering mechanism – amongst surface, volume, and double bounce – in each pixel, which is typically provided by the alpha parameter. If there is no vegetation on an area or field, there is simply surface scattering which can be considerably lower in the cross-polarized channels compared to the co-polarized channel.

The choice of decomposition algorithm employed in this study (the H-alpha dual-pol) – is twofold. As per a dual-pol SAR, it is grounded on the 2 x 2 covariance matrix of complex numbers carried on the SLC Sentinel 1 products, representing the energy of VV and VH polarisation. Thus, it is a straightforward method. Second, the algorithm is available in the open-code software Sentinel 1- Toolbox SNAP.

Albeit motivated to deal with the challenge of vegetated soil, this part of the herein study does not intend to be exhaustive in examining the efficiency of the method to accurately distinguish targets, rather it seeks to explore and expand this method for soil PSF prediction going beyond the traditional employment of polarimetric decomposition technique for land use land cover classification, segmentation of landscapes, and vegetation studies. Thus, the exploratory

assessment of the methods carried out in this Chapter is a starting point for the development of other case studies to ameliorate predictions of soil particle size fractions and soil texture over low- and short vegetated areas.

Notwithstanding being published in a peer-reviewed journal for the remote sensing research community, the outreach of this study is also beneficial to the soil science audience.

5.1.1 Abstract

Data extracted from Synthetic Aperture Radar (SAR) have been widely employed to estimate soil properties. However, these studies are typically constrained to bare soil conditions, as soil information retrieval in vegetated areas remains challenging. Polarimetric decomposition has emerged as a potentially useful method to separate the scattering contributions of different targets (e.g. canopy/leaves and the underlying soil), which is of significance for areas that are near-permanently covered in low-lying vegetation (e.g. grass) like Ireland – the study area for this investigation. Here, we test the surface scattering mechanism, derived from H-alpha dual-pol decomposition, together with other covariates, to estimate percentages of sand, silt, and clay, over vegetated terrain, using Sentinel 1 data (dual-pol C-band SAR). The statistical modelling approaches evaluated – linear regression (LRM) and tree-based regression models (machine learning) – explicitly consider the compositional nature of soil texture. When compared to the models fitted without surface scattering data, results showed that the inclusion of the surface scattering data improved estimates of silt and clay, with the compositional linear regression model, and estimates of sand and silt fractions with different tree-based models. While not without limitations, our study demonstrated that the polarimetric decomposition method, which is typically used for classification and segmentation purposes, could also be used for soil property estimation, broadening the application of this technique in microwave remote sensing studies.

5.2 Introduction

Researching and developing novel methods for mapping soil properties contributes to advances in soil science. Geotechnologies, including remote sensing techniques, play an important role in this task, for example, by exploiting methods and ancillary variables for compositional data interpolation to improve model performance (Wang & Shi, 2017; Wang & Shi, 2018; Mondejar & Tongco, 2019). In turn, the importance of using remote sensing data and techniques to analyse and estimate soil properties relies on their use over large areas, where achieving similar coverage taking field measurements are costly and time-consuming, and may be limited by access to sites. Furthermore, optical- and microwave remote sensing have significantly contributed to the advancement of soil science across a range of different applications, such as the estimation of soil particle size-fractions (Gholizadeha *et al.*, 2018; Bousbih *et al.*, 2019; Domenech *et al.*, 2020; Marzahn & Meyer 2020; Azizi *et al.*, 2023), soil moisture (Jackson & Schmugge, 1989; Barret *et al.*, 2007; Barret *et al.*, 2009; Pratola *et al.*, 2014;

Petropoulos *et al.* 2015), soil carbon content (Gholizadeha *et al.*, 2018; Žižala *et al.*, 2019; Katebikord *et al.*, 2022), soil organic matter (van Wesemae *et al.* 2023; Yang *et al.*, 2023), soil salinity (Metternicht, & Zinck, 2003; Khajehzadeh *et al.*, 2022) and soil mineralogical composition (Coblinski *et al.*, 2021).

In the microwave remote sensing realm, synthetic aperture radar (SAR), such as Sentinel 1, can retrieve biophysical measures of vegetation and soil by the echoes from the Earth's surface through a radar antenna, measured as backscattering. In soils, the sensitivity of SAR relies on the interaction between the transmitted electromagnetic wave and the soil moisture and soil roughness (Jagdhuber, 2012). In the microwave region of the electromagnetic spectrum, natural materials have a dielectric constant ranging from 3-8 when dry, whereas water has a value of approximately 80 (ESA, 2007:25) (Appendix B – Table B5.1). As water has a dielectric constant 10 times that of dry soil, the presence of water in the top few centimetres of unvegetated soil (bare soil) can thereby be detected in radar imagery (ESA, 2007:25). In wet soil, the dielectric constant value is ~ 20 (Barret *et al.*, 2007).

Studies on the importance of soil properties, including soil texture and soil PSF, on the dielectric constant date back to the 1980s, conducted by Schmugge (1980), Wang and Schmugge (1980), Dobson *et al.* (1985), and Jackson (1987) who developed dielectric mixture models for soil. The dielectric constant is related with the bound water in soil surfaces which in turn is dependant on the total surface area of the soil available to the water molecules. Jackson (1987) demonstrated that a sand with a specific soil surface of $1 \text{ m}^2/\text{g}$ would have a much smaller bound water fraction than a clay with a value of $300 \text{ m}^2/\text{g}$, at a microwave frequency of 1.4 GHz (Appendix B – Figure B5.1).

The radar backscatter signal has a different relationship with the soil dielectric (moisture, and physical properties) and surface roughness. For example, Baghdadi *et al.* (2016) indicated that VV and VH intensities are more sensitive to surface roughness (i.e. k-root mean square height) than change in moisture, being more significant for higher incidence angles (45°) (Appendix B – Figure B5.2). Sentinel-1 with Interferometric Wide (IW) swath mode acquisition, which is the product employed in this study, has an incidence angle ranging from 29.1° to 46.0° (ESA, n.d.).

Soil texture is important for remote sensing (either optical or microwave), particularly considering soil mineralogy, since clayey minerals (hematite, goethite, kaolinite, chlorite, and illite to name a few) affect soil reflectance (Coblinski *et al.*, 2021), dielectric properties (permittivity) (Hallikainen *et al.*, 1985; Das & Paul, 2015; Babaeian *et al.*, 2019) and electrical conductivity (Fletcher, 2022). Clayey minerals influence the water dynamics in soil (Coblinski *et al.*, 2021) and both chemical and physical soil properties.

Regarding wave polarisations (e.g. VV, VH), polarimetry is an important concept in SAR applications, which relies on oscillations of the transmitted/received signals. The interaction of the transmitted pulse with the surface scatterers modifies the polarisation of the transmitted signal. Such modification depends on the geometry (size, shape, orientation) and the physical properties

(dielectric content, density) of the scatterers (Jagdhuber, 2012). In the context of soils, such differential propagation of microwaves causes interferometric effects that lead scattering elements on soil to produce phase and coherence variations (De Zan *et al.*, 2014). Polarimetry thereby is a potentially useful technique for retrieving soil properties as it provides information about the physical description of scattering processes (Jagdhuber, 2012). Such a technique builds upon interferometry as it uses the phase and amplitude information of the wave. In the soil context, the differential propagation of electromagnetic waves into the soil causes interferometric effects, meaning that small scattering elements in the soil or moisture gradients can produce phase and coherence variations (De Zan *et al.*, 2014). Due to the unique characteristic of SAR polarimetry, polarimetric decomposition is a common technique to separate scattering processes for use in segmentation and classification studies (Cloude and Pottier, 1996; Freeman and Durden, 1998; Ji and Wu, 2015; Abdel-Hamid *et al.*, 2018; Husman *et al.*, 2021).

To date, the investigation of soil parameters with SAR has largely been constrained to bare soil areas and for retrieving soil moisture or soil roughness (Bousbih *et al.* 2019; Marzahn & Meyer, 2020; Bhogapurapu *et al.*, 2022). Notwithstanding the importance of soil texture and PSF as a proxy for other soil properties estimates and various applications (e.g. agriculture, erosion, civil engineering, etc), the use of SAR to estimate these remain underexploited.

Separating individual scattering contributions between soil and vegetation under vegetated fields is a pivotal task to extracting the soil properties (Jagdhuber, 2012). This is of significance for areas that are near permanently covered in low-lying vegetation (e.g. grass) like Ireland. In areas covered with some vegetation, the SAR backscatter comprises of scattering from soil, vegetation, and the interaction of these two elements (Dou *et al.*, 2022). However, removing the contribution of vegetation to capture only the topsoil surface scattering is challenging (Wang *et al.*, 2017; Marzahn & Meyer, 2020; Bhogapurapu *et al.*, 2022). In this context, polarimetric decomposition emerges as a potential method, as it separates the scattering contributions of different targets (e.g. canopy and underlying soil) (Barret *et al.*, 2009; Jagdhuber, 2012; Jagdhuber *et al.*, 2013; Ji & Wu, 2015; Wang *et al.*, 2017; Mascolo *et al.*, 2021). In optical remote sensing, such technique is similar to the spectral mixing analysis or spectral unmixing methods used to decompose a reflectance or corrected radiance source spectrum into a set or a given endmember spectra or spectral signature, since not all pixel of a natural scene is pure, but a mixture (Somers *et al.*, 2011; Manolakis *et al.*, 2016; Halbgewachs *et al.*, 2022).

Polarimetric decomposition or target decomposition is the process of extracting information about the scattering process and describing scattering properties by partitioning the total power or intensity into relative contributions of different idealised scatterers. The H-alpha Pol Decomposition (H- α decomposition), otherwise known as Cloude-Pottier Decomposition, was developed by Cloude and Pottier (1997) to separate the scattering contributions of targets in radar images. Similar to other polarimetric decomposition methods, it converts the random scattering mechanism into several independent elements that can be related to the physical properties of the

target (Salma *et al.*, 2022; Cloude & Pottier, 1997), whose information is given by the resulting alpha, entropy, and anisotropy parameters. The assumption behind this method is that there is always a dominant scattering mechanism in each pixel or cell, which is typically provided by the alpha parameter. As a target decomposition, the resulting scattering mechanisms can be surface, volume or double-bounce types.

Surface scattering refers to the scattering that occurs on the surface between two different but homogeneous media (e.g. soil surface or topsoil, sea surface) and the direction of scattering depends on the degree of surface roughness. When electromagnetic radiation transmits from one medium to another, volume scattering occurs. Trees or branches, subsurface or deeper soil layers, snow layers, etc. are examples of volume scattering. The surface and the volume components come from the ground surface and the vegetation, respectively (Magagi *et al.*, 2022). In general, single scatterers will have lower alpha values being produced by a rough surface (i.e. seawater, snow surface, ploughed bare soil) which will increase as biomass increases. On a bare, smooth field, alpha will indicate a dominance of single bounce scattering and entropy will remain low. With increasing vegetation cover (e.g. dense vegetation, forests), the polarisation of the scattered wave becomes less predictable and entropy will increase.

While H-alpha Decomposition was originally developed for quad-polarised SAR, the method has been adapted for dual-polarised radar, such as Sentinel 1. Despite inherent limitations due to the reduced number of polarisations (i.e. two channels), dual-polarisation has produced reasonable results in different applications as outlined in the next section. Here, we test the potential importance of surface scattering, derived from a dual-polarimetric decomposition method (H-alpha decomposition), to estimate the content (%) of sand, silt, and clay from in situ topsoil samples from bare-soil and low vegetated cover employing a linear regression and a tree-based model with additional covariates. An intrinsic characteristic of soil particle size-fractions (PSF) – i.e. sand, silt, and clay – is their compositional nature (i.e. the sum of the fractions-components is equal to 100). This requires appropriate multivariate analysis that can handle compositional data in order to avoid the potential for spurious correlation or unrealistic results (Aitchison, 1982; Aitchison, 2005; Niang *et al.*, 2014; Pawlowsky-Glahn & Egozcue, 2015; Filzmoser *et al.* 2018; Todorov, 2021). Moreover, the sum of the components in a predicted soil sample (e.g. 100%) is not guaranteed with standard statistical methods that are designed for unconstrained multivariate data in a Euclidian space (Weiss *et al.* 2017; Amirian-Chakan *et al.*, 2019). We explicitly consider the compositional nature of soil texture in the modelling approach.

We sought to address two key research questions: (i) to what extent polarimetric decomposition with scattering partitioning can be employed to predict soil particle-size fractions, that is, to assess how well the Sentinel 1 pseudo-polarimetric product contributes to the estimates of soil PSF in comparison with the original dual-polarization product SLC-1; and, (ii) to evaluate if the surface scattering mechanism can ameliorate vegetation cover issues for soil particle size prediction on soils under low- and sparse vegetated cover.

It is worth noting that, due to the limited polarisation channels, this work does not intend to be exhaustive in examining the efficiency of the method to accurately distinguish targets, rather it seeks to explore and expand this method for soil PSF prediction according to the research questions. Thus, limitations and future works are highlighted based on our findings.

This paper is organised into the following sections. Section 5.3 briefly presents the theoretical background and some related works. Section 5.4 outlines the methodological procedure and datasets employed, including a description of the study area. Sections 5.5 and 5.6 present the results and discussion, respectively. Finally, Section 5.7 draws some conclusions from the work.

5.3 Theoretical background and related works

5.3.1 Radiometric calibration of SAR image

Prior to employing SAR data, the radar data must be calibrated, since this procedure provides a measure of radar reflectivity of a surface (i.e. the radar backscatter). Radiometric calibration corrects a SAR image so that the pixel values represent backscattering from the Earth surface, being a crucial procedure for a quantitative use of SAR data.

The radiometric corrections are purpose-specific. For interferometric and polarimetric analyses – the latter includes polarimetric decomposition – the output data need to be in complex number format (i.e. real and imaginary bands) to obtain coherence and covariance matrices. For other applications, radiometric calibration converts backscatter intensity as received by the sensor to the normalized radar cross section (sigma nought or σ^0).

Backscattering coefficients are reliant on the local incidence angle and local topography and related to characteristics of the targets, such as physical (geometric and electromagnetic), soil moisture, surface roughness, and dielectric constant (Imperatore & Di Martino, 2023). Abrupt land use land cover changes (e.g. from grass or bare soil to trees) may also affect backscattering coefficients when the backscatter and incidence angle behaviour varies over the image (Freeman, n.d.). The radiometric corrections are thereby also area-specific.

There are three different calibrations required for Sentinel-1 data, depending on the plane reference area used to normalise the backscatter (Figure 5.1). Sigma nought (σ^0) refers to the backscatter returned to the antenna from a unit area on the ground, being related to ground range (i.e. normalised to ground area) and based on ellipsoid (or incidence angle) corrected SAR backscatter. It is useful for flatter terrain and is commonly used to analyse surface scattering and surface properties (Barret *et al.*, 2007; Zhang *et al.*, 2017; Periasamny, 2018; Nasirzadehdizaji *et al.* 2019; Mandal *et al.*, 2020; Qu & Chen, 2020; Salma *et al.*, 2022). Beta nought (β^0) is normalised to a slant range plane area and refers to the radar brightness coefficient, which is the ratio between the power transmitted and received by the antenna. Gamma nought (γ^0) or RTC Gamma nought (Small, 2011) is normalised to an area perpendicular to the line of sight and accounts for terrain variations, advantageous over undulated terrain, mountainous and hilly areas.

An recent attempt to improve Sentinel-1 Radiometrically Terrain Corrected (RTC) to reduce processing and storage through orbital stability is outlined in Navacchi *et al.* (2023).

Thus, for a spaceborne SAR, the σ^0 and the γ^0 backscattering coefficients relies on a flattened earth model and a curved earth model, for the receive and transmit polarizations of the radar (Freeman, n.d.)

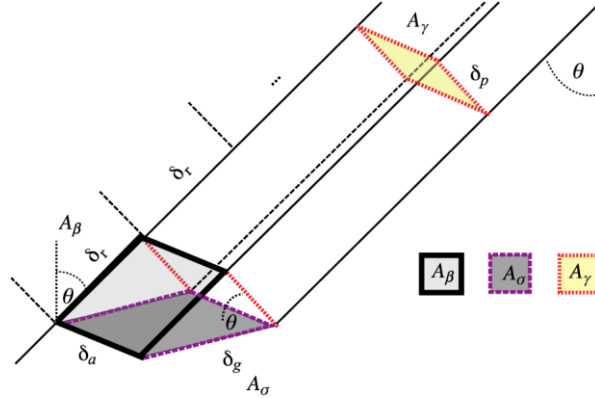


Figure 5.1 Normalization areas for SAR backscatter (Small, 2011). β^0 refers to reflectivity in an area in slant range and measures brightness in SAR images; σ^0 is illumination corrected and measures the power returned to antenna from ground; γ^0 is plane perpendicular to the local look direction and is radiometrically corrected being suitable for volume scatterers.

5.3.2 The H-alpha Dual Pol Decomposition

Polarimetric decomposition is the process of partitioning the total intensity into contributions of different scatterers to extract information about the scattering properties of targets. The H-alpha Pol Decomposition (H- α decomposition), otherwise known as Cloude-Pottier Decomposition, was developed by Cloude and Pottier (1997). Similar to other polarimetric decomposition methods, it converts the random scattering mechanism into several independent elements that can be related to the physical properties of the target (Cloude & Pottier, 1997; Salma *et al.*, 2022). One advantage of this method is the lack of physical constraints imposed by assumptions of a particular statistical distribution (Cloude and Pottier, 1997). The assumption behind this method is that there is always a dominant scattering mechanism in each pixel or cell. Based on this, Cloude and Pottier (1997) employed a three-symbol Bernoulli model to obtain estimates of the average parameters for the dominant target using POLSAR data (AIRSAR) from NASA's Jet Propulsion Laboratory (JPL).

Typically, there are five different target polarimetric statistical-based descriptors: Sinclair Matrix [S], Kennaugh Matrix [K], Target Vectors (k, Ω), Coherence Matrix [T], and Covariance Matrix [C]. The H-alpha Decomposition pertains to the coherence matrix group (Lee and Pottier, 2009), whose coherence matrix allows analysing single scattering mechanisms (i.e. surface scattering) within a pixel as well as their contribution to the total signal. The H-alpha Pol Decomposition method relies on the eigen decomposition of the coherency matrix [T3] into types of scattering processes (the eigenvectors) and their relative magnitudes (the eigenvalues) (Lee and Pottier, 2009). It is an eigenvectors-eigenvalues-based analysis (Equations 5.1 - 5.4) used to examine the intensity (coherency matrix) of the scattering matrix.

The method produces three secondary parameters – Entropy (H) (Equation 5.1), Anisotropy (A) (Equation 5.2) and Alpha angle (α) (Equation 5.3), which contain information on the scattering, from the eigen decomposition of the coherency matrix (Lee and Pottier, 2009).

$$H = \sum_{i=1}^3 -P_i \log_3(P_i), \quad P_i = \lambda_i / \sum \lambda_k \quad \text{Equation 5.1}$$

$$A = (\lambda_2 - \lambda_3) / (\lambda_2 + \lambda_3) \quad \text{Equation 5.2}$$

$$\alpha = - \sum_{i=1}^3 P_i \alpha_i \quad \text{Equation 5.3}$$

where λ_1 , λ_2 , and λ_3 are the eigenvalues of the coherency matrix ($\lambda_1 > \lambda_2 > \lambda_3 > 0$). From this, the corresponding eigenvectors $[\underline{u}_1, \underline{u}_2, \underline{u}_3]^*T$ are expressed as follows:

$$\underline{u}_i = [\cos \alpha_i, \sin \alpha_i \cos \beta_i e^{i\delta}, \sin \alpha_i \sin \beta_i e^{i\gamma}]^T \quad \text{Equation 5.4}$$

where α_i denotes the scattering mechanisms of a target; β_i is the orientation angles; δ , and γ represents the phases.

Since the method was adapted from a quad-pol SAR equation, the parameters and indices in equations 5.1 - 5.4 that refer to four polarisations are reduced to 2 to be suitable for use with two polarisations in a coherence matrix method, since both the scattering matrix and scattering vector of VV-VH or HH-HV SAR, for example, has two elements.

Entropy indicates how random the scattering mechanisms are and lies between 0-1 where H = 0 exhibits a pure target (single scattering element) and polarised surface (Harfenmeister *et al.*, 2021). If entropy is greater than zero, thereby, there is an indication of the presence of two different mechanisms, meaning that the scattering mechanisms are not surface, but a mixture of two different mechanisms. Anisotropy provides information on the relative effect of the second and third mechanisms and ranges between 0-1, being therefore complementary to entropy (Harfenmeister *et al.*, 2021), especially when $H > 0.7$ in full-pol SAR, since $H < 0.7$ the anisotropy is noisy and has limited importance (dos Santos *et al.*, 2009). The Alpha parameter is used to determine the dominant scattering mechanism and varies between 0° and 90° , particularly in full polarisation SAR. It is the main parameter (Cloude & Pottier, 1997) as it provides information related to the surface, volume (e.g. trees) and double-bounce scattering (e.g. urban areas).

Complementary to the polarimetric decomposition, there is the H-alpha plane plot that builds upon the H- α variation for a coherency matrix and consists of nine zones representing different scattering behaviour (Cloude & Pottier, 1997) (Appendix B – Figure B5.3). These zones result from the interplay between Alpha and Entropy. An example of this polarimetric decomposition method applied to an agricultural field is depicted in Appendix B – Figure 5.4a. It is worth noting that the H-alpha plane for dual-pol differ from full-pol in terms of the boundary lines and zone thresholds (Appendix B – Figure 5.4b) as demonstrated by Ji and Wu (2015). A summary of the

physical scattering characteristics of each of the nine zones in the H- α classification plane is found in Appendix B – Table B5.2.

Regarding applications of the method, Jagdhuber *et al.* (2013) underline the success of polarimetric decomposition to separate the vegetation contribution and estimate soil moisture on isolated bare ground components, enabling the interpretation and decomposition of different scattering contributions. The method has also been employed for bio-parameter extraction such as biomass (Suman, 2022), rice monitoring (Koppe *et al.*, 2013) and soil moisture (Magagi *et al.*, 2022). Qu *et al.* (2022) demonstrated the use of polarimetric decomposition from quad-pol SAR data for land cover classification.

In the context of dual-pol SAR, some limitations of the method have been reported in the literature. For example, the closer α is to 0° , the greater the chance of surface scattering. However, some studies demonstrated that it is not possible to distinguish pure surface scattering (i.e. $\alpha = 0^\circ$) from dihedral or double-bounce scattering ($\alpha = 90^\circ$) with S1 data due to reduced polarisation or lack of the remaining quad-pol (e.g. Engelbrecht *et al.*, 2017; Mascolo *et al.* 2021; Salma *et al.* 2022); that is, α values for the ground may not be as low as 0.0 . According to the dual-pol covariance matrices of the surface and dihedral targets, (i.e., $C_{\text{surface}} = [1, 0; 0, 0] = C_{\text{double-bounce}}$), it is not possible to distinguish them. As a result, the scattering mechanisms can not be separated. Cross-polar channel (VH or HV), which is sensitive to vegetation, becomes a poor estimate of volume scattering (Mascolo *et al.*, 2021), particularly under vertical dipole situations, hindering the separation of soil and vegetation. While the α values vary from 0° to 90° in quad-polarization, Salma *et al.* (2022) obtained a maximum α value of approximately 45° , in a study area situated in India, with Sentinel-1 dual-pol. In our study area located in Ireland (Figure 5.2), α values mostly ranged from 0.051 to approximately 39 , for the whole area – Appendix B – Figure B5.5.

Despite inherent limitations due to the reduced number of polarisations, dual-polarisation has been employed in a number of different applications. For example, Ji and Wu (2015) analysed the abilities of dual-polarisations (HH-VV, HH-HV, and HV-VV) from different SAR sensors to discriminate scattering mechanisms, namely surface, volume and double bounce scattering. Despite the lack of cross-polarization, they demonstrated that dual-pol HH-VV could effectively derive the scattering mechanisms; however, the HH-HV and HV-VV could only partially extract low, medium, and high entropy scattering mechanisms due to the lack of co-polarization. Their work builds upon theoretical and experimental results comparing performance with that of fully polarimetric SAR.

Mascolo *et al.* (2022) applied dual-pol decomposition to Sentinel-1 data over different land cover types, including agricultural, forest, urban, and glacial land-ice. They showed that some elements of physical modelling can be used to better exploit information from Sentinel 1 and that the dual-pol decompositions can be used to extract important underlying physical information, provided that the correct interpretation of parameters is used. Notwithstanding the benefits of this

method, they also considered the limitations associated with the use of dual polarisation SAR (e.g. the selective availability of different combinations of VH-VV and HH-HV and their impact on the interpretation of the decomposition outputs and identification of phenomena in an image). These authors also demonstrated the importance of the VH and VV backscattered signal for separating the land-cover types and for the estimation of important crop variables, based on biomass and phenology.

Salma *et al.* (2022) analysed target properties for a range of crop types, including ginger, tobacco, rice, cabbage, and pumpkin crops, using H-A- α dual-polarization decomposition with a temporal series of Sentinel 1 images (spanning 19th January and 27th September 2020). They found variation from the quad-polarisation decomposition, where the alpha values fell between 0° and 90°, whereas the alpha values of dual-polarisation data were found to vary between 7.92° and 45.23°. They also observed that the corresponding pixels on the ground exhibited surface scattering over the H- α plane. Finally, they found that the crop growth scattering mechanism on the H- α plane produced similar results to the temporal analysis.

The high frequency overpass of Sentinel 1 together with the sensitivity of polarimetric parameters (entropy, anisotropy, and alpha) to biophysical properties have increased the use and application of S1 for monitoring biophysical parameters over agricultural fields, as demonstrated by Dave *et al.* (2023) and Harfenmeister *et al.* (2021), and for crop classification (e.g. Husman *et al.*, 2021; Woźniak *et al.*, 2022). An attempt to improve interference between different scattering components in distributed scatterer employing Sentinel 1 is found in Zhang *et al.* (2022).

5.4 Data and methods

5.4.1 Study Region

The study area encompasses a central swath of the Republic of Ireland (Figure 5.2). Agriculture is the dominant land use land cover in Ireland, accounting for 67% of the land area, from which pasture is the main agricultural class (55%) (CORINE, 2018). Grass pasture is typically grazed by cattle on a rotation basis and/or harvested 2-3 times per year to provide a source of overwinter feed. Both management practices maintain a minimum grass height of 4 cm.

Topographically, the study area consist of uplands occurring in the middle-east of the domain with geoforms oriented at W-N, N-E and S-W, whereas the lowlands are mostly present in the mid-west with W-N, E-S and N-E faced geoforms (Figure 5.3). In terms of slope, mountainous areas are found in the northern and western areas, while flatter areas/lower relief are observed in the northern and southeast part of the domain. It is worth noting that the soil samples selected in the modelling stage of this work are mostly located in flatter terrain and lowland areas in the north, south and southeast areas of the study region.

The geomorphological setting affects the rainfall amounts which are typically higher on the west coast. This is due to the predominant maritime airflow off the North Atlantic Ocean that interacts with orography along that coast (Fealy *et al.*, 2011). While Ireland experiences year

round rainfall, the drier seasons are spring and summer while winter and autumn are the wettest, with approximately 260 mm and 350 mm of rain, respectively (Walsh, 2012). The predominant soil texture class in Ireland is Fine Loamy (Creamer *et al.*, 2016).

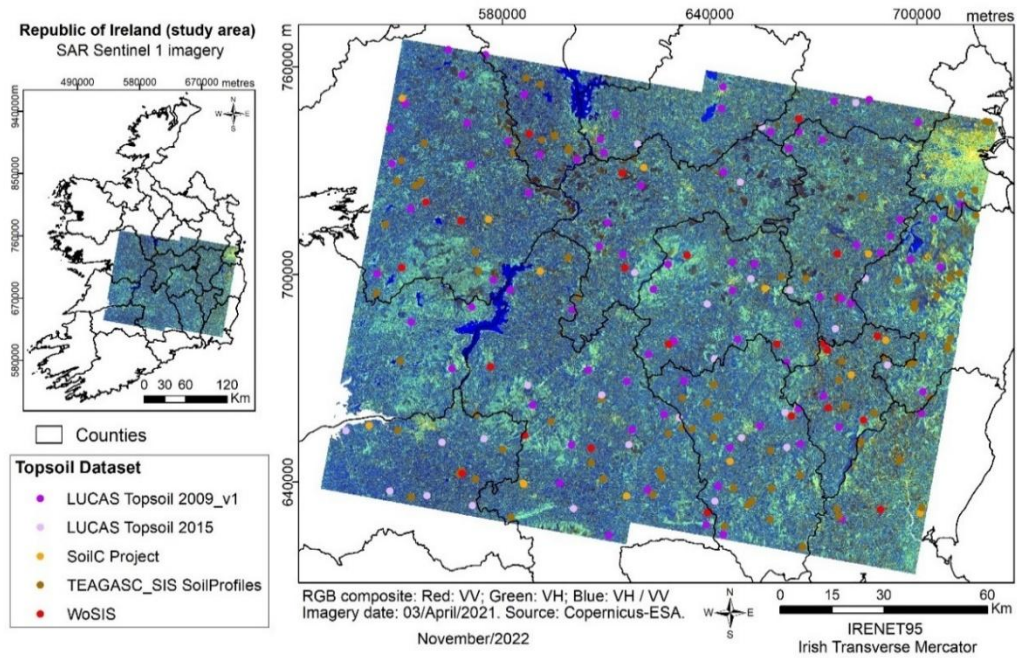


Figure 5.2 The land use-land cover categories from an RGB composite applied to the Sentinel 1 imagery over the study area. Topsoil sampled sites in the study area (points) overlaid on RGB composite using VV and VH polarisations from Sentinel 1-C band. The RGB resulting image highlights some of the land use/land cover categories, such as man-made structures and urban areas (yellow) within which Dublin city is highlighted on the top-right of the figure, water bodies (blue), and bogs/peatlands/bare soil (brown), dense and higher vegetation are displayed in light cyan, and grassland, crops and low vegetation cover are shown in medium and dark shades of cyan.

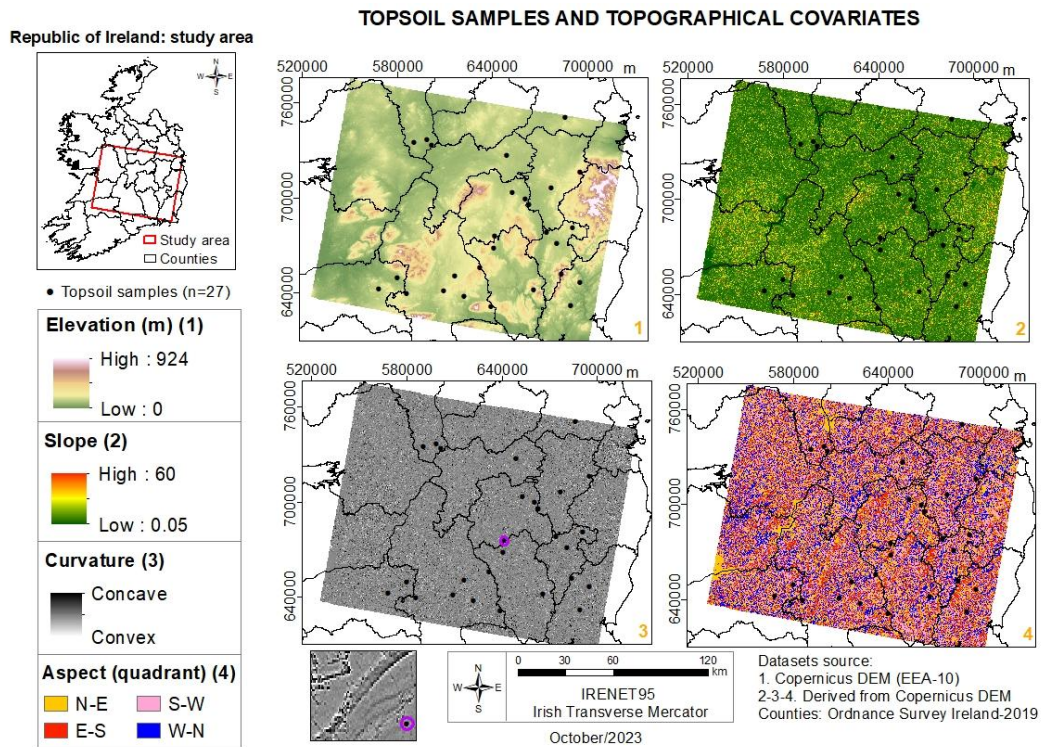


Figure 5.3 Topsoil samples superimposed over topographical (morphometric) covariates images of the study area. The soil samples (n=27) were employed in the models. Elevation data refers to the DEM-EEA-10 (10 m spatial resolution) from which the other covariates were derivate.

5.4.2 Data

5.4.2.1 Soil information

A dataset comprising 235 soil samples from across Ireland and collected by various institutions – Irish Soil Information System, LUCAS Topsoil 2009-2015, Soil Carbon (Kiely *et al.*, 2009), and WoSIS Soil Profile Database in both shallow and deep soil layers, was used to train and validate the models. Only the topmost layer or topsoil (0-15 cm in depth) was selected for use in this study in order to be consistent with the pulsed signals of Sentinel 1. This rationale was based on the relatively lower capability of C-band radar to penetrate into the soil – estimated to be approximately the first 4 cm of the soil layer (Babaeian *et al.* 2019) (Appendix B – Table B5.3) – compared to lower frequency-based sensors, such as L-band SAR. Our approach is consistent with Read *et al.* (2018), who dealt with different legacy databases to predict sand (%) and clay (%) in upper soil profiles from airborne geophysical data. In their work, they constrained soil samples collected from soil profiles up to 30 cm depth for training the model and restricted them to 0-5 cm (topsoil) in external model validation, since gamma rays emanate from topsoil layer (Read *et al.*, 2018).

5.4.2.2 Microwave remote sensing data

Sentinel-1 SAR (C-band, dual polarisation VV+VH, at a centre frequency of 5.405 GHz) was employed in this work as the radar-based data. It was obtained with Interferometric Wide (IW) swath mode acquisition, from the European Space Agency (ESA). Since polarimetric decomposition requires the phase information of the wave, the Single Look Complex (SLC) product was selected. The acquisition date was the 3rd April 2021. This date was selected due to the antecedent conditions (29/03/2021-03/04/2021) – which had no associated rainfall, based on information from the Irish national meteorological service, Met Éireann. Moreover, the rainfall recorded for the preceding month of March, was below the long-term average and a number of meteorological stations within the study domain recorded their driest March since 2012 (Met Éireann, 2021).

The pre-processing procedures applied to the Sentinel 1 imagery – included co-registration (S1-TOPS split as subset; orbit file applying), radiometric calibration to the complex number (i.e. a measure of radar reflectivity of a surface), S1-TOPS Deburst, S1-TOPS Merge, polarimetric speckle filtering using a Boxcar method with a 7×7 window size, and range-doppler terrain correction – were carried out using the Sentinel-1 Toolbox (SNAP-ESA). The complex output was selected for polarimetric processing. For calculation of the radar vegetation index, the backscattering coefficients were converted to sigma nought using the linear scale units. The Copernicus Digital Elevation Model for Europe (ESA EEA-10) at 10 m spatial resolution was used for terrain correction. All datasets were geo-coded to the Irish Transverse Mercator projection.

The radar-based data consisted of (i) VV and VH backscatter coefficients; (ii) the dual-pol radar vegetation index, (Equation 5.5) (Nasirzadehdizaji *et al.*, 2019; Gururaj *et al.*, 2019), and (iii) topsoil surface soil moisture (SSM) measured with Sentinel-1 as part of the Copernicus programme. All data was acquired for the same date. Typically, radar vegetation indexes are useful for measuring random scattering in radar signals and values close to 0 indicate smooth, bare surfaces (Nasirzadehdizaji *et al.*, 2019). The index employed here is an adaptation of the quad-pol RVI to dual-pol SAR and refers to the contribution of volume scattering given by cross-polarized response (Nasirzadehdizaji *et al.*, 2019). It was employed to filter noise from the data and to reduce the vegetation effect.

$$RVI = \frac{4\sigma_{VH}^0}{\sigma_{VV}^0 + \sigma_{VH}^0} \quad \text{Equation 5.5}$$

Sentinel 2 imagery (S2), a product of optical remote sensing, was also obtained for the same date in order to compare the surface scattering from the polarimetric decomposition with soil through an RGB composite (R4G3B2 and R11G3B2).

In order to estimate the effect of soil organic matter (SOM) on the backscatter, we performed a correlation test with SOC and the SAR backscatter. Results are presented in the Appendix B – Figure B5.6 and Appendix B – Table B5.4. From the soil legacy datasets, we used SOC data from the WoSIS Database, since it contains more measurements of this soil property

5.4.2.3 Topographical data

Topographical data, derived from the digital elevation model ESA EEA-10, was used to represent the soil-landscape relationship and an earth-surface characteristic related to soil-surface roughness (micro- and macro-relief of a soil surface), which is a result of variations in soil grains, soil aggregates, soil clods and tillage. In this study, topographic covariates consisted of altitude, slope, aspect and curvature (concave and convex surface), whose values were extracted from the corresponding pixels to the soil samples used in the modelling stage for predicting soil PSF. In the study area, values of altitude and slope for these points, ranged from 42–364 m and 0.16–11 degrees, respectively. For these sampling sites, slopes exhibit low-medium inclinations due to convex curvature and flatten summit, which justify the use of the σ^0 backscatter coefficient.

5.4.3 Methods

5.4.3.1 Dual Polarimetric decomposition

The covariance 2x2 matrix (C2) was generated from the Sentinel 1. The SLC product (Equation 5.6) was converted from the complex output to its covariance matrix representations (Equation 5.8), to perform the polarimetric decomposition. The matrix C2 (Equation 5.8) consists of two real parts (C11, C22), which contains the variances, and two complex parts (C12, C21) that contains the covariance (Salma *et al.* 2022).

The H-alpha Dual Pol Decomposition was further performed using the Sentinel-1 Toolbox available on the Sentinel application Platform (SNAP) in order to obtain the alpha, entropy and anisotropy values. Alpha, the main parameter (i.e. dominant scattering), was selected as a covariate in the statistical modelling as the other two parameters are highly correlated to alpha, especially entropy, with a Pearson's correlation coefficient (r) of 0.95. Thus, they were not considered in order to avoid multi-collinearity.

$$S = \begin{bmatrix} S_{VV} & S_{VH} \\ 0 & 0 \end{bmatrix} \quad \text{Equation 5.6}$$

$$k = [S_{VV} \quad 2S_{VH}]^T \quad \text{Equation 5.7}$$

$$C_2 = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \quad \text{Equation 5.8}$$

where S is the 2×2 complex scattering matrix or backscattering matrix (S_{VH} and S_{VV} are backscattering coefficients) (Qu & Chen, 2020); k is the corresponding scattering vector representation wherein T refers to the transpose of matrix; and C is the dual covariance matrix for S1-SLC (Qu & Chen, 2020).

In the (dual) polarisation matrix, the complex numbers C12 and C21 are conjugate, while C11 and C22 represent the energy of VV polarisation and VH polarisation, respectively. Thus, these complex numbers are displayed in an S1-SLC (Single Look Complex) file as a matrix.

Once the H-alpha Dual-Pol Decomposition was obtained, we carried out a visual inspection by overlaying, the alpha image and the Sentinel 2 image with the aid of RGB composite to identify features such as peat/bogs, water bodies, dense vegetation (forests), bare soil, ploughed soil, and soil with some degree of vegetation cover. Since we sought a distinction between soil and vegetation, only these features are indicated in the analysis.

After the visual inspection, the in situ soil sample locations were overlain on the resulting maps from the H-alpha Dual Pol Decomposition (alpha, entropy and anisotropy images). Those locations with low alpha or surface scattering were selected (blue patches in the alpha image) to extract the pixel values, following Harfenmeister *et al.* (2021), Mascolo *et al.* (2021) and Salma *et al.* (2022). Typically, in agricultural areas, alpha=0° indicates a single scattering event (surface) from smooth soil or a large leaf; alpha=45° indicates more than two scattering events from within a crop canopy (volume).

In our study area, the threshold for alpha and entropy were < 18° (low) and <0.6 (medium), respectively (Mascolo *et al.*, 2021). The closer alpha is to 0°, the more the surface scattering occurrence. However, there is limitation issues in distinguishing pure surface scattering from dihedral or double-bounce scattering with S1, resulting in alpha values > 0.0 for bare ground. In light of such limitations, the aforementioned threshold was set based on Mascolo *et al.* (2021) considering a range values of $0^\circ < \alpha_{VHV} < 22.5^\circ$ and horizontal dipole (Mascolo *et al.*, 2021).

From this, the surface scattering contribution is expected due to either the spacing between plants or direct scattering from the large leaves (Mascolo *et al.*, 2021). Surface scattering can come from vegetation, which was confirmed further from the analysis of the H-alpha plots in our study. Thus, to account for the gaps between plants in some agricultural fields, the threshold were increased from zero to include backscattering from low- and sparse- vegetation. Due to the agricultural field setting, it is difficult to have pure targets (low entropy) values that clearly distinguish them from their surroundings. Thus, they are more distributed targets and fall into the medium range for entropy. To consider the vegetation effect, that is, to reduce the effect of it on the soil backscattering, we employed the dual-pol RVI, which relies on sigma naught backscattering in its formula, in the model.

With this procedure, 27 points were ultimately selected, consisting of bare soil and low or sparse vegetation. Spectral signatures from Sentinel 2 images (obtained from 03/04/2021) were extracted for selected field locations in the study area to compare the behaviour of some features (e.g. soil, vegetation) with the decomposed targets by the H-alpha dual-pol decomposition. This is based on the fact that when there are small contributions from soil backscattering, the vegetation signature is highest in volume scattering. (Guerriero *et al.*, 2013; Salma *et al.*, 2022).

5.4.3.2 Statistical modelling

Tree-based models are alternatives to linear regression models due to their advantages of being relatively robust to overfitting and not requiring standardisation or normalisation of data; they are insensitive to the ranges in the predictor values (Zhang & Shi, 2019). Decision trees are used in classification and regression problems whose predictor space is divided up by recursive partitioning. A widely used extension to decision trees is the random forests (RF) algorithm (Breiman, 2001). Some examples of RF applications to estimate soil properties – without considering soil texture as compositional data – in a remote sensing context are found in Mirzaeitalarposhti *et al.* (2022), Domenech *et al.* (2020), Dotto *et al.* (2020), Cisty *et al.* (2019), Bousbih *et al.* (2019) and Ballabio *et al.* (2016).

Compositional data (CoDa) analysis is mostly performed in terms of log-transformations and relies on log-ratios between the parts or components of one sample. Some of the benefits of applying the log-ratio (linear) transformations are the possibility of using standard unconstrained multivariate analysis on the transformed data. Moreover, the sum of the components equal to 100 (or 1) is guaranteed when the predicted values are back-transformed to the original units (Comas-Cufí & Thió-Henestrosa, 2011). The theoretical foundations are found in Aitchison (1982). As highlighted by Morais and Thomas-Agnan (2021), in simplicial regression (i.e. compositional linear model; Y-LRM) one can use transformations to transport the simplex space S^D into the Euclidean space R^{D-1} to eliminate the simplex constraints problem.

Considering the compositional nature of soil texture (a composition of particle-size fractions), log-ratio transformation methods rooted in symmetry – centred log-ratio

transformation (CLR), and isometric log-ratio transformation (ILR) – were employed in this study because they preserve distances. The ILR transformation was applied to the response variables or Y (sand, silt, and clay) for fitting the compositional Linear Regression Model (Y-LRM) following the default partition method in the CoDaPack software (Comas-Cufí & Thió-Henestrosa, 2011). Due to the sampling size, the model was validated by using the leave-one-out cross validation (LOOCV) method.

The CLR transformation was applied to the response variables to train a tree-based model for different extensions of tree-based algorithms, including, the traditional Random Forests-RF (Breiman, 2001), Local Linear Forests-LLF (Friedberg *et al.*, 2021), Bayesian Additive Regression Trees-BART (Chipman *et al.*, 2010), and SoftBART-SBART (Linero & Yang, 2018). CLR transformation was applied to the response variables prior to model training, since the machine learning modelling does not rely on distributional assumptions. Moreover, the singular covariance matrix resulting from the CLR-transformed data set is not an issue for the Random Forest model. We used the default method built into the CoDaPack software (Comas-Cufí & Thió-Henestrosa, 2011). For comparison with the CoDa approach, a non-CoDa-based regression was carried out for these different tree-based algorithms.

The compositional data approach has increasingly been employed to estimate soil PSF in both non-spatial (Loosvelt *et al.*, 2013; Chappell *et al.* 2019; Deodoro *et al.*, 2023) and explicitly spatial models (Zong & Shi 2017; Zong & Shi, 2018; Zhang & Shi, 2019; Wang & Shi, 2018; Wang & Shi, 2017; Odeh *et al.*, 2003). Here, we employed CoDa in order to satisfy the constraint where the sum of the parts is 100. We employed two different statistical modelling approaches to see the results behaviour with a assumption-based model (LRM) and non-assumption or black-box model (tree-based model).

It is worth noting that, in addition to soil PSF estimates, the corresponding soil texture classes were also obtained. However, they were inferred from the compositional predictions of sand, silt, and clay content rather than directly using a categorical modelling approach. The resultant soil textural classes will also be outlined in the Results section.

5.4.3.3 Evaluation metrics for soil particle size predictions

The performance of the modelling approaches were evaluated using three statistical indicators, as follows: Root Mean Square Error (RMSE) (Equation 5.9), MAE (Mean Absolute Error) (Equation 5.10) and Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970) (Equation 5.11). The latter is a normalised metric that indicates the magnitude of the statistical variances between the residuals and the measured data (Nash and Sutcliffe, 1970). It measures the agreement between observed versus predicted (simulated) data in relation to the 1:1 line according to the following: $NSE = 1$ indicates a great fit of the model to the observed data; $NSE = 0$ indicates that the accuracy of predictions is given by the mean of the observed data; $-\infty < NSE < 0$ shows that the observed mean is predicting better than the model.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad \text{Equation 5.9}$$

where y_i is the observed value of the response variable, \hat{y}_i is the estimated value, and n is the number of observations.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad \text{Equation 5.10}$$

where y_i , \hat{y}_i , and n represents the same descriptors that those of RMSE.

$$NSE = 1 - \left[\frac{\sum_{i=1}^n (Y_i^{obs} - Y_i^{sim})^2}{\sum_{i=1}^n (Y_i^{obs} - \overline{Y^{obs}})^2} \right] \quad \text{Equation 5.11}$$

where Y_i^{obs} is the i^{th} observation, Y_i^{sim} is the i^{th} simulated (predicted) value, $\overline{Y^{obs}}$ is the mean of observed data, and n is the total number of observations.

5.5 Results

The calculated Alpha values (Figure 5.4) indicate that the study area is mostly represented by volume scattering mechanisms (yellow patches) indicating occurrence of vegetation, consistent with the land cover conditions (Figure 5.5). For the entire study area, Alpha values varied from 0.13 to 73.3 (Figure 5.4). For the selected soil samples ($n=27$), corresponding to the surface scattering patches, the obtained alpha values ranged from 5.43 to 18.63, with a mean equal to 13.26 (Table 5.1). Similarly, the entropy values, indicating surface scattering, produced values ranging between 0.19 and 0.66, with a mean equal to 0.53. Alpha and Entropy followed a negative, left-tailed distribution, whereas Anisotropy exhibited a distinct positive, right-tailed distribution.

Table 5.1 Descriptive statistics for the dual-pol decomposition parameters retrieved for the soil samples (N=27)

Parameter	Min.	Max.	Mean	Standard Dev.
Alpha Surface Scattering	5.43	18.63	13.26	3.16
Entropy Surface Scattering	0.19	0.66	0.53	0.11
Anisotropy Surface Scattering	0.66	0.94	0.75	0.07

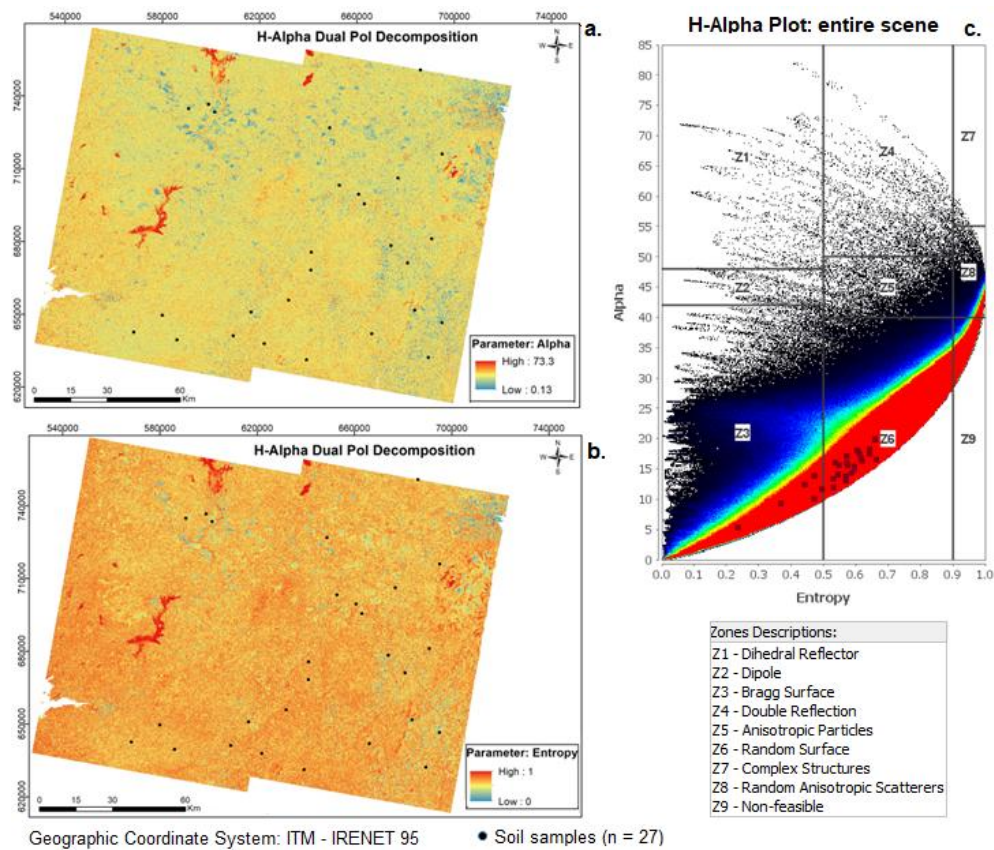


Figure 5.4 Alpha and entropy maps of the entire scene (figures a, b) as retrieved from Sentinel 1 data using the dual H-alpha pol-decomposition. The corresponding H-Alpha plot for the entire scene (figure c). Square dots represent the soil samples selected (n=27). Blue: surface scattering; yellow: volume scattering; red: random and bragg scattering.

The corresponding H-Alpha plot for the study area (Figure 5.4a) shows the contribution of soils and vegetation on radar backscattering. From Tables 5.1 and 5.2, the sample locations (n=27) were consistent with the low entropy surface scattering ($\text{Alpha} \leq 40$ and $H \leq 0.6$). From the H-Alpha plot, the selected points (n=27) typically lay within Z3 and Z6 (Figure 5.4c), defined as bragg and random surface scattering regions, respectively. Points located in Z3 and those located in Z6 in transition to Z3 indicate sparse or very low vegetation cover and soil.

Table 5.2 The typical scattering mechanisms in an H- α feature space

Zone	Colour	Scattering mechanism	Boundaries
3	Red	Low entropy surface scattering	$\text{Alpha} \leq 40$ and $H \leq 0.6$
2	Blue	Low entropy dipole scattering	$40 < \text{Alpha} \leq 46$ and $H \leq 0.6$
1	Orange	Low entropy multiple scattering	$\text{Alpha} > 46$ and $H \leq 0.6$
6	Green	Medium entropy surface scattering	$\text{Alpha} \leq 34$ and $0.6 < H \leq 0.95$
5	Black	Medium entropy vegetation scattering	$34 < \text{Alpha} < 46$ and $0.6 < H \leq 0.95$
4	Cyan	Medium entropy multiple scattering	$\text{Alpha} \geq 46$ and $0.6 < H \leq 0.95$
8	Magenta	High entropy vegetation scattering	$34 < \text{Alpha} < 46$ and $H > 0.95$
7	Yellow	High entropy multiple scattering	$\text{Alpha} > 46$ and $H > 0.95$
9		High entropy surface scattering (non-feasible)	$\text{Alpha} \leq 34$ and $H > 0.95$

Source (Heras, 2015). Adapted by the authors - the zone numbers were adapted to match the H- α plane plot for SAR Sentinel 1.

The alpha values from a subset of the entire study area were visually compared to the Sentinel 2 (optical imagery) RGB composite (R4G3B2) (Figure 5.5). Broadly, blue in the alpha parameter refers to surface scattering; yellow is commonly related to volume scattering; and red colour indicates a double-bounce scattering. This is also consistent with the spectral signatures obtained from Sentinel 2 (Figure 5.5, bottom) of the soil sampling sites depicted in the S2-RGB composite. The spectral curves shows the signature of soil (red) and vegetation (green) for sampling sites under bare soil and vegetated soil, respectively.

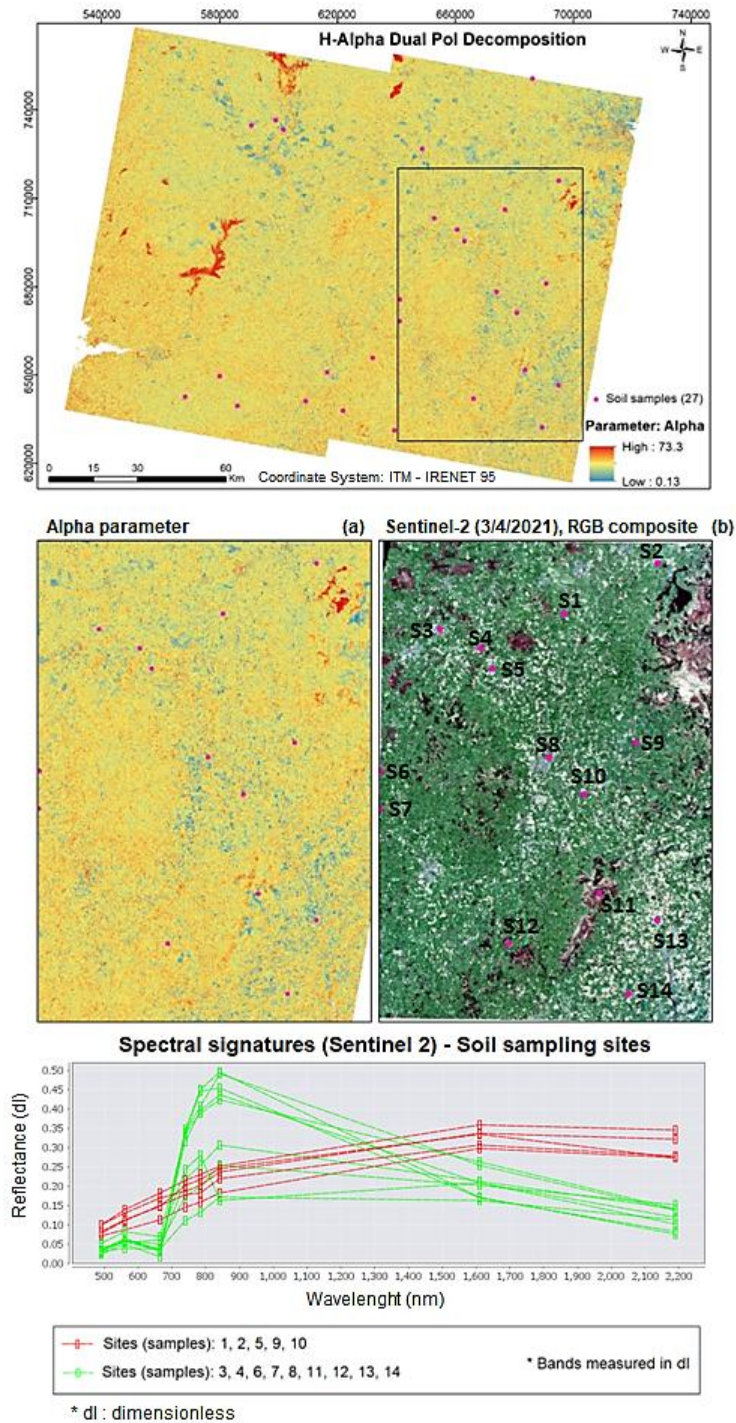


Figure 5.5 A subset of Alpha of the entire scene and a comparison to the RGB composite (R4G3B2) Sentinel 2 (optical imagery) showing a correspondence between the two images (a, b). The square box in the entire scene highlights the location of the subset. Blue colour in figure “a” (surface scattering) depicts bare soil and arable lands (with very low-

or more-sparse vegetation), while yellow colours (volume scattering) correspond to vegetated areas as depicted in figure “b”. A spectral plot (bottom) was obtained from Sentinel 2 imagery for the selected soil sampling sites (14 points), which depicts spectral signatures of vegetation (green lines) and bare soil (red lines).

A more detailed subset of alpha and Sentinel 2-RGB composite (Figures 5.6 and 5.7) illustrates the response under different conditions of vegetation cover, on which surface scattering is depicted in areas with bare soil and very-low vegetation indicating that the decomposition method retrieved soil backscattering under that degree of vegetation cover. These figures were useful to demonstrate the method in a surrounding area of the selected points shown in Figure 5.5. They have a visual purpose only, and examples from these were not included in the models.

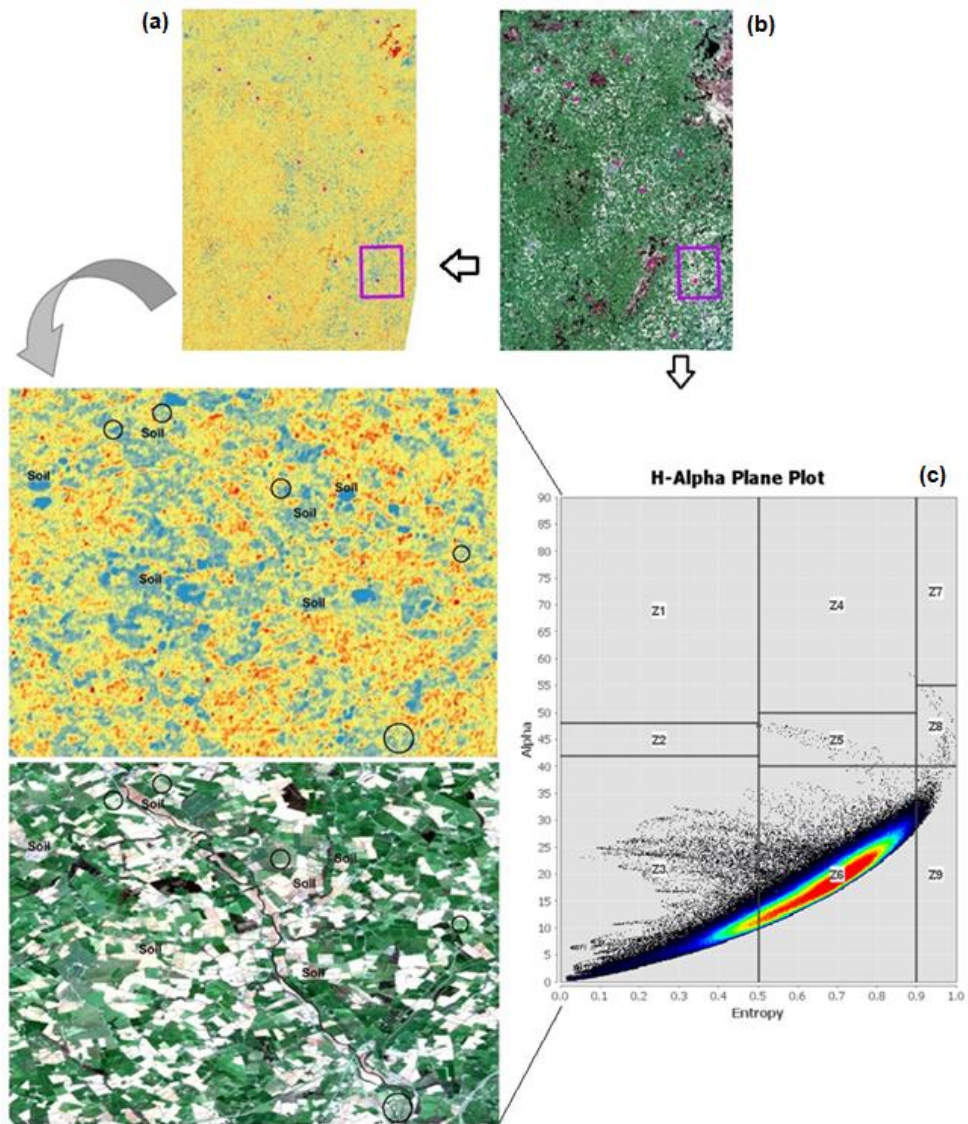


Figure 5.6 A detailed subset of Alpha (a) and Sentinel 2-RGB composite (b) with the corresponding H-alpha plane plot for the subset scene (c). The figure illustrates soils under different conditions of vegetation cover. Circles on the figures were only drawn to highlight low vegetation amongst soil.

Bare soil and low vegetation (i.e., grass) were associated with surface scattering when some pixels samples were extracted from the Alpha image (S1) and the corresponding RGB composite from S2 (Figure 5.7), which could be confirmed by the H- α classification regions in the H- α plane

plot. This is also consistent with the spectral signatures (Figure 5.7) of a representative pixels within each polygon traced over the alpha and RGB images. The red and yellow spectral curves depict bare soil in fields (polygons) 3 and 5, respectively; green and pink curves indicate some degree of vegetation in fields 2 and 4 where soil occurs in the gaps between plants; and grey and blue spectral curves indicating the influence of dried crop residues (e.g. straw) from previous harvest in fields 1 and 6. It also worth noting that moisture content and surface roughness affect the soil spectrum, particularly in the visible spectrum (< 0.750 nm).

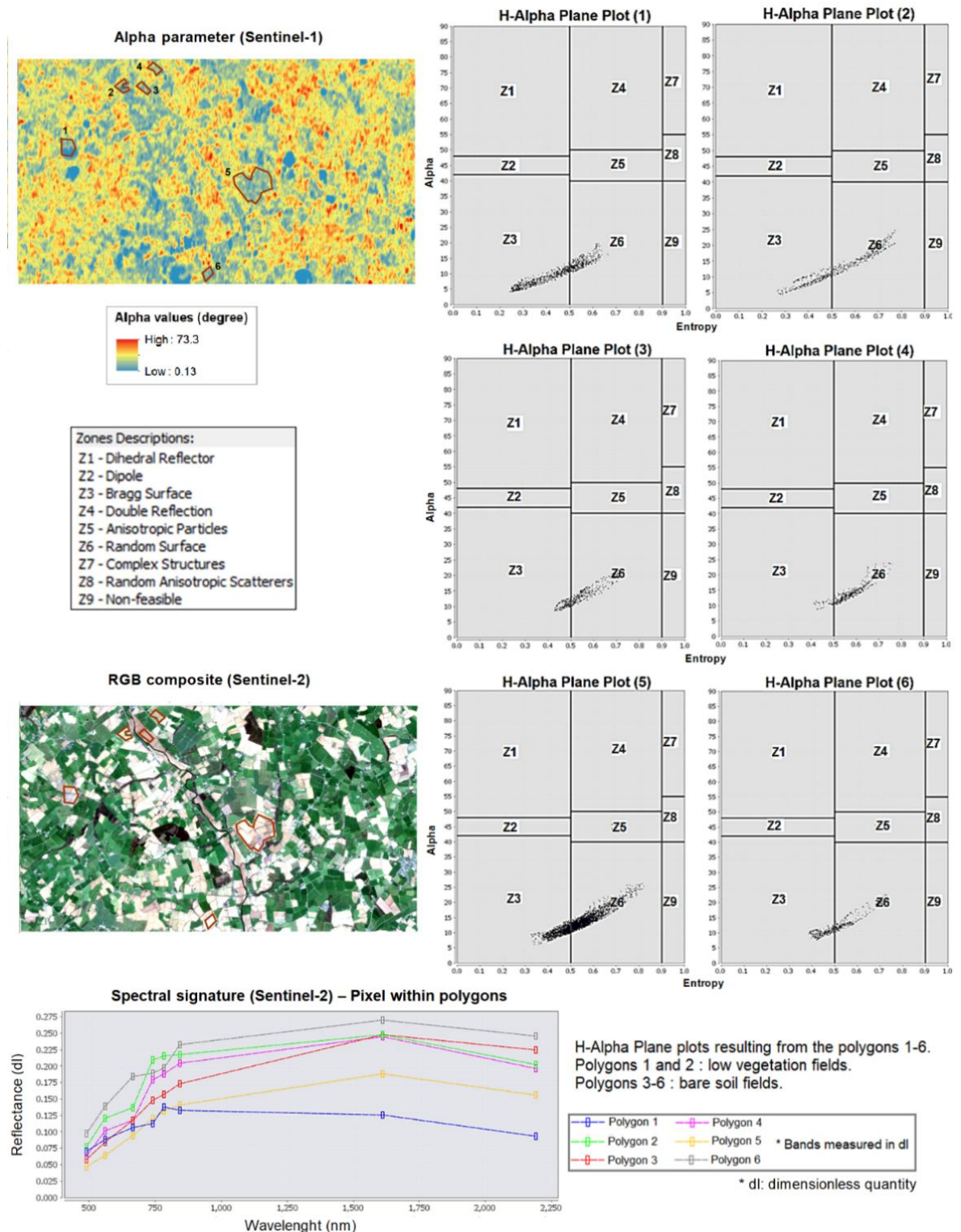


Figure 5.7 H- α plane plots (right) of samples selected over bare soil and low vegetation (left) in the study area. Points (samples) located in Z3 and Z6 in the plots indicate surface scattering mechanisms as brag surface and random surface types, respectively, with low and medium entropy, respectively. In agricultural fields, brag surface may be associated with fallows while random surface could be associated with ploughed fields and fields with short grass (Anup Das, n.d: pp 27, 70, 77). The spectral plot on the bottom shows the spectral signature of a representative pixel within each polygon.

5.5.1 Statistical models

The compositional LRM (Y-LRM) fitted with surface scattering as a covariate yielded an overall R^2 value of 83%. In the model without the surface scattering variable, the overall R^2 value was 78.6%. The NSE values (1:1 line) between observed and predicted samples indicated a improvement for silt and clay (Table 5.3). Fraction clay estimates showed the lowest agreement with the measured data in both modelling approaches (Figures 5.8a and 5.8b). Regarding soil texture classes, the corresponding classes from the soil particle-size fractions predicted were Sand (S), Sandy Loam (SL), Loam (L), Clay Loam (CL), and Silt Loam (SIL) – Figure 5.9b. The model fitted without surface scattering, resulted in more clayey classes, such as Clay (C) and Silty Clay Loam (SICL). Sand class (S) was not observed with this model.

In the tree-based models fitted with CLR transformation applied to the response variables and validated with LOOCV, the inclusion of surface scattering as a covariate improved the estimation of soil particle-size fractions. For example, the error reduced for all PSF in all models and the agreement increased, except for clay, in the RF, BART and SBART models (Table 5.3). The clay fraction estimates showed the lowest agreement with the measured data in all models. The sand fraction was highlighted as the optimum response variable and the LLF and SBART models exhibited the best performance (Figure 5.8c). All the tree-based models yielded the same soil textural classes (Figure 5.9c-5.9f).

For the tree-based models without log-ratio transformation (non- CoDa approach), the inclusion of surface scattering as a covariate also improved the estimation of soil particle-size fractions (Table 5.3). In general, the models demonstrated better performance, from the NSE, RMSE and MAE metrics, between the observed and test samples, except for clay.

Table 5.3 Model evaluation metrics on test data for models validated through LOOCV method. (N=27)

	MODELS WITH SURFACE SCATTERING			MODELS WITHOUT SURFACE SCATTERING		
	Y-LRM (ILR transformation)			Y-LRM (ILR transformation)		
SPSF*	RMSE	MAE	NSE	RMSE	MAE	NSE
Sand	6.4	4.9	0.86	6.1	4.2	0.86
Silt	11.2	7.9	0.17	12.5	8.8	-0.33
Clay	8.0	6.0	-0.01	15.5	9.9	-1.16
	RF (CLR transformation)			RF (CLR transformation)		
SPSF*	RMSE	MAE	NSE	RMSE	MAE	NSE
Sand	10.06	7.41	0.65	10.64	7.50	0.58
Silt	8.21	6.40	0.56	8.52	6.74	0.38
Clay	7.64	5.21	0.08	9.29	7.31	0.22
	BART (CLR transformation)			BART (CLR transformation)		
SPSF*	RMSE	MAE	NSE	RMSE	MAE	NSE
Sand	6.83	4.83	0.84	7.51	5.27	0.79
Silt	7.29	5.81	0.65	7.95	6.35	0.46
Clay	7.03	4.87	0.22	9.11	7.09	0.25
SPSF*	SBART (CLR transformation)			SBART (CLR transformation)		

	RMSE	MAE	NSE	RMSE	MAE	NSE
Sand	3.51	2.45	0.96	4.70	2.65	0.92
Silt	7.41	5.54	0.64	6.90	5.64	0.59
Clay	7.80	5.49	0.05	8.68	6.69	0.32
SPSF*	LLF model (CLR transformation)			LLF model (CLR transformation)		
	RMSE	MAE	NSE	RMSE	MAE	NSE
Sand	3.84	2.97	0.95	3.70	2.57	0.95
Silt	6.61	5.56	0.71	7.10	5.53	0.57
Clay	5.86	4.16	0.42	8.36	6.33	0.37
SPSF*	RF (Non-CoDa approach**)			RF (Non-CoDa approach**)		
	RMSE	MAE	NSE	RMSE	MAE	NSE
Sand	10.84	8.03	0.60	10.75	7.48	0.57
Silt	8.97	7.01	0.47	8.48	6.68	0.38
Clay	7.61	5.35	0.09	9.29	7.45	0.22
SPSF*	BART (Non-CoDa approach**)			BART (Non-CoDa approach**)		
	RMSE	MAE	NSE	RMSE	MAE	NSE
Sand	6.99	5.00	0.84	8.26	5.76	0.75
Silt	7.25	5.79	0.67	8.18	6.20	0.43
Clay	7.31	5.27	0.17	9.30	7.31	0.22
SPSF*	SBART (Non-CoDa approach**)			SBART (Non-CoDa approach**)		
	RMSE	MAE	NSE	RMSE	MAE	NSE
Sand	3.27	1.72	0.96	4.31	2.79	0.93
Silt	6.68	5.07	0.71	6.90	5.56	0.59
Clay	7.26	5.26	0.17	9.24	7.12	0.23
SPSF*	LLF model (Non-CoDa approach**)			LLF model (Non-CoDa approach**)		
	RMSE	MAE	NSE	RMSE	MAE	NSE
Sand	4.06	3.16	0.94	4.26	3.65	0.93
Silt	6.51	5.47	0.72	7.59	5.92	0.5
Clay	5.97	4.43	0.44	8.01	6.44	0.42

RF: Random Forest; **GRF**: Generalized Random Forests; **LLF**: Local Linear Forest; **BART**: Bayesian Additive Regression Tree; **SBART**: SoftBART. Hyperparameters for each tree-based method: default.

*Soil Particle size Fraction. ** Without log-ratio transformation.

In an assessment of Sentinel-1, geophysical and topographical covariates for estimating topsoil particle-size fractions over the same study area, Deodoro *et al.* (2023) did not find a significant correlation between the radar backscatter coefficients (σ_{VV}^0 , σ_{VH}^0) and the topsoil particle-size fractions for either the non-transformed or transformed soil PSF. Han *et al.* (2017) also found a weak correlation between radar backscatter (UAVSAR) and in situ soil PSF along the lower Mississippi river (riverside and landside). Such results indicate that soil PSF cannot be estimated directly from SAR data solely employing VV and VH backscatter values, that is, without covariates. We observed higher values for the association between SAR backscattering and SOC, however, correlations were weak with values ranging from -0.25 to 0.33 and not statistically significant (Appendix B – Table B5.4). A possible explanation for these results, is that soil samples usually go through treatment or sample pre-treatment analytical procedures in laboratory to remove cementing materials such as organic matter or “humus”, calcium carbonate and iron oxides, to obtain the particle-size fractions (Ribeiro *et al.*, 2020).

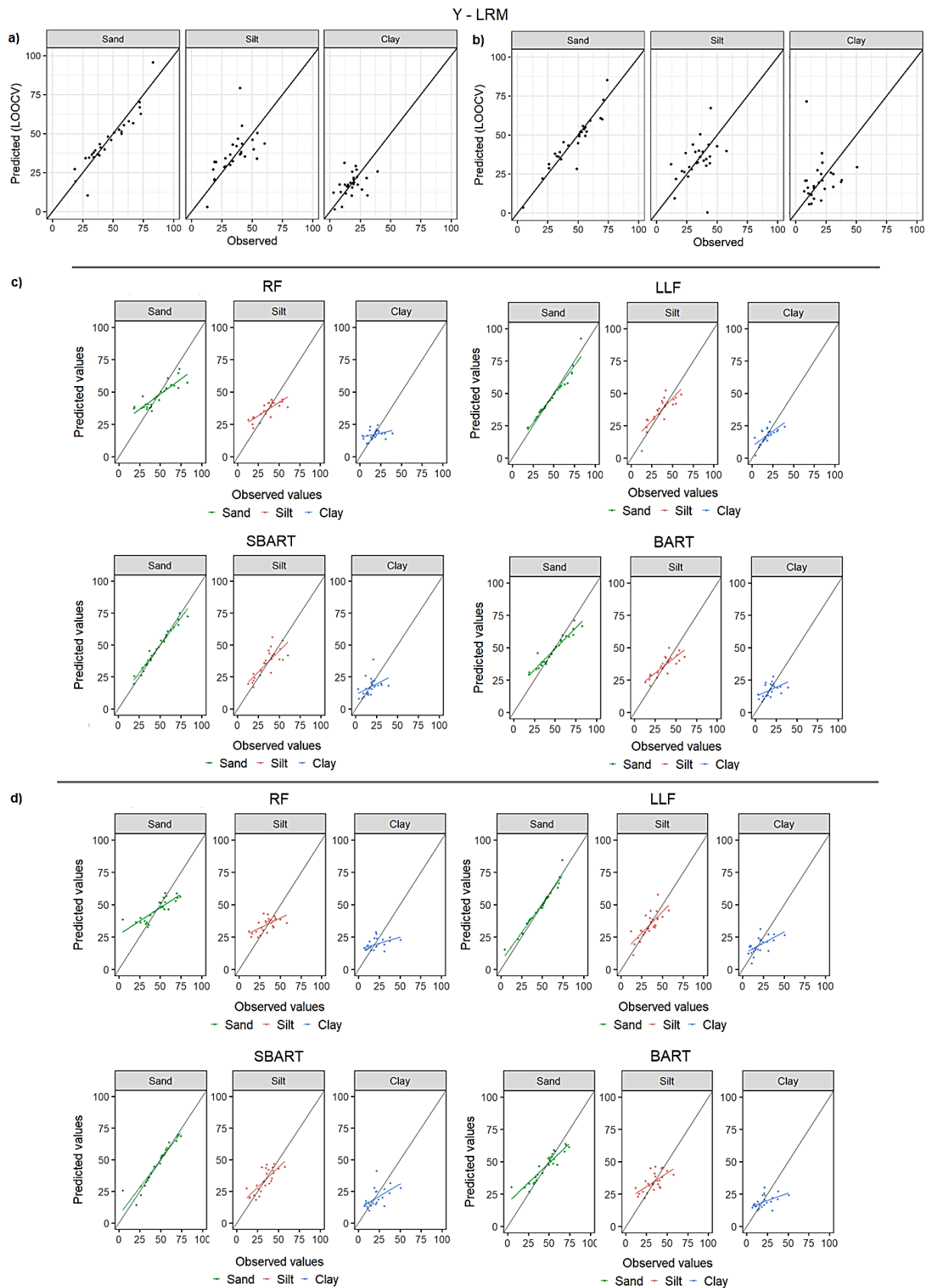


Figure 5.8 Scatterplots depicting the observed and predicted data resulted from the compositional linear model and different tree-based models with the H-alpha dual-pol decomposition applied (surface scattering partition) (a, c) and without surface scattering in the model (b, d). **RF**: Random Forest; **GRF**: Generalized Random Forests; **LLF**: Local Linear Forest; **BART**: Bayesian Additive Regression Tree; **SBART**: SoftBART.

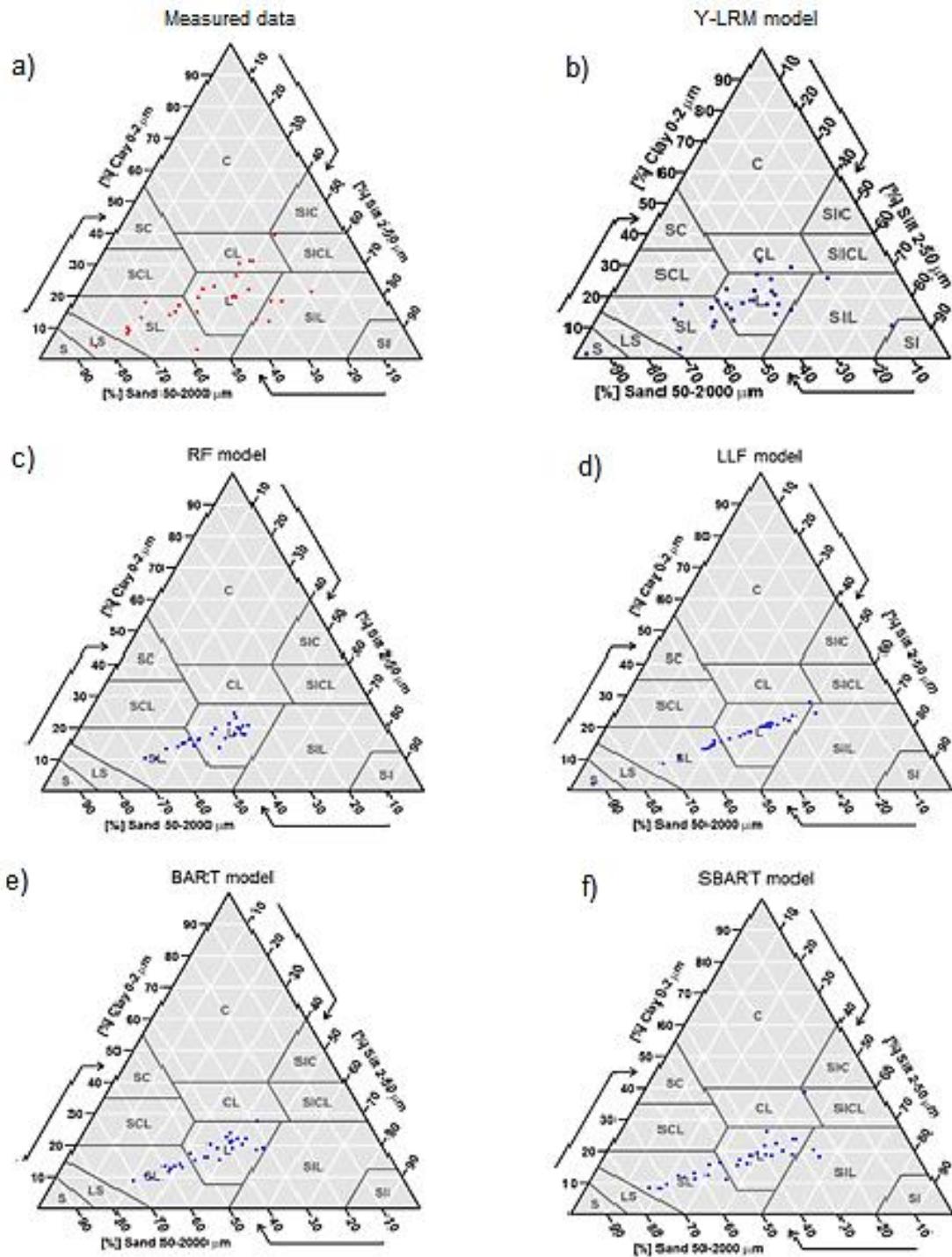


Figure 5.9 Soil ternary (USDA) with the textural classes obtained from the statistical models with surface scattering data included in the models and log-ratio transformation applied to the response variables (CoDa approach). Soil ternary in figure a) refers to the classes obtained directly from the measured data. Soil textural classes inferred from the regression models are depicted in figure (b) with the compositional LRM and in figures (c-f) with the tree-model approaches.

5.6 Discussion

As far as we are aware, no study to date has employed dual-pol decomposition for estimating soil PSF and predicting soil texture. Hence, comparative analyses based on results from other studies that employed this technique for biophysical properties retrieving were used to underpin this discussion section.

The H-alpha Dual-Pol Decomposition method provided, to a certain extent, surface scattering over low vegetation, which may indicate underlying soil. This was observed in our study by comparing the Alpha image and the RGB composite from Sentinel 2 image. Despite using a C-band radar, which has a lower ability than an L-band radar to penetrate (low) vegetation and reach the soil (Babaeian *et al.*, 2019). Typically, cross-polarisation (e.g. VH) is associated with multiple scattering and is useful for volume (multiple) scattering occurrence (Husman *et al.*, 2021). Thus, it may be useful to separate soil and vegetation (Sowter, n.d.). In a study on crop growth analysis in India, Salma *et al.* (2022) used the H- α plane plot derived from S1 and target decomposition to analyse the behaviour of the crop field and soil plough. They found a match between the scattering shown on the H- α plane and field's temporal backscattering during growth period, which means that dual-pol decomposition, to a certain extent, dealt with gap space between soil and plants on the crop fields.

On the other hand, Ji and Wu (2015) observed that HH-HV and HV-VV SARs cannot effectively separate surface, dipole, and multiple scattering mechanisms due to the lack of co-polarization (HH-VV), indicating that co-polarization is vital for extracting scattering mechanisms (Ji and Wu, 2015). They argue that the scatters of most zones strongly diffusing and transferring in the H- α plots for HH-HV and HV-VV SARs. Notwithstanding caveats raised by Mascolo *et al.* (2021) and Ji and Wu (2015), dual-pol decomposition has been employed, for example, for feature identification for classification and for biophysical parameters estimates. In order to improve results and to avoid misinterpretation on scattering and target separation, adaptation of H-alpha planes and cosine distribution for dual-pol SAR could be useful.

The pixel samples of bare soil and low vegetation (i.e. grass) that were extracted from the Alpha parameter image (from Sentinel 1) and the corresponding RGB composite from Sentinel 2 could be confirmed as being surface scattering by the H- α classification regions in the H- α plane plot. These two types of land cover are observed in the plots corresponding to the Bragg and random scattering, which are known as surface scattering regions. This can be also observed in Figure 5.7.

From the results of Alpha and Entropy shown in Table 5.1, we can note that the samples selected (n=27) are more related to the low entropy surface scattering (Alpha \leq 40 and H \leq 0.6). This can be explained by the Bragg scattering associated with surface roughness conditions, which is common in agricultural lands due to plough, hay, or straw. On the other hand, the effect of second mechanism over quasi-bare soil (i.e. sparse vegetation cover), which might be due to the Bragg scattering (also a surface scattering) or double-bounce, needs to be more investigate for dual-pol decomposition.

The dual H- α polarimetric decomposition method was useful to improve the models' performance when the samples were selected over surface scattering mechanisms through alpha information. The compositional linear regression model fitted with ILR transformation yielded better results than the model fitted without the surface scattering mechanism. The surface

scattering predictor improved estimations for silt and clay on the test data. The highest magnitude NSE values for estimations in the test dataset for both scenarios were found for sand, in spite of remaining the same ($NSE = 0.86$). Such agreement results reflected the estimate errors, wherein RMSE and MAE values were reduced for silt and clay. Harfenmeister *et al.* (2021) in a study of crop growth monitoring with Sentinel 1, in Northeast Germany, identified alpha, entropy and anisotropy as important inputs in their model. They found R^2 values from single regression models for the plant height of wheat related to entropy and anisotropy with of 0.64 and 0.61, respectively. With multiple regression models of VH, VV, entropy, and alpha, R^2 values were 0.76, 0.7, 0.7 and 0.69 for plant height, wet biomass, dry biomass, and vegetation water content, respectively. These authors thereby concluded that dual-polarimetric decomposition parameters, derived from Sentinel 1 is capable to provide meaningful input parameters for multiple regression models to improve the prediction of biophysical parameters.

Regarding the tree-based models (CLR and non- CoDa), in the modelling scenario with the surface scattering, this predictor improved the particle size-fraction estimates, particularly for sand and silt fractions regarding the NSE. The estimates error were reduced in all models for all PSF, according to RMSE and MAE values. In general, LLF and SBART models showed the best performance, while sand was the soil fraction with better estimates. Random Forest algorithm yielded the lowest evaluation metrics values amongst the tree-based models. The tree-based models with CLR transformation showed improved metrics in comparison to this modelling approach without log-ratio transformation applied (non- CoDa approach). Likewise for the compositional LRM model, a distinctive behaviour of sand and clay results was also observed with the tree-based models. Moreover, the tree-based models appeared to capture some non-linearity, particularly for clay estimation. The different performances for estimating sand, silt, and clay content (%) with the tree-based models, they could be also explained by the sample size as machine learning methods are data-based methods.

The textural classes and the pattern of the samples displayed in the soil ternary (USDA) in the test data with the Y-LRM were most consistent with the soil ternary in the measured data than those derived from the tree-based models. It is important to note that such classes are derived from the estimations of sand, silt, and clay in the statistical models, thus, they were indirectly obtained. In the soil ternary with the true (observed) data, textural classes are obtained directly from the measured data.

Despite the models' performance, the results obtained from the tree-based models fitted with a non- CoDa approach suggest that they need to be analysed with caution. An important characteristic when dealing with soil texture is the sum of the parts (components or sand, silt and clay) being equal to 100 in each sample and such a constraint has to be guaranteed when such models are employed without transformation applied to the response variables (sand, silt and clay). Moreover, in this statistical modelling approach (standard regression method), compositional data is considered as univariate data or variables, that is, they are independent of

each other. This means that the predictions were obtained separately. Filzmoser *et al.* (2018) highlight that the use of classical statistical methods to deal with compositional data can lead to biased results. In our work, the constraint of the total sum equal to 100 in the estimations for each sample and the response variables being treated as multivariate were taken into account.

The distinctive behaviour of soil particle-size estimations from our results suggests that sand and silt may be more related to surface scattering – sand was invariant only in the linear model. Clay did not respond to the surface scattering as sand and silt did, thereby it seems to be more related to volumetric scattering - recalling that surface and volume are partitions of alpha. In general, clay soil has a little more amount of plants nutrients than sandy soils and plays a vital role in holding plant nutrients and water. Notwithstanding the dry period of Sentinel 1 scenes, the returned radar beam (backscatter intensity) may have captured some remaining moisture from soils and the vegetation structure (i.e. leaves, stems, etc.), since both clay and radar signals are sensitive to moisture content. From this, the soil particle-size fractions estimations reflected in the soil textural ternary wherein more-sandy loam classes resulted from the model with the surface scattering covariate, while more-clayey loam classes were observed from the model without that covariate.

The signal return, or backscattering coefficient (σ^0), is largely dependent on soil texture (Dobson *et al.*, 1981) and soil moisture content because of the dielectric constant (ϵ) of the soil (Das & Paul, 2015). A theoretical explanation on the importance of microwave remote sensing applied to soils relies on the behaviour of soil texture in the microwave spectrum, as the distribution of particle-size fractions controls the content of free water on soils (Dobson *et al.*, 1981). Consequently, the interpretation of SAR images relies on differences between the dielectric constant of the targets and is sensitive to the presence of water.

Regardless of the statistical modelling approaches, the different patterns in the prediction of soil PSF may be related to the fact that clay fractions reflect more the physical and chemical soil properties, while sand and silt fractions are more related to weathering-resistant primary minerals of parent rocks (Niang *et al.*, 2014). Moreover, in microwave remote sensing, soil texture affects the radar-backscattering coefficient as an interplay between free water and bound water (Jackson & Schmugge, 1989; Dobson *et al.*, 1981). Sandy soils are richer in free water than clay soils that preserve more bound water (Srivastava *et al.*, 2006; Das & Paul, 2015). Recalling again that the Sentinel-1 dataset was acquired during a dry period (03/04/2021) inclusive of days before the Sentinel-1 passing over the study area. This is important to be considered in terms of the interaction of free water with the incident signal, which in its turn, contributes to the SAR backscatter. (Das & Paul, 2015).

5.6.1 Strengths and limitations of the work

Regarding the challenging task of estimating soil properties from SAR data and techniques in light of the increased attenuation of the microwave signal over vegetated surfaces (Babaeian *et al.*, 2019), the method applied in this work seemingly ameliorates such an issue.

This research showed that the dual H- α polarimetric decomposition method, which is typically used for classification and segmentation purposes, can also be employed for soil properties prediction from dual-pol C-band SAR such as Sentinel 1 that is freely available for users, broadening the application of this technique in microwave remote sensing studies.

Both modelling approaches showed the same pattern of results for the estimation in the test data – an improvement of silt and clay content (%) with surface scattering as one of the predictors, and sand exhibiting the highest values (although invariant with alpha-surface scattering), while clay the lowest. These findings can lead to learning more about the scattering mechanisms, soil PSF and soil texture in further works.

The performance of the models evaluated here can potentially be improved by involving other field measures such as soil moisture and the dielectric constant of soil. Moreover, a fully-pol H-alpha decomposition can be applied to an L-band SAR or even a dual-pol decomposition with a HH-VV SAR in order to compare the efficiency of the dual-pol decomposition method applied to a C-band SAR for the study area for soil PSF estimates.

The methodological approach used in our work is not without limitations, which can be summarized in two main points. Firstly, the lack of the dielectric constant of the soil in the modelling approaches, which could be useful to analyse and exploit the relationship between sand-silt-clay and microwave backscattering coefficients (σ_{VH}^0 and σ_{VH}^0), delving more deeper into the physical scattering properties and physical soil properties. Such data could be also useful to perform a simulation of extended-Bragg scattering scenario relating alpha, entropy and dielectric values using cosine distributions. This was carried out using full-pol data (Ponnuram and Rao, 2017), but an adaptation of it could be useful for dual-pol VV-VH case to observe scattering angle alpha and entropy at a particular pattern.

Secondly, the relatively small size of data obtained from the method here employed to train and test the model that might lead to an overestimation of sand % and an underestimation of clay %. Nevertheless, the results may reflected the true proportion among the components in the samples (compositional sand, silt, and clay fractions) rather than the sampling size.

Likewise in Ji and Wu (2015) and Mascolo *et al* (2021), we are aware of the inherent limitation of dual polarisation-based SAR for target decomposition, nonetheless, the results are encouraging as the models here evaluated with the surface scattering covariate produced better estimates of soil particle-size fractions than the model without surface scattering data. This indicates the potentiality of the method (H-alpha dual pol-decomposition) to obtain soil information under some degree of vegetation cover (i.e. very low-and sparse vegetation) even with C-band radar.

5.7 Conclusion

In this work, we tested the surface scattering mechanism, derived by H-alpha dual pol-decomposition from Sentinel 1, and additional covariates to predict percentages of compositional sand, silt, and clay (topsoil). Our results reflected the dual-pol S1 limitation, highlighted in the literature, to derive alpha values for surface scattering retrieval under vegetated soil, as the results showed a slight improvement in soil PSF estimates for the model with surface scattering values employed. Notwithstanding that, the method improved results in all the models employed in this study.

The method here applied is an empirical modelling-based approach wherein regression fits between in situ measurements – in this case study the percentages of sand, silt, and clay – and σ^0 . The advantages of it are based on the fact that they are simple and straightforward, and there is no physical basis behind the model (Barret *et al.*, 2009). One shortcoming is that empirical models are usually only valid for the area under investigation (Barret *et al.*, 2009). As a result, in an example of soil texture mapping purposes, the method could be used in an area with similar characteristics to those of our study area, that is, low- and sparse vegetation (e.g. short grass, crops), bare soil, and arable- or fallow lands. Nonetheless, in light of the potential wide applicability of polarimetric decomposition methods to other than classification, the method could be employed to evaluate its usage in a more-dense vegetated area using SAR sensors built on longer wavelengths/shorter frequencies such as L-band and P-band. This could raise an issue in terms of data availability – which is not the case of Sentinel 1 since it is freely available for users. However, there is an increasing effort of space agencies to provide data freely available to everyone. For example, the Argentine SAOCOM (Satélite Argentino de Observación CON Microondas), a fully polarimetric satellite constellation of two spacecrafts that operate a spaceborne L-band SAR system.

In the context of an agricultural field setting, one advantage of empirical modelling is the possibility to differentiate a pixel from various scattering mechanisms (Barret *et al.*, 2009). Our work demonstrated that, under the presence of some degree of vegetation (i.e., a low- or sparse vegetation), target decomposition theorems and Pol-SAR decomposition methods together with radar vegetation indexes such as DpRVI can be used to compensate for this. Particularly for dual-pol SAR without an HH-VV polarisation (Ji and Wu, 2015), as there is a scattering mixing in the H-alpha planes hampering an optimum separation of scatterers. This was observed in our study in the plots from Figure 5.7. It is worth noting that alpha refers to a dominant scattering mechanism in a pixel. Thus, despite of being a target decomposition, the pol-decomposition technique does not separate individual targets in a pixel, but rather dominant, since the radar image pixels contain many scatterers. On the other hand, notwithstanding constraints about dual-pol decomposition, the H-alpha Dual Pol-Decomposition method with S1-C band SAR was able to capture the gaps between soil and vegetation through the dominant targets within a pixel (10 m spatial resolution). This was also demonstrated in Figures 5.6 and 5.7, which was also

corroborated by spectral signatures of a representative pixel within each polygon. From this, dual-pol decomposition remains a potential method for retrieving soil properties and biophysical parameters when employed with caution and meticulousness.

Digital soil mapping generally relies on empirical models to integrate ancillary data influenced by topographical variables derived from a digital elevation model to soil particle-size fractions to understand distribution patterns of soil PSF across an area. (Mondejar & Tongco, 2019). In this context, methods outlined here have the advantage of accounting for the requirements of compositional data such as soil texture. Moreover, regarding the laborious soil sampling associated with soil surveys, this work highlights the possibility of combining model based approaches with more limited in situ sampling methods – especially for sand and silt – for a sample size of 27 observations even splitting data into training and testing.

Thus, the exploratory assessment of the methods carried out in this work is a starting point for the development of other case studies to ameliorate predictions of soil particle-size fractions and soil texture over low-and short vegetated areas, going beyond the traditional polarimetric decomposition technique applied for classification and segmentation of landscapes or for vegetation studies.

Future works on this application – soil PSF and soil texture predictions with SAR techniques – may be beneficial by employing different backscattering coefficients – particularly gamma nought –, other speckle filtering methods, cosine distributions, and soil-specific covariates related to SAR such as dielectric constant, soil roughness, and soil moisture. Moreover, in light of the endeavour of the Sentinel satellite constellation and its benefits (e.g., free availability, temporal-radiometric-spatial resolutions, and orbit stability), joining Sentinel 1 and Sentinel 2 products and their techniques can be also beneficial. For example, combining H-alpha dual-pol decomposition with spectral unmixing method (i.e. spectral mixture analysis using endmember values) to extract values to serve as input in a statistical model, as they are pixel-based analyses. A pure pixel or endmember is considered as such when a pixel contains a reflectance of one land cover/use feature. The aim of the spectral unmixing method is roughly the same as that of target polarimetric decomposition, that is, to distinguish the contributions of the targets in a single pixel since it is not pure due to the randomness of scatterings.

CHAPTER 6 | EVALUATING THE USE OF EXPLICITLY SPATIAL MODELS AND NON-SPATIAL MODELS TO ESTIMATE SOIL PARTICLE SIZE FRACTIONS AND SOIL TEXTURE USING SENTINEL-1 AND COVARIATES

6.1 Context

This Chapter corresponds to the final part of the methodology section of this research. It is rooted in the findings from Chapters 4 and 5, which provide the basis for addressing the following research question that motivated this part of the study: *Spatial or non-spatial models? What is the difference between them in terms of performance, accuracy and issues for estimating soil grain-size fractions?* This Chapter also allowed this study to answer the research question concerned with the application of the methodology in another study area.

A geostatistical modelling approach is employed to capture trends and patterns in the response variables in the sampling sites, which is not possible to achieve with non-spatial approaches. Two spatial models were tested namely Empirical Bayesian Kriging Regression and Cokriging to compare with the non-spatial models previously employed in Chapters 4 and 5.

6.2 Introduction

Soil texture (i.e., proportions of sand, silt, and clay content) is important for understanding other soil properties as it influences soil water retention. Sandy soils typically have a higher amount of free water than clay soils (Srivastava *et al.*, 2006; Das & Paul, 2015), whereas the latter contains more bound water. However, obtaining soil texture and other soil properties using traditional methods (i.e. field sampling, laboratory analysis) remains costly and time-consuming, especially in small scale (larger areas) geographical studies. As a potential low-cost alternative, Synthetic Aperture Radar (SAR) has been employed to estimate soil properties (Petroopoulos *et al.*, 2015; Babaeian *et al.*, 2019; Domenech *et al.*, 2020; Pradipta *et al.*, 2022). Typically, these studies have focussed on the utilisation of SAR for estimating soil moisture or carbon content over bare or sparsely vegetated ground and employ soil texture as a secondary parameter to retrieve these parameters. Few studies have employed SAR for estimating soil particle size fractions (PSF) (the relative proportions of sand, silt and clay), where soils are permanently covered by vegetation, especially by grass or crops (Jagdhuber, 2012, 2013; Harfenmeister *et al.*, 2021; Salma *et al.*, 2022; Dave *et al.* 2023). This may, in part, be due to the complexity that this soil property is compositional in nature (i.e., the sum of the relative proportions is equal to 100%) and challenges around the use and interpretation of radar data.

One of the main parameters related to energy intensity measurements in microwave remote sensors (i.e., the proportion of energy that is returned from the target) is the normalized radar cross section (σ^0), known as the backscattering coefficient, which is a measure of the polarisation intensity (i.e., vertical transmit and receive-VV, vertical transmit horizontal receive-VH). The signal return, or backscattering coefficient (σ^0), is largely dependent on soil texture (Dobson *et*

al., 1981) and soil moisture content, due to the dielectric constant (ϵ) of the soil (Das & Paul, 2015) which is directly related to the presence of water in the soil pore spaces. Soil properties can be estimated using SAR through establishing a relationship between the backscattering coefficient and soil property of interest.

The geostatistical approach has been largely employed to model, estimate, predict and simulate geographical phenomena by means of distance and neighbourhood analyses and based on three main principles or laws of geography that govern geostatistical techniques: spatial autocorrelation (Tobler, 1970), spatial heterogeneity (Goodchild, 2004) and spatial similarity (Zhu *et al.*, 2018). Typically, there are two different methods to predict values at unsampled locations (spatial prediction, spatial interpolation, or spatial modelling) – deterministic (e.g. inverse distance weighting-IDW, global polynomial interpolation-GPI, radial basis function-RBF, local polynomial interpolation-LPI) and non-deterministic (e.g., kriging, cokriging, areal interpolation, empirical Bayesian kriging-EBK) (Singh & Sarma, 2023). Spatial machine learning is another approach, however, it differs from geostatistical-based models in terms of rigid assumptions required such as distribution and stationarity of the target variable (Hengl *et al.*, 2018). Uncertainty of spatial predictions can also be measured with this approach and details of the methods can be found in Hengl *et al.* (2018).

However, a challenge emerges in both spatial and non-spatial modelling when dealing with compositional data (CoDa) like soil texture, due to the inherent nature of the variable (i.e. either a compositional response or compositional predictor), and the prediction/estimate needs to reflect this constraint. Wang & Shi (2017) conclude that to improve spatial interpolation performance, a combination of geospatial methods and the use of ancillary variables is necessary, since many of the traditional methods may not reflect the specific requirements of compositional data for interpolation.

In this context, two spatial models from the geostatistical approach (Empirical Bayesian Kriging Regression and Cokriging), and two non-spatial models (linear regression and tree-based regression) were tested in this study to estimate the percentages of sand, silt and clay on topsoil layer using microwave remote sensing data (Sentinel-1 SAR). The modelling approaches employed explicitly consider the compositional nature of soil texture. This work seeks to address two key research questions: (i) Do spatial models outperform non-spatial models in estimating soil particle size fractions? and, (ii) How, and to what extent, do location and neighbourhood affect the model estimates? The focus of this study is an evaluation on the impact of considering a numerical, but non-geographical dataset and a spatial dataset (e.g. raster, shapefile) in the models for spatial prediction of soil fractions using of SAR information.

The remainder of this chapter is organised as follows. Section 6.3 presents a brief review of the theory and review of relevant literature. Section 6.4 outlines the methods and data used to carry out the research, including the study area. Sections 6.5 and 6.6 present the results of the

modelling approaches and discussion, including a description of the strengths and limitations of the work. Section 6.7 presents the conclusion of the study.

6.3 Background: Spatially explicit models and geostatistics

Spatially explicit models are those that fulfil at least one of the following requirements/tests and any of their combinations (Goodchild, 2001; Janowicz *et al.*, 2020): Invariance test, Representation test, Formulation test, and Outcome test. Nikparvar and Thill (2021) discuss the properties of spatially explicit data that influence the performance of machine learning – spatial dependence, spatial heterogeneity, scale, specific representation of the data (polygon, line, point, regular grid, text), and measurement process. Details on those requirements/tests and discussions can be found in Janowicz *et al.* (2020) and Nikparvar and Thill (2021).

Geostatistical methods builds upon the principles of spatial autocorrelation (nearby things are related), spatial heterogeneity (variance across areas) and spatial similarity (similarity of geographic configurations), referring to the three laws of Geography – Tobler’s Law (Tobler, 1970), Goodchild’s Second Law (Goodchild, 2004), and Zhu’s law (Zhu *et al.*, 2018), respectively. A number of geospatial techniques are based on these laws, including Voronoi mapping (first law), kriging/cokriging (second law). The second law implies that the requirements imposed by Tobler’s Law would not be met over large areas, since the spatial autocorrelation and/or the correlation may vary over space. Delaunay Triangulation results in a Voronoi map, where features that share an edge will be included as neighbours – the proximity diagrams. It is a useful method for obtaining an influential area of a feature (Evans & Jones, 1987) (e.g., a point representing a measure of soil particle size fractions) or geographical phenomenon. Voronoi diagrams, also known as Dirichlet tessellation or Thiessen polygons, express ubiquitous patterns in nature (Dierking *et al.*, 2021). Thus, the Voronoi diagram and its relative or dual graph Delaunay triangulation (Appendix C – Figure C6.1) represent potentially valuable tools for analysing spatial data (Okabe *et al.*, 2000).

Spatial correlation, which declines with distance, is typically measured using semi-variograms. A steeper curve near the origin (nugget effect) means a greater influence of the closest neighbours on the prediction/point of interest. However, spatial models are prone to inductive spatial bias due to spatial autocorrelation (Smith, 2009; Mialon, 2023).

The most common geostatistical method to predict data at unsampled locations is Kriging, which relies on mathematical and statistical models, as well as on the principle of spatial heterogeneity (Goodchild’s second law). These features distinguish Kriging methods from more deterministic methods such as IDW. A deterministic method offers reproducibility and stability as it operates without any randomness (Singh & Sarma, 2023), being a valuable method when and where repeatability, reliability, and precision are essential (Singh & Sarma, 2023). On the other hand, it does not provide uncertainty estimation in predictions. In turn, apart from creating spatial maps, non-deterministic methods seeks to quantify uncertainty in predictions. Further, methods

such as Cokriging (CK) that employ two or more predictor(s) to incorporate spatial correlation between multiple variables can result in improved models and enhance the accuracy of the estimates (Singh & Sarma, 2023).

Cokriging, a variation of Kriging, is employed when covariates are considered in the model. It uses information on both the autocorrelation of the response variable and the cross-correlations between the predictor and the covariates to improve the predictions. By exploiting information on the covariates, this method is commonly used when the variable of interest has a lower sampling density. Both Kriging and CK assume normality and stationarity. Where a trend exists, either an anisotropic variogram or Universal Kriging-Cokriging can be used.

Empirical Bayesian Kriging Regression (EBKR) is an extension of Empirical Bayesian Kriging (EBK), which combines Kriging with regression (Linear Mixed Model-LMM) (Krivoruchko and Gribov, 2019; Gribov and Krivoruchko, 2020), leading to more accurate predictions over a single method. Basically, the dependent variable is a function of the mean (the regression component) and the error term (the semivariogram/covariance component). It is also an alternative method when stationarity is not present in the response variable as it treats local variance separately (subsets) in lieu of variance being constant in the whole area. Thus, EBKR performs Kriging locally.

There is an extensive body of literature on the use of spatial models (geostatistical and machine learning approaches) for soil property estimation and prediction (Table 6.1). Further, Webster and Oliver (2007) highlight the valuable role of geostatistical techniques in solving real-world problems, such as soil properties, for farmers.

Table 6.1 Some related works regarding statistical and/or geostatistical models for soil properties estimation and prediction

Technique			Soil property				Reference
Artificial Neural Network (ANN)			soil moisture				Araya <i>et al.</i> (2021)
Bayes Network			soil texture				Smith & Peng (2009)
Boosted Regression Trees (BRT)			soil moisture				Araya <i>et al.</i> (2021)
Compositional Log-ratio Kriging	Kriging;	Log-ratio	sand, silt, and clay content				Wang & Shi (2018); Wang & Shi (2017)
Geographically Regression		Weighted	organic content; pH	matter; total organic content	organic carbon	carbon	Xu & Zhang (2021); Zeng <i>et al.</i> (2016); Zhang <i>et al.</i> (2011); Wang <i>et al.</i> (2012)
Log-ratio Cokriging			sand, silt, and clay content				Wang & Shi (2018)
Multilayer Perceptron Neural Network (MLP)			soil texture				Smith & Peng (2009)
Multiple Linear Regression			organic content	matter; organic	carbon		Zeng <i>et al.</i> (2016); Zhang <i>et al.</i> (2011)
Naïve Bayes			soil texture				Smith & Peng (2009)
Ordinary Kriging			organic content	matter; organic	carbon		Zeng <i>et al.</i> (2016), Zhang <i>et al.</i> (2011)
Radial Basis Function Neural Network (RBF)			soil texture				Smith & Peng (2009)
Random Forests			soil texture; soil classes; sand, silt, and clay content; soil moisture				Akpa <i>et al.</i> (2014); Araya <i>et al.</i> (2021); Bousbih <i>et al.</i> (2019); Cisty <i>et al.</i> (2019); Dotto <i>et al.</i> (2020); Ließ <i>et al.</i> (2012); Pittman & Hu (2021); Silva <i>et al.</i> (2019)

Regression Kriging			organic matter	Wang <i>et al.</i> (2012)
Relevance Vector Regression (RVR)	Vector	Regression	soil moisture	Araya <i>et al.</i> (2021)
Support Vector Regression (SVM) or Support Vector Regression (SVR)	Vector	Machine	sand, silt, and clay content; soil texture; soil moisture; organic carbon	Araya <i>et al.</i> (2021); Bousbih <i>et al.</i> (2019); Gholizadeh <i>et al.</i> (2018); Silva <i>et al.</i> (2019); Smith & Peng (2009).

As part of the Irish Soil Information System Project, Random Forests, Bayesian belief networks and neural networks have been used to predict soil map units at a 1:250,000 scale. The highest accuracy levels were achieved with Bayesian Belief Networks and Random Forests (Mayr *et al.*, 2013). These findings are consistent with recent work from Deodoro *et al.* (2023), who found that Random Forests outperformed the Linear Regression Model when interaction terms were applied to the latter for estimating sand, silt, and clay content using a non-spatial modelling approach.

Regarding spatial modelling approaches treating soil texture as CoDa, Wang & Shi (2017, 2018) explored Kriging and CK with different log-ratio transformation methods for mapping soil particle size fractions in China. They found that a combination of the interpolators with robust variogram estimators significantly improved the prediction accuracy compared to using standard estimators, highlighting the performance of Ordinary Kriging with CLR log-ratio transformation, and CK with ILR log-ratio transformation.

6.4 Data and Methods

6.4.1 Study Region

The study domain encompasses the central portion of Ireland from E-W direction, representing approximately 24,989 km² (Figure 6.1). The land cover and land use consist predominantly of agricultural areas (including pasture, crops, and arable land), grassland, peat bogs, and manufactured structures, including urban areas. Agriculture is the primary land use-land cover category in Ireland, accounting for 67% of the national land cover according to the CORINE 2018 land cover inventory.

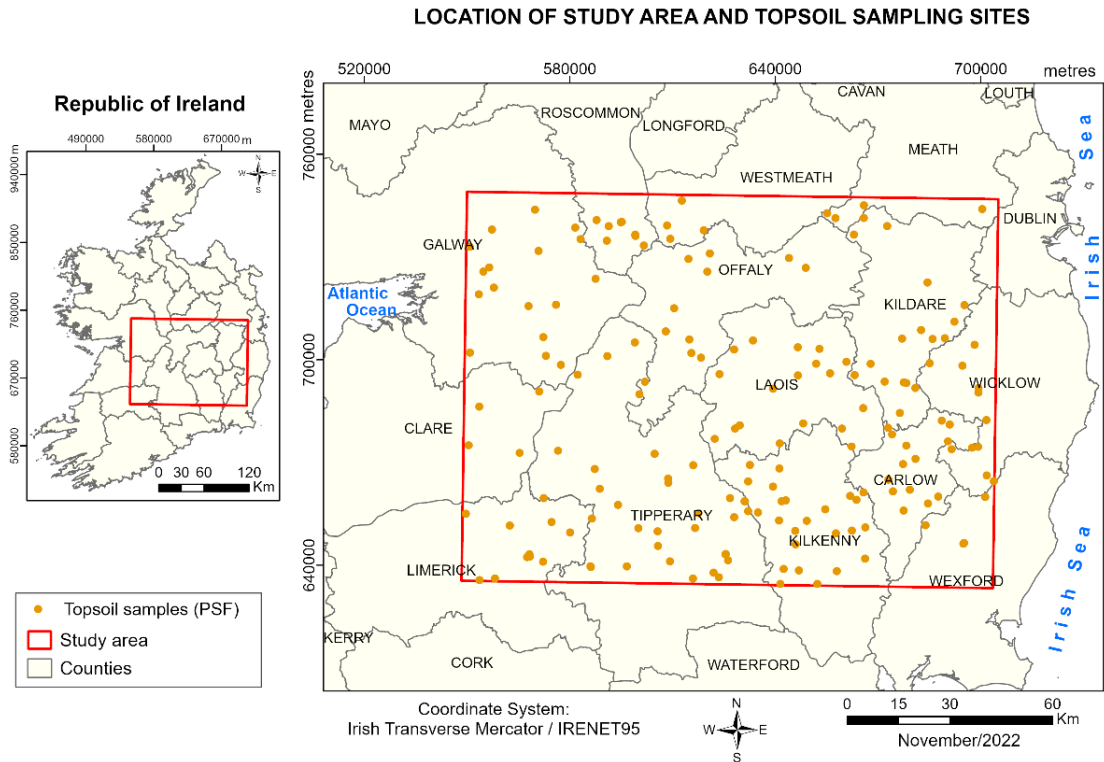


Figure 6.1 Map of the study area with the location of topsoil sampling sites available from different soil surveys and measurement programmes.

Within the area of interest, altitude values range from 923 m on the uplands to 1 m in the flood plains in the south-west. Spring and summer are the driest seasons while autumn and winter are the wettest. Approximately 256-282 mm of rain occurs during March-August whereas this value increases to roughly 369-380 mm during September-February (Curley *et al.*, 2023). The West coast generally exhibits higher rainfall amounts associated with the prevailing westerlies interacting with relief. Regarding soil texture, fine loamy is the predominant soil texture class in Ireland (Creamer *et al.*, 2016), whose landscape is moulded by solid geology (weathered materials) and drift geology (unconsolidated deposits or loose sediments) (Creamer *et al.*, 2018).

6.4.2 Data

6.4.2.1. Soil information

Soil data used to train and validate the prediction models consisted of 235 samples originally extracted from a legacy soil dataset whose sampling was conducted across Ireland, in both shallow and deep soil layers, by various institutions for different survey purposes, as follows:

- Modal and Non-Modal profiles (Creamer *et al.*, 2014) obtained from Teagasc, the Irish Agriculture and Food Development Authority, for the Irish Soil Information System (SIS).
- LUCAS Topsoil 2009 Survey-v1 (Toth *et al.*, 2013; Orgiazzi *et al.*, 2018) and LUCAS Topsoil 2015 Survey (Jones *et al.*, 2020) - soil campaigns carried out by the European Soil Data Centre (ESDAC).

- Soil Carbon (SoilC) (Kiely *et al.*, 2009) – soil survey undertaken by the University College Cork (UCC) and Teagasc.
- WoSIS Soil Profile Database (June 2022)(Ribeiro *et al.*, 2020) - the World Soil Information Service (WoSIS) is a global soil database that stores soil profile data relating to 45 soil properties (physical and chemical).

Only the topmost layer or topsoil (0-15 cm in depth) was selected for use in this study to be consistent with the relatively low capability of C-band radar beams of Sentinel 1 to reach the ground compared to L-band SAR sensors.

6.4.2.2 Microwave remote sensing data

This study employed microwave remote sensing data from the European Space Agency (ESA) Sentinel-1 SAR (C-band) satellite, with Interferometric Wide (IW) swath mode acquisition, Single Look Complex (SLC) product, as radar-based predictors. Sentinel-1 data are available in dual polarisation (VV+VH). The SLC data were obtained with the backscattering coefficients expressed in decibels (db). Co-registration, multi-look, speckle filtering, terrain correction and calibration were carried out using the Sentinel-1 Toolbox (SNAP-ESA). The Copernicus Digital Elevation Model for Europe (ESA EEA-10) at 10 m spatial resolution was used for terrain correction. All datasets were geo-coded to the Irish Transverse Mercator projection.

Radar-based data consist of (i) the backscatter coefficients (backscatter intensity) provided by the VV and VH polarisations measured in sigma nought (σ^0) obtained for the in-situ soil sampling locations, georeferenced; (ii) alpha parameter resulting from the dual-pol decomposition, which varies between 0° and 90° and is used to determine the dominant scattering mechanism; (iii) the dual-pol radar vegetation index (RVI, Eq.6.1) (Nasirzadehdizaji *et al.*, 2019; Gururaj *et al.*, 2019); and, (iv) surface soil moisture (SSM) which refers to the relative water content of the top few centimetres of the soil, measured by Sentinel-1 under the Copernicus programme. Similar to the alpha parameter, RVI considers the vegetation effect on soil backscattering:

$$RVI = \frac{4\sigma_{VH}^0}{\sigma_{VV}^0 + \sigma_{VH}^0} \quad \text{Equation 6.1}$$

The Sentinel-1 dataset (VV, VH, SSM) was acquired for the 3rd of April 2021 - this was selected as April was the month with the lowest recorded rainfall during that year based on information from the Irish national meteorological service, Met Éireann (Met Éireann, 2021). In selecting the date, the antecedent conditions were taken into consideration (29/03/2021-03/04/2021), which had little to no rainfall for the days preceding the date of acquisition.

6.4.2.3 Environmental data

Topographical data derived from the digital elevation model (DEM) ESA EEA-10 was used considering a soil-landscape approach. The topography covariates consist of altitude, slope, aspect and curvature of the slope (concave surface or convex surface). The surface spatial analyst algorithms available in the ArcGIS® toolbox were used to extract these covariates from the DEM.

Topographical covariates (i.e. altitude, slope, aspect, and curvature) and radar covariates derived from Sentinel-1 (i.e. VV and VH backscattering, RVI, and alpha) were employed to account for surface roughness.

6.4.3 Methods

6.4.3.1 Geostatistical exploratory analysis

In both the spatial and non-spatial models evaluated, radar-based data and topography data were employed as predictors. Prior to the prediction modelling, an exploratory geostatistical analysis of the response variables, built upon the approaches of pattern and neighbourhood, was undertaken to understand the spatial structure of the data and to tune the parameters in the models accordingly. Such a procedure is grounded on the fact that spatial interpolation methods rely on weights from surrounding measured values to predict unmeasured locations. Furthermore, when assessing for spatial influences, it is worth considering how regular the sampling units are distributed in space (e.g., regular or gridded, random, dispersed, clustered).

Regarding the pattern analysis, the Global Moran's Index (GMI) and the Average Nearest Neighbour (ANN) were employed (Equations 6.2 and 6.3, respectively).

$$GMI = \frac{n}{S_0} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{i,j} z_i z_j}{\sum_{i=1}^n z_i^2}, \quad S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{i,j} \quad \text{Equation 6.2}$$

where z_i is the deviation of an attribute for feature i from its mean; $w_{i,j}$ is the spatial weight between feature i and j ; n is the number of features; and S_0 is the aggregate of all the spatial weights.

$$ANN = \frac{\bar{D}_o}{\bar{D}_e}, \quad \bar{D}_o = \frac{\sum_{i=1}^n d_i}{n}, \quad \bar{D}_e = \frac{0.5}{\sqrt{n/A}} \quad \text{Equation 6.3}$$

where \bar{D}_o is the observed mean distance between each feature and its nearest neighbour; \bar{D}_e is the expected mean distance for the features given in a random pattern; d_i denotes the distance between feature i and its nearest neighbouring feature; n is the total number of features; and A corresponds to the area value (specified).

The GMI was calculated to assess the spatial autocorrelation within the data for both the response variables and the covariates. It ranges from -1 to 1, wherein positive values occur when similar values cluster together (positive spatial autocorrelation); negative values denote dissimilarity amongst the data, that is, dissimilar values are next to each other (negative spatial autocorrelation); and zero value denotes no spatial autocorrelation. It is worth noting that autocorrelation violates the core principles of classical statistics, that is, observations are independent of each other. The ANN was useful to find the distribution pattern of the data. It

measures the distance between each feature and its nearest neighbour's location. If the ANN index is less than 1, the pattern exhibits clustering; if greater than 1, the trend is toward dispersion.

For the neighbourhood analysis, the Distance Band and Delaunay Triangulation methods were used because they are appropriate when dealing with point features. The former is a critical distance within which each feature is included as neighbours. Here, the distance found was 20,000 m, based on the distribution pattern of the data. We selected Manhattan distance as a method to account for outliers and spatial discontinuities (e.g., topography, landforms).

6.4.3.2 Prediction modelling approaches and interpolation

Spatial Models

▪ Co-kriging (CK):

Simple CK (Eq. 6.4) was employed based on the exploratory analysis of the data (known mean):

$$Z_1(s) = \mu + \epsilon_1(s), Z_2(s) = \mu + \epsilon_2(s) \quad \text{Equation 6.4}$$

where Z_1 and Z_2 are the random variables located in (s) ; μ is the known and constant mean; ϵ_1 and ϵ_2 are random error terms. Cokriging assumes two models and stationarity of the data (mean and variance constant).

Variograms used in Cokriging (variance-based cross-variograms, Eq. 6.5) are estimated from data and explain the spatial variability of soil PSF (Wang & Shi, 2018):

$$\hat{\gamma}_{u,v}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \{z_u(x_i) - z_u(x_i + h)\} \{z_v(x_i) - z_v(x_i + h)\}, \quad \text{Equation 6.5}$$

where $\hat{\gamma}$ is the estimated variability; h is the lag or distance; $N(h)$ are pairs of sample locations grouped by h ; i^{th} is observation located at $x_i + h$ included in the summation; $(z(x_j), j=1, 2, \dots, n)$ refers to the data used to estimate the cross-variogram; z_u and z_v are the actual variates at spatial locations u and v (Wang & Shi, 2018).

Simple CK was also employed to be consistent with the EBKR, which relies on Simple Kriging for the interpolation step. For lag size, the value of 3,000 m was used to fit the cross-variogram, which corresponds to half of the largest distance between any two points, based on the exploratory geostatistical analysis and the in situ soil sampling distribution across the study area. The number of lags used was 20, based on the exploratory analysis and following the rule of thumb of finding half of the largest distance between any two points and dividing it by the lag size. Here, the largest distance was approximately 27,000 m. Since a trend was observed for sand and clay in the exploratory analysis, this study considered anisotropy for these fractions (199° SSW-NNE, and 134° SE-NW, respectively). After testing different empirical semi-variogram models for the response variables – stable, circular, spherical, exponential and gaussian, rotational quadratic, and k-bessel – the stable model was selected based on fit and ability to capture the

spatial structure of the data. The model was validated with the leave-one-out cross validation method (LOOCV).

- **EBK regression (EBKR)**

For the EBKR model (Eq. 5.6) (Krivoruchko & Gribov, 2019), an exponential semi-variogram with a minimum percent of variance equal to 95, subset size 36 (to account for local influence and local variance) and 100 simulations were selected in this study.

$$z_1 = y(s_i) + \varepsilon_i, \quad i = \overline{1 \dots K} \quad \text{Equation 6.6}$$

where z_1 is the measured value at the observed location s_i ; $y(s)$ is the Gaussian process under study at the location s_i ; ε_i is the Gaussian measurement error; and K is the number of measurements.

The EBKR model was also validated with the leave-one-out cross-validation method (LOOCV). In attempting to handle non-linearity, which was observed in the results from the CK and EBKR models, this study applied polynomial functions (cubic and 6-degrees) and splines over curvature, and interactions terms over all the covariates. Only the statistically significant interactions were employed as covariates to re-run the EBKR and CK models (Appendix C – Table C6.1).

Non-spatial models

For comparison purposes, a non-spatial linear modelling approach was also carried out. Considering the compositional nature of soil texture (a composition of particle size fractions), log-ratio transformation methods rooted on symmetry were employed in this study because they preserve distances. The isometric log-ratio transformation (ILR) was applied to the response variables or Y (sand, silt, and clay) for fitting the compositional Linear Regression Model (Y -LRM) following the default partitioning built in the CoDaPack software (Comas-Cufí & Thió-Henestrosa, 2011). The centred log-ratio transformation was applied to the response variables to train a tree-based model (Deodoro *et al.*, 2023). The total sum equal to 100 (or 1) of the PSF components is guaranteed when the predicted values are back-transformed to the original units (Comas-Cufí & Thió-Henestrosa, 2011). The CLR transformation was applied to the response variables to fit the tree model-based regression (Random Forests model-RF). Details about both CLR and ILR transformations are presented in Appendix C – Models. Interaction terms in the compositional LRM were also applied to account for the interplay between the predictors and the response variables as RF and geostatistical models do, whose interactions between influential processes are already embedded in the algorithms.

The geostatistical analysis and the modelling approaches were performed using the free and open source software GeoDA (Anselin, 2004), and the extensions Geostatistical Analyst and Spatial Statistics available in ArcGIS Pro®. The non-spatial RF model was performed with R – package “Ranger” (Wright & Ziegler, 2017)

6.4.4 Evaluation metrics for soil particle size predictions

Root Mean Square Error (RMSE) (Eq. 6.7), MAE (Mean Absolute Error) (Eq. 6.8) and Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970) (Eq. 6.9) were calculated to evaluate the performance of the models validated with LOOCV. NSE measures the magnitude of the residual variance compared to the measured data variance (Nash and Sutcliffe, 1970) and indicates how well the graphs of observed versus predicted (simulated) data fits the 1:1 line. A value equal to 1 corresponds to a perfect match of the model to the observed data; NSE = 0 indicates that predicted data and observed data have same accuracy; $-\infty < \text{NSE} < 0$ denotes a poor performance of the model with observed mean being a better predictor than the model.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad \text{Equation 6.7}$$

where y_i is the actual value of the dependent variable, \hat{y}_i is the predicted value of the dependent variable, and n is the number of observations

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad \text{Equation 6.8}$$

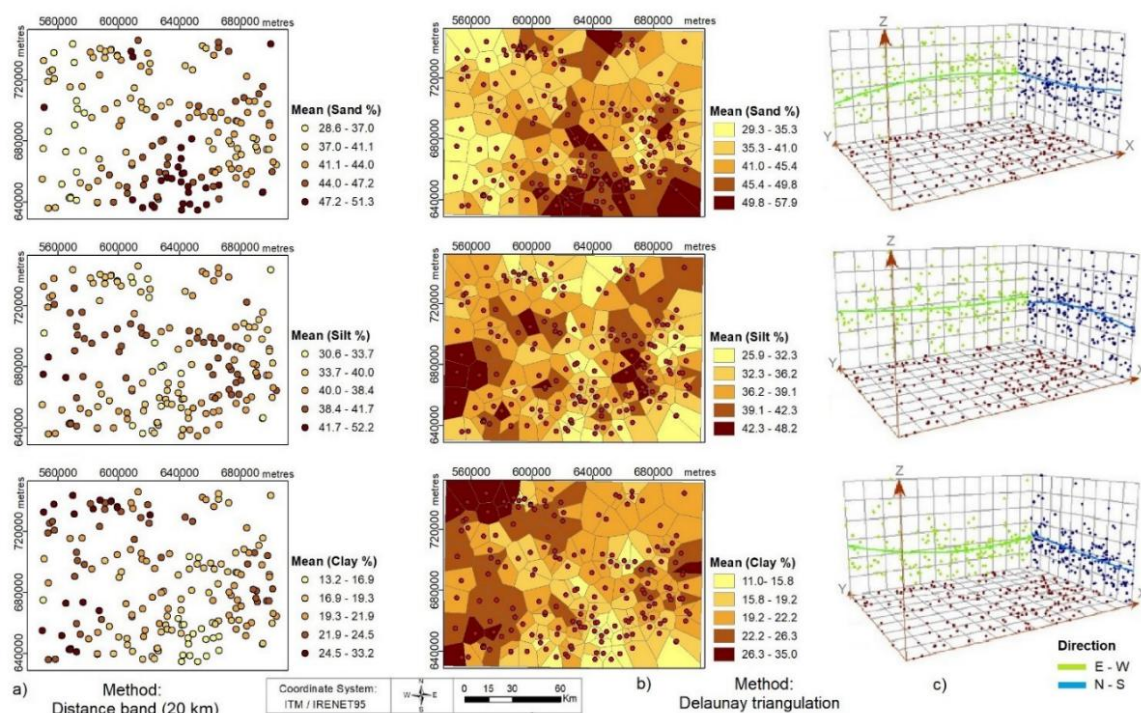
where y_i , \hat{y}_i , and n represents the same descriptors that those of RMSE.

$$NSE = 1 - \left[\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \right] \quad \text{Equation 6.9}$$

where y_i is the i th observation, \hat{y}_i is the i th predicted value, \bar{y} is the mean of observed data, and n is the total number of observations.

6.5 Results

The distribution of the response variables could be approximated by a normal distribution for sand and silt whereas it is skewed to the right for clay (Appendix C – Figure C6.2). Based on a grouping of the mean values, sand and clay fractions (content measured %) showed geographical patterns in terms of the highest values (Figure 6.2a-b). The former exhibits larger values ($> 45\%$) in the S-SE portions of the study area, while larger values of clay content ($> 22\%$) occur in western region (NW-SW). Larger values of silt content ($> 39\%$) tended to W-E, although its pattern is not clear – both the lowest and the highest values are dispersed across the area. The trend analysis plots (Figure 6.2 and Appendix C – Figure C6.3) confirm the pattern distributions of soil PSF. These results indicate that, locally, the data sampling represents a non-stationary process, which is observed in the Voronoi diagrams depicting local variance and local variation (Appendix C – Figure C6.4). These results is consistent with Singh and Sarma (2023) who highlight that in soil properties the spatial correlation may vary across the region, consequently, hypothesis of stationarity may not hold, as well as the Tobler's Law. This means that the second law of geography (spatial heterogeneity) was achieved, especially due to large area, being aligned with kriging approach.



* The range of absolute values (as measured, %): **Sand**: 3 - 83 ; **Silt**: 12 - 66 ; **Clay**: 3 - 59

Figure 6.2 Neighbourhood analysis of response variables (measured data, content %) according to a) Distance Band and b) Delauney Triangulation (Voronoi tessellation) methods. A three-dimensional perspective of the data is depicted in "c", wherein the locations of sample points are plotted on the x, y plane. The x, z plane and the y, z plane are scatterplots of the values projected onto such planes and represent the trend directions (blue and green colours). Above each sample point, the value is given by the height of a stick in the z-dimension.

The average nearest neighbour ratio was greater than 1 (Table 6.2), thus the pattern exhibits dispersion of the data, which is consistent with the distribution pattern given by the Global Moran Index-GMI (Table 6.3). The GMI also indicated a lack of spatial autocorrelation for the response variables (See also Appendix C – Figure C6.17). Regarding the covariates, there is a clustered pattern for VH, surface soil moisture and aspect. Only the soil moisture and aspect covariates are autocorrelated or spatially dependent.

Table 6.2 Distance band from neighbour

Data	Distance band (at least one neighbour) *			
	Minimum (metres)	Maximum (metres)	Average (metres)	Nearest Neighbour Ratio
Soil samples	164	32,932	6,312	1.28

*Manhattan distance.

Table 6.3 Spatial autocorrelation: Global Moran Index

Soil PSF*	Global Index	Moran	Distribution Pattern	P-value	Significance level	Spatially autocorrelated (or dependent)
Sand	-0.0500		Dispersed	0.01	0.10	No
Silt	-0.0578		Dispersed	0.00	0.01	No
Clay	-0.0022		Random	0.84	-	No
Sand.clr	-0.0499		Dispersed	0.01	0.05	No
Silt.clr	-0.0551		Dispersed	0.00	0.01	No
Clay.clr	-0.0022		Random	0.85	-	No
ilr.1	-0.0022		Random	0.85	-	No
ilr.2	-0.0669		Dispersed	0.00	0.01	No
Covariates*						
VV	-0.0049		Random	0.98	-	No
VH	0.0798		Clustered	0.17	0.05	No

RVI	0.0256	Random	0.39	-	No
Curvature	-0.0701	Dispersed	0.07	0.10	No
Surface soil moisture (SSM)	0.2469	Clustered	0.00	0.01	Yes
Alpha	-0.0346	Random	0.42	-	No
Altitude	-0.0193	Random	0.70	-	No
Slope	0.0327	Random	0.28	-	No
Aspect	0.9600	Clustered	0.00	0.01	Yes

Source: ArcGIS output. Distance method: Manhattan; Spatial relationships method: Inverse Distance.

*Related to soil samples (observed data).

Incremental distance, measuring the intensity of spatial clustering for each increasing distance (small scale or large area), showed that only clay (measured data) exhibited clustering at two distances, approximately 7,500 m and 17,500 m (Appendix C – Figure C6.5). This means that the distance of 7.5 km could be employed as a more-local scale of analysis for clay, which is the first statistically significant peak at a significance level of 0.10. For sand and silt, results indicated potential multiple spatial processes (no clear spatial pattern) operating at a different distance in the study area.

6.5.1 Spatial modelling

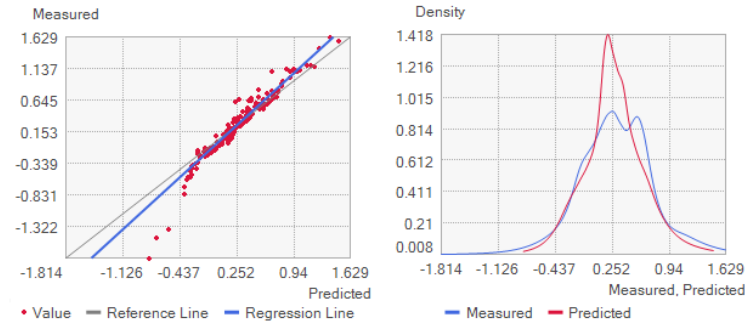
Results showed different behaviour for EBKR and CK. In terms of estimated values, CK yielded better results according to the model evaluation metrics (Table 6.4) and the agreement between observed and predicted data (Figure 6.3). However, the statistical diagnostics exhibited a pattern in the errors for predicted sand, associated with the curvature predictor (Appendix C – Figures C6.10-12). This was not observed in the EBKR error plots (Appendix C – Figures C6.7 and C6.8). The use of the CoDa approach did not improve the predictions in the models.

Table 6.4 Evaluation metrics resulted from the spatial modelling approaches validated with LOOCV

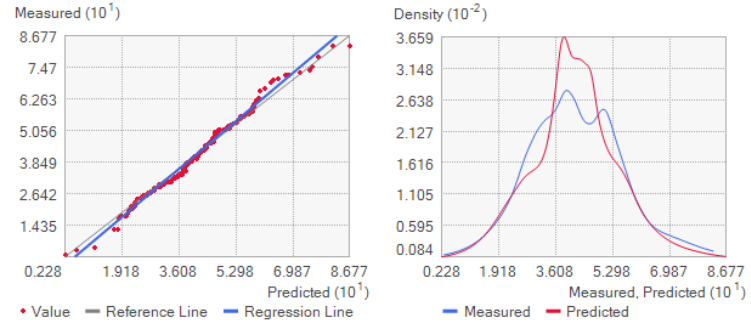
Modelling approach (Validation method: LOOCV)	Soil PSF	Model evaluation metrics		
		RMSE	MAE	NSE
Model: EBK regression (non-Coda)	Sand	14.1	11.1	0.2
	Silt	10.0	8.0	0.2
	Clay	8.0	6.0	0.3
Model: EBK regression with 2 nd order interaction terms applied to the covariates (non-CoDa)	Sand	14.7	11.4	0.1
	Silt	10.7	8.4	0.1
	Clay	8.3	6.1	0.3
Model : EBK regression Y-compositional (clr transformation)	Sand	14.1	11.2	0.2
	Silt	10.2	8.1	0.2
	Clay	8.5	6.2	0.2
Model : Cokriging (non-CoDa)	Sand	2.0	1.5	0.9
	Silt	6.9	5.1	0.6
	Clay	6.6	4.8	0.5
Model : Cokriging with 2 nd order interaction terms applied to the covariates (non-CoDa)	Sand	3.6	2.7	0.9
	Silt	7.8	6.0	0.5
	Clay	6.6	4.9	0.5
Model : Cokriging Y-compositional (clr transformation)	Sand	3.3	2.4	0.9
	Silt	7.0	5.3	0.6
	Clay	6.9	5.1	0.5

COKRIGING

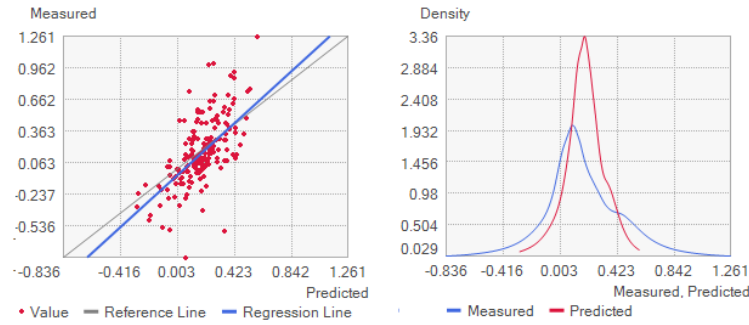
Sand.clr



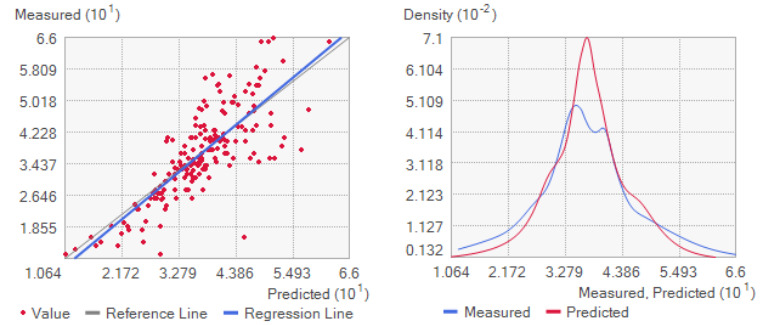
Sand



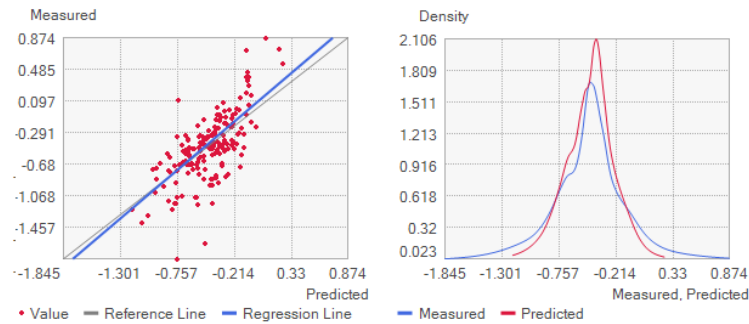
Silt.clr



Silt



Clay.clr



Clay

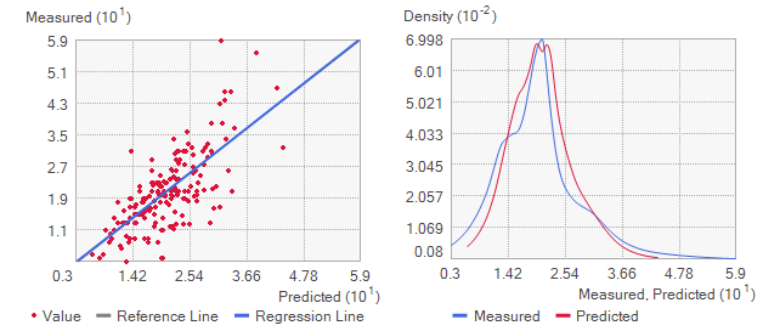
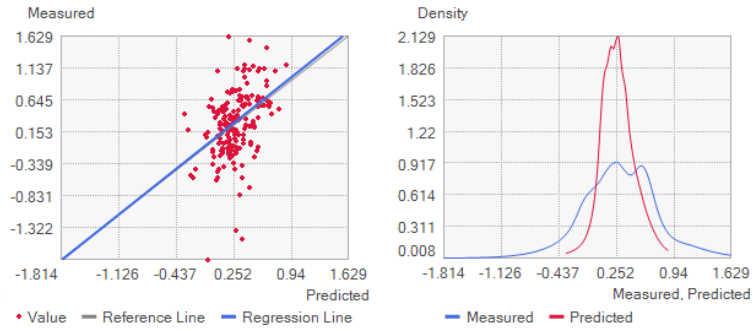


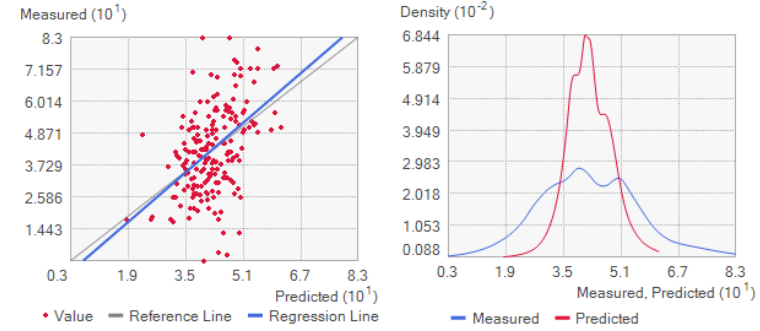
Figure 6.3 Scatterplots of observed and predicted values resulting from the CK approach with and without CLR transformation validated with LOOCV method. Model without interaction terms applied to the covariates.

EBKR

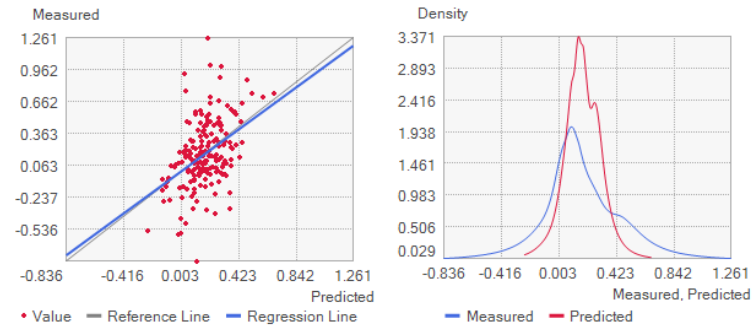
Sand.clr



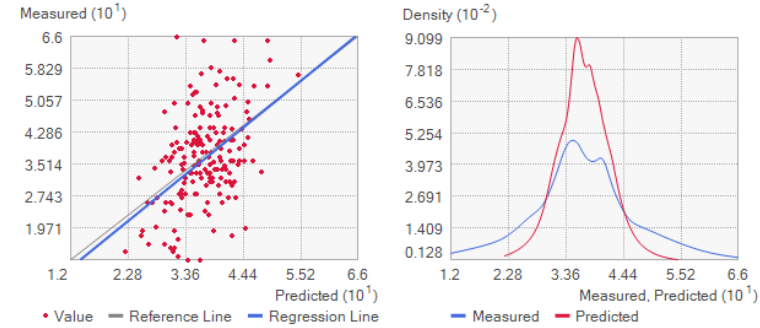
Sand



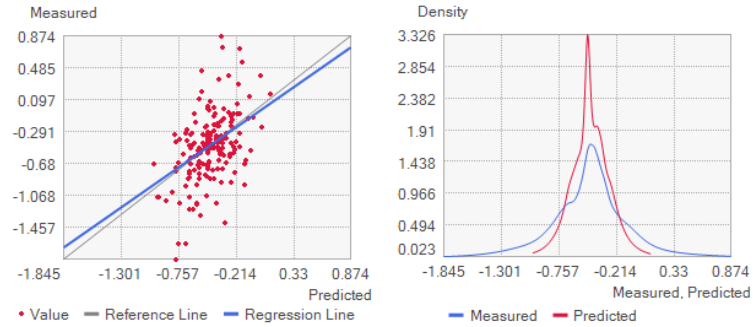
Silt.clr



Silt



Clay.clr



Clay

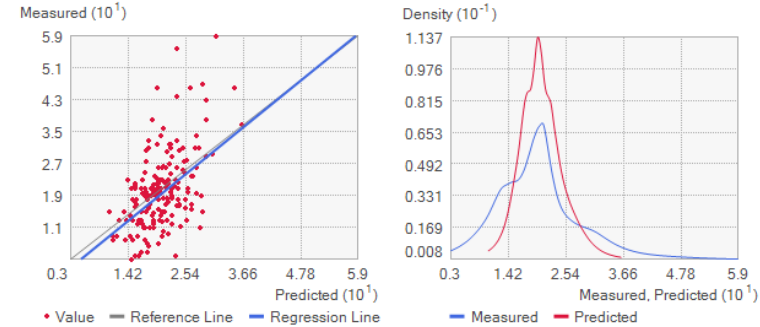


Figure 6.4 Scatterplots of observed and predicted values resulting from the EBKR model with- and without CLR transformation validated with LOOCV method. Model without interaction terms applied to the covariates.

Conversely, in terms of surface predictions (interpolated surface), EBKR visually showed better results, as the corresponding prediction maps (Figure 6.5) were found to follow the pattern and trend in the observed (measured) data (Figure 6.2), notwithstanding the scatterplots of observed and predicted values (Figure 6.4). The CK was not a good approach in terms of surface predictions (interpolation or geovisualisation) (Figure 6.6). The CoDa approach was useful for capturing the trend in the measured data in the spatial models regarding interpolation.

A nugget effect, which could mean either measurement error or microscale variation, was not observed in the semi-variograms (Appendix C – Figures C6.6 and C6.9). The later could be the case since the exploratory analysis plots and the semi-variograms showed no autocorrelation between pairs over short distances.

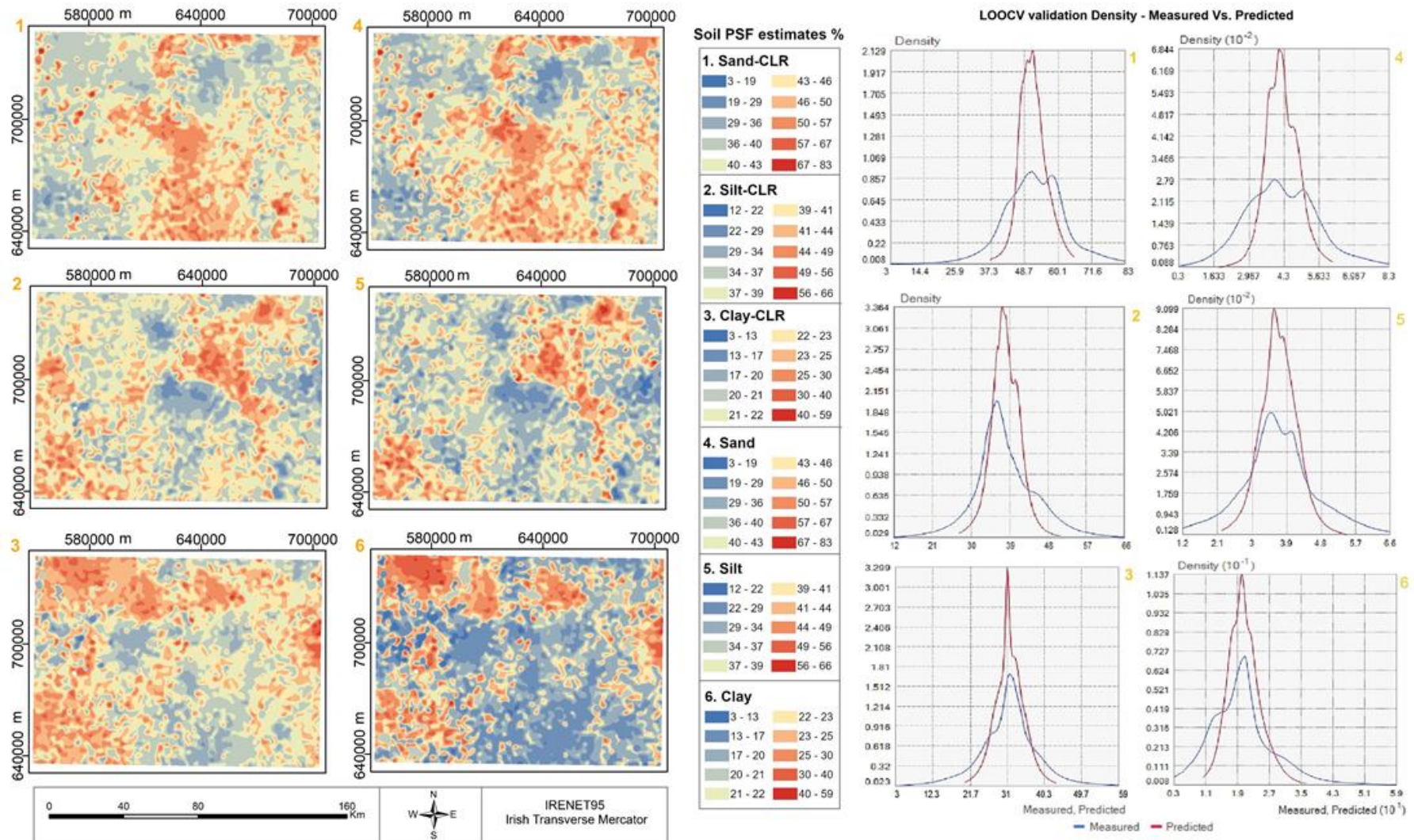


Figure 6.5 Surface predictions resulting from the EBKR approach with and without CLR transformation (1-3 and 4-5, respectively) and corresponding LOOCV validation plots.

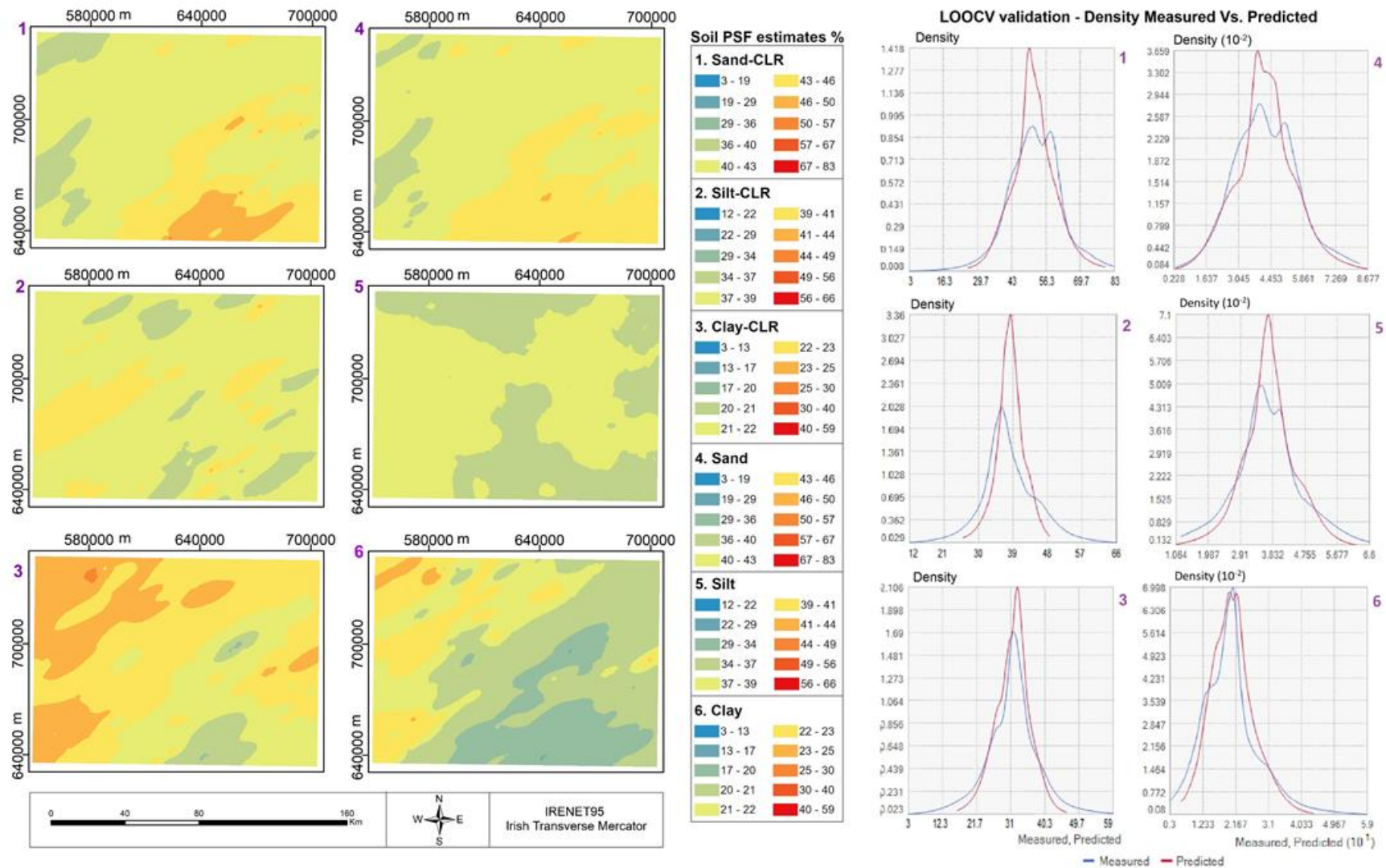


Figure 6.6 Surface predictions resulting from the CK approach with- and without CLR transformation (1-3 and 4-5, respectively) and corresponding LOOCV validation plots.

6.5.2 Non-spatial model

For the non-spatial model approaches (numerical models), the compositional LRM without 2-order interaction terms applied to the response variables yielded better results, according to both error-based and agreement-based metrics (Table 6.5). On the other hand, these results were similar to those of compositional RF, particularly for clay and sand fractions. In general, the use of the CoDa approach improve the predictions in the numerical models.

Table 6.5 Evaluation metrics resulted from the non-spatial modelling approaches validated with LOOCV

Modelling approach (Validation method: LOOCV)	SPSF*	Model evaluation metrics		
		RMSE	MAE	NSE
Model: Y-compositional LRM (ILR transformation; without 2-order interaction terms)	Sand	4.12	3.37	0.94
	Silt	7.08	5.44	0.62
	Clay	7.00	5.02	0.46
Model: Y-compositional LRM (ILR transformation; with 2-order interaction terms)	Sand	4.88	3.35	0.91
	Silt	9.08	6.45	0.37
	Clay	8.93	6.16	0.13
Model : Y-compositional RF (CLR transformation)	Sand	4.28	2.91	0.93
	Silt	5.55	5.52	0.59
	Clay	7.13	5.27	0.44
Model: Non-CoDa approach Simple LRM (univariate responses)	Sand	7.10	5.35	0.81
	Silt	7.92	6.10	0.52
	Clay	7.36	5.33	0.41

* SPSF: soil particle size fraction

6.5.3 Soil texture classes – all models

In general, soil texture classes obtained indirectly from the statistical models were similar for both the spatial and non-spatial models (Figure 6.7), except for EBKR, which differed in terms of the number of classes and the pattern display (non-linear) of the samples, assessed via the ternary diagram. The CK approach exhibited a similar distribution of soil properties in the ternary diagram (Figures 6.7f and 6.7g) to those of the non- spatial LRM fitted without 2nd order interaction terms (Figures 6.7b and 6.7d).

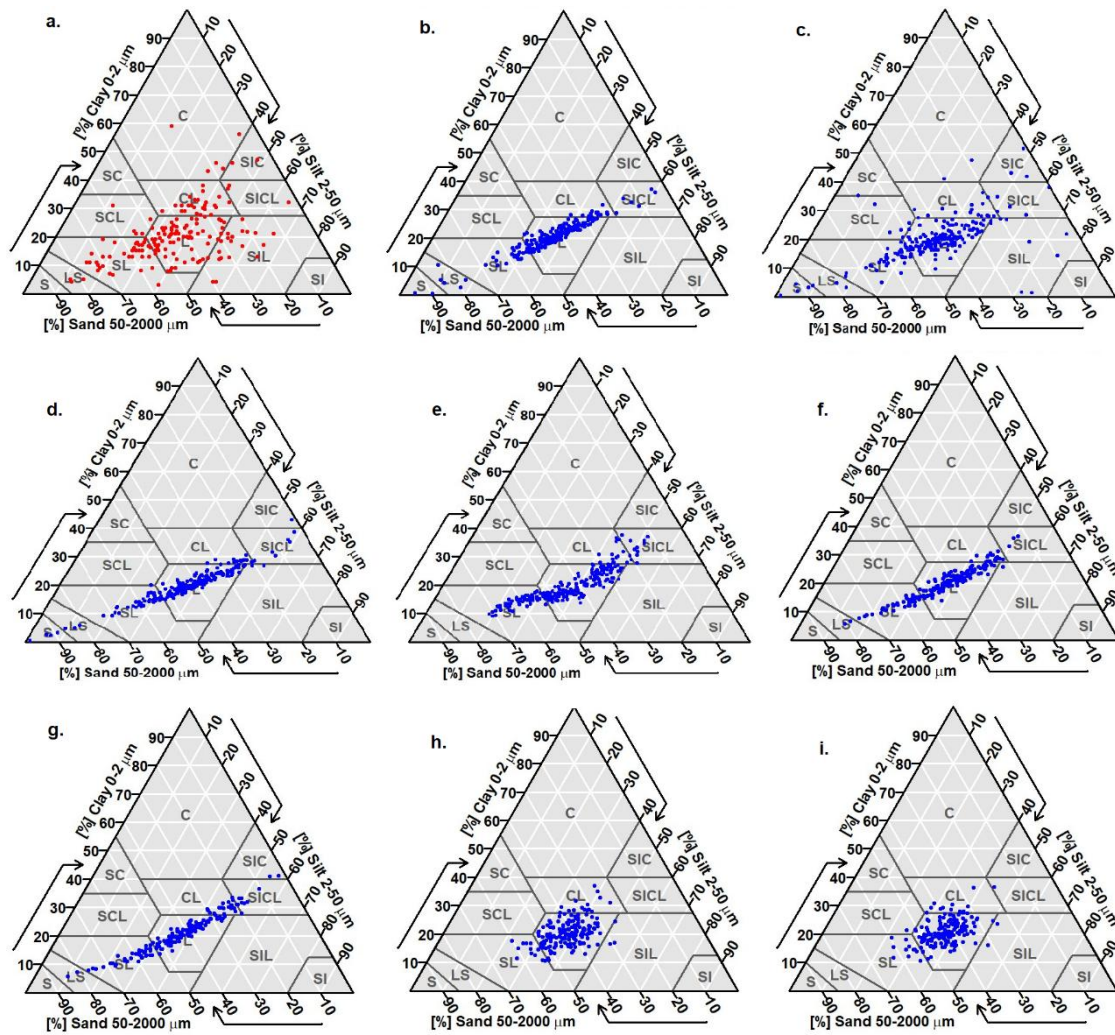


Figure 6.7 Observed Soil texture classes displayed in the USDA ternary and the resulting classes from both non-spatial (b-e) and spatial models (f-i). **a:** Observed (measured) data; **b:** LRM model, without log-ratio transformation and without interaction terms; **c:** LRM with ILR transformation and interaction terms; **d:** LRM with ILR transformation, without interaction terms; **e:** RF model with CLR transformation; **f:** CK model with CLR transformation; **g:** CK without CLR transformation; **h:** EBKR model with CLR transformation; **i:** EBKR without CLR transformation.

6.6 Discussion

Exploratory analysis showed dissimilarity in lieu of similarity in the data concerning spatial structure of the response variables and covariates in light of the negative GMI values. The semi-variograms could be considered as a dissimilarity function as the variance of the difference between pairs of observations increases with distance. A dispersed spatial pattern often reflects some type of repelling process between high and low values, which is consistent with soil PSF. For example, in this study, sand% ranges from 6-83.

The sampling size and density might have impacted the GMI values that produced zero and negative values, since configurations of samples (i.e., gridded, random, and clustered) can affect the degree of autocorrelation in data. The lack of stationarity in the data may also be related to both density and configuration sampling. The lack of spatial correlation amongst soil PSF in this study also indicates a lack of spatial dependence, that is, observations are independent of one another. This does not violate the main assumption in traditional statistics. However, it conflicts with a core tenet of geostatistics concerning the first Tobler's Law (spatial autocorrelation). In spite of that, it is worth noting that EBKR relies on classical statistics (i.e., LMM) in its regression part to perform the prediction. Furthermore, it appears that the second law of geography

(Goodchild's Second Law), which is a key of Kriging methods, holds due to spatial heterogeneity and the size of the study area (large area).

Notwithstanding the geostatistical results, the spatial models fitted were useful for spatially estimating soil PSF at unmeasured sites and for capturing spatial structure at larger distances. Moreover, CK and EBKR rely on linear regression for making predictions or interpolation. Whilst regional scale is the geographical analysis unit of the study area, the incremental spatial autocorrelation indicated a scale for analysing and mapping clay at a smaller unit (large scale) based on the spatial processes revealed by the peaks for clustering. (Appendix C – Figure C6.5). The geostatistical exploratory analysis undertaken here was important for fitting the cross-semi variograms and performing the CK and EBKR models. Moreover, since trend was observed and autocorrelation was not detected, the rationale for fitting a spatial model was to model the influence of spatial patterns in the response variables. Furthermore, while non-stationarity is the characteristic of the measured soil PSF according to Voronoi maps, CK was employed to compare methods. From this, the main finding in this study is that CK was associated with observed data whereas EBKR was related to their neighbours used to fit the local models (subsets) being useful for surface predictions (interpolation). Again, it is worth noting the sampling configuration of the data (dispersed) that may have influenced the results. Would a more-dense grid lead to different interpretations?

Regarding estimated values (PSF %), CK yielded better predictions for sand, silt, and clay than EBK regression, according to the NSE, RMSE and MAE metrics. From this, CK was a good model for predicting locally. Notwithstanding these results, it is important to note that the NSE values for CK substantially varied amongst PSF (e.g. 0.5-0.9) in comparison to more constant values from EBKR (e.g. 0.2-0.3). From the error diagnostics, EBKR exhibited more consistent plots for the predicted values and the standardised error as no pattern was observed in the graph for CoDa and non-CoDa approaches. Conversely, a pattern was noted only in the sand plot with CK model, associated with the predictor curvature (Appendix C – Figures C6.10-12). For clay, in terms of surface prediction (interpolation), CK yielded good result. This indicates that curvature has a nonlinear relationship and that the sand prediction might be biased with the CK model. In attempting to handle such a non-linearity, polynomial functions (cubic and 6-degrees) were applied over curvature for estimate sand Appendix C – Figure C6.13), however without success (Appendix C – Figure C6.14). Splines function were also employed, but they also were not efficient for addressing this.

Second-order interaction terms were particularly important for estimating sand and silt with CK, whereas they were not good with the EBKR model (Table 6.4 and Appendix C – Figures C6.15 and C6.16). They were good for clay with both CK and EBKR. Although interactions led to a slight overfitting for sand estimates, they improved the model's diagnostics. Results indicated that anisotropy should not be considered when interactions terms are employed in the models. This was clear for silt surface predictions whose semi-variograms were isotropic (no trend). This is an essential trade-off in spatial models, as interpolation is aimed.

Curvature was the most frequent covariate present in the interactions for sand; curvature and alpha, for silt; and slope and alpha for clay. Alpha accounts for microwave information from the surface and vegetation. RVI was another radar-based covariate present in the interactions, which accounts for vegetation. These are important pointers since they are related to surface moisture and present in clay and silt models interactions.

Moreover, they are sensitive to moisture. Interestingly, the surface soil moisture (SSM) and aspect covariates showed spatial dependence (spatial autocorrelation), and EBKR and CK models captured it through the semi-variograms. The spatial relationship between radar variables, soil moisture, and topography (relief conditions) was clear from the results, especially for clay whose surface prediction maps reflected these relationships.

Concerning the map of surface predicted without interactions, the EBKR model was a better approach to capturing trends and patterns, particularly with CLR transformation, as the interpolated surface was similar to the Voronoi maps and neighbourhood analysis of the measured data. Cokriging is not optimal as it assumes stationarity. Empirical Bayesian Kriging (EBK) is useful for dealing with non-stationarity as it performs Kriging geographically apart to capture underlying processes in different spaces, treating local variance separately in lieu of global variance. From this, EBKR appears to capture the pattern and trend of the response variables at small scale (larger areas), since it predicts locally by modelling each semi-variogram for each subset. This finding is consistent with Wadoux *et al* (2020) who also found that the Bayesian calibration enabled to capture model input, initial state, parameter and structural uncertainty at a river basin scale in Switzerland. EBKR also deals with non-stationarity, which is an intrinsic characteristic in agricultural areas due to the complex variability of the soil properties (Gribov and Krivoruchko, 2020). Moreover, one way to deal with the bias-variance (or accuracy-precision) trade-off is to use mixture models, which is the case of the EBKR model on its regression stage.

In terms of the comparison between the predicted (interpolated) and observed (in situ) data, the spatial models were useful for predicting clay, which did not occur with the non-spatial models in terms of [numerical] estimates. It would appear that the spatial models captured some relationship between clay and the predictors employed. This might be related to the fact that spatial modelling seeks to find and interpret spatial relationships, patterns, trends, and variations within a geographical area. This was noticed in this study as results from the exploratory analysis indicated patterns and trends in the response variables. Furthermore, clay was the only fraction to show autocorrelation over relatively short distances – a minimum of 7.5 km. The models also seemly capture that.

Regarding soil texture (categorical), CK was more consistent with the observed classes in the USDA ternary as CK yielded better estimates in terms of PSF% than EBKR. Ternary diagrams are a function of numerical values of the parts (compositions). Unlike soil PSF, the CoDa approach had no impact on soil texture since the classes obtained indirectly by numerical models (inferences) were similar with and without log-ratio transformation applied to the soil fractions. Hence, unlike PSF (continuous), the compositional nature of the response variables is not an issue when dealing with soil texture (categorical).

The CoDa approach may be important for digital soil mapping because the sum of the soil PSF is equal to 100% with that technique. Nonetheless, results showed that treating soil texture as compositional data was not worthwhile for the estimates in the spatial models as it did not lead to any substantial improvement in terms of numerical PSF content %, except for the interpolated surface (spatial predictions). Conversely, the CoDa approach resulted in more significant findings in the non-spatial models. Amirian-Chakan *et al.* (2019) also achieved such finding in their study to investigate the effect of soil texture on the spatial estimation of available

soil water capacity and the total amount of irrigation water using log-ratio transformations methods (ALR, CLR, ILR). They observed that no significant difference between transformation method in the estimates.

Notwithstanding the findings in the spatial model with CoDa, since there was no effect of it in the models, such an approach is important for soil PSF to guarantee the component unit (i.e. sum equal to 100%) in a predicted sample.

6.6.1 Strengths and limitations of the work

This work demonstrated the relevance of comparing two contrasting modelling approaches for estimating soil PSF – more statistical-based (non-spatial models) and more geographical-based (spatial models) – providing a general basis for mapping soil PSF and soil texture, particularly in Ireland.

Another highlight regarding the potential relevance of this study to soil properties mapping builds upon the fact that the models indicated soil particle size estimation as more spatial pattern processes-based rather than spatial autocorrelation processes, following the spatial configuration of the measured data used.

Nevertheless, this work is not without limitations. One caveat is related to the interplay between sample size and area size under investigation for conducting a geostatistical approach as this type of analysis is reliant on a grid configuration. A second limitation is the lack of the dielectric constant of the soil as a covariate in the models, which could be useful to understand the interactions between sand-silt-clay and SAR signal through backscattering coefficients (σ_{VV}^0 and σ_{VH}^0) due to interstice water content into soils, thereby improving the results in the modelling approaches.

6.7 Conclusion

The findings demonstrate the relevance of considering the geographical location (spatial condition) and the compositional nature of soil texture in both statistical modelling approaches because of the constraint sum of the parts (sand, silt, and clay), which may provide the basis for performing digital soil mapping of soil texture. Due to the complex nature of spatial information, a spatial inductive bias, which inherently occurs as a result of a certain type of spatial structure present in the data such as spatial heterogeneity and spatial autocorrelation (collinearity), is incorporated into spatial model's realms. It is thereby important to know the foundation and formulae behind them to choose the proper approach for an application.

The explicitly spatial models did not outperform the non-spatial models in estimating soil PSF in terms of numerical estimates. On the other hand, when seeking surface predictions, EBKR provided good results for soil PSF mapping, particularly regionally (small scale), while CK was better for local estimates. Concerning clay fraction, both the spatial models here evaluated were found to be good approaches for surface prediction. Moreover, spatial direction should be considered with caution in semi-variograms when interaction terms between covariates are employed.

Since tree-based models do not rely on any prior assumptions, notwithstanding a black-box modelling approach, their spatial approaches such as the spatial Random Forest could be employed to estimate PSF, with the covariates used here, to deal with linearity to improve surface predictions. Bearing in mind the trade-off between model evaluation metrics and model diagnostics. In relation to non-spatial approaches, when

interactions between variables matter in a study for estimating PSF, the RF model could be employed (Chapter 4).

Regarding the effects of location and neighbourhood in the models, in light of the sample configuration, autocorrelation could not be detected in the semi-variograms at short distances. Neighbourhood was hence measured as a dissimilarity function, which was observed in the surface prediction obtained from the CK and EBKR models and in the trend analysis.

Radar-based covariates and topography-based covariates can be employed in prediction models to estimate topsoil PSF as relief conditions may account for roughness surface in radar signals to a certain extent.

Concerning the presented research approach, this work addressed elements from the interconnected knowledge fields of soil, geomorphology, remote sensing, and geographical information systems. For example, with the aid of topographical covariates, this work stressed the role of process-based analysis rather than pedogenesis of topsoil (regarding both soil grainsize fractions estimation and soil texture prediction), particularly because relief conditions influence remote sensing imagery. Regarding statistical modelling approaches for estimating soil properties, depending on the research purposes and the research questions and whether geographical location matters, researchers can decide between a spatial/geostatistical and a non-spatial modelling approach. Importantly, this work provides a methodological framework to inform the design of costly and labour intensive in situ soil surveys (areas of large uncertainty estimates in the models are those areas that may require detailed soil surveys). It also provides a means to estimate soil properties over large spatial areas and generate new data products for use in hydrological, land surface (e.g. Noah LSM models), climate and other model-based approaches that currently employ coarse global scale soil texture products. The methods outlined here are also relevant to locations where optical remotely sensed data is limited and/or where vegetation cover is present throughout the year.

Thus, the main contribution of this chapter is that in case predicting soil textural classes is the aim of the research, the non-spatial model, especially using radar and geophysical covariates could be a better approach. Spatial models could be useful for obtaining the distribution of sand, silt, and clay estimates, and to indicate areas that may require detailed soil surveys.

7.1 Introduction

In addition to the results and findings, this chapter provides a discussion on the limitations of the data, models, and methods employed in this study; the role of geographical scales in the models and, consequently, for the estimates; and a brief discussion on the implications of big/open data from remote sensing to this research.

7.2 Results and findings from the methods employed

Although a discussion of the results and findings was presented in each of the three methodology chapters, this section aims to bind them to bring cohesiveness to the work. To achieve this, the research questions and objectives presented in the introductory part are here resumed with a discussion on how and to what extent they were met with the methods employed.

- *How effectively can SAR data be employed to retrieve particle size fractions over topsoil in areas like Ireland?*

Extensive areas of permanent vegetation (grassland) and a dominant climatic factor in which precipitation exceeds evaporation (O’Sullivan *et al.*, 20180), leading to permanent soil moisture, affects radar backscattering due the dielectric constant and clayey soils (Dobson *et al.*, 1981; Hallikainen *et al.*, 1985; Jackson and Schmugge, 1989; Srivastava *et al.*, 2006; Das & Paul, 2015). Such land use/land cover and climate context emerge as a challenge for employing SAR data in areas like Ireland, particularly for soil analysis. On the other hand, remote sensors can capture information from soils over agricultural fields, through the gaps between the crops. Such trade-off between climate and land cover factors thereby was beneficial for employing polarimetric decomposition method in this research to deal with vegetated topsoil.

Concerning the evaluation of SAR data being employed to retrieve particle size fractions over topsoil in the study area, key findings from the modelling approaches highlighted that SAR data were unable to predict soil particle size fractions without the aid of covariates, since the models evaluated in this study did not identify their importance as main covariates (Chapter 4. DOI 10.1111/ejss.13414). Typically, cross polarisation (VH) measurements are significantly lower than the co-polarisation (VV) (Ulaby *et al.*, 1978; Ulaby *et al.*, 1979; Sano *et al.*, 2020), and this was observed over the study area wherein the VV intensity values (σ^0) ranged from 0.01 to 0.31, whereas for VH polarisation the backscatter coefficient varied between 0.02 and 0.07 (Chapter 4. DOI 10.1111/ejss.13414). In relation to the topography covariates being identified as important predictors and machine learning models (e.g. RF) not performing well in some situations, these general findings are consistent with Schönbrodt-Stitt *et al.* (2021) and Mirzaeitalarposhti *et al.* (2022), who also applied Sentinel-1 data and topography covariates to estimate soil particle-size fractions and soil moisture (Chapter 4. DOI 10.1111/ejss.13414).

Nonetheless, in light of the challenging environmental setting for the use of optical remote sensing and having soil permanently covered by vegetation, some strategies can be highlighted to predict soil PSF from Sentinel 1-SAR (Chapter 4. DOI 10.1111/ejss.13414):

(i) The linear model (compositional Y-LRM), without 2nd-order interaction terms, was found to be the optimum approach; where it is necessary to consider interactions between variables, the RF approach should be employed.

(ii) To obtain the soil textural classes from the estimates of PSF (sand%, silt %, clay%) – instead of predicting the classes using a classification method – the compositional Y-LRM with interaction terms applied to the covariates is useful as the classes were more consistent with the ternary diagram of the measured data than those obtained without interaction. This finding may be useful for subsequent use in models, which require soil textural class and not soil textural components (i.e. sand, silt, and clay individually).

(iii) LOOCV is a better validation method over randomly splitting data for dealing with small sampling size (e.g. n=235).

(iv) Employing geophysical data with SAR data did not improve the model estimates of soil particle fractions; the inclusion of geophysical covariates resulted in poor model estimates. However, the use of geophysical data was found to result in soil textural classes that better matched the textural classes derived from the measured data.

These findings are consistent with Niang *et al.* (2014), in their study in Rouville County, near Montreal, Canada, who concluded that using SAR polarimetric data extracted from a RADARSAT-2 image as covariates (monopolarization SAR channels and the alpha/entropy/anisotropy/ polarimetric parameters extracted from polarimetric decomposition methods) was useful for digital mapping of soil surface texture in Rouville area.

▪ *Is there a correlation between the backscattering coefficient and sand, silt, and clay fractions?*

According to the findings from Chapter 4, no linear or direct relationships were found between the backscattering coefficients (σ_{VV}^0 , σ_{VH}^0) and the soil particle size fractions in the models evaluated. Notwithstanding the CoDa log-transformations slightly improved the correlation coefficients, but the values remained low. This is consistent with Gorrab *et al.* (2015) who argue that soil texture has only a very limited influence on radar signals, and is difficult to directly retrieve one component of soil texture (e.g., clay) from such data.

Despite their low importance, the VV and VH covariates (i.e. backscattering of σ_{VV}^0 and σ_{VH}^0) were found to be related to sand, silt, and clay in the compositional linear regression models. For example, it was observed an increase in VV resulting in a decrease in the clay estimate and an increase on sand; an increase in VH led to an increase in the clay estimate. Such findings are aligned with some results from the literature reviewed, such as in Han *et al.* (2017) who found a weak correlation in a study investigating the association between radar backscatter – using UAVSAR data – and in situ soil property measurements on an approximately 3 km long section of earthen levees along the lower Mississippi river (riverside and landside). While UAVSAR is an aerial Earth Observation (EO) platform, Sentinel 1 is an orbital-based EO platform (about 700 km altitude) and thus more distant from the targets. This means a loss of energy equal to the power of 4 (Wolff, 2007)

according to the radar range equation (Figure 4.1). Notwithstanding that, the statistical behaviour between the backscattering coefficient and soil texture in both studies was found similar.

Furthermore, Niang *et al.* (2014) found that using a nonlinear support vector regression provided the best prediction accuracy compared with the geostatistical interpolation techniques employed and the linearly dependent assumption, in their study for digital mapping of soil surface texture with RADARSAT-2.

The possible reasons for a lack of a linear relationship between the backscattering coefficient and soil texture in this research are: (i) soil textural configuration that affects soil dielectric properties, regarding water-holding capacity (Hallikainen *et al.*, 1985; Das & Paul, 2015); (ii) pre-treatment stage of soil laboratory analysis wherein organic matter and moisture are removed from the soil samples, whose laboratory results may not be consistent with the signal backscattering informed in the SAR data; and (iii) the challenge of matching the date on which samples were collected in the field and Sentinel 1 was passing over the site to reflect environmental conditions and or land use/land cover; (iv) soil sampling design and pixel image size. Items (ii), (iii), and (iv) are more sensitive when dealing with legacy data.

Moreover, in their study, Gorrab *et al.* (2015) found a relationship between the measured decrease in intensity of backscattered radar signals and the fraction of clay present in the soil, with radar data acquired during rainfall months. A similar tendency was also noticed in this research for clay content between 40% - 60% under the meteorological conditions for Sentinel-1 image acquired (6 days of no rainfall before the acquisition date on the 3rd of April 2021 and the rainfall for the preceding month of March was below the long-term average). That is, the backscattering coefficient slightly decreased when the content of clay became higher. The results here may be influenced by the presence of soil moisture, as precipitation exceeds evaporation in Ireland (O'Sullivan *et al.* 2018), and the relation between free water (sandy soils) and bound water in soils (clayey soils) (Dobson *et al.*, 1981; Jackson and Schmugge, 1989; Srivastava *et al.*, 2006; Das & Paul, 2015), since soils having predominately clay-sized particles retain a much higher volume of water as a result of the predominance of micro-pores (Gorrab *et al.*, 2015). These factors affect radar backscattering as a function of soil texture as highlighted in Chapter 2. Furthermore, goodness of fit metrics were found to vary depending on the soil fraction. This denotes the complexity of soil texture for statistical modelling due to the magnitude (size) of the soil fractions and the mineral constitution, as well as the reliance of the signal intensity from the SAR beam on how clayey or sandy the soil is (Chapter 4, DOI 10.1111/ejss.13414).

- *To what extent can polarimetric decomposition predict soil particle size fractions? Can surface scattering ameliorate vegetation cover issues for soil particle size estimation?*

The calculated Alpha values indicate that the study area is mostly represented by volume scattering mechanisms indicating an occurrence of vegetation, consistent with the land cover conditions found on the Sentinel-1 image acquisition date. H-alpha dual pol-decomposition was found to be useful to separate the SAR backscattering from vegetation and soil in terms of surface and volume scattering in the study area, as demonstrated in Chapter 5 (DOI 10.1016/j.jag.103742), when compared to Sentinel-2 imagery. Thus, the H-alpha Dual-Pol Decomposition method provided, to a certain extent, surface scattering over low vegetation,

which may indicate underlying soil, despite using a C-band radar, which has a lower ability than an L-band radar to penetrate (low) vegetation and reach the soil (Babaeian *et al.*, 2019).

Salma *et al.* (2022) also found a match between the scattering shown on the H- α plane (from Sentinel-1) and field's backscattering when analysing the behaviour of the crop field with its growth, which means that dual-pol decomposition, to a certain extent, dealt with gap space between soil and plants on the crop fields.

On the other hand, Ji and Wu (2015) observed that HH-HV and HV-VV SAR could not effectively separate surface, dipole, and multiple scattering mechanisms due to the lack of co-polarization (HH-VV), indicating that co-polarization is vital for extracting scattering mechanisms (Ji & Wu, 2015). They argue that the scatters of most zones strongly diffuse and transfer in the H- α plots for HH-HV and HV-VV SARs (Chapter 5. DOI 10.1016/j.jag.103742).

The surface scattering mechanism improved estimates in the models in this research, nevertheless, under low magnitude. One possible reason for this may relate to the fact that the polarimetric decomposition method employed is an adaptation of the method originally developed for quad- or full polarisation. Hence, the lack of another co-polarisation (HH) and cross-polarisation (HV) in Sentinel-1, mainly HH, somewhat limited the target decomposition. The results reflected the dual-pol S1 limitation, highlighted in the literature, to derive alpha values for surface scattering retrieval under vegetated soil. A second potential reason is owing to the lack of the dielectric constant of the soil in the modelling approaches, which could be useful to analyse and exploit the relationship between sand-silt-clay and the microwave backscattering coefficients (σ_{VH}^0 and σ_{HV}^0), delving deeper into the physical scattering properties and physical soil properties.

Notwithstanding the caveats raised by Ji and Wu (2015) and Mascolo *et al.* (2021), dual-pol decomposition has been employed, for example, for feature identification, for classification and for biophysical parameter estimation (Chapter 5. DOI 10.1016/j.jag.103742). For example, Harfenmeister *et al.* (2021) identified alpha, entropy and anisotropy as important inputs in their model, in their study of crop growth monitoring with Sentinel 1 in Northeast Germany. Moreover, Suman (2022) concluded that decomposition parameters were more meaningful to forest structure and feature identification than biomass information, in their study on land feature detection and biomass estimation over tropical deciduous forest in India, using ALOS-PALSAR.

▪ *What is the effect of treating soil particle size as compositional data into the models?*

Chapter 4 demonstrated that, based on the results from the prediction models, it was worthwhile to explicitly treat soil PSF as compositional data in the statistical analysis and the models, for the non-spatial modelling approach. This is consistent with the findings of Loosvelt *et al.* (2013) and Chappell *et al.* (2019). Results from Chapter 4 also demonstrated that the compositional linear model slightly outperformed the tree-based model.

Conversely, from the spatial modelling approach, results showed that treating soil PSF as compositional did not improve the model estimates. Amirian-Chakan *et al.* (2019) highlight similar findings in their study to investigate the effect of soil texture on the spatial estimation of available soil water capacity and the total amount of irrigation water using log-ratio transformations methods (alr, clr, ilr). They observed that no

significant difference was found between transformation methods in the estimates. The compositional data approach has increasingly been employed to estimate soil particle size fractions in explicitly spatial models, as demonstrated by Odeh *et al.* (2003), Wang and Shi (2017), Wang and Shi (2018), and Zhang and Shi (2019).

Notwithstanding the findings in the spatial model with CoDa, since there was no effect of it in the models, such an approach is important for soil PSF to guarantee the component unit (i.e. sum equal to 100%) in a predicted sample. Thus, considering a geo-spatial perspective in the modelling approaches, including compositional formulation might be a useful approach to correctly estimating statistical parameters.

- *Can soil texture classes be indirectly obtained from the quantitatively predictive models?*

From the non-spatial models (Chapter 4), it was found that soil textural classes can be obtained from the estimates of sand (%), silt (%), and clay (%) using a regression model. This was observed with the compositional Y-LRM with geophysical covariates and interaction terms applied to the covariates since the soil textural classes modelled were more consistent with the textural classes derived from the measured data. This may be due to the structure of this model that captured the relative contribution of each grain component (sand, silt, and clay) in the samples, following the intrinsic characteristic or principle of soil texture, and some relationship with the geophysical covariates. In the spatial models (Chapter 6), such a result was not observed.

This finding is also likely due to the effect of soil texture on radiometric signals. Read *et al.* (2018) found that gamma radiometric thorium (Th) and potassium (K) both had strong relationships with clay (%) and sand (%). Cattle *et al.* (2003) found that topsoils with high clay contents were distinguishable from sandy soils using radiometric Th and K. However, findings from this research require further exploration of the geophysical data and soil PSF, since such covariates contributed to obtaining more consistent classes (soil texture; categorical) rather than estimates (soil PSF % ; continuous or numerical) in the regression models with interactions terms (Chapter 4, DOI 10.1111/ejss.13414).

The caveat notwithstanding, this finding may be useful for subsequent use in models requiring soil textural class and not soil textural parts (i.e. sand, silt, and clay separately).

- *Should we employ spatial or non-spatial models to estimate soil grainsize fractions? What are the difference between them in terms of performance, accuracy and potential issues?*

Concerning the first part of this research question, the response relies on the work's purpose, as geographical location or geo-coordinates may not be crucial to achieving the aims and objectives of the phenomenon to be investigated or the problem to be solved. Thus, this research provided the basis for both contexts (geographical and non-geographical).

Considering the first research scenario, results from the geostatistical exploratory analysis outlined in Chapter 6 indicate that, locally, the data sampling represents a non-stationary process, which is observed in the Voronoi diagrams depicting local variance and local variation (Appendix C – Figure C6.4). These results are consistent with Singh and Sarma (2023) who highlight that in soil properties the spatial correlation may vary across the region. This means that the second law of geography (spatial heterogeneity) was achieved, due to large area.

Despite Wang & Shi (2017) concluding that to improve spatial interpolation performance, a combination of geospatial methods and the use of ancillary variables is necessary, many of the traditional methods may not reflect the specific requirements of compositional data for interpolation; findings in Chapter 6 highlighted that the spatially explicit models did not outperform the non-spatial models in estimating soil PSF in terms of the model's performance in light of the evaluation metrics. In general, non-spatial models yielded higher accuracies. Niang *et al.* (2014) also found that using a regression model (support vector regression-SVR) was better than the geostatistical interpolation techniques (ordinary kriging-OK, and regression kriging) employed in their study for digital mapping of soil surface texture with RADARSAT-2 in Canada. Their results showed that SVR produced the best prediction accuracy, since RMSE values were reduced by 18% for sand, 17% for silt, and 35% for clay, leading to an improvement of the digital mapping.

Here, the spatial cokriging exhibited a quasi-similar accuracy to that of the non-spatial models. Further, while non-spatial models were better for sand estimates, the spatial approaches showed better estimates for clay. On the other hand, when seeking surface predictions (interpolation), EBKR provided good results for soil PSF mapping, particularly regionally (small scale), while CK was better for local estimates. The EBKR model was a better approach to capturing trends and patterns, particularly with CLR transformation, as the interpolated surface was similar to the Voronoi maps and neighbourhood analysis of the measured data. Cokriging is not optimal as it assumes stationarity. Wadoux *et al* (2020) also found that the Bayesian calibration captured initial state, parameter and structural uncertainty at a river basin scale in Switzerland. EBKR also deals with non-stationarity, which is an intrinsic characteristic in agricultural areas due to the complex variability of the soil properties (Gribov and Krivoruchko, 2020). A general issue is related to the relatively small sample size to train and test both modelling approaches, whereas the non-statistically randomness of the soil survey data ("free sampling") could be an issue for the spatial models.

- *To what extent or under what environmental conditions can the method be reproduced in another study area?*

Addressing this research question is twofold, centred on the statistical modelling approaches and the remote sensing. Regarding the former, the methods used in the non-spatial models for estimating soil particle size fractions and predicting soil texture in the topsoil can be employed in areas other than Ireland. The strategies highlighted to predict soil PSF from Sentinel 1-SAR with non-spatial models, from Chapter 4, can also be applied to aid this endeavour.

The spatial modelling framework can also be replicated in other areas provided that parametrisation is tailored according to the size of the area under investigation and the geographical/cartographical scale, as well as exploratory geostatistical analysis being carried out. As geographical coordinates matter, spatial models usually differentiate behaviours, patterns, trends, and predictions according to spatial locations. This is especially valid when dealing with variograms. Further, the methods here employed can be used to create soil texture maps at field scale (greater cartographical scale) – provided that there is more highly spatially resolved data available – and larger areas (smaller cartographical scale). EBKR model is an option to account for non-stationarity in the variables. Instead of fitting a single model to the entire data, it uses local models to capture

small-scale effects over larger areas. Therefore, combining the local surfaces together to get the final surface this method can generate soil PSF/soil texture maps at both greater- and smaller scales.

Geophysical data were not good predictors for estimating PSF in topsoil. However, the use of geophysical covariates resulted in improved estimates of the soil textural classes. Electromagnetic, magnetic and radiometric signals may be more related to deeper soil layers, due to their relationship with the underlying rock, while topsoil may be influenced by topography conditions such as slope and curvature in terms of surficial materials deposits (geomorphological process) (Chapter 4, DOI 10.1111/ejss.13414). The study region investigated, as an extensive landscape, is relatively heterogeneous concerning these geomorphological variables. Interpretation of radiometric data where parent materials of transported sediments have different origins is more challenging, which suggests a natural limit on soil textural interpretation of soil radiometric data (Read *et al.* 2018; Wilford *et al.*, 1997). The findings from this research highlight the need for further investigation of radiometric data and soil properties in the study region (Chapter 4, DOI 10.1111/ejss.13414).

From a remote sensing standpoint, the method employed for dealing with vegetated soil, using Sentinel-1, can also be applied to areas other than of this research. Due to the physical properties targeted by remote sensing techniques and the size of the study area, only cautions concerning pre-processing SAR data, such as speckle filter, window size, co-registration, and calibration are required. Terrain derivatives from (finer) digital elevation models (DEM) may relate to physical processes in soil that are of great importance for SAR. Further, when geomorphological conditions vary significantly (from flatter to mountainous relief), gamma naught (Small, 2011) is recommended as calibration method for the backscattering coefficient, as demonstrated in Navacchi *et al.* (2023).

It is also worth noting that, in case the objective of the research is an explanation of the underlying soil process, apart from seeking pattern detection/prediction, it is worthwhile considering variables related to soil forming factors from the CLORPT² model (Jenny, 1941) and the SCORPAN³ model (McBratney *et al.*, 2003) for non-spatial and spatial modelling approaches, respectively.

7.3 Limitations of the data, models, and methods employed in this study

Apart from informing treatment for data, models, and methods, the subsequent topics in this chapter are also useful to provide an understanding of uncertainty in the estimates, or to a certain extent, to justify them.

7.3.1 Concerning field data from the soil legacy surveys

It was important to look up information in reports, field spreadsheets, and metadata accompanying the legacy survey. Reusing data from different soil surveys usually means dealing with surveys being carried out with different soil sampling methods at variable soil layers/horizons, and varied laboratory analysis methods employed to obtain proportions of sand, silt, and clay fractions – they lack a common standardisation. Such issues are more problematic when dealing with deeper soil layers. In the topsoil case, the limitation of commonly available soil data for remote sensing application is related to the pre-treatment stage where organic

² Climate, Organisms, Parent Material, Time and Topography.

³ Soil (s), climate (c), organisms (o), relief (r), parent material (p), age (a), and spatial location (n).

matter and moisture are removed from the soil samples and whose laboratory results may therefore not be consistent with the signal backscattering informed in the SAR data – for the case of a satellite passing over at the same date of the field survey. Soil sampling methods differed between spade (e.g. LUCAS) and auger (e.g. LUCAS, SoilC-2008, SIS-Teagasc), whilst laboratory analysis methods varied from the pipette (SoilC2008, SIS-Teagasc) and laser diffraction (LUCAS). Notwithstanding, Fernández-Ugalde et al. (2020) demonstrated that the spade and gouge auger methods produced similar results for all soil properties with the LUCAS topsoil dataset, and concluded that, in general, the relation and average magnitude of the differences for clay, silt, and sand contents between the two sampling methods were satisfactory. Thus, despite differences in methods, they are consistent in considering topsoil only for the C-band SAR application in this research. Moreover, the use of Sentinel-1 data acquired in drier meteorological conditions (both on date and prior passage of satellite over the study area) justifies the image's date selected for this research.

7.3.2 Concerning Sentinel-1 data

Typically, in remote sensing, it is known that variability of environmental and terrain conditions influence both proximal and remote sensors due to weather and/or meteorological aspects. Since object(s) are located far away from the sensor, approximately 800 km, the target's Signal-to-Noise-Ratio (SNR) might be equal to 1. Concerning the total of observations ($n = 235$) used in the models fitted in this thesis, that is, 235 pixels, the average of SNR ratio in the image corresponding to VH polarisation is 0.8 with the preprocessing methods employed in this study, whereas this ratio is equal to 0.6 for VV polarisation.

Sentinel-1 is a C-band SAR operating with dual polarisation (VV and VH), which poses some limitations. Definition of geometry of surface in microwave remote sensing is a function of SAR wavelength. SAR is sensitive to roughness, but the definition of a rough or a smooth surface varies according to the use of a C, L or X band, for example. Water content may change the spectral fingerprint or backscattering, while surface roughness may lower the reflectance values due to micro shading (Ogen *et al.*, 2022) that can be detected or not depending on the radar band. Polarisation is another limitation for better target distinction or scattering mechanism separation in a pixel when working with Sentinel 1 data. A major limitation, in terms of decision on datasets, was the use of a single sensor and a single satellite imagery at a specific date to be applied in a single study area, notwithstanding the use of a time series of Sentinel 1 satellite images or a comparison of different SAR platforms (and wavelengths) and locations in Ireland might lead to an enhancement of the work. Such procedures would have also been useful for results validation and demonstrating reproducibility.

Even when employing SAR with longer wavelengths, data and methods are not without limitations. Gorraeb *et al.* (2015) identified the number of neighbouring pixels needed to minimize the effects of local terrain heterogeneities and speckle in the radar signals when employing a X-band radar (TerraSAR-X), by suggesting a small number (i.e., 3×3). Ultimately, notwithstanding the experiments being designed in bare soils, Gorraeb *et al.* (2015) concluded that it should also be possible to implement the proposed algorithm on surfaces with vegetation cover. They also conclude that the approach could also be applied to Sentinel-1 data.

As Gorraeb *et al.* (2015) highlight, soil texture has only a very limited influence on radar signals; therefore, it is difficult to directly retrieve one component of soil texture from such data. Direct determination of soil

texture through SAR images is a difficult task because its textural configuration as compositional data poses a change in soil dielectric properties, regarding its water-holding capacity (Hallikainen *et al.*, 1985; Das & Paul, 2015). As a result, environmental and topographical covariates are needed for making predictions.

A limitation of dealing with soil legacy data (field data) together with remote sensing data is also pointed out in terms of the challenge of matching the date on which samples were collected in the field and Sentinel-1 was passing over the site to reflect environmental conditions and or land use/land cover. Therefore, looking up technical and methodological information in metadata from both datasets was the procedure taken in this study to alleviate such limitations.

7.3.3 Concerning models

Despite this research employing different statistical models in terms of their structure/nature (traditional statistics and geostatistical) and methods (linear model, machine learning or tree-based models, and semi-variogram-based models), the research is not without limitations on this matter.

The primary limitation relates to the number of observations (field sampling from the soil legacy dataset) used to train and test the models as well as the sampling design and/or spatial distribution of samples available for the study area. The criteria used to select samples were that they needed to be consistent with the capability of the C-band SAR signal to reach and interact with the soil, such as shallow topsoil (15 cm depth), and that the surveys addressed the analysis of soil particle size fractions. In turn, these pre-selection procedures resulted in a relatively small sample size in light of the study area size (i.e. low density of points) and in a sparse and randomly samples distribution (i.e., not centred-gridded). The sampling design is important as it affects map accuracy owing to the number spatial observations used to calibrate a model for spatial prediction (Wadoux, 2019). In geostatistical methods, the strength of determined autocorrelation depends on the density and pattern of sampling, and on the variogram characteristics (Brus & Gruijter, 1997). Autocorrelation should be strong to increase the efficiency (Brus & Gruijter, 1997). On the other hand, soil sampling is costly and labour intensive activity at large areas (small scale). Notwithstanding that, the dataset subset accounted for the heterogeneity of the study area as depicted by the geostatistical methods used. The sampling strategy for the training data set is essential to cover the heterogeneity of the phenomena of interest, while also bringing efficiency because collecting spatial samples is an expensive undertaking. (Nikparvar & Thill, 2021).

Second, regarding both spatial and non-spatial models, while linear models have statistical assumptions to be maintained, which somewhat cause stability in the results, machine learning methods such as random forests rely upon optimisation of tuning that governs the variables from the bootstrapped data to randomly grow each tree, which at certain extent poses a randomness effect in the models. Typically, statistical modelling algorithms for digital soil mapping, particularly machine learning-based, are mostly able to make accurate predictions rather than to get better understanding of mechanistic soil processes (Wadoux *et al.*, 2019; 2020b). Furthermore, spatial data have distinctive properties that separated them from non-spatial data, such as spatial dependence, spatial heterogeneity, and scale (Nikparvar & Thill, 2021), with the latter playing a pivotal role in framing the model. Therefore, this research addressed this limitation by demonstrating similarities and differences in results from the modelling approaches.

Third, the accuracy-precision (bias-variance) trade-off, that is, accuracy (and bias) can be improved by selecting local information, whereas precision (and variance) can be achieved by selecting information from a comparatively larger space. One way of resolving the trade-off is to use mixture models (Ting *et al.*, 2011), as in the EBKR model approach that combines linear model and kriging. Errors can come from different measurements reflecting spatial-temporal environmental conditions, such as in laboratory, field, and spaceborne. Sources such as pixel and objects can also cause errors (Ogen *et al.*, 2022). On the other hand, this is compensated by the fact that the study area in this PhD research encompasses the central portion of Ireland from E-W coast, representing approximately 24,989 km² and covering variation in terms of physical factors such as climate patterns, lithology and soil types occurring in the country. This means that the models can be extended to other parts of Ireland, such as north and south. It is worth noting that heterogeneity of soil grain size fractions was observed by the geostatistical analysis.

8.1 Introduction

This Chapter commences by informing how this study has advanced the science for soil mapping. It also provides suggestions and directions for future research and how might the approaches developed assist in the design of future soil sampling surveys – in Ireland or elsewhere. Finally, this chapter highlights the societal relevance and impact of the research beyond the scholarly audience, for example, for soil policy.

8.2 How this study has advanced the research for Geography and soil science

This research was undertaken by addressing frontier topics in Physical Geography such as soil mapping, remote sensing, and spatial analysis on emerging themes/concepts (statistical machine learning, compositional data analysis-CoDa), thus, outcomes of this work has a broader audience and potential impact. From the experiences learned by conducting this research, this work also addressed epistemological topics and raised questions of relevance to Geography as a discipline and science.

In the context of soil science, Chapter 4 developed a framework to estimate soil particle size fractions and predict soil texture classes with non-spatial models. A methodological workflow depicting the data, the processing steps performed for the SAR Sentinel-1, and the modelling approaches is outlined. Chapter 6 provides a methodological framework to inform the design of in situ soil surveys for areas of large uncertainty estimates in the models and provides a means to estimate soil properties over large spatial areas and generate new data products for use in other model-based approaches that require soil texture products. Further, depending on the research purposes and the research questions and whether geographical coordinates matter, a researcher can decide between a spatial/geostatistical and a non-spatial modelling approach for estimating soil properties. This work also offered insights into the decision making required to consider (or not consider) soil grainsize and soil texture as compositional data in the models, given their intrinsic compositional nature commonly demonstrated in soil ternary diagrams. Hence, the option of incorporating pedological knowledge in both spatial and non-spatial models.

In the context of soil geography or soilscales, this work is focused on topsoil, and hence focused on process in lieu of genesis (i.e. soil forming factors), which is different from the well-known CLORPT model (Jenny, 1941) and SCORPAN model (McBratney *et al.*, 2003) for non-spatial and spatial modelling approaches, respectively. This is justified by the radar signal from a C-band SAR being restricted to a shallower surface layer, hence, having candidate covariates that influence remote sensing imagery, such as surface moisture and topographical conditions since they account for physical processes in soil for multi-scale variation. Therefore, after testing the approaches presented here, this work indicated a simpler model for both modelling approaches for the sake of interpretability of the model, similar in Wadoux *et al.* (2021) with regards to parsimony of models to handle the reality for the objective of research (i.e. explanation of the underlying soil process or pattern detection/prediction).

Concerning the remote sensing aspect, this research extended the application of remote sensing data, especially from Sentinel-1 and Sentinel-2, to soil texture prediction. This topic will be of significant interest

to the agri-SAR community, since remote sensing usually provides more rapid monitoring of environmental processes than a traditional field survey. From Chapter 5, the analysis and findings are a starting point to the use of radar polarimetry for studying the physical properties of soils, which contribute to a pivotal advancement in this subject. Further, the methods outlined here are also relevant to locations where optical remotely sensed data is limited and/or where vegetation cover is present throughout the year, addressing a significant gap in the existing literature. Therefore, underpinned by the research questions and driven by the objectives outlined, this PhD research fostered the employment of microwave remote sensing to soil analysis and predictions, especially soil grainsize and soil texture.

Within the Irish context, in addition to addressing a significant gap in the existing literature, this research sought to address a lack of studies that employ SAR data to estimate soil particle size fractions and prediction of soil texture classes.

8.3 Potential areas and approaches for future research

The exploratory assessment of the methods carried out in this work provides a robust basis for the development of other case studies to augment the accuracy of soil particle size fractions estimates and soil texture prediction over low-and short vegetated areas, broadening the polarimetric decomposition technique traditionally applied for classification and segmentation of landscapes and for vegetation studies. The methods outlined here are also relevant to locations where optical remotely sensed data is limited (e.g. by climatic/meteorological conditions) and/or where vegetation cover is present throughout the year.

Soil PSF and soil texture predictions with SAR techniques – may benefit from employing different backscattering coefficients – particularly gamma nought – other speckle filtering methods, cosine distributions, and soil-specific covariates related to SAR such as dielectric constant, soil roughness, and soil moisture. Moreover, a fully-pol H-alpha decomposition can be applied to an L-band SAR or even a dual-pol decomposition with a HH-VV SAR in order to compare the efficiency of the dual-pol decomposition method applied to a C-band SAR for soil PSF estimates.

In light of the endeavour of the Sentinel satellite constellation and its benefits (e.g., free availability, temporal-radiometric-spatial resolutions, and orbit stability), fusing Sentinel 1 and Sentinel 2 products together with their data processing techniques may also be beneficial. For example, combining H-alpha dual-pol decomposition with spectral unmixing method (i.e. spectral mixture analysis using endmember values) to extract values to serve as input in a statistical model, as they are pixel-based analyses. The aim of the spectral unmixing method is approximately the same as that of target polarimetric decomposition.

Regarding statistical modelling approaches for estimating soil properties, depending on the research purposes and the research questions, this study provided future works with a general basis for selecting a spatial/geostatistical or a non-spatial modelling approach. It also provides a means to estimate soil properties over large spatial areas and generate new data products for use in hydrological, land surface (e.g. Noah LSM models), climate and other model- based approaches that currently employ coarse global scale soil texture products.

8.4 Informing the design of future soil sampling surveys

As shown previously, the discussion about the last research question – *To what extent or under what environmental conditions can the method be used in another study area?* – highlighted some suggestions for the replicability of the methods. The following paragraphs bring up other pointers to assist in the design of future soil sampling surveys.

In the digital soil mapping context, the methods outlined in this study provided some directions for accounting for the requirements of compositional data such as soil texture. Moreover, regarding the laborious soil sampling associated with soil surveys, this work highlights the possibility of combining model based approaches with more limited in situ sampling methods for a sample size varying from 27 and 256 observations, in a non-gridded sampling design (i.e. the distance between sampling points does not have to be kept at a constant).

Applying a classical variance formula to sample data from a regular grid when employing geostatistical methods is not recommended since statistical and experimental bias may be introduced in the models (Brus & Gruijter, 1997), which, in turn, over-estimates the population variance and hence the classical estimation variance (Brus & Gruijter, 1997). Thus, this research informs directions for opting for either a model-based approach, as it describes processes on data, or a design-based approach that describes the population or the response variable itself. Considering that, in the former, the soil-forming process of a particular property in the study area is modelled as a stochastic process (Brus & Gruijter, 1997). Therefore, the approaches developed here could also assist in the design of future soil sampling surveys that focus on design-based soil sampling strategies (configuration of sample locations), since the spatial models rely on regionalised variables theory for interpolation in space. In this regard, also useful is the original work of McBratney and Webster (1983), in which they discuss how many observations are needed for regional estimation of soil properties.

To conclude, as sampling is the basis of mapping, this work provides a methodological framework to inform the design of costly and labour-intensive in situ soil surveys (areas of large uncertainty estimates in the models are those areas that may require detailed soil surveys). This is of great importance since a rationale for new data collection is usually required, particularly for funding bodies (EPA, 2022).

8.5 Wider societal relevance of the research

Unless erosion occurs exposing top and deeper soil layers, soils are a less visible natural resource, as they are either mostly covered by vegetation, or sealed due to urban development and infrastructure construction, leading to impermeability. Occasionally, they are also a material that can be disposed of or placed somewhere for environmental remediation. In this context, this doctorate study highlights the importance of soils as a natural resource and supports the Sustainable Development Goals and the development of ongoing- and future soil policy at both the national (Ireland) and European level by demonstrating methodological alternatives to obtain soil particle size fractions and soil texture at local and regional geographic scales. For example, the non-spatial modelling approach was useful to predict soil textural classes, whereas the spatial modelling approach can be used for estimating sand, silt, and clay fractions; moreover, using spatially estimates approach, areas of large uncertainty estimates in the models are those areas that may require detailed soil surveys. In turn, this

work has the potential opportunity to serve the public good in a way of deriving soil properties estimates through available resources (e.g., soil data legacy and satellite data) in the context of costly soil surveys, and providing additional information to guide decisions on soil policies. Particularly for the agri-sector in Ireland, as agriculture (including pasture and arable land) is the primary land use-land cover category in the country.

Therefore, having measured soil data available is crucial to making informed decisions and fulfilling policies around efficient land use management, soil management and sustainable development, and soil spatial distribution maps are often preferred to point source information. (Gök & Gürbüz, 2020). The research here sought to bring together the available measured soil data for Ireland, providing a compiled national data set of soil data from different existing soil surveys. Further, mostly important, the compiled dataset addresses soil properties other than soil texture and soil PSF (e. g., chemical parameters, carbon content, moisture, bulk density, etc.), broadening the potential use and application of such data for different purposes.

Furthermore, the trade-off amongst different levels of spatial scale of a soil property is another significant consideration as the demand for soil properties may range from local, regional, national or continental scales. For example, the demand for soils to sequester carbon applies ultimately at the global scale with management necessary at the national scale, however, soil information (and carbon sequestration) is supplied at a local scale by private ownership and management (O’Sullivan, Creamer & Schulte, 2018). In this context, this work provides a methodological framework to inform the design of in situ soil surveys for areas with large uncertainty estimates indicated by the models. In turn, estimates over small geographic scales or large areas would reciprocally benefit from the refined model.

8.6 Policy relevance

Notwithstanding awareness of the importance of soils for society at large, there is no specific or dedicated legislation on soils at the European Union and national levels, meaning that the existing soil policy is fragmented under environmental policies that indirectly approach soil protection (e.g. water, agriculture, forestry, waste, and chemicals) (EPA, 2022). Nevertheless, there is a monitoring framework currently undergoing at the EU level to make soil health monitoring obligatory, the ‘Soil Monitoring Law’, whose draft is currently under review by Member States.

This research is also timely in light of the European and Irish policy initiatives for soils such as “A Soil Deal for Europe 2021-2030 (Mission Soil)”, the Soil Monitoring Law, and “A Signpost for Soil Policy in Ireland 2021-2030”. Rooted in research and innovation, the former aims to tackle societal challenges towards healthy soils by targeting two out of five mission areas related to soils (Adaptation to Climate Change and Soil Health and Food) (EN, 2024). One of the actions planned in the Mission Soil that is straightly related to this study is developing a harmonised framework for soil monitoring in Europe. The directive on soil monitoring and resilience stresses the importance of improved monitoring, development of remote sensing solutions and enhanced data collection for more details on the state of soils; and envisages enhanced access to soil data and information for land managers, policymakers and other stakeholders (European Commission, 2023).

At national level, the research has direct relationship with the findings and highlights from the report *A Signpost for Soil Policy in Ireland* (EPA, 2022) under its three strands targets – Human Impacts, Soil

Characteristics, and New National Programs (particularly soil monitoring) – and potentially contributes to recommendations for soil policy. For example, the report points out a recommendation concerning sample size and data collected (4.2.4 Data Quality), data repeatability and reproducibility, generation of data (i.e. “new large datasets on Irish soils that can be mined by future studies to better understand trends, patterns and processes affecting Irish soils”. EPA, 2022:16), which were addressed in this thesis. One of the main issues currently encountered in Irish soil data is the uneven geographical coverage of data and the uneven soil properties currently analysed (i.e., focus on pH and nitrogen), notwithstanding soils have many other essential physical and chemical soil parameters. The compiled dataset from this PhD work can help to approach such issues by indicating areas that need more data and informing decisions on soil parameters to be analysed. This is important so that the national policy is aligned with the EU policy, particularly in terms of soil research themes targeted in both policies. In light of a compiled dataset, a recommendation that emerges from this PhD thesis is the use of sampling protocols for comparability and to facilitate data reuse. The contemporary aspect of this study concerning soil policy development for societal demands for healthy soil and soil data also dovetail with the EU Green Deal, Soil Fertility, Carbon Farming (national and EU levels). Concerning comparability and data reuse, an example is the Irish Soils Information System (SIS) project that developed comprehensive protocols for harmonising soil mapping in Europe, resulting in a comprehensive database of the soil types existing in Ireland with a national map of soil associations at a scale of 1:250,000.

BIBLIOGRAPHY

- Abdel-Hamid, A., Dubovyk, O., El-Magd, I., Menz, G. (2018). Mapping Mangroves Extents on the Red Sea Coastline in Egypt using Polarimetric SAR and High-Resolution Optical Remote Sensing Data. *Sustainability*. <https://doi.org/10.3390/su10030646>
- Abdul Munnaaf, M.; Nawar, S. & Mouazen, A.M. (2019). Estimation of Secondary Soil Properties by Fusion of Laboratory and On-Line Measured Vis–NIR Spectra. *Remote Sensing*, 11(23), 2819. <https://doi.org/10.3390/rs11232819>
- Ai, L.; Sun, S.Y.; Li, S.G.; Ma, H.Z. (2021). Research progress on the cooperative inversion of soil moisture using optical and SAR remote sensing. *Remote Sensing of Natural Resources*, 33, 10–18.
- Aitchison, J. (1982). The Statistical Analysis of Compositional Data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 44(2), 139-177. Retrieved July 13, 2021, from <http://www.jstor.org/stable/2345821>
- Aitchison, J. (1986). *The statistical analysis of compositional data. Monographs on Statistics and Applied Probability*. Chapman & Hall Ltd., London (UK) [Reprinted in 2003 with additional material by The Blackburn Press], 1986, 416 p.
- Aitchison, J. (2005). *A Concise Guide to Compositional Data Analysis*. Available at: <https://eprints.gla.ac.uk/259608/> (accessed 19 Dez 2022).
- Akpa, S. I. C., Odeh, I. O. A., Bishop, T. F. A. & Hartemink, A. E. (2014). Digital Mapping of Soil Particle-Size Fractions for Nigeria. *Soil Science Society of America Journal*, 78 (6), 1953-1966, doi:10.2136/sssaj2014.05.0202
- Amirian-Chakan, A.; Minasny, B.; Taghizadeh-Mehrjardi, R.; Akbarifazli, R.; Darvishpasand, Z.; Khordehbin, S. (2019). Some practical aspects of predicting texture data in digital soil mapping. *Soil and Tillage Research*, 194, e104289, <https://doi.org/10.1016/j.still.2019.06.006>
- Anselin L (2004). GeoDaTM 0.9.5-i Release Notes. <http://geodacenter.github.io/docs/geoda095i.pdf>
- Anup Das (n.d.). *Polarimetric SAR Data Analysis an Applications*. Available at https://vedas.sac.gov.in/vedas/downloads/ertd/SAR/L_5_Polarimetric_SAR_Data_Analysis_and_Applications_Dr_Anup_Das.pdf (accessed 25 July 2023).
- Araya, S. N., Fryjoff-Hung, A., Anderson, A., Viers, J. H., & Ghezzehei, T. A. (2021). Advances in soil moisture retrieval from multispectral remote sensing using unoccupied aircraft systems and machine learning techniques. *Hydrology and Earth System Sciences*. 25, 2739-2758. <https://doi.org/10.5194/hess-25-2739-2021>
- Arrouays, D., McBratney, A., Minasny, B., Hempel, J., Heuvelink, G., MacMillan, R., Hartemink, A., Lagacherie, P. & McKenzie, N. (2014). *The Global Soil Map project specifications*. In: Proceedings of the First GlobalSoilMap Conference, October 2013. Orleans, France: CRC Press. doi:10.1201/b16500-4. Available at: https://www.researchgate.net/publication/283156396_The_GlobalSoilMap_project_specifications (accessed 12 July 2021)
- ASF – Alaska Satellite Facility (n.d.). SAR Frequently Asked Questions. In: *What is SAR? SAR: The Power Tool of Remote Sensing*. <https://asf.alaska.edu/information/sar-information/what-is-sar/> Accessed 22 January 2024.
- Aslam, B., Maqsoom, A., Alaloul, W. S., Musarat, M. A., Jabbar, T., & Zafar, A. (2021). Soil erosion susceptibility mapping using a GIS-based multi-criteria decision approach: Case of district Chitral, Pakistan, *Ain Shams Engineering Journal*, Vol 12 (2), 1637-1649, <https://doi.org/10.1016/j.asej.2020.09.015>
- Atkinson, P. M. (2013). Downscaling in remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 22:106–114, <https://doi.org/10.1016/j.jag.2012.04.012>
- Attema, E. P.W. & Ulaby, F. T. (1978) Vegetation modeled as a water cloud, *Radio Science-AGU*, 13, 357–364, 1978. <https://doi.org/10.1029/RS013i002p00357>
- Azizi, K., Garosi, Y., Ayoubi, S., Tajik, S. (2023) Integration of Sentinel-1/2 and topographic attributes to predict the spatial distribution of soil texture fractions in some agricultural soils of western Iran. *Soil and Tillage Research*, 229-105681, <https://doi.org/10.1016/j.still.2023.105681>
- Babaeian, E., Sadeghi, M., Jones, S. B., Montzka, C., Vereecken, H. & Tuller, M. (2019). Ground, proximal, and satellite remote sensing of soil moisture. *Reviews of Geophysics*, 57. <https://doi.org/10.1029/2018RG000618>
- Baghdadi, N., Choker, M., Zribi, M., Hajj, M.E., Paloscia, S., Verhoest, N.E., Lievens, H., Baup, F. and Mattia, F. (2016). A new empirical model for radar scattering from bare soil surfaces. *Remote Sensing*, 8(11), p.920. <https://doi.org/10.3390/rs8110920>

- Ballabio, C., Panagos, P. & Montanarella L. (2016). Mapping topsoil physical properties at European scale using the LUCAS database (2016). *Geoderma*, 261, pp.110-123. <https://doi.org/10.1016/j.geoderma.2015.07.006>
- Barrett, B., Dwyer, N., Whelan, P., & Bartlett, D. (2007). *Soil moisture determination in southern Ireland using an ASAR time series*. In: Proceedings of Envisat Symposium, April 2007. Montreux, Switzerland: European Space Agency (Special Publication) ESA SP-636. Available at: <https://earth.esa.int/envisatsymposium/proceedings/sessions/4D1/460228BB.pdf> (accessed 06 Jun 2021).
- Barrett, B.W., Dwyer, E. & Whelan, P. (2009). Soil Moisture Retrieval from Active Spaceborne Microwave Observations: An Evaluation of Current Techniques. *Remote Sensing*, 1, 210-242. <https://doi.org/10.3390/rs1030210>
- Bashir, O. *et al.* (2024). Mathematical vs. machine learning models for particle size distribution in fragile soils of North-Western Himalayas. *Journal of Soils and Sediments*, 24, 2294–2308, <https://doi.org/10.1007/s11368-024-03820-y>
- Baveye, P. C., Baveye J., Gowdy, J. (2016). Soil “Ecosystem” Services and Natural Capital: Critical Appraisal of Research on Uncertain Ground. *Frontiers in Environmental Science*, volume 4, <https://www.frontiersin.org/articles/10.3389/fenvs.2016.00041>
- Bayle, P., Bayle, A., Janson, L., & Mackey, L. (2020). Cross-validation confidence intervals for test error. *Advances in Neural Information Processing Systems*, 33, 16339-16350, <https://doi.org/10.48550/arXiv.2007.12671>
- Beale, J., Snapir, B., Waine, T., Evans, J. & Corstanje, R (2019). The significance of soil properties to the estimation of soil moisture from C-band synthetic aperture radar. *Hydrology and Earth System Sciences*. Preprint. <https://doi.org/10.5194/hess-2019-294>
- Benedetto, F. & Tosti, F. (2013). GPR spectral analysis for clay content evaluation by the frequency shift method. *Journal of Applied Geophysics*, 97, 89–96. <https://doi.org/10.1016/j.jappgeo.2013.03.012>
- Besson, A., Cousin, I, Samouëlian, A., Boizard, H. and Richard, G. (2004). Structural heterogeneity of the soil-tilled layer as characterized by 2D electrical resistivity surveying. *Soil and Tillage Research* 79:239-249. DOI:10.1016/j.still.2004.07.012.
- Bezák, N., Borrelli, P., Matjaž M., Auflič M., & Panagos, P. (2024). Towards multi-model soil erosion modelling: An evaluation of the erosion potential method (EPM) for global soil erosion assessments. *Catena*, 234, 107596. DOI:10.1016/j.catena.2023.107596
- Bhogapurapu, N.; Dey, S.; Mandal, D.; Bhattacharya, A.; Karthikeyan, L.; McNairn, H.; Rao, Y. (2022). Soil moisture retrieval over croplands using dual-pol L-band GRD SAR data. *Remote Sensing of Environment*, 271, 112900. <https://doi.org/10.1016/j.rse.2022.112900>
- Bogale, A. (2021). Morphometric analysis of a drainage basin using geographical information system in Gilgel Abay watershed, Lake Tana Basin, upper Blue Nile Basin, Ethiopia. *Applied Water Science* 11, 122. <https://doi.org/10.1007/s13201-021-01447-9>
- Bole, A., Wall, A., Norris, A. (2014). Target Detection – Chapter 3. In: *Radar and ARPA Manual*, Butterworth-Heinemann, 3rd edition, 139-213 pp, ISBN 9780080977522, <https://doi.org/10.1016/B978-0-08-097752-2.00003-9>
- Bolten, J.D.; Lakshmi, V.; Njoku, E.G. (2033). Soil moisture retrieval using the passive/active L-and Sband radar/radiometer. *IEEE Trans. Geosci. Remote Sens.* 41, 2792-2801.
- Bongers, A., Torres J.L. (2024). Low-Earth Orbit Faces a Spiraling Debris Threat. Scientific American. Springer Nature America, Inc. Available at <https://www.scientificamerican.com/article/low-earth-orbit-faces-a-spiraling-debris-threat/> (accessed 24 Apr 2024).
- Borgeaud, M.; Wegmueller, U. (1996). On the use of ERS SAR interferometry for the retrieval of geoand biophysical information. In *Proceedings of the 'Fringe 96' Workshop on ERS SAR Interferometry*, ESA SP-406, Zurich, Switzerland, pp. 83-94.
- Bourgeau-Chavez, L.L., Riordan, K., Powell, R.B., Miller, N., & Nowels, M. (2009). Improving Wetland Characterization with Multi-Sensor, Multi-Temporal SAR and Optical/Infrared Data Fusion. In: *Advances in Geoscience and Remote Sensing*. Gary Jedlovec (editor). Tech Publishers, India. ISBN 978-953-307-005-6, pp.679-708. Available at <https://www.intechopen.com/chapters/9542> (accessed 12 October 2021)
- Bousbih, S.; Zribi, M.; Pelletier, C.; Gorraeb, A.; Lili-Chabaani, Z.; Baghdadi, N.; Aissa, N.B. & Mougenot, B. (2019). Soil Texture Estimation Using Radar and Optical Data from Sentinel-1 and Sentinel-2. *Remote Sensing*, 11(13), 1520; <https://doi.org/10.3390/rs11131520>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5-32. doi:10.1023/A:1010933404324
- Brown, S. C. M., Quegan, S., Morrison, K., Bennett, J. C. & Cookmartin, G. (2003). High-resolution measurements of scattering in wheat canopies- implications for crop parameter retrieval. *IEEE*

- Transactions on Geoscience and Remote Sensing*, 41, 1602–1610. <https://doi.org/10.1109/TGRS.2003.814132>
- Brus, D.J.; & Gruijter, J.J. de. (1997) Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion), *Geoderma*, 80 (1–2), 1 – 44, [https://doi.org/10.1016/S0016-7061\(97\)00072-4](https://doi.org/10.1016/S0016-7061(97)00072-4)
- Byers, M., Wright, E., Boley, A. *et al.* (2022). Unnecessary risks created by uncontrolled rocket reentries. *Nature Astronomy*, 6, 1093–1097, <https://doi.org/10.1038/s41550-022-01718-8>
- Byrne, K. A., Lanigan, G., Creamer, R., & Renou-Wilson, F. (2018). Soils and Carbon Storage. In: Creamer, R. & O’Sullivan, L. (eds). *The Soils of Ireland*. World Soils Book Series. Springer, Cham. Available at: <https://doi.org/10.1007/978-3-319-71189-8> (accessed 4 April 2021)
- Cattle, S. R., Meakin, S. N., Ruszkowski, P., & Cameron, R. G. (2003). Using radiometric data to identify aeolian dust additions to topsoil of the Hillston district, western NSW. *Australian Journal of Soil Research*, 41(8), 1439–1456. <https://doi.org/10.1071/SR03026>
- CCRS (n.d) – Canada Centre for Remote Sensing. Canadian Space Agency. *Fundamentals of Remote Sensing*. Available at: https://www.nrcan.gc.ca/sites/www.nrcan.gc.ca/files/earthsciences/pdf/resource/tutor/fundam/pdf/fundamentals_e.pdf (accessed 17 May 2021).
- Certini, G., Scalenghe, R. (2023). The crucial interactions between climate and soil. *Science of the Total Environment*, 856-Part 2, 159169, <https://doi.org/10.1016/j.scitotenv.2022.159169>
- Chappell, M. A., Seiter, J. M., West, H. M., Durham, B. D., Porter, B. E., & Price, C. L. (2019). Building geochemically based quantitative analogies from soil classification systems using different compositional datasets. *PLoS One*, 14(2), e0212214. <https://doi.org/10.1371/journal.pone.0212214>
- Chen, H. & Kasilingam, D. (2005). *New blind source separation technique for removing vegetation bias in polarimetric SAR interferometer measurements*. In: Proceedings of IEEE International Geoscience and Remote Sensing Symposium, November 2005. Seoul, Korea (South): IEE. pp. 4860-4862, <https://doi.org/10.1109/IGARSS.2005.1526762>
- Chipman, H. A., George, E.I., McCulloch, R.E. (2010). Bart: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4(1), 266–298, <https://www.jstor.org/stable/27801587>
- Cisty, M., Soldanova, V. & Cyprich, F. (2019). Random forest based reclassification of soil texture for hydrological modelling. *Geophysical Research Abstracts*, 21, EGU2019-14320. Available at <https://meetingorganizer.copernicus.org/EGU2019/EGU2019-14320.pdf> (accessed 24 May 2021)
- Cloude, S.R., Pottier, E. (1996). A review of target decomposition theorems in radar polarimetry. *IEEE Transactions on Geoscience Remote Sensing*, 34 (2), 498-518, <https://doi.org/10.1109/36.485127>
- Cloude, S. R. & Pottier, E.(1997). An entropy based classification scheme for land applications of polarimetric SAR. *IEEE Transactions on Geoscience Remote Sensing*, 35 (1), 68–78. DOI. 10.1109/36.551935
- Coblinski, J. A., Inda, A. V., Demattê, J. A. M., Dotto, A. C., Gholizadeh, A. & Giasson, E. (2021). Identification of Minerals in Subtropical Soils with Different Textural Classes by VIS-NIR-SWIR Reflectance Spectroscopy. *Catena*. 203. 105334. <https://doi.org/10.1016/j.catena.2021.105334>
- Comas-Cufí, M., Thió-Henestrosa, S. (2011). *CoDaPack 2.0: a stand-alone, multi-platform compositional software*. In: Egozcue JJ, Tolosana-Delgado R, Ortego MI, eds. *CoDaWork'11: 4th International Workshop on Compositional Data Analysis*. Sant Feliu de Guíxols; 2011.
- Couclelis, H. (1986). Artificial intelligence in geography: conjectures on the shape of things to come. *The Professional Geographer*, 380 (1), 1–11. doi:10.1111/j.0033-0124.1986.00001.x
- Creamer, R., Simo, I., Reidy, B., Carvalho, J., Fealy, R. M., Hallett, S., Jones, R., Holden, A., Holden, N., Hannam, J., Massey, P.A., Mayr, T., McDonald, E., O'Rourke, S., Sills, P., Truckell, I., Zawadzka J., & Schulte, R. (2014). *Irish Soil Information System*. Synthesis Report. Report No. 130. Environmental Protection Agency, Johnstown Castle, Wexford, Ireland.
- Creamer, R. E., Simo, I., O’Sullivan, L., Reidy, B., Schulte, R. P. O., Fealy, R. M. (2016). *Irish Soil Information System: Soil Property Maps. Report N. 204*. EPA Research Programme 2014–2020. Teagasc, Wexford, Ireland.
- Creamer, R. & O’Sullivan, L. eds. (2018). *The Soils of Ireland*. World Soils Book Series. Springer, Cham. Available at: Available at: <https://doi.org/10.1007/978-3-319-71189-8> (accessed 4 April 2021)
- Cui, H.Z.; Jiang, L.M.; Lu, Z.; Wang, G.; Wang, J. (2017). Improvement and Validation of QP Model. with Dual-Channel Soil. Moisture Retrieval Algorithm in Genhe, China. In *Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Fort Worth, TX, USA, 23–28 July 2017.

- Curley, M., Coonan, B., Ruth, C.E. and Ryan, C. (2023). Ireland's Climate Averages 1991-2020. *Climatological Note No. 22*. Met Éireann, Ireland.
- Cybèle, C., Daniel, H., Jean-Daniel, S., Frédérik, D., Audrey, M. (2022). Climate Change Increases the Severity and Duration of Soil Water Stress in the Temperate Forest of Eastern North America. *Frontiers in Forests and Global Change*, vol 5, e879382 doi: 10.3389/ffgc.2022.879382, <https://doi.org/10.3389/ffgc.2022.879382>
- Das, K. & Paul, P. K. (2015). Present status of soil moisture estimation by microwave remote sensing. *Cogent Geoscience*, 1(1), DOI: 10.1080/23312041.2015.1084669
- Dave, R.B., Saha, K., Kushwaha, A., Vithalpur, M., Nidhin P., Murugesan, A. (2023). Analysing the potential of polarimetric decomposition parameters of Sentinel-1 dual-pol SAR data for estimation of rice crop biophysical parameters. *Journal of Agrometeorology*, 25(1), 105–112. <https://doi.org/10.54386/jam.v25i1.2039>
- Davies, T. (2019). Algorithms and artificial intelligence. In: Davies, T., Walker, S., Rubinstein, M., & Perini, F. (Eds.). *The State of Open Data: Histories and Horizons*. Cape Town and Ottawa: African Minds and International Development Research Centre.
- De Zan, F., Parizzi, A., Prats-Iraola, P. & López-Dekker, P. (2014). A SAR Interferometric Model for Soil Moisture. *IEEE Transactions on Geoscience and Remote Sensing*, 52(1), 418-425. <https://doi.org/10.1109/TGRS.2013.2241069>
- Deodoro, S.C., Moral, R. de A.; Fealy, R.; McCarthy, T., & Fealy, R. (2023). An assessment of Sentinel 1 SAR, geophysical and topographical covariates for estimating topsoil particle size fractions. *European Journal of Soil Science.*, 74(5), e13414, <https://doi.org/10.1111/ejss.13414>
- Dey, A. (2016). Machine Learning Algorithms: A Review. *International Journal of Computer Science and Information Technologies*, 7 (3), 2016, 1174-1179. Available at <http://ijcsit.com/docs/Volume%207/vol7issue3/ijcsit2016070332.pdf> (accessed 4 April 2021)
- Dharumarajan, S., Hegde, R., & Singh, S.K. (2017). Spatial prediction of major soil properties using Random Forest techniques - A case study in semi-arid tropics of South India, *Geoderma Regional*, Volume 10, 154-162, <https://doi.org/10.1016/j.geodrs.2017.07.005>
- Dierking I, Flatley A and Greenhalgh, D (2021). Voronoi patterns in liquid crystal textures. *Journal of Molecular Liquids* 335 e116553. <https://doi.org/10.1016/j.molliq.2021.116553>.
- dos Santos, J. R., Narvaes, I. S., S., I., A. Graca, P. M. L. A., Gonçalves, F. G. (2009). Polarimetric Responses and Scattering Mechanisms of Tropical Forests in the Brazilian Amazon. *Advances in Geoscience and Remote Sensing. InTech*. <http://dx.doi.org/10.5772/8340>
- Dobson, M. C. & Ulaby, F. T. (1981). Microwave backscatter dependence on surface roughness, soil moisture, and soil texture: Part III-soil tension. *IEEE Transactions on Geoscience and Remote Sensing*, 19, 51–61. doi:10.1109/TGRS.1981.350328
- Dobson, M. C., Kouyate, F. & Ulaby, F. T. (1984). A reexamination of soil textural effects on microwave emission and backscattering. *IEEE Transactions on Geoscience and Remote Sensing*, GE-22(6), 530-536. doi: 10.1109/TGRS.1984.6499165.
- Dobson, M. C., Ulaby, F. T., Hallikainen, M. T., & El-Rayes, M. A. (1985). Microwave dielectric behavior of wet soil-Part II: Dielectric mixing models. *IEEE Transactions on Geoscience and Remote Sensing*, 23, 35–46. doi:10.1109/TGRS.1985.289498
- Dobson, M. C. & Ulaby, F. T. (1986). Active microwave soil moisture research. *IEEE Transactions on Geoscience and Remote Sensing*, 24, 23–36. doi:10.1109/TGRS.1986.289585
- Dobson, M.; Pierce, L.; Sarabandi, K.; Ulaby, F. & Sharik, T. (1992). Preliminary analysis of ERS-1 SAR for forest ecosystem studies. *IEEE Transactions on Geoscience and Remote Sensing*, 30(2), 203-211. doi: 10.1109/36.134071.
- Domenech, M. B.; Amiotti, N. M., Costa, J. L. & Castro-Franco, M. (2020). Prediction of topsoil properties at field-scale by using C-band SAR data. *International Journal of Applied Earth Observation and Geoinformation*, Vol.93, 102197, <https://doi.org/10.1016/j.jag.2020.102197>
- Dotto, A. C., Demattê, J. A. M., Rossel, R. A. V. & Rizzo, R. (2020). Soil environment grouping system based on spectral, climate, and terrain data: a quantitative branch of soil series. *SOIL*, 6, 163–177, <https://doi.org/10.5194/soil-6-163-2020>
- Dou, Q., Xie, Q., Peng, X., Lai, K., Wang, J., Lopez-Sanchez, J.M., Shang, J., Shi, H., Fu, H., Zhu, J. (2022). Soil moisture retrieval over crop fields based on two-component polarimetric decomposition: a comparison of generalized volume scattering models. *Journal of Hydrology*, 615, A, e128696, <https://doi.org/10.1016/j.jhydrol.2022.128696>

- Dubois, P., van Zyl, J. & Engman, T. (1995). Measuring soil moisture with imaging radars. *IEEE Transactions on Geoscience and Remote Sensing*, 33, 915–926. doi:10.1109/36.406677
- Eaton, J.M., McGoff, N.M., Byrne, K.A., Leahy, P. & Kiely, G. (2008). Land cover change and soil organic carbon stocks in the Republic of Ireland 1851–2000. *Climatic Change*, 91, 317–334. <https://doi.org/10.1007/s10584-008-9412-2>
- Egozcue, J.J. & Pawlowsky-Glahn, V. (2006). Compositional data in the geosciences. *Geological Society, London. Special Publications: Simplicial geometry for compositional data*, 264, 145–159, Available at: <https://www.lyellcollection.org/doi/abs/10.1144/gsl.sp.2006.264.01.11> (accessed 06 February 2023)
- Emmet-Booth, J.P., Forristal, P.D., Fenton, O., Bondi, G., Wall, D.P., Creamer, R.E. & Holden, N.M. (2018). The Soil Structure ABC – A practical guide to managing soil structure. Available at: <https://www.teagasc.ie/media/website/environment/soil/The-soil-structure-ABC.-A-practical-guide-to-managing-soil-structure.pdf> (accessed 10 April 2024)
- EN – European Commission (2024). Horizon Europe Work Programme 2023–2025. Missions and Cross-cutting Activities. European Commission Decision C(2024) 2371 of 17 April 2024. Available at: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/wp-call/2023-2024/wp-12-missions_horizon-2023-2024_en.pdf (accessed 01 May 2024)
- Engelbrecht, J.; Theron, A.; Vhengani, L.; Kemp, J. (2017). A Simple Normalized Difference Approach to Burnt Area Mapping Using Multi-Polarisation C-Band SAR. *Remote Sensing*, 9, 764. <https://doi.org/10.3390/rs9080764>
- EPA-Environmental Protection Agency (2013). *Ireland National Inventory*. Report. EPA, Johnstown Castle, Ireland.
- ESA – European Space Agency (2007). *EVISAT ASAR Product Handbook*, Issue 2.2, 27 February 2007. Available at: <https://earth.esa.int/eogateway/documents/20142/37627/ASAR-Product-Handbook.pdf> (accessed 25 October 2023).
- ESA – European Environment Agency (2019). *Soil and United Nations Sustainable Development Goals*. Available at: <https://www.eea.europa.eu/signals/signals-2019-content-list/infographics/soil-and-united-nations-sustainable/view> (accessed 26 Jan 2024)
- ESA – European Space Agency, Earth Observation College-EO (2021). *Basic Principles of Radar Backscatter*. Available at <https://eo-college.org/courses/basic-principles-of-radar-backscatter/#learndash-course-content> (accessed 26 February 2022)
- ESA – European Space Agency (n.d.). *RADAR and SAR Glossary. Table 5.2*. ESA earth net online. Available at <https://earth.esa.int/handbooks/asar/CNTR5-2.html> (accessed 17 May 2021)
- ESA – European Space Agency (n.d.). *Sentinel Online. Interferometric Wide Swath*. <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-1-sar/acquisition-modes/interferometric-wide-swath> (accessed 28 October 2023).
- European Commission (2023). Directive on Soil Monitoring and Resilience. *Questions and Answers on a Directive on Soil Monitoring and Resilience*. Available at https://ec.europa.eu/commission/presscorner/detail/en/qanda_23_3637 (accessed 2 May 2024).
- European Commission, Directorate-General for Research and Innovation, Veerman, C., Pinto Correia, T., Bastioli, C. *et al.* (2020). Caring for soil is caring for life: ensure 75% of soils are healthy by 2030 for food, people, nature and climate. Report of the Mission board for Soil health and food. Publications Office. Available at <https://data.europa.eu/doi/10.2777/821504> (accessed 24 April 2024)
- Evans D. G. & Jones S. M. (1987). Detecting Voronoi (area-of-influence) polygons. *Mathematical Geology* 19 523–537. <https://doi.org/10.1007/BF00896918>
- FAO (1985)-Food and Agriculture Organization. Soil and Water. In: *Irrigation Water Management: Training Manual No.1-Introduction to Irrigation*. Available at <https://www.fao.org/3/r4082e/r4082e03.htm> (accessed 24 Jan 2024).
- FAO (n.d.)-Food and Agriculture Organization. Soil structure. Available at https://www.fao.org/fishery/docs/CDrom/FAO_Training/FAO_Training/General/x6706e/x6706e07.htm (accessed 10 April 2024).
- Fealy, R., Bruyère, C., Duffy, C.(2011). *Regional Climate Model Simulations for Ireland for the 21st Century*. EPA Research Report 2011-CCRP-MS-2.2 Environmental Protection Agency, Wexford, Ireland. Available at https://www.epa.ie/publications/research/climate-change/Research_Report_244.pdf (accessed 3 November 2023).
- Fernández-Ugalde, O., Jones, A., Meuli, R.G. (2020). Comparison of sampling with a spade and gouge auger for topsoil monitoring at the continental scale. *European Journal of Soil Science*, 71(2): 137–150. <https://bsssjournals.onlinelibrary.wiley.com/doi/epdf/10.1111/ejss.12862>,

- Filzmoser, P., Hron, K., Templ, M. (2018). Compositional Data as a Methodological Concept. In: *Applied Compositional Data Analysis*. Springer Series in Statistics. Springer, Cham. https://doi.org/10.1007/978-3-319-96422-5_1
- Fletcher, R. (2022) Temporal Comparisons of Apparent Electrical Conductivity: A Case Study on Clay and Loam Soils in Mississippi. *Agricultural Sciences*, 13(8), 936-946. doi: 10.4236/as.2022.138058.
- Freeman, A. & Durden, S. L. (1998). A three-component scattering model for polarimetric SAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 36(3), 963-973, doi: 10.1109/36.673687.
- Freeman, A. (nd). Radiometric calibration of SAR image data. International Society for Photogrammetry and Remote Sensing. *ISPRS Proceedings, XXIX, part 1*, 2012. Available at https://www.isprs.org/proceedings/xxix/congress/part1/212_xxix-part1.pdf (accessed 23 October 2023).
- Frelih-Larsen, A., Riedel, A., Hobeika, M., Scheid, A., Gattinger, A., Niether, W., & Siemons, A. (2022). *Role of soils in climate change mitigation*. Climate change 56/2022, Ressortforschungsplan of the Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection. Project No. (FKZ) 3721 42 502 0 Report No. (UBA-FB) FB000883/ENG, Interim report, German Environment Agency. Available at <http://www.umweltbundesamt.de/publikationen> (accessed 14 Jan 2024).
- Friedberg, R., Tibshirani, J., Athey, S., Wager, S. (2021). Local Linear Forests. *Journal of Computational and Graphical Statistics*, 30(2), 503-517, DOI: 10.1080/10618600.2020.1831930
- Fung, A.K.; Li, Z.; Chen, K.S. (1992). Backscattering from a randomly rough dielectric surface. *IEEE Trans. Geosci. Remote Sens.* 30, 356-369.
- Gao et al. (2022). Multi-frequency radiometer-based soil moisture retrieval and algorithm parameterization using in situ sites. *Remote Sensing of Environment*, 279, 113113, <https://doi.org/10.1016/j.rse.2022.113113>
- Geological Survey Ireland. *Quaternary Geological Map of Ireland*. 2017. Scale:1:500,000. Available at: <https://www.gsi.ie/en-ie/data-and-maps/Pages/Quaternary.aspx> (accessed 20 December 2022).
- Gerke, H. H., Vogel, H-J, Weber, T. K. D., van der Meij, W. M., Scholten, T. (2022). 3–4D soil model as challenge for future soil research: Quantitative soil modeling based on the solid phase. *Journal of Plant Nutrition and Soil Science*, 185:720–744. <https://doi.org/10.1002/jpln.202200239>
- Gholizadeh, A., Žižala, D., Aberioon, M., C. & Borůvka, L. (2018). Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. *Remote Sensing of Environment*, 218, 89–103. <https://doi.org/10.1016/j.rse.2018.09.015>
- Gök, G., Gürbüz, O.A. (2020). Application of geostatistics for grid and random sampling schemes for a grassland in Nigde, Turkey. *Environmental Monitoring and Assessment*, 192, e300, <https://doi.org/10.1007/s10661-020-08281-7>
- Goudie, A. S. (1986). The integration of human and physical geography. *Transactions of the Institute of British Geographers*, 11(4): 454-458.
- Goodchild, M. F. (2001). Issues in spatially explicit modeling. In: D. C. Parker, T. Berger, and S. M. Manson, (eds). *Agent-based models of land-use and land-cover change report and review of an international workshop*, October 47. Irvine, CA, 12–15.
- Goodchild, M. F. (2004). The validity and usefulness of laws in geographic information science and geography. *Annals of the Association of American Geographers*, 94: 300-303. <https://doi.org/10.1111/j.1467-8306.2004.09402008.x>
- Google Developers (2024). Google Machine Learning Education. Machine Learning Glossary. Available at <https://developers.google.com/machine-learning/glossary#f> (accessed 15 Mai 2024).
- Gorrab, A., Zribi, M., Baghdadi, N., Lili-Chabaane, Z. & Mougenot, B. (2014). Multi-frequency analysis of soil moisture vertical heterogeneity effect on radar backscatter. In *Advanced Technologies for Signal and Image Processing (ATSIP)*, 1st International Conference, March 2014. (pp. 379–384). 1653 Sousse, Tunisia: IEEE. <http://dx.doi.org/10.1109/ATSIP.2014.6834640>
- Gorrab, A., Zribi, M., Baghdadi, N., Mougenot, B., Fanise, P. & Chabaane, Z. L. (2015). Retrieval of Both Soil Moisture and Texture Using TerraSAR-X Images. *Remote Sensing*, 7(8):10098-10116. <https://doi.org/10.3390/rs70810098>
- Graham, A. J., & Harris, R. (2003). Extracting biophysical parameters from remotely sensed radar data: a review of the water cloud model. *Progress in Physical Geography: Earth and Environment*, 27(2), 217–229. <https://doi.org/10.1191/0309133303pp378ra>
- Grandjean, G., Cousin, I., Seger, M., Thiesson, J., Lambot, S., Van Wesemael, B., Stevens, A., Samyn, K., Bitri, A., Bernardie, S. (2009). *From geophysical parameters to soil characteristics*. Report N°BRGM/FP7-DIGISOIL-D2.1, 52 pages. Available at: https://esdac.jrc.ec.europa.eu/public_path/Digisoil-D2.1.pdf (accessed 18 December 2022)

- Gribov A and Krivoruchko K (2020). Empirical Bayesian kriging implementation and usage. *Science of the Total Environment* 722 e137290. <https://doi.org/10.1016/j.scitotenv.2020.137290>
- Grote, K., Hubbard, S. & Rubin, Y. (2003). Field-scale estimation of volumetric water content using ground-penetrating radar ground wave techniques. *Water Resources Research*. 39(11), 1321. <https://doi.org/10.1029/2003WR002045>
- Grote, K., Anger, C., Kelly, B., Hubbard, S. & Rubin, Y. (2010). Characterization of Soil Water Content Variability and Soil Texture using GPR Groundwave Techniques. *Journal of Environmental and Engineering Geophysics*, 15(3), 93–110. <http://dx.doi.org/10.2113/JEEG15.3.93>
- Guerriero, L., Pierdicca, N., Pulvirenti, L., Ferrazzoli, P. (2013). Use of satellite radar bistatic measurements for crop. *Remote Sensing*, 5(2), 864–890. <https://doi.org/10.3390/rs5020864>
- Guo, J.; Liu, J.; Ning, J.F.; Han, W.T. (2019). Construction and validation of soil moisture retrieval model in farmland based on Sentinel multi-source data. *Trans. Chin. Soc. Agric. Eng. (Trans. CSAE)*, 35, 71–78, (In Chinese with English abstract).
- Gururaj, P., Umesh, P., & Shetty, A. (2019). Assessment of spatial variation of soil moisture during maize growth cycle using SAR observations. In: *Remote Sensing for Agriculture, Ecosystems, and Hydrology XXI* (Vol. 11149, pp. 372-379). SPIE.
- Hajnsek, I.; Alvarez-Perez, J.L.; Papathanassiou, K.P.; Moreira, A.; Cloude, S.R. (2003). Surface parameter estimation using interferometric coherences at different polarisations. In *POLinSAR Workshop on Applications of SAR Polarimetry and Polarimetric Interferometry*, ESA-ESRIN, Frascati, Italy, January 14-16.
- Halbgewachs, M.; Wegmann, M.; da Ponte, E. (2022). A Spectral Mixture Analysis and Landscape Metrics Based Framework for Monitoring Spatiotemporal Forest Cover Changes: A Case Study in Mato Grosso, Brazil. *Remote Sensing*, 14, 1907. <https://doi.org/10.3390/rs14081907>
- Hallikainen, M. T., Ulaby, F. T., Dobson, M. C., El-rayes, M. A., & Wu, L. K. (1985). Microwave dielectric behavior of wet soil-Part 1: Empirical models and experimental observations. *IEEE Transactions on Geoscience and Remote Sensing*, 23, 25–34. doi:10.1109/TGRS.1985.289497
- Han, D., Vahedifard, F., Aanstoos, J. V. (2017). Investigating the correlation between radar backscatter and in situ soil property measurements. *International Journal of Applied Earth Observation and Geoinformation*. (57)136-144. <https://doi.org/10.1016/j.jag.2016.12.018>.
- Harfenmeister, K.; Itzerott, S.; Weltzien, C.; Spengler, D. (2021). Agricultural Monitoring Using Polarimetric Decomposition Parameters of Sentinel-1 Data. *Remote Sensing*, 13(4), 575. <https://doi.org/10.3390/rs13040575>
- Harvard (n.d.). *Rayleigh's Criterion*. Harvard Natural Sciences-Lecture Demonstrations. Harvard University. <https://sciencedemonstrations.fas.harvard.edu/presentations/rayleighs-criterion> (accessed 19 Jan 2024)
- Hastie, T., Tibshirani, R., & Friedman, F (2009). *The Elements of Statistical Learning Data Mining, Inference, and Prediction*. Springer Series in Statistics, Springer, 2nd ed., textbook. <https://link.springer.com/book/10.1007/978-0-387-84858-7> (accessed 21 Jan 2024).
- Hengl T, Nussbaum M, Wright MN, *et al.* (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *Peer Journal* 6:e5518. <https://doi.org/10.7717/peerj.5518>
- Heras, A. B. I. O. (2015). *Decomposition and unsupervised segmentation of dual-polarized polarimetric SAR data using fuzzy entropy and coherency clustering method*. Master dissertation in Telecommunication Engineering. Universitat Autònoma de Barcelona. Escola d'Enginyeria. Available at <https://ddd.uab.cat/record/133501?ln=en> (accessed 14 Dec 2023).
- Hess, D. & McKnight, T. L. (2017). *Physical Geography: A Landscape Appreciation*. NJ: Pearson. 12 ed.
- Holah, N.; Baghdadi, N.; Zribi, M.; Bruand, A.; King, C. (2005). Potential of ASAR/ENVISAT for the characterization of soil surface parameters over bare agricultural fields. *Remote Sensing of Environment*, 96, 78-86.
- Hou, D., Bolan, N. S, Tsang, D. C. W., Kirkham, M. B., O'Connor, D. (2020). Sustainable soil use and management: An interdisciplinary and systematic approach. *Science of the Total Environment*, 729:138961. doi: 10.1016/j.scitotenv.2020.138961.
- Hubbard, S. S., J. E. Peterson, Jr., E. L. Majer, P. T. Zawislanski, K. H. Williams, J. Roberts, & F. Wobber (1997). Estimation of permeable pathways and water content using tomographic radar data. *The Leading Edge*, 16, 1623-1628, 1997. Available at <http://www.geo.umass.edu/faculty/wclement/dielec.html> (accessed 06 October 2021)
- Husman, S. de R., Van der Sanden, J. J., Lhermitte, S., Eleveld, M.A. (2021). Integrating intensity and context for improved supervised river ice classification from dual-pol Sentinel-1 SAR data. *International Journal*

- of *Applied Earth Observation and Geoinformation*, 101, e102359, <https://doi.org/10.1016/j.jag.2021.102359>
- Hyvärinen, A. & Oja, E. (1997). A Fast Fixed-Point Algorithm for Independent Component Analysis. *Neural Computation*, 9(7), 1483-1492. <https://doi.org/10.1162/neco.1997.9.7.1483>
- Ibrahim, G.R.F., Rasul, A., & Abdullah, H. (2023). Improving Crop Classification Accuracy with Integrated Sentinel-1 and Sentinel-2 Data: A Case Study of Barley and Wheat. *Journal of Geovisualization and Spatial Analysis*, 7(22), <https://doi.org/10.1007/s41651-023-00152-2>
- Ilic, J. (2001). Wood: Electrical Properties. Editor(s): K.H. Jürgen Buschow, Robert W. Cahn, Merton C. Flemings, Bernhard Ilshner, Edward J. Kramer, Subhash Mahajan, Patrick Veyssière. In: *Encyclopedia of Materials: Science and Technology*, Elsevier, 9629-9633, ISBN 9780080431529, <https://doi.org/10.1016/B0-08-043152-6/01744-7>
- Imperatore, P.; Di Martino, G. (2023). SAR Radiometric Calibration Based on Differential Geometry: From Theory to Experimentation on SAOCOM Imagery. *Remote Sensing*, 15, 1286. <https://doi.org/10.3390/rs15051286>
- Jackson, Thomas. J. (1987). Effects of soil properties on microwave dielectric constants. *Transportation Research Record*, 1119, 126-131. Available at <https://onlinepubs.trb.org/Onlinepubs/trr/1987/1119/1119-016.pdf> (accessed 2 November 2023)
- Jackson, T. J. & Schmugge, T. J. (1989). Passive microwave remote-sensing system for soil moisture: Some supporting research. *IEEE Transactions on Geoscience and Remote Sensing*, 27, 225-235. doi:10.1109/36.20301
- Jagdhuber, T. (2012). *Soil Parameter Retrieval under Vegetation Cover Using SAR Polarimetry*. PhD thesis. University of Potsdam, Potsdam, Germany.
- Jagdhuber, T., Hajnsek, I., Bronstert, A. & Papathanassiou, K. P. (2013). Soil moisture estimation under low vegetation cover using a multiangular polarimetric decomposition. *IEEE Transactions on Geoscience and Remote Sensing*, 51(4), 2201-2215. <https://doi.org/10.1109/TGRS.2012.2209433>
- Jagdhuber, T., Hajnsek, I. & Papathanassiou, P. K. (2015). An iterative generalized hybrid decomposition for soil moisture retrieval under vegetation cover using fully polarimetric SAR. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(8), 3911-3922. <https://doi.org/10.1109/JSTARS.2014.2371468>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2023). *An Introduction to Statistical Learning with Applications in R*. Springer Texts in Statistics, Springer, 2nd ed., pp. 615.
- Janowicz, K., Gao, S., McKenzie, G., Hu, Y., & Bhaduri, B. (2020). GeoAI: spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond. *International Journal of Geographical Information Science*, 34:4, 625-636, DOI: 10.1080/13658816.2019.1684500
- Jenny, H. (1941). *Factors of soil formation: A system of quantitative pedology*. New York, NY: McGrawHill, USA.
- Ji, K., & Wu, Y. (2015). Scattering Mechanism Extraction by a Modified Cloude-Pottier Decomposition for Dual Polarization SAR. *Remote Sensing* 7(6), 7447-7470. <https://doi.org/10.3390/rs70607447>
- Jiang, L.M.; Cui, H.Z.; Wang, G.X.; Yang, J.; Wang, J.; Pan, F.; Su, X.; Fang, X. (2020). Progress on Remote Sensing of Snow, Surface Soil Frozen/Thaw State and Soil Moisture. *Remote Sens. Technol. Appl.*, 35, 1237-1262.
- Jones, A., Fernández-Ugalde, O., Scarpa, S. LUCAS 2015 (2020). *Topsoil Survey. Presentation of dataset and results. EUR 30332 EN*. Publications Office of the European Union: Luxembourg. 2020, ISBN 978-92-76-21080-1, doi:10.2760/616084, JRC121325
- Kaya, F., Başayığit, L., Keshavarzi, A., Francaviglia, R. (2022). Digital mapping for soil texture class prediction in northwestern Türkiye by different machine learning algorithms, *Geoderma Regional*, 31, e00584, <https://doi.org/10.1016/j.geodrs.2022.e00584>
- Karam, M.A.; Fung, A.K.; Lang, R.H.; Chauhan, N.S. (1992). A microwave scattering model for layered vegetation. *IEEE Trans. Geosci. Remote Sens.*, 30, 767-784.
- Kasilingam, D., Nomula, M. & Cloude, S. (2002). A technique for removing vegetation bias from polarimetric SAR interferometry. *IEEE International Geoscience and Remote Sensing Symposium*, 2, 1017-1019. doi: 10.1109/IGARSS.2002.1025761.
- Katebikord, A., Sadeghi, S.H. & Singh, V.P. (2022). Spatial modeling of soil organic carbon using remotely sensed indices and environmental field inventory variables. *Environmental Monitoring and Assessment*, 194, 152. <https://doi.org/10.1007/s10661-022-09842-8>

- Khajehzadeh, M., Afzali, S. F., Honarbakhsh, A. *et al.* (2022) Remote sensing and GIS-based modeling for predicting soil salinity at the watershed scale in a semi-arid region of southern Iran. *Arabian Journal of Geosciences*, 15, 423, <https://doi.org/10.1007/s12517-022-09762-4>
- Kiely, G., Ms. McGoff, N. M., Eaton, J. M., Xu, X., Leahy, P., Carton, O. (2009). *SoilC - Measurement and Modelling of Soil Carbon Stocks and Stock Changes in Irish Soils*. EPA STRIVE Programme 2001-2007. STRIVE Report. SoilC Final Report – June 12, 2009.
- Kitchin, R. and Dodge, M. (2011). *Code/Space: Software and Everyday Life*. MIT Press, Cambridge.
- Kitchin, R. (2017) Thinking critically about and researching algorithms. *Information, Communication and Society*, 20(1): 14-29. <https://doi.org/10.1080/1369118X.2016.1154087>
- Klotzsche, A., Jonard, F., Looms, M. C., van der Kruk, J. & Huisman, J. A. (2018). Measuring soil water content with ground penetrating radar: A decade of progress. *Vadose Zone Journal*, 17(1). <https://doi.org/10.2136/vzj2018.03.0052>
- Kong, X.; Dorling, S.R. (2008). Near-surface soil moisture retrieval from ASAR Wide Swath imagery using a Principal Component Analysis. *International Journal of Remote Sensing*, 29, 2925-2942.
- Koppe, W., Gnyp, M. L., Hütt, C., Yao, Y., Miao, Y., Chen, X., Bareth, G. (2013). Rice monitoring with multi-temporal and dual-polarimetric TerraSAR-X data, *International Journal of Applied Earth Observation and Geoinformation*, 21, p568-576, <https://doi.org/10.1016/j.jag.2012.07.016>
- Kovačević, M., Bajat, B. & Gajić, B. (2010). Soil type classification and estimation of soil properties using support vector machines. *Geoderma*, 154 (3-4), 340–347. DOI:10.1016/j.geoderma.2009.11.005
- Krishnan, N., Manav, T.P., Geethanjali, A.N., Kumar, D., Seenu, N., & Balaji, G. (2022). Design and Development of Drone. *IEEE Bombay Section Signature Conference (IBSSC)*, Mumbai, India, 2022, pp. 1-6, doi: 10.1109/IBSSC56953.2022.10037366.
- Krivoruchko K and Gribov A (2019). Evaluation of empirical Bayesian kriging. *Spatial Statistics* 32: e100368, <https://doi.org/10.1016/j.spasta.2019.100368>
- Kumari, N., & Mohan, C. (2021). Basics of Clay Minerals and Their Characteristic Properties. In: Morari, do N., G. (ed.), *Clay and Clay Minerals*. IntechOpen. <http://dx.doi.org/10.5772/intechopen.97672>
- Lal, R., Monger, C., Nave, L., & Smith, P. (2021). The role of soil in regulation of climate. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1838) <http://doi.org/10.1098/rstb.2021.0084>
- Lauriault, T. P., Craig, B. L., Taylor, F. P., Peter, L.(2007). Today's Data are Part of Tomorrow's Research: Archival Issues in the Sciences. *Archivaria* 64, 123–179. Available at <https://archivaria.ca/index.php/archivaria/article/view/13156> (accessed 05 July 2021)
- Lee, J-S., Pottier, E., 2009. *Polarimetric Radar Imaging: From Basics to Applications*. 1ed. CRC Press, New York. <https://doi.org/10.1201/9781420054989>
- Lee, J.S.; Hoppel, K. (1992). Principal components transformation of multifrequency polarimetric SAR imagery. *IEEE Trans. Geosci. Remote Sens.*, 30, 686-696.
- Ließ, M., Glaser, B. & Huwe, B. (2012). Uncertainty in the spatial prediction of soil texture: Comparison of regression tree and Random Forest models. *Geoderma*, 170 (15), 70-79. <https://doi.org/10.1016/j.geoderma.2011.10.010>
- Linero, A. R., Yang, Y. (2018). Bayesian tree ensembles that adapt to smoothness and sparsity. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 80(5), 1087-1110, <https://doi.org/10.1111/rssb.12293>
- Liu, Chang-an; Chen, Zhong-xin; Shao, Yun; Chen, Jin-song; Hasi, Tuya; Pan, Hai-zhu (2019). Research advances of SAR remote sensing for agriculture applications: A review. *Journal of Integrative Agriculture*, 18 (3), 506-525, [https://doi.org/10.1016/S2095-3119\(18\)62016-7](https://doi.org/10.1016/S2095-3119(18)62016-7)
- Loosvelt, L., Vernieuwe, H., Pauwels, V. R. N., De Baets, B., & Verhoest, N. E. C. (2013). Local sensitivity analysis for compositional data with application to soil texture in hydrologic modelling. *Hydrology and Earth System Sciences*, 17, 461–478, <https://doi.org/10.5194/hess-17-461-2013>
- Magagi, R.; Jammali, S.; Goïta, K.; Wang, H.; Colliander, A. (2022). Potential of L- and C- Bands Polarimetric SAR Data for Monitoring Soil Moisture over Forested Sites. *Remote Sensing*, 14, 5317. <https://doi.org/10.3390/rs14215317>
- Mahaney, W. & Ramanathan, Krishnamani. (2003). Understanding geophagy in animals: Standard procedures for sampling soils. *Journal of chemical ecology*. 29(7). DOI: 10.1023/A:1024263627606
- Mahnke, M., & Uprichard, E. (2014). Algorithming the algorithm. In: R. König & M. Rasch (Eds.), *Society of the query reader: Reflections on web search*, pp. 256–270. Amsterdam: Institute of Network Cultures.
- Mandal, D. *et al.* (2020). Dual polarimetric radar vegetation index for crop growth monitoring using sentinel-1 SAR data. *Remote Sensing of Environment*, 247, 111954, <https://doi.org/10.1016/j.rse.2020.111954>

- Manolakis, D., Lockwood, R., Cooley, T. (2016). Spectral Mixture Analysis. In *Hyperspectral Imaging Remote Sensing: Physics, Sensors, and Algorithms*, pp. 443-493. Cambridge: Cambridge University Press. doi:10.1017/CBO9781316017876.010
- Marques F, P. C. (2022). Confidence intervals for the random forest generalization error, *Pattern Recognition Letters*, 158, 171-175, ISSN 0167-8655, <https://doi.org/10.1016/j.patrec.2022.04.031>.
- Martín-Fernández, J.A., Barceló-Vidal, C., Pawlowsky-Glahn, V. (1998). A critical approach to non-parametric classification of compositional data. In: *Advances in Data Science and Classification* (eds A. Rizzi, M. Vichi and H.H. Bock), Springer, Berlin.
- Martín-Fernández, J.A. (2019). Comments on: Compositional data: the sample space and its structure by Egozcue, J.J. and Pawlowsky-Glahn, V., *TEST* 28 (2019), no. 3, 653-657.
- Martinelli, G., & Gasser, M-O. (2022). Machine learning models for predicting soil particle size fractions from routine soil analyses in Quebec. *Soil Science Society of America Journal*, 86(6), 1509-1522, <https://doi.org/10.1002/saj2.20469>
- Marzahn, P., & Meyer, S. (2020). Utilization of multi-temporal microwave remote sensing data within a Geostatistical regionalization approach for the derivation of soil texture. *Remote Sensing*, 12(16), 2660. <https://doi.org/10.3390/rs12162660>
- Mascolo, L., Cloude, S. R., Lopez-Sanchez, J. M. (2022). Model-Based Decomposition of Dual-Pol SAR Data: Application to Sentinel-1. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1-19, e5220119, <https://doi.org/10.1109/TGRS.2021.3137588>
- Masoudi, M., Richards, D.R. ,& Tan, P.Y. (2024). Assessment of the Influence of Spatial Scale and Type of Land Cover on Urban Landscape Pattern Analysis Using Landscape Metrics. *Journal of Geovisualisation and Spatial Analysis* 8(8). <https://doi.org/10.1007/s41651-024-00170-8>
- Mayr, TR, Zawadzka JE, Corstanje R, *et al.* (2013). *Predictive Mapping*. ISIS Project Technical Report No.ISIS_WP2_D3.1, 78pp. Cranfield University and Teagasc
- McBratney, A. B., Santos, M. M., & Minasny, B. (2003). On digital soil mapping. *Geoderma*, 117, 3–52, [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)
- McClatchie, M., O'Carroll, E., & Reilly, E. (2015). *NRA Palaeo-environmental Sampling Guidelines Retrieval, analysis and reporting of plant macro-remains, wood, charcoal, insects and pollen from archaeological excavations*. National Rouds Authority, Ireland. Available at https://www.tii.ie/technical-services/archaeology/resources/NRA_Palaeo-environmental-Sampling_Guidelines_rev-4.pdf (accessed 12 Jan 2024).
- McDonald, N. T., Watson, C. J., Lalor, S. T. J., Laughlin, R. J., Wall, D. P. (2014). Evaluation of Soil Tests for Predicting Nitrogen Mineralization in Temperate Grassland Soils. *Soil Science Society of America Journal*, 78(3):1051–1064 doi:10.2136/sssaj2013.09.0411
- Met Éireann-The Irish Meteorological Service. (2021). *Climate Statement for March 2021*. Available at: <https://www.met.ie/climate-statement-for-march-2021> (accessed 19 June 2023).
- Metternicht, G. I.; Zinck, J. A (2003). Remote sensing of soil salinity: potentials and constraints. *Remote Sensing of Environment*, 85(1), 1-20, [https://doi.org/10.1016/S0034-4257\(02\)00188-8](https://doi.org/10.1016/S0034-4257(02)00188-8).
- Mialon G (2023). On Inductive Biases for Machine Learning in Data Constrained Settings. *Computer Science [cs]*. Ecole Normale Supérieure (ENS), Paris, FRA. Available at https://hal.science/tel-03999078/file/On_Inductive_Biases_for_Machine_Learning_in_Data_Constrained_Settings.pdf (accessed 10 October 2023)
- Minasny, B., McBratney, A. B. & Lark, R. M. (2008). Digital soil mapping technologies for countries with sparse data infrastructures. In: Hartemink, A.E.; McBratney, A.B.; Mendonça Santos, M.D.L. (eds). *Digital Soil Mapping with Limited Data*. Springer, London, p. 15–30
- Mironov, V. L. (2004). Bound water spectroscopy for the soils with varying mineralogy. *IEEE International Geoscience and Remote Sensing Symposium*, Anchorage, AK, USA, 5: 3478-3480, doi: 10.1109/IGARSS.2004.1370456.
- Mirzaeitalarposhti, R.; Shafizadeh-Moghadam, H.; Taghizadeh-Mehrjardi, R.; Demyan, M.S. (2022). Digital Soil Texture Mapping and Spatial Transferability of Machine Learning Models Using Sentinel-1, Sentinel-2, and Terrain-Derived Covariates. *Remote Sensing*, 14, 5909. <https://doi.org/10.3390/rs14235909>
- Molin, J. P., & Faulin, G. D. C. (2013). Spatial and temporal variability of soil electrical conductivity related to soil moisture. *Scientia Agricola*, 70(1), 01-05. <https://doi.org/10.1590/S0103-90162013000100001>
- Mondejar, J. P. & Tongco, A. F. (2019). Estimating topsoil texture fractions by digital soil mapping - a response to the long outdated soil map in the Philippines. *Sustainable Environment Research*, 29:31. <https://doi.org/10.1186/s42834-019-0032-5>

- Monteiro *et al.* (2023). Digital mapping of soil classes in Southeast Brazil: environmental covariate selection, accuracy, and uncertainty. *Journal of South American Earth Sciences*, 132, 104640, <https://doi.org/10.1016/j.jsames.2023.104640>
- Montzka, C., Rötzer, K., Bogen, H. R., Sanchez, N. & Vereecken, H. (2018). A new soil moisture downscaling approach for SMAP, SMOS, and ASCAT by predicting sub-grid variability. *Remote Sensing*, 10(3), 427. <https://doi.org/10.3390/rs10030427>
- Morais & Thomas-Agnan (2021). Impact of Covariates in Compositional Models and Simplicial Derivatives. *Austrian Journal of Statistics*, 50, 1-15. <http://www.ajs.or.at/doi:10.17713/ajs.v50i2.1069>
- Moran, M. S., Alonso, L., Moreno, J. F., Mateo, M. P. C., De la Cruz, D. F., & Montoro, A. (2012). A RADARSAT-2 Quad-Polarized Time Series for Monitoring Crop and Soil Conditions in Barrax, Spain. *IEEE Transactions on Geoscience and Remote Sensing*, 50(4), 1057-1070. doi: 10.1109/TGRS.2011.2166080
- Morandeira, N.S.; Barber, M.E.; Grings, F.M.; Ahern, F.; Kandus, P.; Brisco, B. Response of Multi-Incidence Angle Polarimetric RADARSAT-2 Data to Herbaceous Vegetation Features in the Lower Paraná River Floodplain, Argentina. *Remote Sens.* 2021, 13, 2518. <https://doi.org/10.3390/rs13132518>
- Morellos A., Pantazi, X-E., Moshou, D., Alexandridis, T., Whetton, R., Tziotzios, G., Wiebensohn, J., Bill, R., & Mouazen, A. M. (2016). Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. *Biosystems Engineering*, 152, Pages 104-116. <https://doi.org/10.1016/j.biosystemseng.2016.04.018>
- Motia, S., & Reddy, S.R.N. (2021). *Journal of Physics*, Conference Series: 1950 012037.
- Mouazen, A. M. & Shi, Z. (2021). Estimation and Mapping of Soil Properties Based on Multi-Source Data Fusion. *Remote Sensing*, 13(5), 978. <https://doi.org/10.3390/rs13050978>
- Müller, B., Bernhardt, M., Jackisch, C., & Schulz, K. (2016). Estimating spatially distributed soil texture using time series of thermal remote sensing – a case study in central Europe. *Hydrology and Earth System Sciences*, 20, 3765–3775, <https://doi.org/10.5194/hess-20-3765-2016>
- Narayan, U.; Lakshmi, V.; Njoku, E.G. (2004). Retrieval of soil moisture from passive and active L/S band sensor (PALS) observations during the Soil Moisture Experiment in 2002 (SMEX02). *Remote Sensing of Environment*, 92, 483-496.
- NASA EarthData (n.d.). *What is Synthetic Aperture Radar?* Available at : <https://www.earthdata.nasa.gov/learn/backgrounders/what-is-sar> (accessed 22 Jan 2024).
- Nash, J. E., & Sutcliffe, J. V. (1970). River Flow Forecasting through Conceptual Model. Part 1-A Discussion of Principles. *Journal of Hydrology*, 10, 282-290. [http://dx.doi.org/10.1016/0022-1694\(70\)90255-6](http://dx.doi.org/10.1016/0022-1694(70)90255-6)
- Nasirzadehdizaji, R., Sanli, F.B., Abdikan, S., Cakir, Z., Sekertekin, A., Ustuner, M. (2019). Sensitivity Analysis of Multi-Temporal Sentinel-1 SAR Parameters to Crop Height and Canopy Coverage. *Applied Sciences*, 9(4):655. <https://doi.org/10.3390/app9040655>
- Navacchi, C.; Cao, S.; Bauer-Marschallinger, B.; Snoeij, P.; Small, D.; Wagner, W. (2023). Utilising Sentinel-1's Orbital Stability for Efficient Pre-Processing of Radiometric Terrain Corrected Gamma Nought Backscatter. *Sensors*, 23, 6072. <https://doi.org/10.3390/s23136072>
- Niang, M. A., Nolin, M. C., Jégo, G. & Perron, I. (2014). Digital Mapping of Soil Texture Using RADARSAT-2 Polarimetric Synthetic Aperture Radar Data. *Soil Science Society of America Journal*, 78:673–684. doi:10.2136/sssaj2013.07.0307
- Nikparvar, B., & Thill, J.-C. (2021). Machine Learning of Spatial Data. *ISPRS International Journal of Geo-Information*, 10, 600. <https://doi.org/10.3390/ijgi10090600>
- Nolan, M.; Fatland, D.R.; Hinzman, L. (2003). DInSAR measurement of soil moisture. *IEEE Trans. Geosci. Remote Sens.*, 41, 2802-2813.
- O'Geen, A. T. (2013). Soil Water Dynamics. *Nature Education Knowledge* 4(5):9. Available at <https://www.nature.com/scitable/knowledge/library/soil-water-dynamics-103089121/> (accessed 13 Jan 2024).
- Oh, Y.; Sarabandi, K.; Ulaby, F.T. (1992). An empirical model and an inversion technique for Radar scattering from bare soil surfaces. *IEEE Trans. Geosci. Remote Sens.* 30, 370–381.
- O'Hara, R. (2019). *The application of Earth Observation for mapping soil saturation and the extent and distribution of artificial drainage on Irish farms*. PhD thesis. National University of Ireland, Maynooth.
- O'Neill, P.; Chauhan, N.; Jackson, T. (1996). Use of active and passive microwave remote sensing for soil moisture estimation through corn. *International Journal of Remote Sensing*, 17, 1851-1865.
- O'Sullivan, L., McConnell, B., Scanlon, R., Walsh, S., Creamer, R. (2018). Soil Formation. In: Creamer, R. & O'Sullivan, L. eds. (2018). *The Soils of Ireland*. World Soils Book Series. Springer, Cham. Available at: <https://doi.org/10.1007/978-3-319-71189-8> (accessed 4 April 2021)

- O'Sullivan, L., Creamer, R., & Schulte, R. (2018). Functional Land Management: A Framework for Soil Policy Formation. In: Creamer and L. O'Sullivan (eds.), *The Soils of Ireland*, World Soils Book Series, https://doi.org/10.1007/978-3-319-71189-8_21
- O'Sullivan, L., Creamer, R. (2018). Landscapes: An Introduction. In: Creamer, R., O'Sullivan, L. (eds). *The Soils of Ireland*. World Soils Book Series. Springer, Cham. https://doi.org/10.1007/978-3-319-71189-8_5
- Odeh, I. O. A., Todd, A. J., & Triantafilis, J. (2003). Spatial prediction of soil particle-size fractions as compositional data. *Soil Science*, 168. <https://doi.org/10.1097/01.ss.0000080335.10341.23>
- Ogen, Y.; Denk, M.; Glaesser, C., & Eichstaedt, H. (2022). A novel method for predicting the geochemical composition of tailings with laboratory field and hyperspectral airborne data using a regression and classification-based approach. *European Journal of Remote Sensing*, 55:1, 453-470, <https://doi.org/10.1080/22797254.2022.2104173>
- Okabe A, Boots B, Sugihara K, *et al.* (2000). Definitions and Basic Properties of Voronoi Diagrams. In: Okabe *et al.* *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. John Wiley & Sons Ltd., pp43-112. DOI:10.1002/9780470317013
- Openshaw, S., & Openshaw, C. (1997). *Artificial intelligence in geography*. 1st ed. New York, NY: John Wiley & Sons, Inc. ISBN 0471969915.
- Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A., Fernández-Ugalde, O. (2018). LUCAS Soil, the largest expandable soil dataset for Europe: A review. *European Journal of Soil Science*, 69(1): 140-153, DOI: 10.1111/ejss.12499
- OSCAR – Observing Systems Capability Analysis and Review Tool (n.d.). *Satellite: SCLP. Satellite details*. Available at: <https://space.oscar.wmo.int/satellites/view/scip> (accessed 22 January 2024).
- Padarian, J., Minasny, B., & McBratney, A. B. (2020). Machine learning and soil sciences: a review aided by machine learning tools, *SOIL*, 6, 35–52, <https://doi.org/10.5194/soil-6-35-2020>
- Panagos, P., Van Liedekerke, M., Jones, A. & Montanarella, A. (2012). European Soil Data Centre: Response to European policy support and public data Requirements. *Land Use Policy*, 29, 329–338
- Patel, K.F., Fansler, S.J., Campbell, T.P. *et al.* (2021). Soil texture and environmental conditions influence the biogeochemical responses of soils to drought and flooding. *Commun Earth Environ*, 2, 127. <https://doi.org/10.1038/s43247-021-00198-4>
- Patera, R. P. (2008). Hazard analysis for uncontrolled space vehicle reentry. *Journal of Space Crafts and Rockets*, 45(5), 1031–1041, <https://doi.org/10.2514/1.30173>
- Patton, N. R., Lohse, K.A., Godsey, S. E. *et al.* (2018). Predicting soil thickness on soil mantled hillslopes. *Nature Communications* 9, 3329. <https://doi.org/10.1038/s41467-018-05743-y>
- Pawlowsky-Glahn, V., Egozcue, J. J. & Tolosona-Delgado, R. (2015). *Modeling and Analysis of Compositional Data*, Wiley, 2015, 272 p.
- Pearson, K. (1897). *Mathematical contributions to the theory of evolution. on a form of spurious correlation which may arise when indices are used in the measurements of organs*. Proc. R. Soc. 60 (1897), 489-498.
- Pellicer, X. M., Lee, M., Meehan, R., & Verbruggen, K. (2018). Physiographic Units Map of Ireland. Geological Survey Ireland. Document Information, February 2018.
- Periasamy, Shoba (2018). Significance of dual polarimetric synthetic aperture radar in biomass retrieval: An attempt on Sentinel-1. *Remote Sensing of Environment*, 217, 537-549, <https://doi.org/10.1016/j.rse.2018.09.003>
- Petropoulos G. P, Ireland G. & Barrett B (2015). Surface soil moisture retrievals from remote sensing: Current status, products & future trends. *Physics and Chemistry of the Earth*, 83–84, p.36–56. <https://doi.org/10.1016/j.pce.2015.02.009>
- Pichierri, M., Hajnsek, I., Zwieback, S. & Rabus, B. (2018). On the potential of Polarimetric SAR Interferometry to characterize the biomass, moisture and structure of agricultural crops at L-, C- and X Bands. *Remote Sensing of Environment*, 204, 596–616. <http://dx.doi.org/10.1016/j.rse.2017.09.039>
- Pittman, R. & Hu, B. (2021). Improvement of Soil Texture Classification with LiDAR Data. In: *IEEE International Geoscience and Remote Sensing Symposium*, September-October 2020. Waikoloa, HI, USA: IEEE, pp. 5018-5021, doi: 10.1109/IGARSS39084.2020.9324152.
- Ponnurangam, G.G. and Rao, Y.S. (2017). Evaluation of different orientation angle distributions within the X-Bragg scattering model for bare soil moisture estimation. *International Journal of Remote Sensing*, 38(15), 4379-4395. <https://doi.org/10.1080/01431161.2017.1320447>
- Poppiel, R.R.; Lacerda, M.P.C.; Rizzo, R.; Safanelli, J.L.; Bonfatti, B.R.; Silvero, N.E.Q.; Demattê, J.A.M. (2020). Soil Color and Mineralogy Mapping Using Proximal and Remote Sensing in Midwest Brazil. *Remote Sensing*, 12, 1197. <https://doi.org/10.3390/rs12071197>

- Pradipta, A.; Soupios, P.; Kourgialas, N.; Doula, M. ; Dokou, Z.; Makkawi, M.; Alfarhan, M.; Tawabini, B.; Kirmizakis, P., & Yassin, M. (2022). Remote Sensing, Geophysics, and Modeling to Support Precision Agriculture—Part 1: Soil Applications. *Water*, 14(7):1158. <https://doi.org/10.3390/w14071158>
- Pratola, C., Barrett, B., Gruber, A., Kiely, G. & Dwyer, E. (2014). Evaluation of a Global Soil Moisture Product from Finer Spatial Resolution SAR Data and Ground Measurements at Irish Sites. *Remote Sensing*, 6, 8190–8219; doi:10.3390/rs6098190
- Pratola, C., Barrett, B., Gruber, A. & Dwyer, E. (2015). Quality Assessment of the CCI ECV Soil Moisture Product Using ENVISAT ASAR Wide Swath Data over Spain, Ireland and Finland. *Remote Sensing*, 7, 15388–15423; doi:10.3390/rs71115388
- Qu, J.; Qiu, X.; Wang, W.; Wang, Z.; Lei, B.; Ding, C. A. (2022). Comparative Study on Classification Features between High-Resolution and Polarimetric SAR Images through Unsupervised Classification Methods. *Remote Sensing*, 14, 1412. <https://doi.org/10.3390/rs14061412>
- Qu, Y.; Zhao, W.; Yuan, Z.; Chen, J.(2020). Crop Mapping from Sentinel-1 Polarimetric Time-Series with a Deep Neural Network. *Remote Sensing*, 12, 2493. <https://doi.org/10.3390/rs12152493>
- Rahimzadeh-Bajgiran, P., Berg, A. A., Champagne, C. & Omasa, K. (2013). Estimation of soil moisture using optical/thermal infrared remote sensing in the Canadian Prairies. *ISPRS Journal of Photogrammetry and Remote Sensing*, 83, 94–103. <https://doi.org/10.1016/j.isprsjprs.2013.06.004>
- Rao, K. S., Rao, P. V. N., Rao, Y. S., Chandra, G., Raju, C. S. & Bapat, M. V. (1988). A study on the effect of the soil texture on passive microwave remote sensing. *Journal of Indian Society of Remote Sensing*, 16, 55–63 (1988). <https://doi.org/10.1007/BF02998739>
- Read, C. F., Duncan, D. H., Ho, C. Y. C., White, M., & Vesk, P. A. (2018). Useful surrogates of soil texture for plant ecologists from airborne gamma-ray detection. *Ecology and Evolution*, 8:1974–1983. <https://doi.org/10.1002/ece3.3417>
- Renou-Wilson, F. (2018). Peatlands. In: Creamer R., O’Sullivan L. (eds). *The Soils of Ireland*. World Soils Book Series. Springer, Cham. 11-37. Available at: <https://link.springer.com/book/10.1007/978-3-319-71189-8> (accessed 17 Jan 2024)
- Reynolds, W. D., Bowman, B. T., Drury, C. F., Tan, C. S. & Lu, X. (2002). Indicators of good soil physical quality: Density and storage parameters. *Geoderma*, 110 (1), 131–146. [https://doi.org/10.1016/S0016-7061\(02\)00228-8](https://doi.org/10.1016/S0016-7061(02)00228-8)
- Ribeiro E., Batjes, N. H., van Oostrum, A. J. M. (2020). World Soil Information Service (WoSIS)-Towards the standardization and harmonization of world soil profile data. Procedures manual 2020, Report 2020/01, ISRIC - World Soil Information, Wageningen. <http://dx.doi.org/10.17027/isric-wdc-2020-01>
- Richards, J.A. (2009). The Imaging Radar System. In: Remote Sensing with Imaging Radar. Signals and Communication Technology. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-02020-9_1
- Rignot, E.J.M.; van Zyl, J.J. (1993). Change detection techniques for ERS-1 SAR data. *IEEE Trans. Geosci. Remote Sensing*, 31, 896–906.
- Roberts, D., Wilford, J. & Ghattas, O. (2019). Exposed soil and mineral map of the Australian continent revealing the land at its barest. *Nature Communications*, 10, 5297. <https://doi.org/10.1038/s41467-019-13276-1>
- Rossel *et al.* (2016). A global spectral library to characterize the world's soil. *Earth-Science Reviews*, 155, 198–230, <https://doi.org/10.1016/j.earscirev.2016.01.012>
- Rossiter, D.G. (1999). Soil Geographic Databases SOL.SGDB. Lecture Notes. Soil Science Division International Institute for Aerospace Survey & Earth Sciences (ITC), Version 2, February 1999. Available at https://www.css.cornell.edu/faculty/dgr2/_static/files/pdf/SoilGeographicDataBases.pdf (accessed 18 July 2024)
- Rossiter, D.G., Poggio, L., Beaudette, D., & Libohova, Z. (2022). How well does digital soil mapping represent soil geography? An investigation from the USA. *SOIL*, 8, 559–586, <https://doi.org/10.5194/soil-8-559-2022>
- Salma, S., Keerthana, N., Dodamani, B.M. (2022). Target decomposition using dual-polarization sentinel-1 SAR data: Study on crop growth analysis. *Remote Sensing Applications: Society and Environment*, 28, e100854, <https://doi.org/10.1016/j.rsase.2022.100854>
- Sano, E. E.; Matricardi, E. A. T. & Camargo, F. F. (2020). State-of-the-art of the Radar Remote Sensing: Fundamentals, Sensors, Image Processing, and Applications. *Revista Brasileira de Cartografia*, 72, 1458–1483. <https://doi.org/10.14393/rbcv72nespecial50anos-56568>
- Santos, M. C. P., Carbonera, M., Rosina, P., Schuster, A. J., Pavei, D. D., Hatté, C., Souza, Á. S. De, Campos, J., Lourdeau, A. (2021). Holocene settlement, stratigraphy and chronology at the site of Uruguai 1-sector

- 1, Foz do Chapecó archaeological area, South Brazil, *Journal of Archaeological Science: Reports*, 39, 103113, <https://doi.org/10.1016/j.jasrep.2021.103113>
- SAS – Statistical Analysis System-SAS (n.d.). *Machine Learning. What it is and why it matters*. Available at: https://www.sas.com/en_us/insights/analytics/machine-learning.html (accessed 01 July 2021)
- Saurette, D. Daniel (2022). Comparing direct and indirect approaches to predicting soil texture class. *Canadian Journal of Soil Science*, 102(4): 835-851. <https://doi.org/10.1139/cjss-2022-0040>
- Sayão, V. M. & Demattê, J. A. M. (2018). Soil texture and organic carbon mapping using surface temperature and reflectance spectra in Southeast Brazil. *Geoderma Regional*, 14, e00174. <https://doi.org/10.1016/j.geodrs.2018.e00174>
- Schaefer, F. K. (1953). Exceptionalism in Geography: A Methodological Examination. *Annals of the Association of American Geographers*, 43(3), 226–249. <https://doi.org/10.2307/2560876>
- Schmugge, Thomas J. (1980). Effects of Texture on Microwave Emission from Soils. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. GE-18, pp. 353-361.
- Schumann, G. & Moller, D. (2015). Microwave remote sensing of flood inundation. *Physics and Chemistry of the Earth*, Parts A/B/C. 83, 84-95. <https://doi.org/10.1016/j.pce.2015.05.002>
- Schönbrodt-Stitt S., Ahmadian N., Kurtenbach M., Conrad C., Romano N., Bogen, H.R, Vereecken, H. & Nasta, P. (2021) Statistical Exploration of SENTINEL-1 Data, Terrain Parameters, and in-situ Data for Estimating the Near-Surface Soil Moisture in a Mediterranean Agroecosystem. *Front. Water* 3:655837. doi: 10.3389/frwa.2021.655837
- Seaver, N. (2013). Knowing Algorithms. *Media in Transition*, 8, Cambridge, MA. Available at <http://nickseaver.net/papers/seaverMiT8.pdf>. (accessed 05 July 2021)
- Sheldon, N. D. & Tabor, N. (2009). Quantitative paleoenvironmental and paleoclimatic reconstruction using paleosols. *Earth-science reviews*, 95(1-2), 1-52. <https://doi.org/10.1016/j.earscirev.2009.03.004>
- Shi, J.C.; Jiang, L.M.; Zhang, L.X.; Chen, K.S.; Wigneron, J.P.; Chanzy, A. (2005). A Parameterized Multi-frequency-polarization Surface Emission Model. *IEEE Trans. Geosci. Remote Sens.* 43, 2831–2841.
- Shi, W., Zhang, M. (2023). Progress on spatial prediction methods for soil particle-size fractions. *Journal of Geographical Sciences*, 33, 1553–1566. <https://doi.org/10.1007/s11442-023-2142-6>
- Siad, Si M. (2023). Advances in SAR-Based Soil Moisture Retrieval Techniques, Applications, and Challenges. *Humanities Commons*. DOI: 10.17613/pr06-fy12. <https://doi.org/10.17613/pr06-fy12>
- Silva, E. B; Giasson, E.; Dotto, A.C; Ten Caten, A.; Demattê, J. A. M.; Bacic, I. L. Z. & Veiga, M. A. (2019). Regional legacy soil dataset for prediction of sand and clay content with VIS-NIR-SWIR in Southern Brazil. *Revista Brasileira de Ciência do Solo*, 2019;43:e0180174. <https://doi.org/10.1590/18069657rbcs20180174>
- Silvero, M.E.Q., Demattê, J.A.M, Minasny, B., Rosin, N.A., Nascimento, J.G., Albarracín, H.S.R., Bellinaso, H., Gómez, A.M.R. (2023). Chapter Three - Sensing technologies for characterizing and monitoring soil functions: A review. Editor(s): Sparks, D.L. *Advances in Agronomy*, Academic Press, 177:125-168, <https://doi.org/10.1016/bs.agron.2022.08.002>.
- Singh, S., & Sarma, K. (2023). Exploring Soil Spatial Variability with GIS, Remote Sensing, and Geostatistical Approach. *Journal of Soil, Plant and Environment*, 2(1), 79-99, <https://doi.org/10.56946/jspae.v2i1.186>
- Sleeman A; McConnell B; Gatley S. (2004). *Understanding Earth Process, Rocks and the Geological History of Ireland. A companion to the 1:1,000,000 scale bedrock geological map of Ireland*. Geological Survey of Ireland.
- Small, D. (2011). Flattening gamma: Radiometric terrain correction for SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 49, 3081–3093.
- Smith, T. E. (2009). Estimation Bias in Spatial Models with Strongly Connected Weight Matrices. *Geographical Analysis*, 41: 307-332. <https://doi.org/10.1111/j.1538-4632.2009.00758.x>
- Smith, T. R. (1984). Artificial intelligence and its applicability to geographical problem solving. *The Professional Geographer*, 360 (2), 147–158. doi:10.1111/j.0033-0124.1984.00147.x
- Smith, D. & Peng, W. (2009). *Machine learning approaches for soil classification in a multi-agent deficit irrigation control system*. In: *IEEE International Conference on Industrial Technology*. 2009. Churchill. Available at: <http://hdl.handle.net/102.100.100/111470?index=1> (accessed 17 June 2021)
- Somers, B. Asner, G. P; Tits, L.; Coppin, P. (2016). Endmember variability in Spectral Mixture Analysis: A review. *Remote Sensing of Environment*, 115(7), 1603-1616, <https://doi.org/10.1016/j.rse.2011.03.003>
- Sowter, Andrew (n.d.). *Introduction to Radar. Landmap Geoknowledge*. Available at <https://learningzone.rspoc.org.uk/index.php/Learning-Materials/Radar-Imaging/Image-Interpretation-Polarisation> (accessed 3 August 2023)

- Srivastava, H. S., Patel, P., & Navalgund, R. R. (2006). How far SAR has fulfilled its expectation for soil moisture retrieval? SPIE Digital Library, 64100, 1–12. In *Microwave Remote Sensing of Atmosphere and Environment-II*, AE107. doi: 10.1117/12.693946
- Srivastava, S.K.; Jayaraman, V. (2001). Relating interferometric signature of repeat pass ERS-1 SAR signals to dynamic land cover changes. *Acta Astronaut.* 48, 37-44.
- Sugiyama, M. (2016). *Introduction to Statistical Machine Learning*, Elsevier Inc., ISBN 978-0-12-802121-7. <https://doi.org/10.1016/C2014-0-01992-2>
- Suleymanov, A.; Gabbasova, I.; Komissarov, M.; Suleymanov, R.; Garipov, T.; Tuktarova, I.; Belan, L. (2023). Random Forest Modeling of Soil Properties in Saline Semi-Arid Areas. *Agriculture*, 13, 976. <https://doi.org/10.3390/agriculture13050976>
- Suman, S. (2022). H/A/α Polarimetric Decomposition Of Dual Polarized Alos Palsar For Efficient Land Feature Detection And Biomass Estimation Over Tropical Deciduous Forest. *Geography, Environment, Sustainability*, 3(15), 37-46, <https://DOI-10.24057/2071-9388-2021-095>
- Taghizadeh-Mehrjardi, R., Mahdianpari, M., Mohammadimanesh, F., Behrens, T., Toomanian, N., Scholten, T., & Schmidt, K. (2020). Multi-task convolutional neural networks outperformed random forest for mapping soil particle size fractions in central Iran. *Geoderma*, 376, 114552.
- Tannous, O. & Kasilingam, D. (2009). Independent component analysis of polarimetric SAR data for separating ground and vegetation components. *IEEE International Geoscience and Remote Sensing Symposium*, pp. IV-93-IV-96. <https://doi.org/10.1109/IGARSS.2009.5417366>
- Thoma, D. P., Moran, M. S., Bryant, R., Rahman, M., Holifield-Collins, C. D., Skirvin, S., Sano, E. E., & Slocum, K. (2006). Comparison of four models to determine surface soil moisture from C-band radar imagery in a sparsely vegetated semiarid landscape. *Water Resource Research*, 42, W01418, doi:10.1029/2004WR003905.
- Thoma, D. P., Moran, M. S., Bryant, R., Rahman, M. M., Holifield Collins, C. D., Keefer, T. O., Noriega, R., Osman, I., Skirvin, S. M., Tischler, M. A., Bosch, D. D., Starks, P. J. & Peters-Lidard, C. D. (2008) Appropriate scale of soil moisture retrieval from high resolution radar imagery for bare and minimally vegetated soils, *Remote Sensing of Environment*, 112(2): 403-414. <https://doi.org/10.1016/j.rse.2007.06.021>
- Thomas, A., Cosby, B. J., Henrys, P., & Emmett, B. (2020). Patterns and trends of topsoil carbon in the UK: Complex interactions of land use change, climate and pollution. *Science of The Total Environment*, 729, <https://doi.org/10.1016/j.scitotenv.2020.138330>
- Ting, J-A; Vijaykumar, S., & Schaal, S. (2011). Locally Weighted Regression for Control. In Sammut, C.; Webb, G.I. (eds.). *Encyclopedia of Machine Learning*. Springer: New York, p. 615. <https://doi.org/10.1007/978-0-387-30164-8>
- Tobin, B., Sara, Vero, S. E., Fenton, O., O’Sullivan, L., & Tuohy, P. (2018). Soils and Productivity. In: Creamer, R. & O’Sullivan, L. eds. (2018). *The Soils of Ireland*. World Soils Book Series. Springer, Cham. Available at: <https://doi.org/10.1007/978-3-319-71189-8> (accessed 4 April 2021)
- Tobler, W. R. (1987). Measuring Spatial Resolution. *Proceedings, Land Resources Information Systems Conference*, Beijing, pp. 12-16.
- Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography* 46:234–240. <https://doi.org/10.2307/143141>
- Todorov, V. (2021). Monitoring Robust Estimates for Compositional Data. *Austrian Journal of Statistics*. Vol. 50, 16-37. <http://dx.doi.org/10.17713/ajs.v50i2.1067>
- Tóth, T., Jones, A. & Montanarella, L. (2013). The LUCAS topsoil database and derived information on the regional variability of cropland topsoil properties in the European Union. *Environmental Monitoring and Assessment*, 185:7409–7425 DOI 10.1007/s10661-013-3109-3
- Ture, M., Muller, M., Hodgson, J. (2020). *Tellus Interpretation Guide to Airborne Geophysical Data. Report 2020*. Geological Survey Ireland. Available at: <https://www.gsi.ie/en-ie/publications/Pages/Tellus-Interpretation-Guide-.aspx> (accessed 2 December 2022)
- UIUC (2010). *Effects of Wavelength on the ability to detect an object*. University of Illinois at Urbana-Champaign. [http://ww2010.atmos.uiuc.edu/\(Gh\)/guides/rs/rad/basics/wvl.rxml](http://ww2010.atmos.uiuc.edu/(Gh)/guides/rs/rad/basics/wvl.rxml) (accessed 19 Jan 2024).
- Ulaby, F. T.; Batlivala, P. P. & Dobson, M. C. (1978). Microwave Backscatter Dependence on Surface Roughness, Soil Moisture, and Soil Texture: Part I-Bare Soil. *IEEE Transactions on Geoscience Electronics*, 16(4), 286-295. doi: 10.1109/TGE.1978.294586.
- Ulaby, F. T.; Bradley, G. A. & Dobson, M. C. (1979). Microwave Backscatter Dependence on Surface Roughness, Soil Moisture, and Soil Texture: Part II-Vegetation-Covered Soil. *IEEE Transactions on Geoscience Electronics*, 17(2), 33-40. doi: 10.1109/TGE.1979.294626.

- Ulaby, F.T., Moore, R.K., & Fung, A.K. (1982) Radar Remote Sensing and Surface Scattering and Emission Theory. *Microwave Remote Sensing Active and Passive*, I, 624 p.
- UNSW – University of New South Wales (2007). TerraGIS Soil. Available at http://www.terragis.bees.unsw.edu.au/terragis_soil/sp_particle_size_fractions.html (accessed 13 Jan 2024).
- Urso, L., Petermann, E., Gnädinger, F., & Hartmann, P. (2023). Use of random forest algorithm for predictive modelling of transfer factor soil-plant for radiocaesium: A feasibility study. *Journal of Environmental Radioactivity*, vol. 270, 107309, <https://doi.org/10.1016/j.jenvrad.2023.107309>
- USDA/NSCS (n.d.). *Basic Soil Properties*. United States Department of Agriculture. https://www.nrcs.usda.gov/sites/default/files/2023-03/Soils2023_Day1_BasicSoilProperties.pdf (accessed 16 Jan 2024).
- van Wesemael, B.; Chabrilat, S.; Sanz Dias, A.; Berger, M.; Szantoi, Z. (2023). Remote Sensing for Soil Organic Carbon Mapping and Monitoring. *Remote Sensing*, 15, 3464. <https://doi.org/10.3390/rs15143464>
- van Zyl, J., & Kim, Y. (2010). *Synthetic Aperture Radar Polarimetry*. JPL Space Science and Technology Series, Joseph H. Yuen, Series Editor, Jet Propulsion Laboratory California Institute of Technology-USA, December 2010.
- Wadoux, A.M.J.-C. (2019). *Sampling design optimization for geostatistical modelling and prediction*. PhD Thesis. Wageningen University, Wageningen, the Netherlands.
- Wadoux AMJ, Heuvelink GBM, Uijlenhoet R, de Bruin S (2020). Optimization of rain gauge sampling density for river discharge prediction using Bayesian calibration. *PeerJ* 8:e9558. doi: 10.7717/peerj.9558. PMID: 32821535; PMCID: PMC7396144
- Wadoux, A-C.; Samuel-Rosa, A.; Poggio, L., Mulder, V.L. (2020b). A note on knowledge discovery and machine learning in digital soil mapping. *European Journal of Soil Science*, 71:133-136, <https://doi.org/10.1111/ejss.12909>
- Wadoux, A.M.J.-C.; Román-Dobarco, M.; McBratney, A.B. (2021). Perspectives on data-driven soil research. *European Journal of Soil Science*, 72:1675-1689, <https://doi.org/10.1111/ejss.13071>
- Walsh, S. (2012). *A summary of climate averages for Ireland 1981-2010. Climatological Note no.14*. Met Éireann. Dublin, May 2012. Available at: <https://www.met.ie/climate-ireland/SummaryClimAvgs.pdf> (accessed 22 March 2023).
- Wang D-C, Zhang G-L, Zhao M-S, Pan X-Z, Zhao Y-G, Li D-C, *et al.* (2015). Retrieval and Mapping of Soil Texture Based on Land Surface Diurnal Temperature Range Data from MODIS. *PLoS ONE*, 10(6): e0129977. <https://doi.org/10.1371/journal.pone.0129977>
- Wang, H.; Magagi, R.; Goita, K. (2017). Comparison of different polarimetric decompositions for soil moisture retrieval over vegetation covered agricultural area. *Remote Sensing of Environment*, 199, 120–136, <http://dx.doi.org/10.1016/j.rse.2017.07.008>
- Wang, K., Zhang, Z. & Li, W. (2012). Comparison of Geographically Weighted Regression and Regression Kriging for Estimating the Spatial Distribution of Soil Organic Matter. *GIScience & Remote Sensing*, 49 (6), 915–932. <https://doi.org/10.2747/1548-1603.49.6.915>
- Wang, J. R. (1980). The dielectric properties of soil-water mixtures at microwave frequencies. *Radio Science*, 15 (5), 977-985. doi: 10.1029/RS015i005p00977
- Wang, J. R., Schmugge, T. J. (1980). An Empirical Model for the Complex Dielectric Permittivity of Soils as a Function of Water Content. *IEEE Trans Geosci Remote Sens*, GE-18, 288-295. DOI:10.1109/TGRS.1980.350304
- Wang Y, Zhao H, Fan J, Wang C, Ji X, Jin D, Chen J. (2023). A Review of Earth's Surface Soil Moisture Retrieval Models via Remote Sensing. *Water*. 15(21):3757. <https://doi.org/10.3390/w15213757>
- Wang, Z., Shi, W. (2017). Mapping soil particle-size fractions: A comparison of compositional kriging and log-ratio kriging. *Journal of Hydrology*, 546: 526-541, <https://doi.org/10.1016/j.jhydrol.2017.01.029>
- Wang, Z., Shi, W. (2018). Robust variogram estimation combined with isometric log-ratio transformation for improved accuracy of soil particle-size fraction mapping. *Geoderma*, Vol 324, 56-66, <https://doi.org/10.1016/j.geoderma.2018.03.007>
- Webster, R., & Oliver, M. A. (2007). *Geostatistics for Environmental Scientists*. John Wiley & Sons, Chichester.
- Weiss, S., Xu, Z.Z., Peddada, S. *et al.* (2017) Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* 5, 27. <https://doi.org/10.1186/s40168-017-0237-y>
- Wellock, M. (2011). *The Impact of Afforestation on the Carbon Stocks of Irish Soils*. PhD Thesis. National University of Ireland-Cork, Cork, Republic of Ireland.

- Wig, E., Michaelides, R., Zebker, H. (2022). High-Resolution Measurement of Soil Moisture from Insar Phase Closure. *IGARSS IEEE International Geoscience and Remote Sensing Symposium*, Kuala Lumpur, Malaysia, 2022, pp. 919-922, doi: 10.1109/IGARSS46834.2022.9884835.
- Wilford, J. R., Bierwirth, P. N., & Craig, M. A. (1997). Application of airborne gamma-ray spectrometry in soil/regolith mapping and applied geomorphology. *AGSO Journal of Australian Geology and Geophysics*, 17(2), 201–216.
- Woodhouse, I. (2021). Basic Principles of Radar Backscatter. Available at <https://eo-college.org/courses/basic-principles-of-radar-backscatter/#learndash-course-content> (accessed 17 February 2025).
- Wolff, C. (2007). *Ground penetrating radar*. Available at <https://www.radartutorial.eu/02.basics/Ground%20penetrating%20radar.en.html#this> (accessed 26 February 2022)
- Woźniak, E., Rybicki, M., Kofman, W., Aleksandrowicz, S., Wojtkowski, C., Lewiński, S., Bojanowski, J., Musiał, J., Milewski, T., Slesiński, P., Łaczyński, A. (2022). Multi-temporal phenological indices derived from time series Sentinel-1 images to country-wide crop classification. *International Journal of Applied Earth Observation and Geoinformation*, 107 e102683, <https://doi.org/10.1016/j.jag.2022.102683>
- Wright, M. N., & Ziegler, A. (2017). ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software*, 77(1), 1–17. <https://doi.org/10.18637/jss.v077.i01>
- Xu, H.; Xu, D.; Chen, S.; Ma, W. & Shi, Z. (2020). Rapid Determination of Soil Class Based on Visible-Near Infrared, Mid-Infrared Spectroscopy and Data Fusion. *Remote Sensing*, 12(9), 1512. <https://doi.org/10.3390/rs12091512>
- Xu, H. & Zhang, C. (2021). Investigating spatially varying relationships between total organic carbon contents and pH values in European agricultural soil using geographically weighted regression. *Science of The Total Environment*, 752, <https://doi.org/10.1016/j.scitotenv.2020.141977>
- Yamaguchi, Y.; Moriyama, T.; Ishido, M. & Yamada, H. (2005). Four-component scattering model for polarimetric SAR image decomposition. *IEEE Transactions on Geoscience Remote Sensing*, 43(8), 1699–1706. DOI: 10.1109/TGRS.2005.852084
- Yang, C.; Yang, L.; Zhang, L.; Zhou, C. (2023). Soil organic matter mapping using INLA-SPDE with remote sensing based soil moisture indices and Fourier transforms decomposed variables, *Geoderma*, 437, 116571, <https://doi.org/10.1016/j.geoderma.2023.116571>
- Yang, L.P.; Liu, F.; Li, Y.F.; Liu, J.; Li, G.Q.; Jin, M. (2019). Simulation of backscattering characteristics of bare surface based on the AIEM model in arid areas. *J. Lanzhou Univ. Nat. Sci.*, 55, 176–182.
- Zeng, C. *et al.* (2016). Mapping soil organic matter concentration at different scales using a mixed geographically weighted regression method. *Geoderma*, 281, 69–82. <https://doi.org/10.1016/j.geoderma.2016.06.033>
- Zhang, C., Tang, Y., Xu, X. & Kiely, G. (2011). Towards spatial geochemical modelling: Use of geographically weighted regression for mapping soil organic carbon contents in Ireland. *Applied Geochemistry*, 26, 1239–1248. doi:10.1016/j.apgeochem.2011.04.014
- Zhang, X., Dao, L., Zhang, C., Morrison, L., Hong, B., Zhang, H., Gan, Y. (2018). Mapping the spatial distribution of soil depth in a grassland ecosystem with the aid of ground penetrating radar and GIS (Northwestern Sichuan, China). *Grassland Science*, 64, 217–225. <https://doi.org/10.1111/grs.12201>
- Zhang, Y., Yang, B., Liu, X., Wang, C. (2017). Estimation of rice grain yield from dual-polarization Radarsat-2 SAR data by integrating a rice canopy scattering model and a genetic algorithm. *International Journal of Applied Earth Observation and Geoinformation*, 57, 75-85, <https://doi.org/10.1016/j.jag.2016.12.014>
- Zhang, M., Shi, W. (2019). Systematic comparison of five machine-learning methods in classification and interpolation of soil particle size fractions using different transformed data. *Hydrology and Earth System Sciences. Discussions [preprint]*. <https://doi.org/10.5194/hess-2018-584>
- Zhang, Z., Jiang, X., Wang, M., Liu, X. Wang, Q., Li, R. (2022). A phase-decomposition-based polarimetric coherence optimization method. *International Journal of Applied Earth Observation and Geoinformation*, 110, e102771, <https://doi.org/10.1016/j.jag.2022.102771>
- Zhou, Y., Wu, W. , Wang, H., Zhang, X., Yang, C., & Liu, H. (2022). Identification of Soil Texture Classes Under Vegetation Cover Based on Sentinel-2 Data With SVM and SHAP Techniques, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 3758-3770, doi: 10.1109/JSTARS.2022.3164140.
- Zhu A-Xing, Lu G, Liu J, *et al.* (2018). Spatial prediction based on Third Law of Geography. *Annals of GIS*, 24(4), 225–240, DOI: 10.1080/19475683.2018.1534890

- Žižala, D.; Minařík, R.; Zádorová, T. (2019). Soil Organic Carbon Mapping Using Multispectral Remote Sensing Data: Prediction Ability of Data with Different Spatial and Spectral Resolutions. *Remote Sensing*, 11, 2947. <https://doi.org/10.3390/rs11242947>
- Zong Wang; Wenjiao Shi (2017). Mapping soil particle-size fractions: A comparison of compositional kriging and log-ratio kriging, *Journal of Hydrology*, Vol. 546, 526-541, <https://doi.org/10.1016/j.jhydrol.2017.01.029>
- Zong Wang, Wenjiao Shi (2018). Robust variogram estimation combined with isometric log-ratio transformation for improved accuracy of soil particle-size fraction mapping, *Geoderma*, Vol. 324, 56-66. <https://doi.org/10.1016/j.geoderma.2018.03.007>
- Zribi, M., Kotti, F., Lili-Chabaane, Z., Baghdadi, N., Ben Issa, N. *et al.* (2012). Soil texture estimation over a semiarid area using TerraSAR-X radar data. *IEEE Geoscience and Remote Sensing Letters*, 9 (3), 353-357. 10.1109/LGRS.2011.2168379. ird-00679652
- Zwieback, S. Hensley, S., Hajnsek, I. (2015). Assessment of soil moisture effects on L-band radar interferometry, *Remote Sensing of Environment*, 164, 77-89, <https://doi.org/10.1016/j.rse.2015.04.012>

Appendix A

This supplementary material refers to chapter 4 (DOI 10.1111/ejss.13414), and is available at <https://doi.org/10.1111/ejss.13414>

Appendix A – Figures

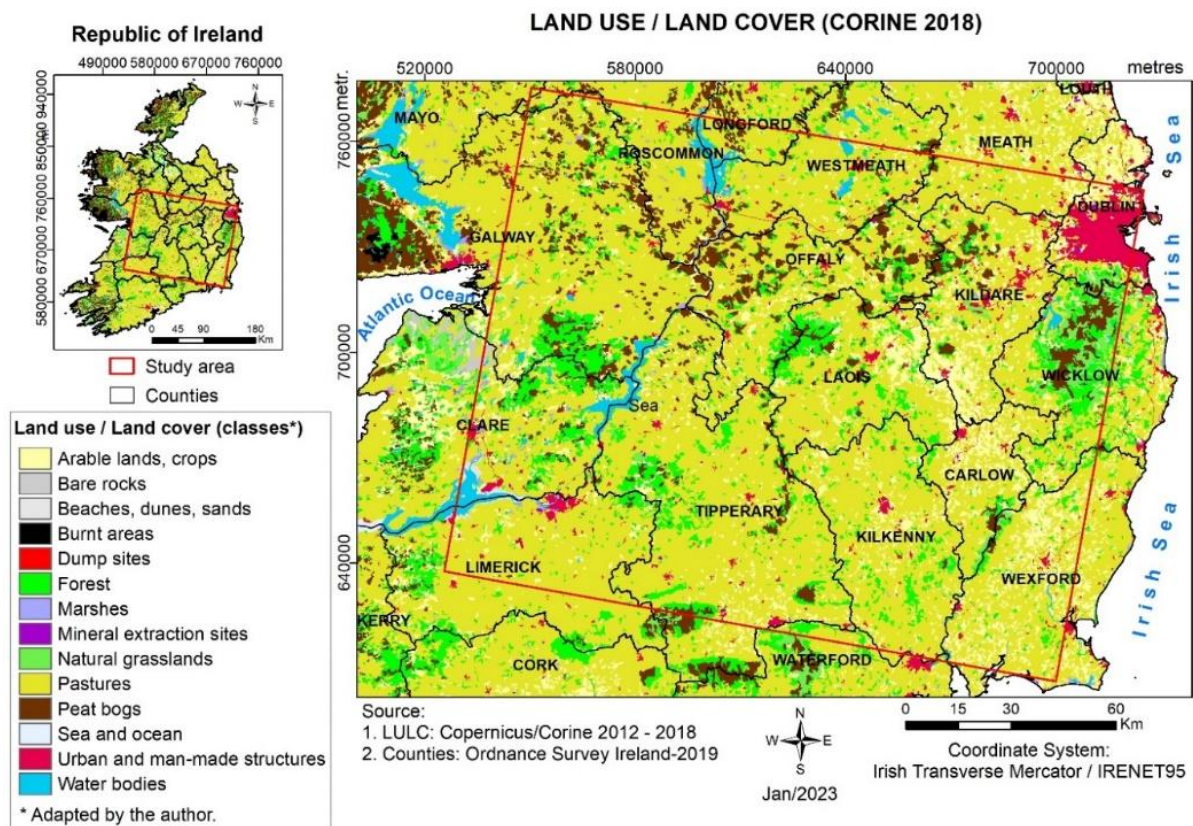


Figure A4.1 Land use and land cover categories for the study domain as derived from the CORINE Land Cover data (2018).

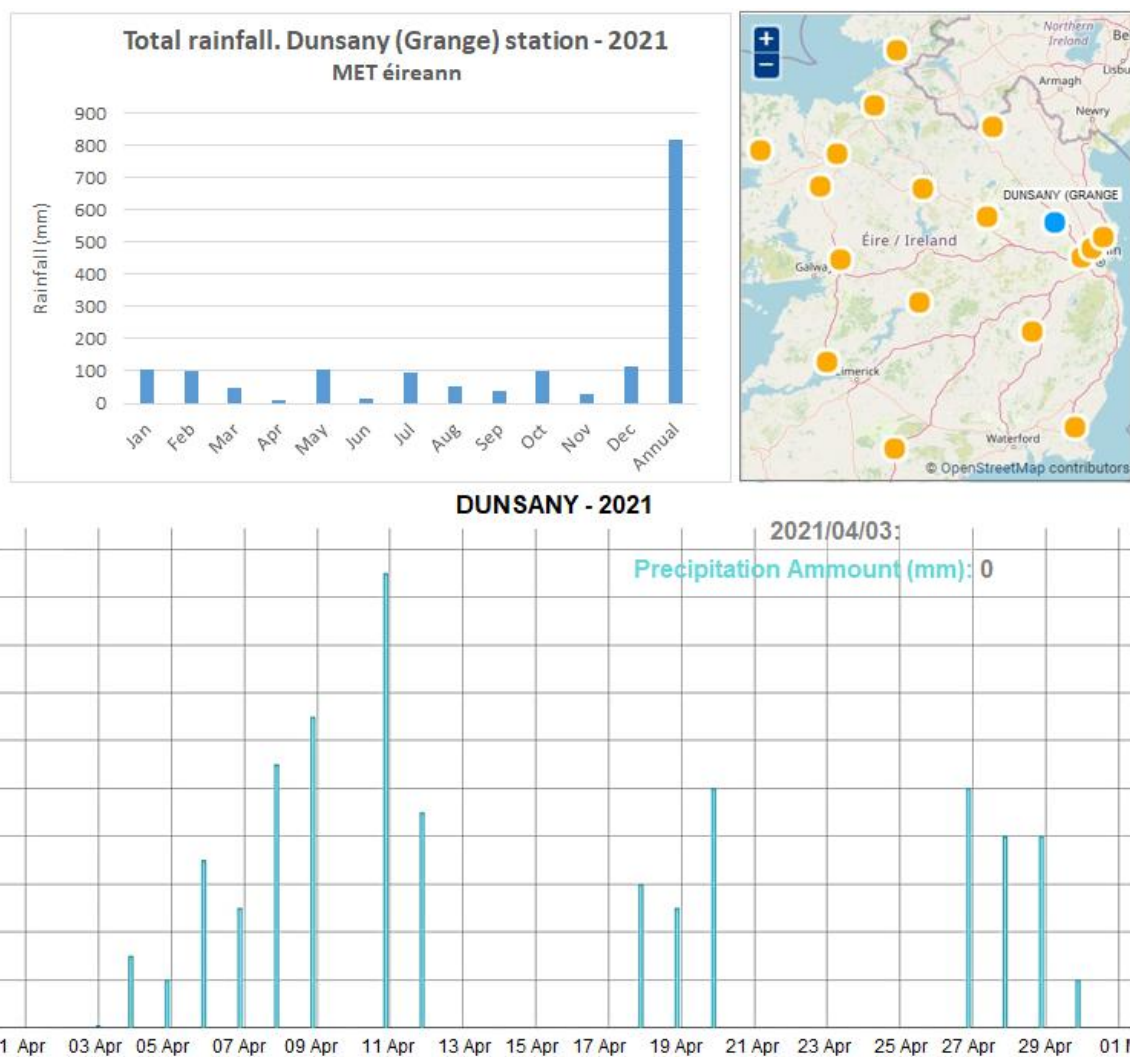


Figure A4.2 Precipitation registered in Dunsany for 2021 (top) and during the Sentinel 1 SAR imagery captured on early April/2021. Source: The Irish Meteorological Service. Met Éireann. Available at: <https://www.met.ie/climate/available-data/historical-data>

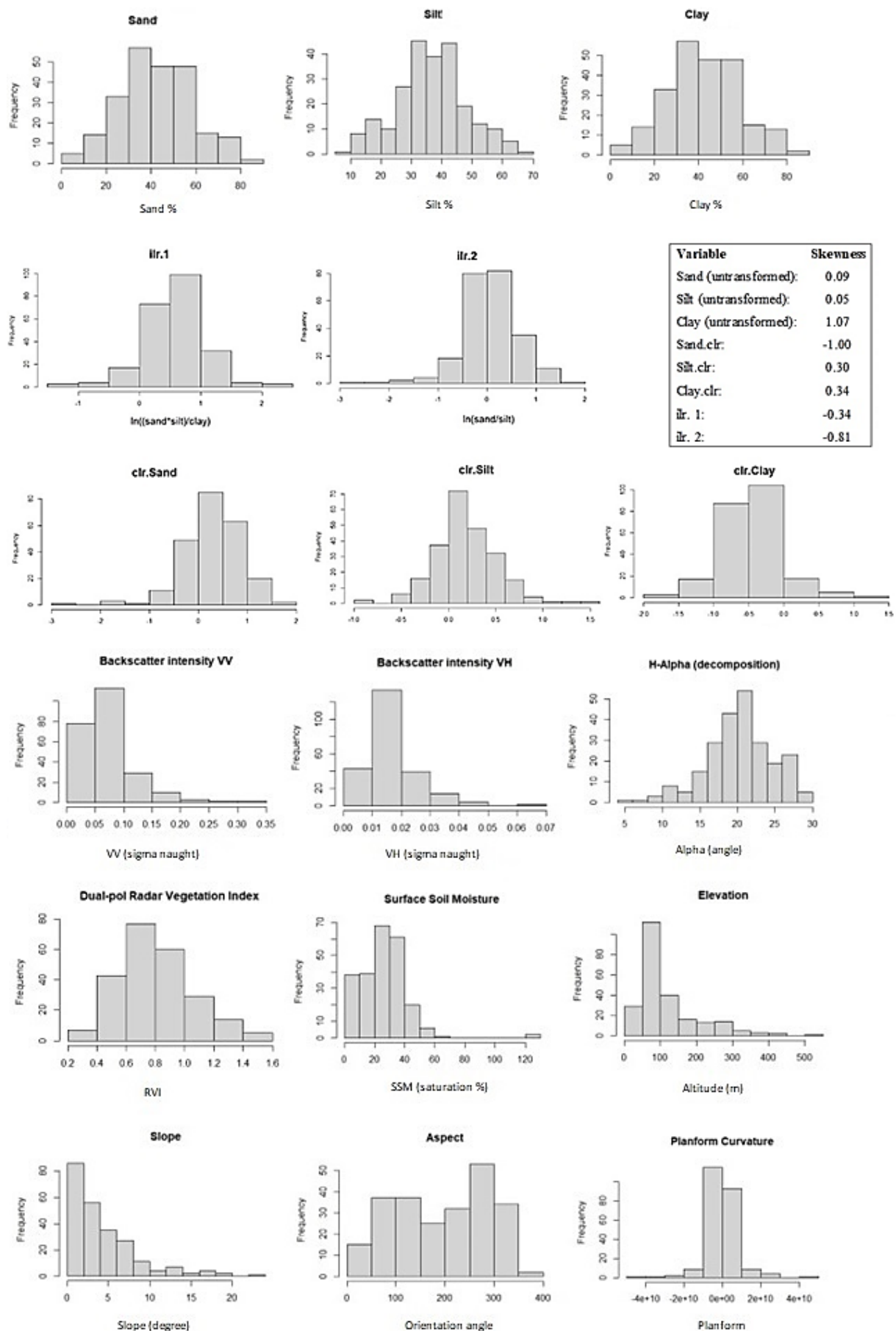


Figure A4.3 Histograms of the response variables (untransformed and transformed; first three rows) and the predictor variables (last three rows). The coefficients ilr.1 and ilr.2 were obtained from the transformation of the proportions of sand, silt and clay.

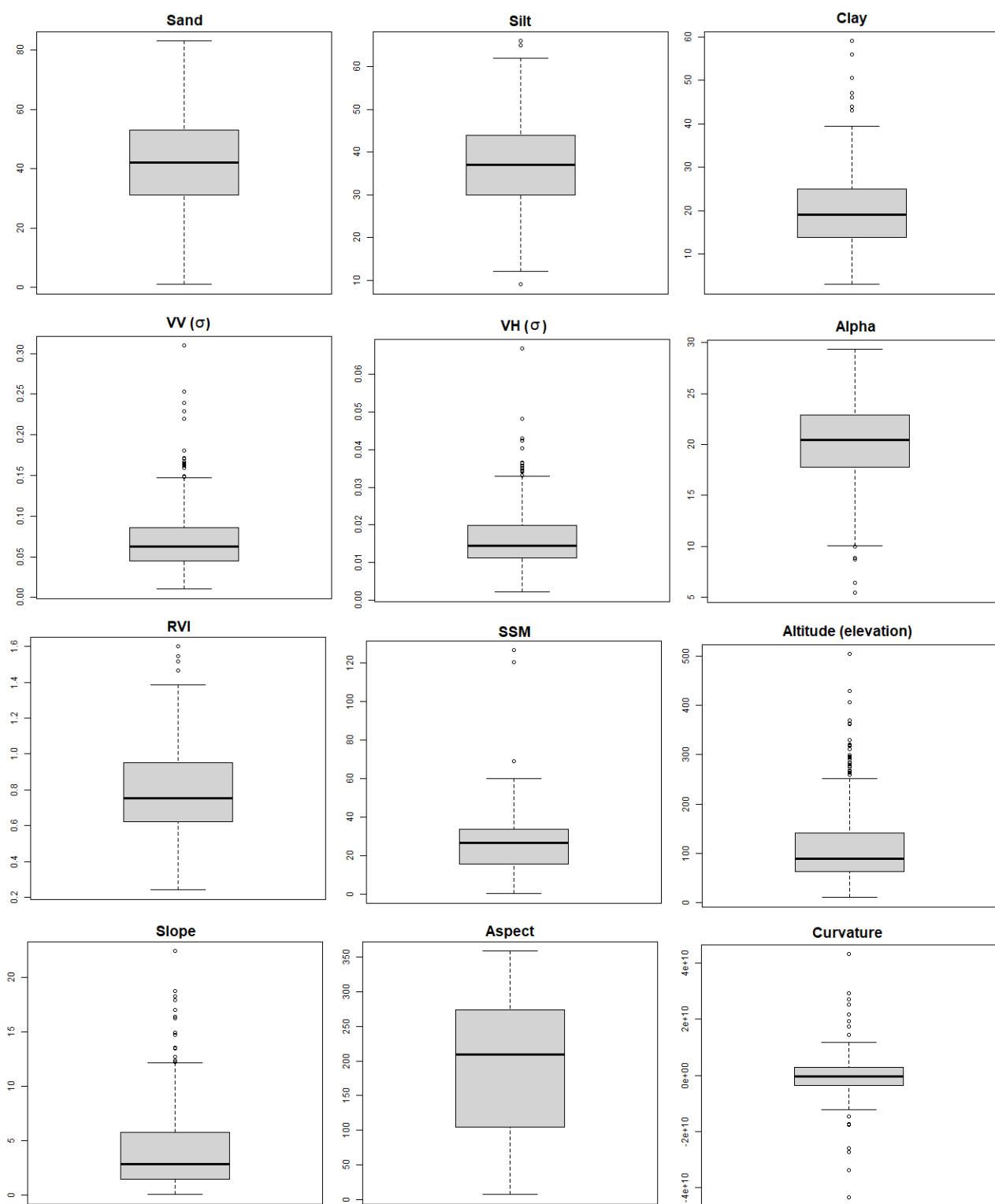


Figure A4.4 Boxplots of the response variables (Sand, Silt and Clay; untransformed) and the predictor variables (remaining boxplots).

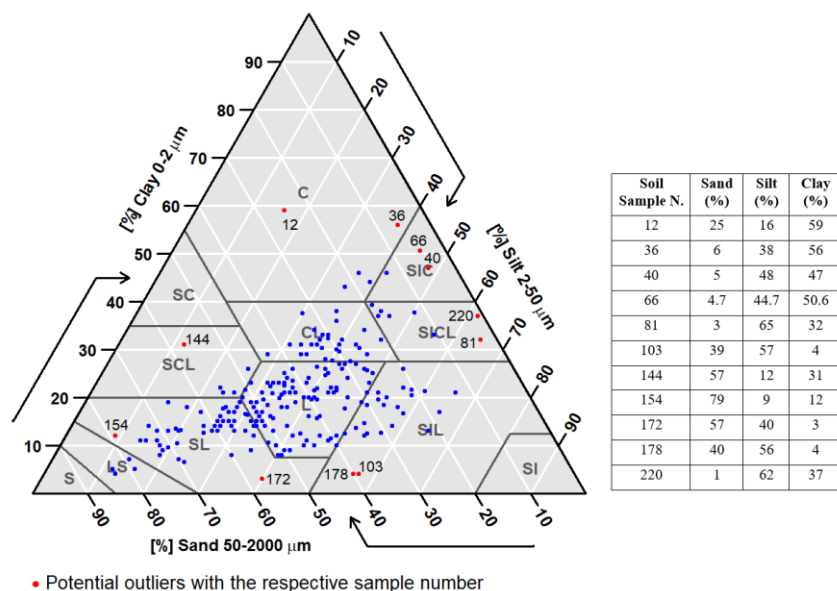


Figure A4.5 Ternary of soil texture (USDA/NCSS) for the observed data (measured, $n = 235$) in the study area. Eleven potential outliers (red dots) were identified in the compositional data according to the Atypicality Index (indicated in red).

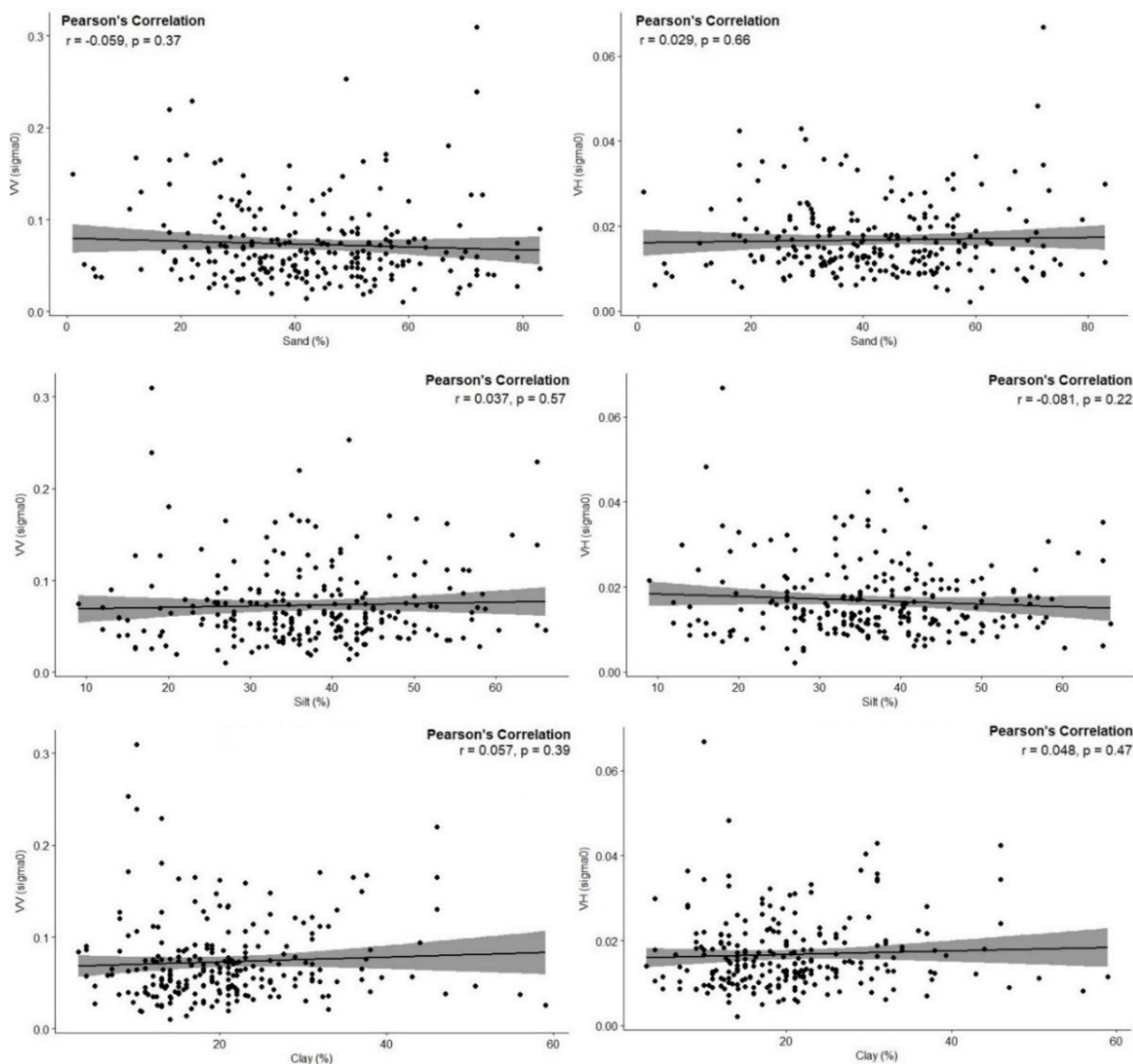


Figure A4.6 Correlations between the radar backscattering coefficients (VV, VH) and the soil particle-size fractions without log-transformation methods applied to them. P refers to the p-value.

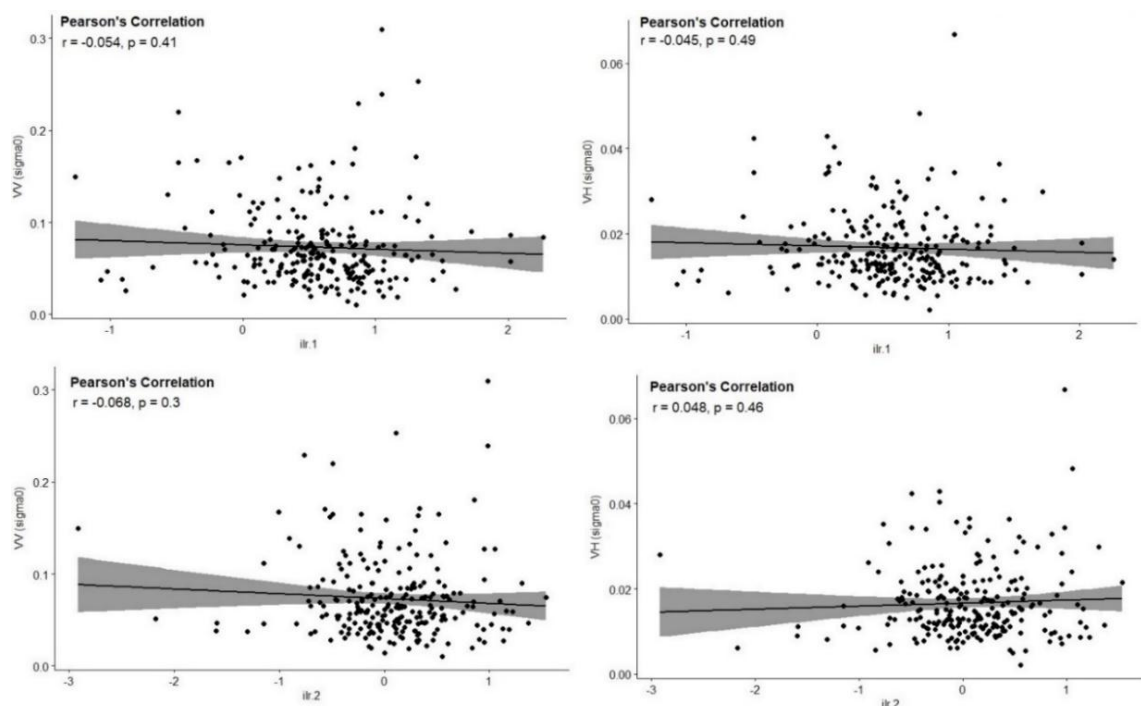


Figure A4.7 Correlations between the radar backscattering coefficients (VV, VH) and the soil particle-size fractions under olr-coordinates (ilr log-transformation). P refers to p-value.

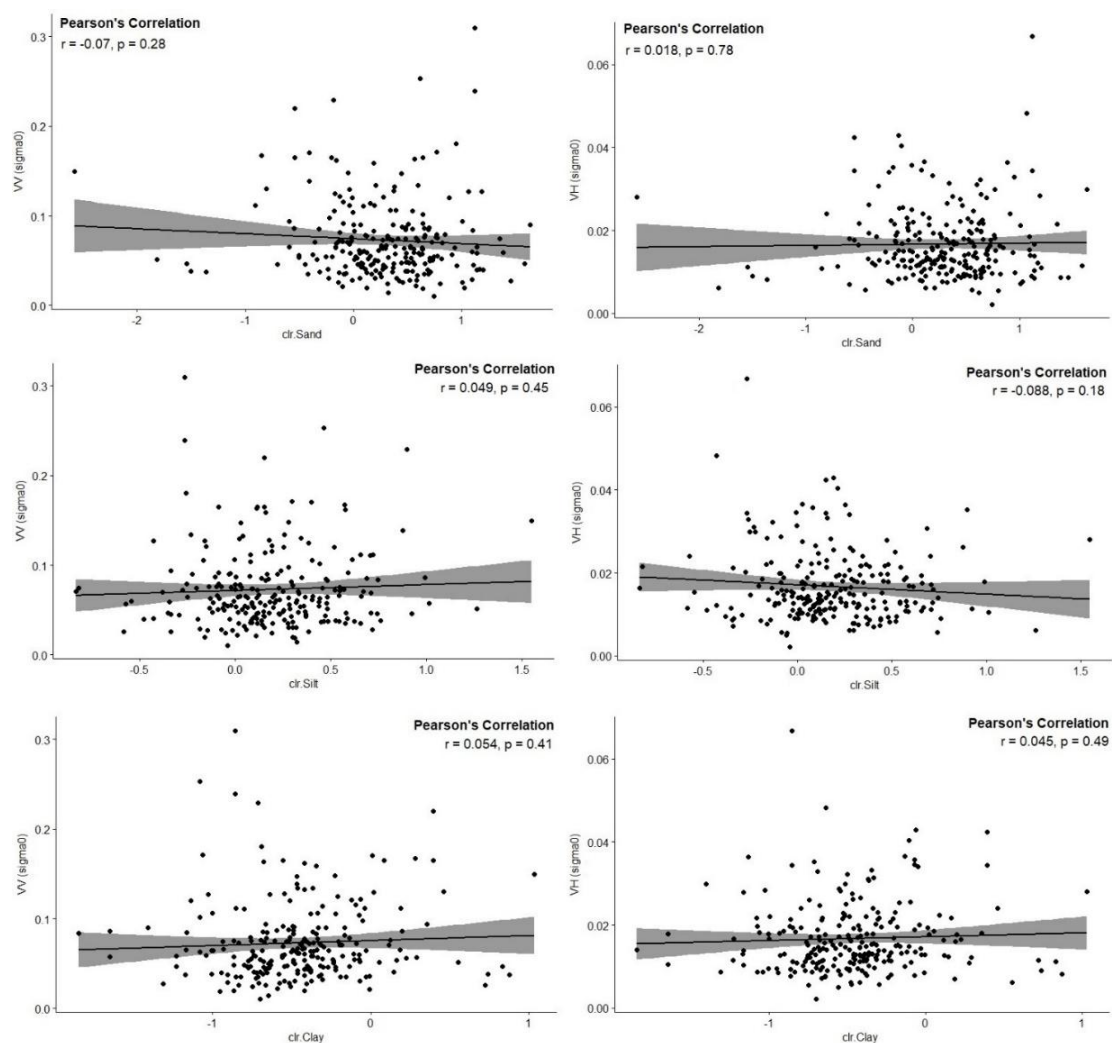
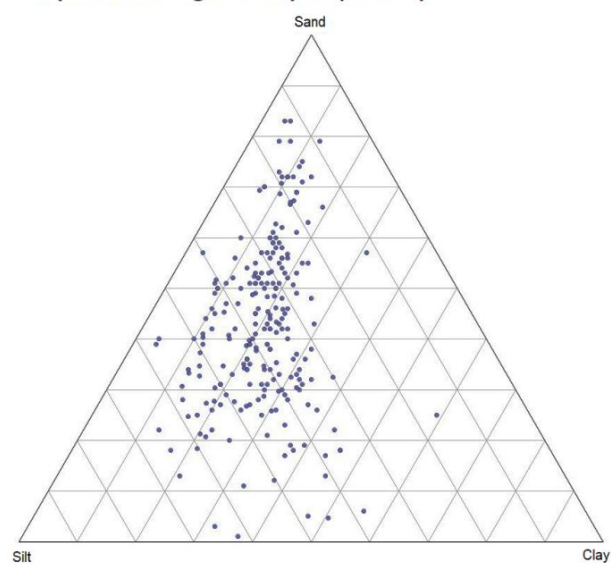


Figure A4.8 Correlations between the radar backscattering coefficients (VV, VH) and the soil particle-size fractions with clr log-transformation applied to them. P refers to p-value.

Compositional original samples (n = 235)



Fitted compositional samples

Compositional LRM
Overall R² = 70.74

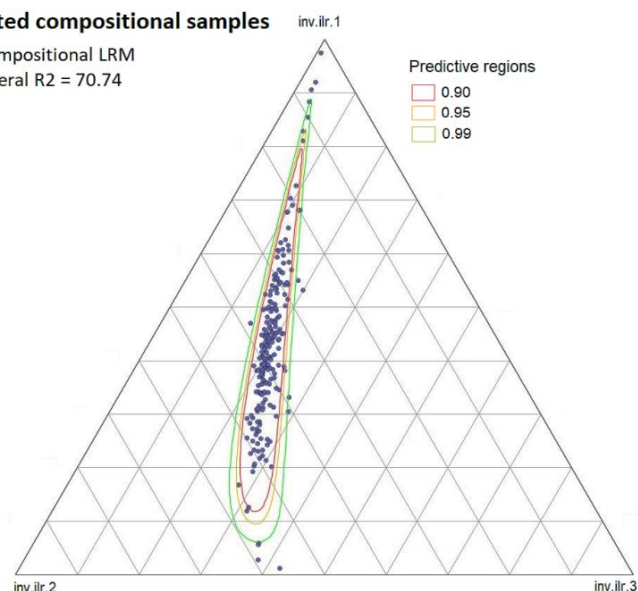


Figure A4.9 Compositional original samples (left) compared to the compositional fitted samples (right) in a ternary diagram. Note: inv.ilr.1=sand, inv.ilr.2=silt, and inv.ilr.3=clay.

Compositional coefficients - LRM

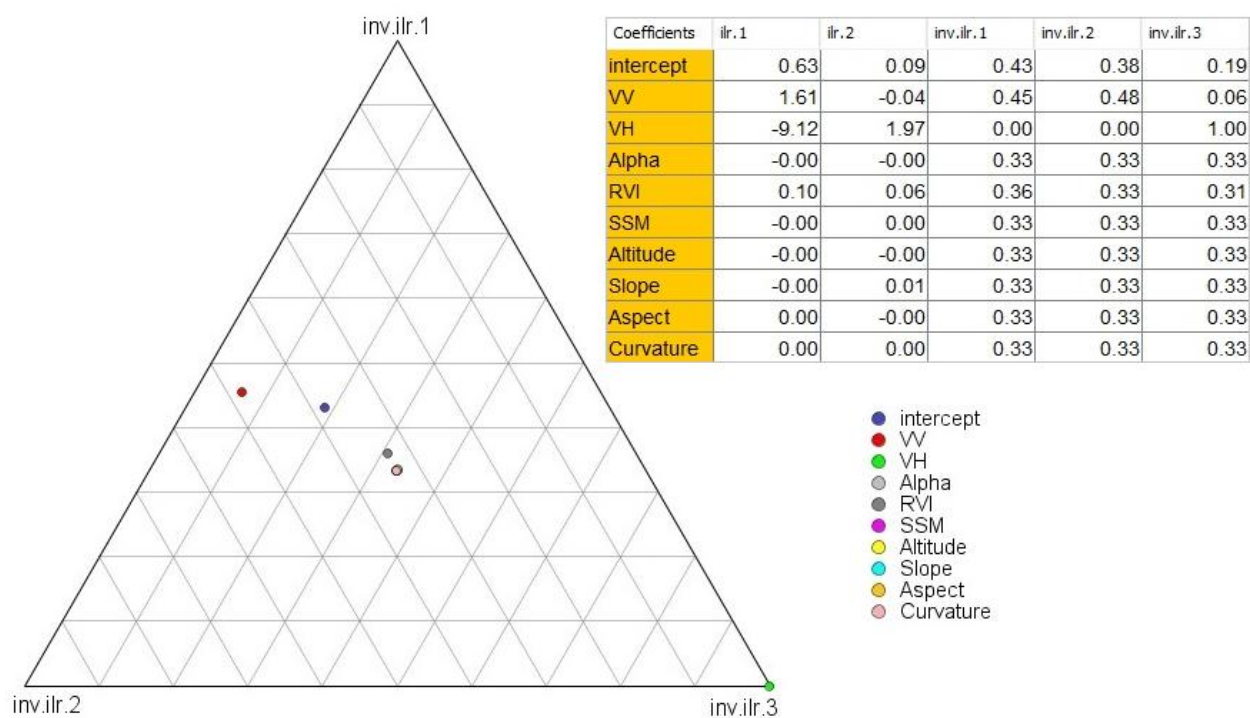
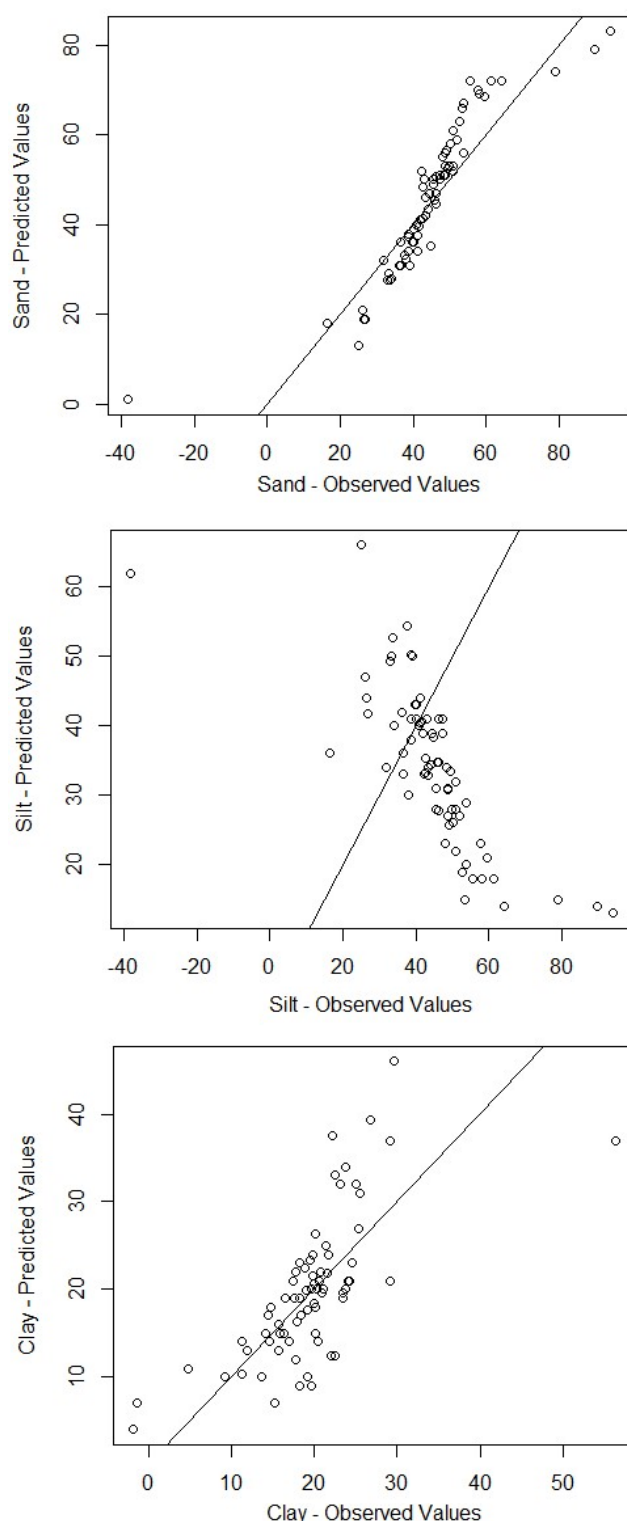


Figure A4.10 The compositional coefficients of the Y-compositional linear regression model. Approach 1. Note: inv.ilr.1=sand, inv.ilr.2=silt, and inv.ilr.3=clay. The covariates whose coefficients values are equal to 0.33 and 0.00-0.001 (Alpha, SSM, Altitude, Slope, Aspect, and Curvature) are superimposed on/closer to each other in the diagram.

LRM (no transformation) - Observed. Vs. Predicted Plot (Test data)



Soil particle size predictions (no transformation; test data)

Sample	Sand pred.	Silt pred.	Clay pred.	Total (%)	Sample	Sand pred.	Silt pred.	Clay pred.	Total (%)
4	40	40	21	100	125	47	47	19	114
5	60	60	11	130	127	34	34	23	91
11	58	58	12	128	131	58	58	15	131
23	43	43	20	107	134	51	51	24	126
24	49	49	17	116	137	42	42	20	104
25	38	38	22	98	140	53	53	15	120
31	43	43	20	105	142	36	36	25	98
32	37	37	23	96	145	34	34	24	91
34	33	33	25	90	148	48	48	18	114
43	40	40	21	101	150	39	39	22	99
49	79	79	5	162	151	51	51	14	116
51	46	46	19	111	152	46	46	20	113
52	45	45	20	110	159	49	49	18	115
55	33	33	24	90	161	49	49	19	117
60	41	41	22	104	162	38	38	22	98
61	39	39	21	99	164	40	40	24	105
62	25	25	29	79	165	54	54	16	124
72	46	46	18	109	167	51	51	18	120
73	49	49	20	118	175	48	48	16	113
76	46	46	18	110	179	45	45	18	109
78	42	42	21	104	180	50	50	16	116
83	39	39	23	101	181	43	43	20	105
91	27	27	29	82	182	94	94	2	190
92	52	52	15	118	192	43	43	20	105
98	27	27	27	81	202	39	39	22	99
101	54	54	18	125	212	32	32	24	87
104	41	41	21	103	214	43	43	17	104
105	46	46	20	112	220	38	38	56	133
110	90	90	1	181	223	42	42	20	104
112	64	64	11	140	224	49	49	14	112
113	36	36	25	98	228	26	26	25	77
117	41	41	20	102	230	54	54	16	124
118	50	50	16	116	231	16	16	30	62
120	47	47	18	113	233	61	61	9	132

RMSE (test data):

Sand: 8

Silt: 27

Clay: 6

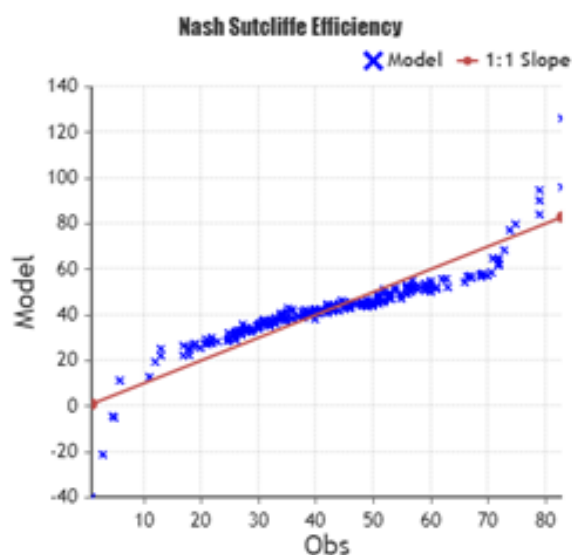
R2 (test data):

Sand: 0.8

Silt: 0.6

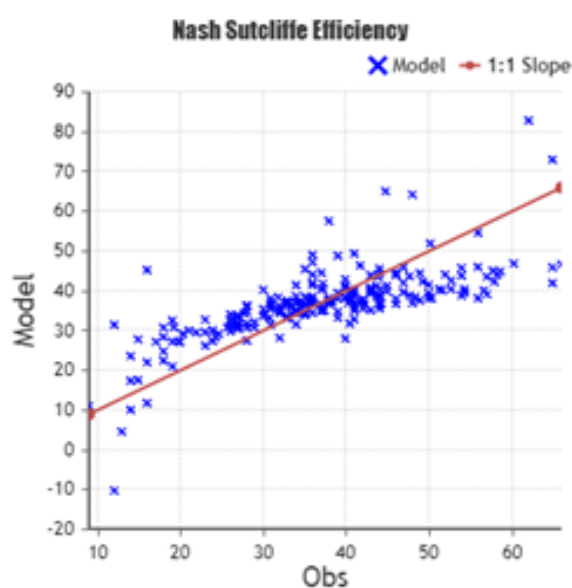
Clay: 0.5

Figure A4.11 Scatterplots and metric values from the test model with the simple linear regression without log-ratio transformation methods applied to the response variables. Validation method: Data split. Train 75% and test 25%. Reference line in the plots refers to the 1:1 agreement line.

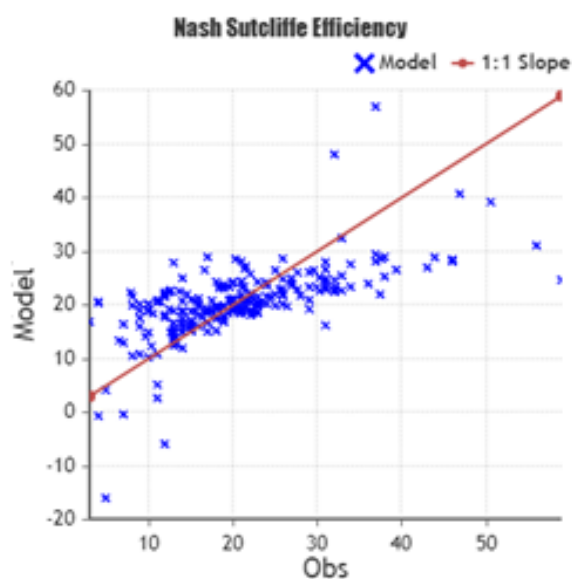


Linear Regression Model - LOOCV
No transformation (Non-CoDa)

Prediction: **Sand**
RMSE: 7.10
MAE: 5.35
NSE: 0.81



Prediction: **Silt**
RMSE: 7.92
MAE: 6.10
NSE: 0.52



Prediction: **Clay**
RMSE: 7.36
MAE: 5.33
NSE: 0.41

Figure A4.12 Scatterplots and metric values from the Linear Regression model without log-ratio transformation methods applied to the response variables. Validation method: LOOCV. Reference line in the plots refers to the 1:1 agreement line. Plots obtained from AgriMetSoft (2019). Available on: <https://agrimetsoft.com/calculators/Nash%20Sutcliffe%20model%20Efficiency%20coefficient>

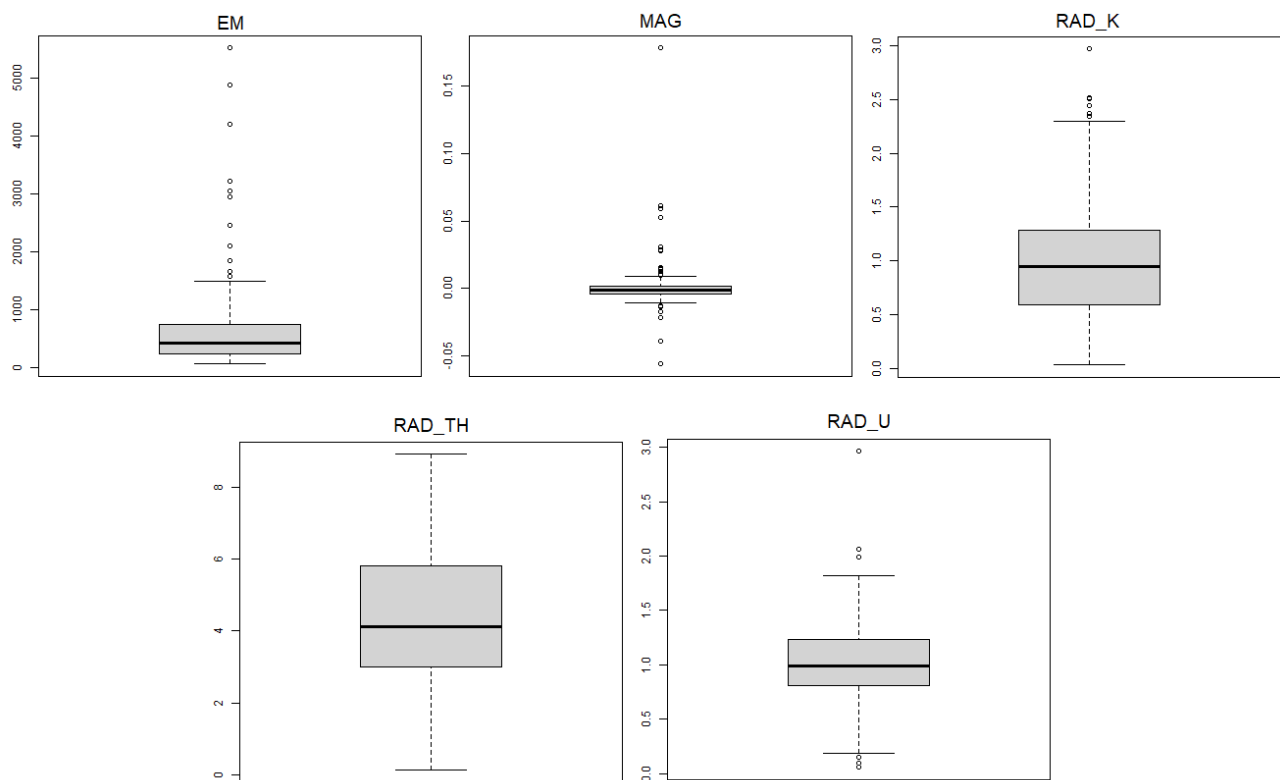


Figure A4.13 Boxplots of the geophysical variables.

Compositional coefficients - LRM (Model 2)

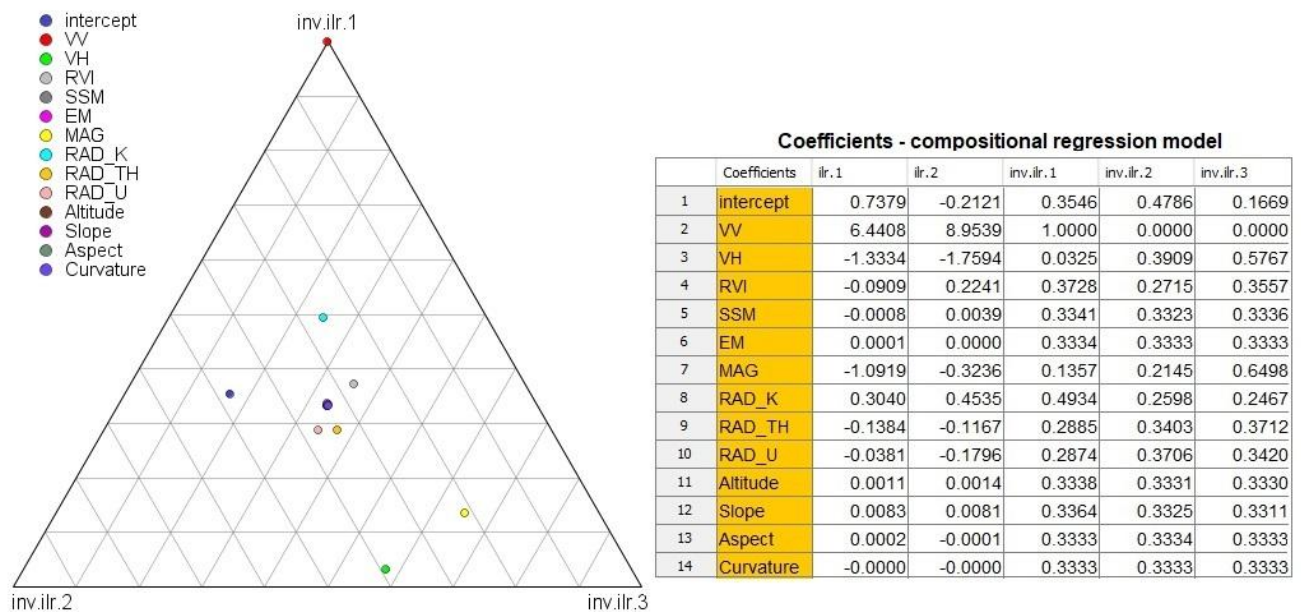


Figure A4.14 The compositional coefficients of the Y-compositional linear regression model. Approach 2. Note: inv.ilr.1=sand, inv.ilr.2=silt, and inv.ilr.3=clay.

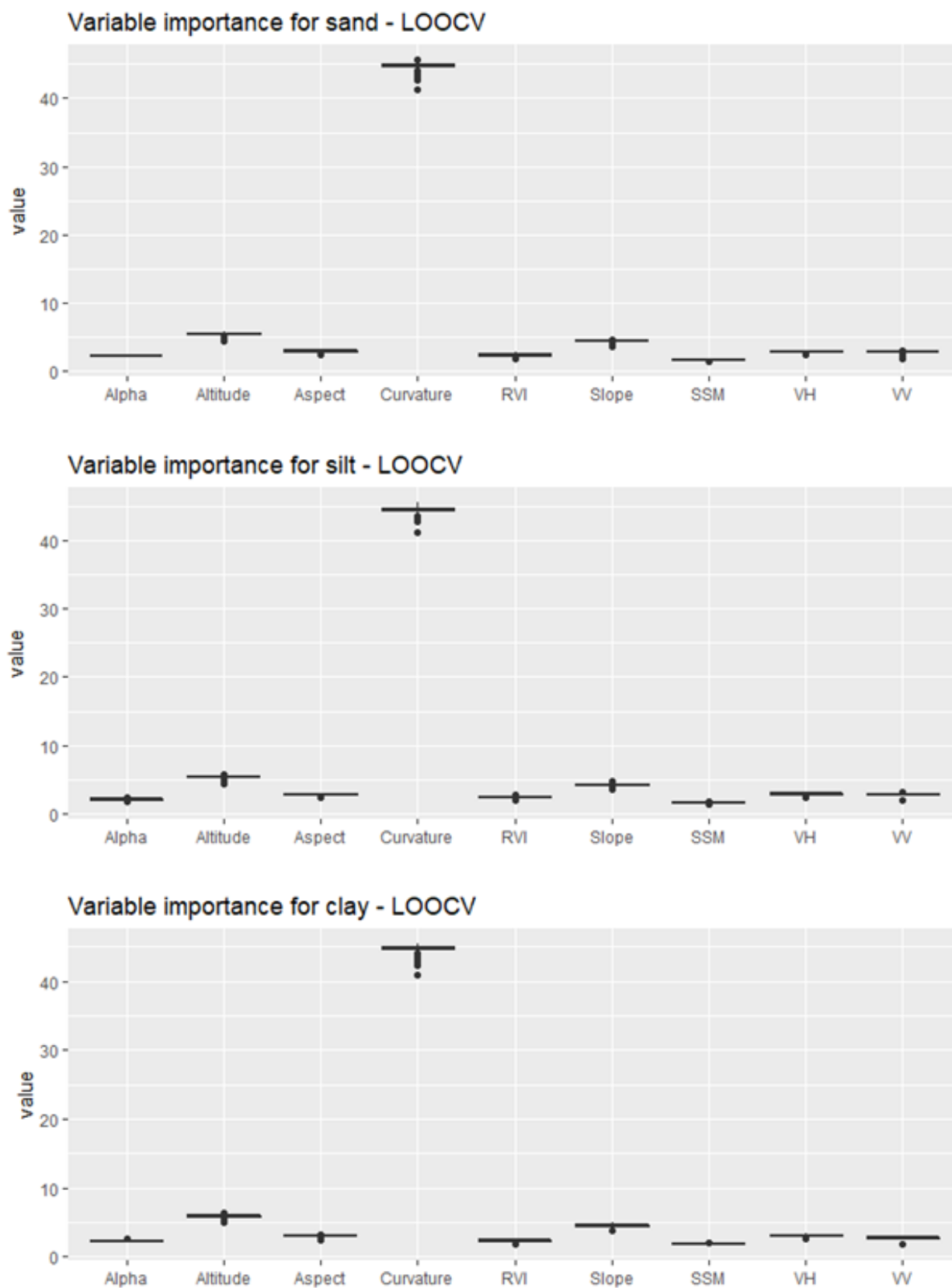


Figure A4.15 Importance of covariates used for predicting soil particle-size fractions (without geophysical data, model 1) according to impurity method, which is a measure of the variance of the responses for regression in Random Forest (Ranger package).

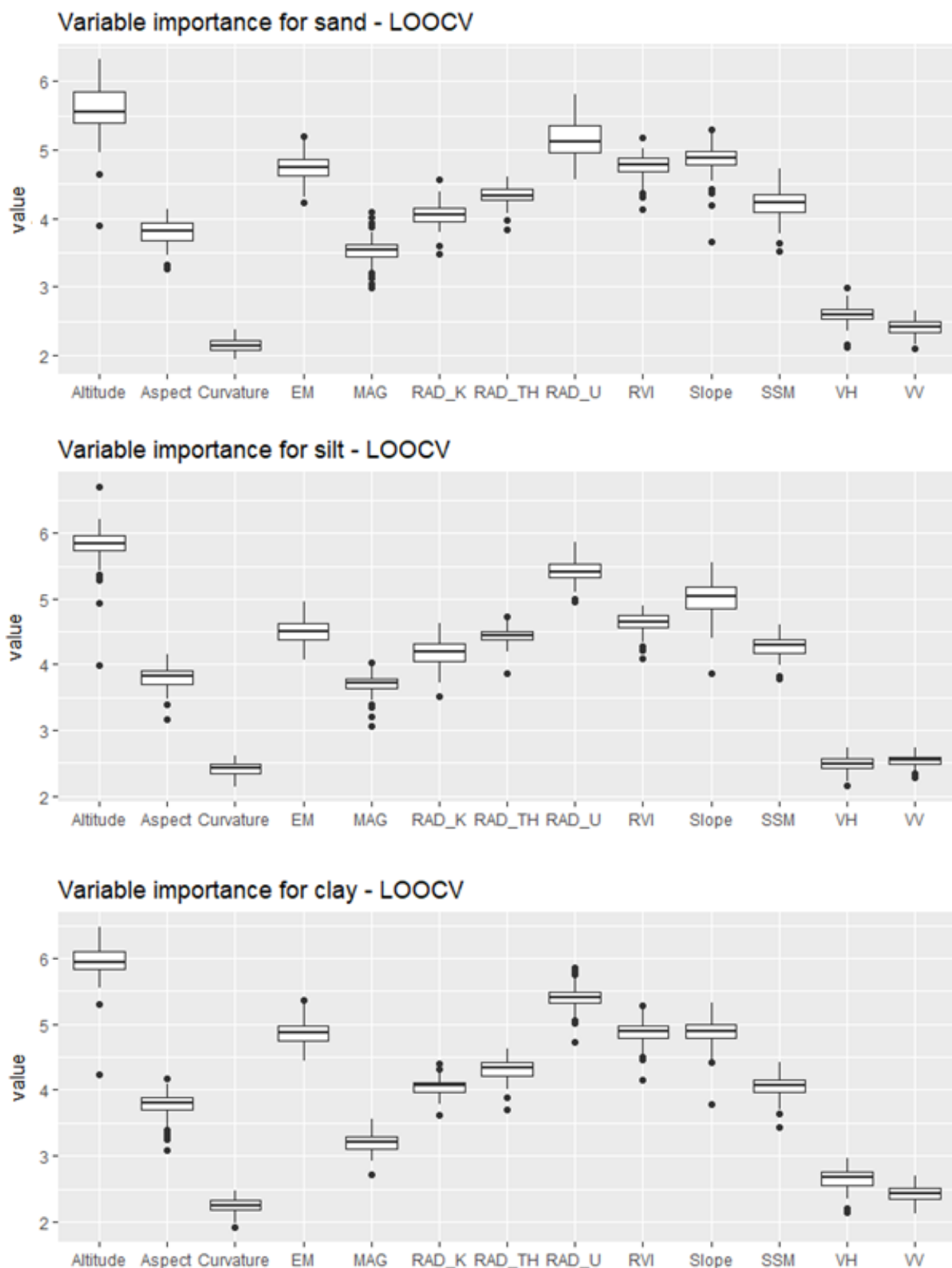


Figure A4.16 Importance of covariates used for predicting soil particle-size fractions (with geophysical data, model 2) according to impurity method, which is a measure of the variance of the responses for regression in Random Forest (Ranger package).

Appendix A – Tables

Table A4.1 Convergence test for RF model. Parameter: Ntree.
Model fitted with CLR transformation applied to the response variables (sand, silt, and clay)

Ntree*	MSE (OOB)*	R squared (OOB)*
10	0.08401456 = 0.08	0.7523555 = 0.75
25	0.07589822 = 0.08	0.7762795 = 0.78
40	0.06709056 = 0.07	0.8022413 = 0.80
50	0.06282882 = 0.06	0.8148034 = 0.81
100	0.06177383 = 0.06	0.8179131 = 0.82
200	0.05952412 = 0.06	0.8245444 = 0.82
300	0.0629968 = 0.06	0.8143082 = 0.81
400	0.0632253 = 0.06	0.8136347 = 0.81
500	0.06306809 = 0.06	0.8140981 = 0.81
600	0.06296497 = 0.06	0.814402 = 0.81
700	0.06340646 = 0.06	0.8131007 = 0.81
1000	0.0646994 = 0.06	0.8092896 = 0.81

Table A4.2 Convergence test for RF model. Parameter: Mtry
Model fitted with CLR transformation applied to the response variables (sand, silt, and clay)

Mtry*	MSE (OOB)*	R squared (OOB)*
1	0.163097 = 0.16	0.5192491 = 0.52
2	0.09559036 = 0.10	0.7182342 = 0.72
3	0.06306809 = 0.06	0.8140981 = 0.81
4	0.04067339 = 0.04	0.8801096 = 0.88
5	0.03312238 = 0.03	0.9023672 = 0.90
6	0.02534686 = 0.03	0.9252866 = 0.93
7	0.02268349 = 0.02	0.9331373 = 0.93
8	0.01965665 = 0.02	0.9420593 = 0.94
9	0.01936281 = 0.02	0.9429254 = 0.94

Table A4.3 Predictions dataset from the Compositional Y-LRM (Model Approach 1-non-geophysical data), with ILR transformation applied to the response variable and validated with LOOCV. Soil particle-size fractions predicted and back-transformed, showing the 100% sum in each sample.

Sample	inv.ilr.1	inv.ilr.2	inv.ilr.3	Total	Sand_pred (%)	Silt_pred (%)	Clay_pred (%)	Total
1	0.5344338	0.3120655	0.1535007	1	53.44338	31.20655	15.35007	100
2	0.4299537	0.3834417	0.1866046	1	42.99537	38.34417	18.66046	100
3	0.495171	0.3350433	0.1697857	1	49.5171	33.50433	16.97857	100
4	0.3836812	0.4129596	0.2033591	1	38.36812	41.29596	20.33591	100
5	0.6013873	0.2762673	0.1223455	1	60.13873	27.62673	12.23455	100
6	0.3191671	0.4378268	0.243006	1	31.91671	43.78268	24.3006	100
7	0.5035314	0.3288967	0.1675719	1	50.35314	32.88967	16.75719	100
8	0.5583203	0.3021144	0.1395653	1	55.83203	30.21144	13.95653	100
9	0.4407989	0.372522	0.1866791	1	44.07989	37.2522	18.66791	100
10	0.4369701	0.3728117	0.1902183	1	43.69701	37.28117	19.02183	100
11	0.6018825	0.2752144	0.1229031	1	60.18825	27.52144	12.29031	100
12	0.2778096	0.4806433	0.2415471	1	27.78096	48.06433	24.15471	100
13	0.4445729	0.3755268	0.1799003	1	44.45729	37.55268	17.99003	100
14	0.860297	0.0939602	0.0457429	1	86.0297	9.396017	4.574287	100
15	0.5416701	0.2949294	0.1634005	1	54.16701	29.49294	16.34005	100
16	0.4734795	0.3503296	0.176191	1	47.34795	35.03296	17.6191	100
17	0.3163371	0.4430517	0.2406112	1	31.63371	44.30517	24.06112	100
18	0.6431664	0.2370704	0.1197632	1	64.31664	23.70704	11.97632	100
19	0.3861937	0.4035912	0.2102151	1	38.61937	40.35912	21.02151	100
20	0.2866181	0.4625968	0.2507852	1	28.66181	46.25968	25.07852	100

21	0.4347668	0.3780775	0.1871557	1	43.47668	37.80775	18.71557	100
22	0.3872364	0.3965414	0.2162222	1	38.72364	39.65414	21.62222	100
23	0.4294499	0.3790129	0.1915372	1	42.94499	37.90129	19.15372	100
24	0.4913436	0.3404893	0.1681671	1	49.13436	34.04893	16.81671	100
25	0.3624203	0.4133003	0.2242794	1	36.24203	41.33003	22.42794	100
26	0.3398677	0.4356826	0.2244498	1	33.98677	43.56826	22.44498	100
27	0.5255585	0.3127173	0.1617242	1	52.55585	31.27173	16.17242	100
28	0.3410568	0.4260969	0.2328463	1	34.10568	42.60969	23.28463	100
29	0.5511669	0.3055257	0.1433075	1	55.11669	30.55257	14.33075	100
30	0.2340943	0.4815838	0.2843218	1	23.40943	48.15838	28.43218	100
31	0.4087964	0.39715	0.1940536	1	40.87964	39.715	19.40536	100
32	0.3423934	0.4372513	0.2203552	1	34.23934	43.72513	22.03552	100
33	0.3872903	0.398048	0.2146617	1	38.72903	39.8048	21.46617	100
34	0.309882	0.4453289	0.244789	1	30.9882	44.53289	24.4789	100
35	0.4068194	0.395019	0.1981616	1	40.68194	39.5019	19.81616	100
36	0.123077	0.5755463	0.3013767	1	12.3077	57.55463	30.13767	100
37	0.2505039	0.4894782	0.260018	1	25.05039	48.94782	26.0018	100
38	0.4497904	0.3354557	0.2147539	1	44.97904	33.54557	21.47539	100
39	0.4251222	0.3841425	0.1907353	1	42.51222	38.41425	19.07353	100
40	0.0592939	0.5828373	0.3578689	1	5.929386	58.28373	35.78689	100
41	0.4511465	0.3688911	0.1799624	1	45.11465	36.88911	17.99624	100
42	0.4732206	0.3434096	0.1833698	1	47.32206	34.34096	18.33698	100
43	0.3941793	0.3985435	0.2072773	1	39.41793	39.85435	20.72773	100
44	0.6760449	0.2227439	0.1012112	1	67.60449	22.27439	10.12112	100
45	0.8328868	0.1183581	0.048755	1	83.28868	11.83581	4.875504	100
46	0.5510001	0.301856	0.1471439	1	55.10001	30.1856	14.71439	100
47	0.4265679	0.3834062	0.1900259	1	42.65679	38.34062	19.00259	100
48	0.3566612	0.4203347	0.2230041	1	35.66612	42.03347	22.30041	100
49	0.8136802	0.1284593	0.0578605	1	81.36802	12.84593	5.786047	100
50	0.2862214	0.4679489	0.2458297	1	28.62214	46.79489	24.58297	100
51	0.4412665	0.3685835	0.19015	1	44.12665	36.85835	19.015	100
52	0.4022022	0.3964936	0.2013042	1	40.22022	39.64936	20.13042	100
53	0.405789	0.3959616	0.1982494	1	40.5789	39.59616	19.82494	100
54	0.3743885	0.4148488	0.2107626	1	37.43885	41.48488	21.07626	100
55	0.306264	0.4539751	0.2397609	1	30.6264	45.39751	23.97609	100
56	0.4301253	0.3767177	0.193157	1	43.01253	37.67177	19.3157	100
57	0.5913908	0.281784	0.1268252	1	59.13908	28.1784	12.68252	100
58	0.3803308	0.4182702	0.201399	1	38.03308	41.82702	20.1399	100
59	0.6765187	0.2224512	0.1010301	1	67.65187	22.24512	10.10301	100
60	0.3932605	0.3945829	0.2121566	1	39.32605	39.45829	21.21566	100
61	0.3718429	0.4161178	0.2120393	1	37.18429	41.61178	21.20393	100
62	0.2207372	0.4906244	0.2886383	1	22.07372	49.06244	28.86383	100
63	0.4965894	0.3348583	0.1685522	1	49.65894	33.48583	16.85522	100
64	0.2282148	0.5026345	0.2691507	1	22.82148	50.26345	26.91507	100
65	0.9837223	0.0119222	0.0043554	1	98.37223	1.192224	0.435544	100
66	0.0571202	0.5914291	0.3514507	1	5.712017	59.14291	35.14507	100

67	0.3736955	0.410506	0.2157985	1	37.36955	41.0506	21.57985	100
68	0.4415894	0.3635415	0.1948691	1	44.15894	36.35415	19.48691	100
69	0.5098213	0.3329708	0.1572079	1	50.98213	33.29708	15.72079	100
70	0.4113291	0.3856611	0.2030099	1	41.13291	38.56611	20.30099	100
71	0.3696199	0.3941764	0.2362037	1	36.96199	39.41764	23.62037	100
72	0.4498821	0.368006	0.1821119	1	44.98821	36.8006	18.21119	100
73	0.5017559	0.3125826	0.1856615	1	50.17559	31.25826	18.56615	100
74	0.4582664	0.3564836	0.1852499	1	45.82664	35.64836	18.52499	100
75	0.4450935	0.3687936	0.1861129	1	44.50935	36.87936	18.61129	100
76	0.4444706	0.3726339	0.1828955	1	44.44706	37.26339	18.28955	100
77	0.5470701	0.3001511	0.1527788	1	54.70701	30.01511	15.27788	100
78	0.4078259	0.3858478	0.2063263	1	40.78259	38.58478	20.63263	100
79	0.4244529	0.3643937	0.2111534	1	42.44529	36.43937	21.11534	100
80	0.5649124	0.2880697	0.1470179	1	56.49124	28.80697	14.70179	100
81	0.0279788	0.5874124	0.3846089	1	2.797879	58.74124	38.46089	100
82	0.2538295	0.4667851	0.2793854	1	25.38295	46.67851	27.93854	100
83	0.3611347	0.4092745	0.2295908	1	36.11347	40.92745	22.95908	100
84	0.3832753	0.3940441	0.2226806	1	38.32753	39.40441	22.26806	100
85	0.581395	0.2674273	0.1511777	1	58.1395	26.74273	15.11777	100
86	0.529502	0.3222895	0.1482085	1	52.9502	32.22895	14.82085	100
87	0.4572027	0.3693761	0.1734212	1	45.72027	36.93761	17.34212	100
88	0.3709965	0.4210648	0.2079387	1	37.09965	42.10648	20.79387	100
89	0.450551	0.3429986	0.2064503	1	45.0551	34.29986	20.64503	100
90	0.4349207	0.372049	0.1930304	1	43.49207	37.2049	19.30304	100
91	0.2453657	0.4722251	0.2824092	1	24.53657	47.22251	28.24092	100
92	0.5384412	0.3164578	0.145101	1	53.84412	31.64578	14.5101	100
93	0.2666665	0.4753242	0.2580093	1	26.66665	47.53242	25.80093	100
94	0.2728257	0.4730994	0.2540749	1	27.28257	47.30994	25.40749	100
95	0.365843	0.4046533	0.2295037	1	36.5843	40.46533	22.95037	100
96	0.3751854	0.393646	0.2311687	1	37.51854	39.3646	23.11687	100
97	0.5296888	0.3129514	0.1573598	1	52.96888	31.29514	15.73598	100
98	0.2345904	0.4991398	0.2662698	1	23.45904	49.91398	26.62698	100
99	0.2548403	0.4787292	0.2664305	1	25.48403	47.87292	26.64305	100
100	0.493839	0.3487559	0.1574051	1	49.3839	34.87559	15.74051	100
101	0.5641864	0.275774	0.1600396	1	56.41864	27.5774	16.00396	100
102	0.4108041	0.3736677	0.2155281	1	41.08041	37.36677	21.55281	100
103	0.3913754	0.3989011	0.2097236	1	39.13754	39.89011	20.97236	100
104	0.380199	0.4087163	0.2110847	1	38.0199	40.87163	21.10847	100
105	0.46251	0.3488656	0.1886244	1	46.251	34.88656	18.86244	100
106	0.4931439	0.3306296	0.1762265	1	49.31439	33.06296	17.62265	100
107	0.3584513	0.4166047	0.2249441	1	35.84513	41.66047	22.49441	100
108	0.6249241	0.2499196	0.1251563	1	62.49241	24.99196	12.51563	100
109	0.4755165	0.341596	0.1828876	1	47.55165	34.1596	18.28876	100
110	0.8903062	0.0780184	0.0316755	1	89.03062	7.801839	3.167546	100
111	0.3775158	0.4120098	0.2104744	1	37.75158	41.20098	21.04744	100
112	0.6893028	0.2085421	0.102155	1	68.93028	20.85421	10.2155	100

113	0.3829836	0.3762302	0.2407862	1	38.29836	37.62302	24.07862	100
114	0.3838951	0.4148067	0.2012982	1	38.38951	41.48067	20.12982	100
115	0.3348914	0.3896242	0.2754843	1	33.48914	38.96242	27.54843	100
116	0.3405604	0.4284037	0.2310359	1	34.05604	42.84037	23.10359	100
117	0.3915675	0.4042741	0.2041585	1	39.15675	40.42741	20.41585	100
118	0.4980995	0.3429216	0.1589788	1	49.80995	34.29216	15.89788	100
119	0.5409088	0.2890306	0.1700606	1	54.09088	28.90306	17.00606	100
120	0.4638342	0.3535961	0.1825697	1	46.38342	35.35961	18.25697	100
121	0.4283147	0.3807699	0.1909154	1	42.83147	38.07699	19.09154	100
122	0.4526918	0.3478756	0.1994326	1	45.26918	34.78756	19.94326	100
123	0.4393539	0.3554094	0.2052366	1	43.93539	35.54094	20.52366	100
124	0.59451	0.270578	0.134912	1	59.451	27.0578	13.4912	100
125	0.4786166	0.3329031	0.1884803	1	47.86166	33.29031	18.84803	100
126	0.4412049	0.3670859	0.1917091	1	44.12049	36.70859	19.17091	100
127	0.3093308	0.456051	0.2346183	1	30.93308	45.6051	23.46183	100
128	0.4168458	0.38779	0.1953642	1	41.68458	38.779	19.53642	100
129	0.2738086	0.4760805	0.2501109	1	27.38086	47.60805	25.01109	100
130	0.6179547	0.2610558	0.1209895	1	61.79547	26.10558	12.09895	100
131	0.6092938	0.2519092	0.138797	1	60.92938	25.19092	13.8797	100
132	0.2081191	0.48832	0.3035609	1	20.81191	48.832	30.35609	100
133	0.4154467	0.3952656	0.1892877	1	41.54467	39.52656	18.92877	100
134	0.535157	0.2655533	0.1992897	1	53.5157	26.55533	19.92897	100
135	0.4022867	0.3914367	0.2062766	1	40.22867	39.14367	20.62766	100
136	0.4233702	0.3845238	0.192106	1	42.33702	38.45238	19.2106	100
137	0.4147402	0.3863525	0.1989073	1	41.47402	38.63525	19.89073	100
138	0.3411387	0.4200032	0.2388581	1	34.11387	42.00032	23.88581	100
139	0.2616803	0.4779694	0.2603503	1	26.16803	47.79694	26.03503	100
140	0.5585773	0.3024721	0.1389506	1	55.85773	30.24721	13.89506	100
141	0.2459455	0.4835903	0.2704642	1	24.59455	48.35903	27.04642	100
142	0.3496716	0.4034655	0.2468628	1	34.96716	40.34655	24.68628	100
143	0.3124728	0.4437582	0.2437689	1	31.24728	44.37582	24.37689	100
144	0.5418722	0.3054128	0.152715	1	54.18722	30.54128	15.2715	100
145	0.3107755	0.4516182	0.2376063	1	31.07755	45.16182	23.76063	100
146	0.3393163	0.437632	0.2230516	1	33.93163	43.7632	22.30516	100
147	0.2936307	0.4681377	0.2382315	1	29.36307	46.81377	23.82315	100
148	0.4960921	0.3325589	0.171349	1	49.60921	33.25589	17.1349	100
149	0.4250508	0.3904037	0.1845455	1	42.50508	39.04037	18.45455	100
150	0.3849811	0.3969533	0.2180656	1	38.49811	39.69533	21.80656	100
151	0.4932798	0.353194	0.1535263	1	49.32798	35.3194	15.35263	100
152	0.4521191	0.3533362	0.1945448	1	45.21191	35.33362	19.45448	100
153	0.433801	0.3664312	0.1997679	1	43.3801	36.64312	19.97679	100
154	0.9137992	0.0650073	0.0211934	1	91.37992	6.500734	2.119341	100
155	0.4489513	0.3458049	0.2052438	1	44.89513	34.58049	20.52438	100
156	0.4030943	0.3859507	0.2109551	1	40.30943	38.59507	21.09551	100
157	0.3857924	0.3937567	0.2204509	1	38.57924	39.37567	22.04509	100
158	0.5686868	0.2896014	0.1417118	1	56.86868	28.96014	14.17118	100

159	0.4971246	0.3304571	0.1724183	1	49.71246	33.04571	17.24183	100
160	0.5538019	0.2994661	0.1467321	1	55.38019	29.94661	14.67321	100
161	0.5226416	0.3038593	0.1734991	1	52.26416	30.38593	17.34991	100
162	0.3651171	0.4200768	0.2148061	1	36.51171	42.00768	21.48061	100
163	0.4211834	0.3876188	0.1911977	1	42.11834	38.76188	19.11977	100
164	0.3928082	0.3760011	0.2311907	1	39.28082	37.60011	23.11907	100
165	0.5459679	0.2977658	0.1562662	1	54.59679	29.77658	15.62662	100
166	0.3899411	0.3734638	0.2365952	1	38.99411	37.34638	23.65952	100
167	0.5427807	0.2913539	0.1658653	1	54.27807	29.13539	16.58653	100
168	0.5810239	0.2756249	0.1433512	1	58.10239	27.56249	14.33512	100
169	0.3138382	0.4256952	0.2604665	1	31.38382	42.56952	26.04665	100
170	0.3566567	0.4097126	0.2336307	1	35.66567	40.97126	23.36307	100
171	0.3929068	0.4018153	0.2052779	1	39.29068	40.18153	20.52779	100
172	0.5558712	0.273986	0.1701428	1	55.58712	27.3986	17.01428	100
173	0.4810286	0.3280023	0.190969	1	48.10286	32.80023	19.0969	100
174	0.4492908	0.3703715	0.1803377	1	44.92908	37.03715	18.03377	100
175	0.4810699	0.3521282	0.166802	1	48.10699	35.21282	16.6802	100
176	0.2502114	0.4589958	0.2907928	1	25.02114	45.89958	29.07928	100
177	0.2234078	0.490172	0.2864203	1	22.34078	49.0172	28.64203	100
178	0.4023506	0.3938267	0.2038227	1	40.23506	39.38267	20.38227	100
179	0.4525554	0.3687694	0.1786753	1	45.25554	36.87694	17.86753	100
180	0.5332621	0.3138508	0.152887	1	53.32621	31.38508	15.2887	100
181	0.4424551	0.3610167	0.1965283	1	44.24551	36.10167	19.65283	100
182	0.9252884	0.0498354	0.0248762	1	92.52884	4.98354	2.487616	100
183	0.3567375	0.4244976	0.2187649	1	35.67375	42.44976	21.87649	100
184	0.3750521	0.4156458	0.2093022	1	37.50521	41.56458	20.93022	100
185	0.4254543	0.3782432	0.1963025	1	42.54543	37.82432	19.63025	100
186	0.4701543	0.3595847	0.170261	1	47.01543	35.95847	17.0261	100
187	0.5072499	0.3328677	0.1598824	1	50.72499	33.28677	15.98824	100
188	0.3223104	0.4353937	0.2422959	1	32.23104	43.53937	24.22959	100
189	0.6060596	0.2652126	0.1287278	1	60.60596	26.52126	12.87278	100
190	0.2039476	0.5154	0.2806524	1	20.39476	51.54	28.06524	100
191	0.2988132	0.4301102	0.2710766	1	29.88132	43.01102	27.10766	100
192	0.4515245	0.3563547	0.1921208	1	45.15245	35.63547	19.21208	100
193	0.3458385	0.4026711	0.2514904	1	34.58385	40.26711	25.14904	100
194	0.2988213	0.4689339	0.2322448	1	29.88213	46.89339	23.22448	100
195	0.4700227	0.3496875	0.1802897	1	47.00227	34.96875	18.02897	100
196	0.439991	0.3543823	0.2056267	1	43.9991	35.43823	20.56267	100
197	0.4043588	0.4005576	0.1950836	1	40.43588	40.05576	19.50836	100
198	0.3390533	0.4301726	0.2307741	1	33.90533	43.01726	23.07741	100
199	0.1268996	0.5594287	0.3136717	1	12.68996	55.94287	31.36717	100
200	0.3830238	0.4071552	0.209821	1	38.30238	40.71552	20.9821	100
201	0.3542674	0.4324381	0.2132945	1	35.42674	43.24381	21.32945	100
202	0.3836367	0.3999441	0.2164192	1	38.36367	39.99441	21.64192	100
203	0.327743	0.4332294	0.2390276	1	32.7743	43.32294	23.90276	100
204	0.3422621	0.4244172	0.2333208	1	34.22621	42.44172	23.33208	100

205	0.4985421	0.2973461	0.2041117	1	49.85421	29.73461	20.41117	100
206	0.4522547	0.364936	0.1828093	1	45.22547	36.4936	18.28093	100
207	0.3058603	0.4069056	0.2872341	1	30.58603	40.69056	28.72341	100
208	0.309528	0.4453057	0.2451664	1	30.9528	44.53057	24.51664	100
209	0.6757091	0.2058236	0.1184673	1	67.57091	20.58236	11.84673	100
210	0.7267773	0.181522	0.0917007	1	72.67773	18.1522	9.170066	100
211	0.4455552	0.3576573	0.1967875	1	44.55552	35.76573	19.67875	100
212	0.3366688	0.4393998	0.2239315	1	33.66688	43.93998	22.39315	100
213	0.1940924	0.5218979	0.2840097	1	19.40924	52.18979	28.40097	100
214	0.4474112	0.3791504	0.1734385	1	44.74112	37.91504	17.34385	100
215	0.5480102	0.2790511	0.1729387	1	54.80102	27.90511	17.29387	100
216	0.3905483	0.4195634	0.1898883	1	39.05483	41.95634	18.98883	100
217	0.2325896	0.4851665	0.282244	1	23.25896	48.51665	28.2244	100
218	0.4594408	0.360706	0.1798533	1	45.94408	36.0706	17.98533	100
219	0.3434655	0.4400076	0.2165269	1	34.34655	44.00076	21.65269	100
220	0.0123563	0.5576411	0.4300026	1	1.235634	55.76411	43.00026	100
221	0.4021117	0.3862853	0.2116031	1	40.21117	38.62853	21.16031	100
222	0.2936155	0.4758988	0.2304857	1	29.36155	47.58988	23.04857	100
223	0.4658378	0.3462236	0.1879386	1	46.58378	34.62236	18.79386	100
224	0.52068	0.3291993	0.1501208	1	52.068	32.91993	15.01208	100
225	0.2835579	0.4837798	0.2326623	1	28.35579	48.37798	23.26623	100
226	0.2162399	0.4981532	0.2856069	1	21.62399	49.81532	28.56069	100
227	0.1734384	0.5591927	0.2673689	1	17.34384	55.91927	26.73689	100
228	0.2583398	0.4984817	0.2431785	1	25.83398	49.84817	24.31785	100
229	0.5212059	0.3323616	0.1464325	1	52.12059	33.23616	14.64325	100
230	0.6006511	0.2596429	0.139706	1	60.06511	25.96429	13.9706	100
231	0.2085428	0.5237379	0.2677194	1	20.85428	52.37379	26.77194	100
232	0.2716358	0.4530501	0.2753141	1	27.16358	45.30501	27.53141	100
233	0.7012482	0.2062833	0.0924685	1	70.12482	20.62833	9.24685	100
234	0.460353	0.3720499	0.1675971	1	46.0353	37.20499	16.75971	100
235	0.6711582	0.2060707	0.122771	1	67.11582	20.60707	12.2771	100

Table A4.4 Predictions dataset from the Random Forest Model 2, with CLR transformation applied to the response variable and validated with LOOCV. Soil particle-size fractions predicted and back-transformed, showing the 100% sum in each sample.

Sample	inv.clr.1	inv.clr.2	inv.clr.3	Total	Sand_pred (%)	Silt_pred (%)	Clay_pred (%)	Total
1	0.5786021	0.2609257	0.1604722	1	57.860208	26.092573	16.047219	100
2	0.4431764	0.3858696	0.170954	1	44.317636	38.586962	17.095402	100
3	0.5106524	0.3314248	0.1579229	1	51.065237	33.142476	15.792288	100
4	0.3026779	0.4849303	0.2123918	1	30.267794	48.493028	21.239178	100
5	0.625781	0.2312552	0.1429638	1	62.578101	23.125522	14.296377	100
6	0.3862645	0.3831923	0.2305432	1	38.62645	38.319229	23.054321	100
7	0.5154479	0.3266944	0.1578578	1	51.544788	32.669437	15.785775	100
8	0.5615597	0.2839731	0.1544672	1	56.155975	28.397306	15.44672	100
9	0.4444906	0.3805915	0.1749179	1	44.449059	38.059148	17.491793	100
10	0.4761513	0.3555143	0.1683344	1	47.615128	35.551427	16.833444	100
11	0.6100871	0.2599136	0.1299993	1	61.008708	25.991358	12.999934	100
12	0.2819855	0.4673876	0.2506269	1	28.198553	46.738762	25.062685	100
13	0.4922756	0.3592074	0.1485169	1	49.227563	35.920743	14.851695	100
14	0.7169372	0.187373	0.0956898	1	71.693722	18.737299	9.5689784	100
15	0.5185383	0.3106173	0.1708445	1	51.853826	31.061725	17.084449	100
16	0.52129	0.325607	0.153103	1	52.128999	32.560699	15.310302	100

17	0.2890024	0.4051579	0.3058397	1	28.90024	40.515789	30.58397	100
18	0.6436774	0.2270884	0.1292342	1	64.367741	22.70884	12.923419	100
19	0.3492634	0.4471534	0.2035832	1	34.926341	44.715339	20.35832	100
20	0.2919736	0.4550039	0.2530225	1	29.197359	45.500389	25.302252	100
21	0.4653361	0.369205	0.1654589	1	46.533606	36.920501	16.545894	100
22	0.3428204	0.436101	0.2210786	1	34.28204	43.610097	22.107863	100
23	0.4323171	0.3840478	0.1836351	1	43.231713	38.40478	18.363507	100
24	0.5234451	0.323518	0.153037	1	52.344506	32.351799	15.303696	100
25	0.3352092	0.4421782	0.2226125	1	33.520925	44.217822	22.261253	100
26	0.3059572	0.4740542	0.2199886	1	30.595718	47.40542	21.998862	100
27	0.5423864	0.2837509	0.1738626	1	54.238644	28.375093	17.386263	100
28	0.3061023	0.4316541	0.2622437	1	30.610227	43.165407	26.224366	100
29	0.5728338	0.2800461	0.1471201	1	57.283375	28.004613	14.712012	100
30	0.2505822	0.4427155	0.3067023	1	25.058224	44.271546	30.67023	100
31	0.4205779	0.4012855	0.1781367	1	42.057789	40.128546	17.813666	100
32	0.2622737	0.4897464	0.2479799	1	26.227373	48.974637	24.79799	100
33	0.333359	0.4433277	0.2233133	1	33.335902	44.332768	22.33133	100
34	0.2894461	0.4511748	0.2593791	1	28.944612	45.117477	25.937911	100
35	0.4163437	0.4040022	0.179654	1	41.634375	40.400223	17.965402	100
36	0.1648581	0.5040559	0.3310861	1	16.485807	50.405585	33.108608	100
37	0.253954	0.4956634	0.2503826	1	25.395405	49.56634	25.038255	100
38	0.5924483	0.2835844	0.1239674	1	59.244826	28.358438	12.396736	100
39	0.4378876	0.3884338	0.1736786	1	43.788764	38.843379	17.367857	100
40	0.1719919	0.4732725	0.3547356	1	17.199191	47.327245	35.473563	100
41	0.4804646	0.3574657	0.1620698	1	48.046457	35.746567	16.206976	100
42	0.5093461	0.3420043	0.1486497	1	50.934608	34.200425	14.864967	100
43	0.4137083	0.4234658	0.1628259	1	41.37083	42.346582	16.282587	100
44	0.6433017	0.2313981	0.1253002	1	64.330172	23.139811	12.530018	100
45	0.6689524	0.2052185	0.125829	1	66.895245	20.521851	12.582904	100
46	0.5712615	0.284361	0.1443775	1	57.126146	28.436104	14.43775	100
47	0.4568296	0.3836742	0.1594963	1	45.682956	38.367417	15.949627	100
48	0.3251462	0.4377944	0.2370595	1	32.514617	43.779436	23.705947	100
49	0.6764427	0.1988657	0.1246916	1	67.644271	19.886571	12.469158	100
50	0.2781328	0.4544888	0.2673785	1	27.813276	45.448877	26.737846	100
51	0.4605057	0.3614287	0.1780656	1	46.05057	36.142871	17.806558	100
52	0.2891197	0.4876074	0.2232729	1	28.911971	48.760743	22.327286	100
53	0.414913	0.404013	0.181074	1	41.491299	40.4013	18.107401	100
54	0.3164024	0.4692656	0.214332	1	31.64024	46.926558	21.433201	100
55	0.3032747	0.4312215	0.2655038	1	30.32747	43.122147	26.550383	100
56	0.4381374	0.367847	0.1940155	1	43.813745	36.784702	19.401553	100
57	0.633353	0.2324952	0.1341517	1	63.335305	23.249525	13.41517	100
58	0.4283491	0.4058216	0.1658294	1	42.834909	40.582156	16.582935	100
59	0.6495212	0.239812	0.1106668	1	64.952119	23.9812	11.066681	100
60	0.3765503	0.4220008	0.2014489	1	37.655032	42.200078	20.14489	100
61	0.3326953	0.4482672	0.2190375	1	33.26953	44.82672	21.903751	100
62	0.2042334	0.4378255	0.3579412	1	20.423336	43.782548	35.794115	100
63	0.509637	0.3449091	0.1454539	1	50.963696	34.490915	14.54539	100
64	0.1998268	0.4979437	0.3022295	1	19.982677	49.794372	30.22295	100
65	0.6995525	0.2088015	0.091646	1	69.955253	20.880145	9.1646016	100
66	0.1363273	0.5178447	0.345828	1	13.632725	51.78447	34.582805	100
67	0.3307564	0.4331112	0.2361324	1	33.075643	43.311116	23.61324	100
68	0.4565699	0.3629863	0.1804438	1	45.656992	36.298626	18.044382	100
69	0.5483971	0.3014471	0.1501558	1	54.839711	30.144713	15.015577	100
70	0.3845564	0.4158664	0.1995772	1	38.455639	41.586641	19.957721	100
71	0.4006	0.3771186	0.2222815	1	40.059996	37.711857	22.228147	100
72	0.4846894	0.3566341	0.1586765	1	48.468938	35.663413	15.867649	100
73	0.5162418	0.3229897	0.1607685	1	51.624184	32.298967	16.076849	100
74	0.5018226	0.3276531	0.1705243	1	50.182261	32.76531	17.052429	100
75	0.4750233	0.3600199	0.1649568	1	47.502331	36.00199	16.495679	100
76	0.4732541	0.3557609	0.170985	1	47.32541	35.576087	17.098503	100
77	0.5863571	0.2607092	0.1529337	1	58.635708	26.070923	15.293369	100
78	0.4242867	0.4229239	0.1527893	1	42.428675	42.292392	15.278934	100
79	0.4332557	0.4021438	0.1646005	1	43.325572	40.214381	16.460047	100
80	0.557052	0.2856212	0.1573268	1	55.705196	28.562121	15.732683	100
81	0.1148893	0.515857	0.3692537	1	11.488929	51.585705	36.925367	100
82	0.2789657	0.4497804	0.2712538	1	27.896572	44.978044	27.125384	100
83	0.2996491	0.422774	0.2775769	1	29.964911	42.277402	27.757687	100

84	0.3703642	0.412043	0.2175927	1	37.036423	41.204302	21.759275	100
85	0.6424803	0.2520866	0.1054331	1	64.248025	25.208664	10.543311	100
86	0.560947	0.2867901	0.1522629	1	56.094702	28.679006	15.226293	100
87	0.4902835	0.3652177	0.1444987	1	49.028355	36.521773	14.449872	100
88	0.2742436	0.5061289	0.2196275	1	27.42436	50.612889	21.962751	100
89	0.4663238	0.3499164	0.1837598	1	46.632379	34.99164	18.375981	100
90	0.4934504	0.3464153	0.1601343	1	49.345038	34.64153	16.013432	100
91	0.2587975	0.4632927	0.2779097	1	25.879755	46.32927	27.790975	100
92	0.5488414	0.3050987	0.1460599	1	54.884141	30.509868	14.605991	100
93	0.2736273	0.4568479	0.2695247	1	27.362734	45.684794	26.952472	100
94	0.2753175	0.4684683	0.2562143	1	27.531747	46.846827	25.621426	100
95	0.3419432	0.4206377	0.237419	1	34.194323	42.063773	23.741904	100
96	0.4067966	0.3888723	0.2043311	1	40.679659	38.887228	20.433113	100
97	0.5865678	0.2550952	0.1583371	1	58.656777	25.509517	15.833707	100
98	0.2431654	0.4959277	0.2609069	1	24.316537	49.592768	26.090695	100
99	0.2524312	0.4801421	0.2674268	1	25.243116	48.014207	26.742677	100
100	0.5012569	0.348024	0.1507191	1	50.125691	34.802399	15.07191	100
101	0.6473173	0.2304104	0.1222724	1	64.731727	23.041035	12.227238	100
102	0.4513542	0.3799945	0.1686512	1	45.135423	37.999452	16.865125	100
103	0.3818062	0.4221309	0.1960629	1	38.180616	42.213093	19.606291	100
104	0.3491951	0.4350853	0.2157196	1	34.919513	43.508528	21.571959	100
105	0.5013322	0.3286649	0.1700029	1	50.13322	32.86649	17.00029	100
106	0.5289162	0.3001567	0.1709271	1	52.891619	30.015673	17.092709	100
107	0.3290183	0.4261149	0.2448668	1	32.901829	42.611492	24.486679	100
108	0.6324818	0.2183957	0.1491224	1	63.248184	21.839571	14.912245	100
109	0.5117104	0.3212514	0.1670382	1	51.171036	32.125141	16.703824	100
110	0.7074671	0.2016969	0.090836	1	70.746706	20.169689	9.083605	100
111	0.3457688	0.4344977	0.2197336	1	34.576876	43.449766	21.973358	100
112	0.6740733	0.2074899	0.1184368	1	67.407331	20.74899	11.843679	100
113	0.3542784	0.4200872	0.2256345	1	35.427835	42.008716	22.563449	100
114	0.3396698	0.4614866	0.1988436	1	33.966977	46.148662	19.884361	100
115	0.2945158	0.4485608	0.2569234	1	29.451582	44.856082	25.692336	100
116	0.3401854	0.4167766	0.243038	1	34.018542	41.677659	24.303799	100
117	0.4249886	0.3965735	0.1784379	1	42.498856	39.657351	17.843793	100
118	0.4640476	0.3654711	0.1704813	1	46.404759	36.54711	17.048131	100
119	0.5669791	0.2746066	0.1584144	1	56.697906	27.460655	15.841439	100
120	0.5084839	0.330159	0.1613571	1	50.848386	33.015902	16.135712	100
121	0.4363456	0.3896097	0.1740447	1	43.634561	38.960965	17.404474	100
122	0.4610062	0.3896172	0.1493766	1	46.100622	38.961722	14.937657	100
123	0.4512703	0.3423143	0.2064154	1	45.127031	34.231428	20.641541	100
124	0.6094005	0.2408896	0.1497099	1	60.940051	24.088956	14.970993	100
125	0.4931415	0.3612695	0.145589	1	49.314153	36.12695	14.558897	100
126	0.4529813	0.3722744	0.1747443	1	45.29813	37.22744	17.47443	100
127	0.2915315	0.4840811	0.2243874	1	29.153153	48.408109	22.438738	100
128	0.4373473	0.3922016	0.1704511	1	43.73473	39.220156	17.045115	100
129	0.2794491	0.4563905	0.2641604	1	27.94491	45.63905	26.41604	100
130	0.6413787	0.2331045	0.1255169	1	64.137866	23.310446	12.551688	100
131	0.6908166	0.2010142	0.1081693	1	69.081657	20.101415	10.816928	100
132	0.3325457	0.445039	0.2224153	1	33.254571	44.503899	22.24153	100
133	0.4346777	0.3957958	0.1695264	1	43.467773	39.579584	16.952642	100
134	0.5313797	0.3370258	0.1315945	1	53.137967	33.702583	13.15945	100
135	0.4285191	0.36813	0.2033509	1	42.851912	36.812995	20.335092	100
136	0.449954	0.3717812	0.1782648	1	44.995405	37.178116	17.82648	100
137	0.4275235	0.4355535	0.1369229	1	42.752353	43.555352	13.692295	100
138	0.3190044	0.4395328	0.2414628	1	31.900441	43.953283	24.146275	100
139	0.2568972	0.4530671	0.2900357	1	25.689722	45.306707	29.003571	100
140	0.5668128	0.2868255	0.1463617	1	56.681276	28.682551	14.636173	100
141	0.2432201	0.461335	0.2954449	1	24.32201	46.133499	29.544491	100
142	0.324084	0.4073708	0.2685451	1	32.408403	40.737082	26.854515	100
143	0.3087832	0.4401129	0.251104	1	30.878319	44.011285	25.110396	100
144	0.5633391	0.2911836	0.1454773	1	56.33391	29.118359	14.547732	100
145	0.2759359	0.4628667	0.2611974	1	27.593592	46.286665	26.119743	100
146	0.3358966	0.4538007	0.2103027	1	33.58966	45.380067	21.030274	100
147	0.2633548	0.4735759	0.2630693	1	26.335482	47.357591	26.306927	100
148	0.5549528	0.2796423	0.1654049	1	55.495282	27.964226	16.540492	100
149	0.4767815	0.3786165	0.144602	1	47.678153	37.861648	14.460199	100
150	0.34323	0.4152974	0.2414726	1	34.322997	41.529739	24.147264	100

151	0.5164188	0.3392247	0.1443565	1	51.641882	33.922469	14.435649	100
152	0.4644066	0.3417598	0.1938337	1	46.440656	34.175979	19.383365	100
153	0.4374786	0.3891051	0.1734163	1	43.747861	38.910511	17.341628	100
154	0.5792005	0.273702	0.1470975	1	57.920046	27.370204	14.709749	100
155	0.5098421	0.3200538	0.1701041	1	50.984212	32.005375	17.010412	100
156	0.3841272	0.4027998	0.2130731	1	38.412716	40.279977	21.307307	100
157	0.3892312	0.4012519	0.2095169	1	38.923124	40.125188	20.951688	100
158	0.5353898	0.2883124	0.1762978	1	53.538979	28.831236	17.629785	100
159	0.5203952	0.3117191	0.1678857	1	52.039522	31.171908	16.788569	100
160	0.5844304	0.2663134	0.1492562	1	58.443039	26.63134	14.925621	100
161	0.5685409	0.2676169	0.1638421	1	56.854095	26.761693	16.384212	100
162	0.3107189	0.4727167	0.2165644	1	31.071887	47.271672	21.656441	100
163	0.5051389	0.3283745	0.1664866	1	50.513889	32.837447	16.648664	100
164	0.384478	0.3759759	0.2395461	1	38.4478	37.597592	23.954608	100
165	0.5437674	0.2964156	0.159817	1	54.376736	29.641559	15.981705	100
166	0.3383068	0.4105494	0.2511438	1	33.830684	41.054941	25.114375	100
167	0.5985624	0.2535074	0.1479302	1	59.856239	25.350737	14.793024	100
168	0.6060627	0.2398924	0.1540449	1	60.606269	23.989239	15.404492	100
169	0.3002996	0.4369637	0.2627367	1	30.02996	43.696371	26.273668	100
170	0.3077253	0.4507018	0.2415729	1	30.772533	45.070178	24.157289	100
171	0.404419	0.4320153	0.1635657	1	40.441901	43.201527	16.356572	100
172	0.5746926	0.2616495	0.1636579	1	57.469265	26.164945	16.36579	100
173	0.5095023	0.3519648	0.1385329	1	50.950229	35.196479	13.853293	100
174	0.479717	0.3618224	0.1584606	1	47.971699	36.182245	15.846056	100
175	0.501208	0.3419819	0.1568101	1	50.120799	34.198193	15.681008	100
176	0.2358833	0.4637749	0.3003419	1	23.588326	46.377486	30.034188	100
177	0.2621624	0.4604164	0.2774212	1	26.21624	46.041639	27.742121	100
178	0.4160524	0.4110344	0.1729132	1	41.60524	41.103444	17.291316	100
179	0.4919414	0.3648156	0.143243	1	49.194139	36.48156	14.324301	100
180	0.5708594	0.2822052	0.1469353	1	57.085945	28.220522	14.693534	100
181	0.4516049	0.3498952	0.1984999	1	45.160492	34.98952	19.849987	100
182	0.6760494	0.20086	0.1230906	1	67.604942	20.085996	12.309062	100
183	0.3374936	0.4301792	0.2323272	1	33.749364	43.017916	23.23272	100
184	0.3325654	0.4752253	0.1922093	1	33.25654	47.522535	19.220926	100
185	0.4904243	0.3332442	0.1763315	1	49.042431	33.324424	17.633145	100
186	0.4753346	0.3530268	0.1716386	1	47.533462	35.30268	17.163859	100
187	0.565727	0.282618	0.151655	1	56.572704	28.261797	15.165499	100
188	0.3054998	0.4441734	0.2503268	1	30.549982	44.417341	25.032677	100
189	0.651691	0.2207999	0.1275092	1	65.169096	22.079988	12.750916	100
190	0.2037281	0.5084792	0.2877927	1	20.372809	50.847923	28.779268	100
191	0.2895488	0.4513022	0.259149	1	28.954877	45.130224	25.914899	100
192	0.4865477	0.3444493	0.169003	1	48.65477	34.444928	16.900302	100
193	0.3325336	0.4107126	0.2567538	1	33.253361	41.071261	25.675378	100
194	0.2215437	0.5011947	0.2772616	1	22.154368	50.119471	27.726162	100
195	0.5245448	0.3158389	0.1596164	1	52.454478	31.583885	15.961636	100
196	0.4774154	0.3572993	0.1652853	1	47.741541	35.72993	16.528529	100
197	0.4103107	0.4325739	0.1571154	1	41.031066	43.257392	15.711542	100
198	0.3007159	0.4644675	0.2348167	1	30.071586	46.446746	23.481668	100
199	0.1874477	0.4383807	0.3741716	1	18.74477	43.838075	37.417156	100
200	0.3168337	0.4480886	0.2350776	1	31.683373	44.808863	23.507764	100
201	0.29884	0.4916639	0.2094961	1	29.884	49.16639	20.94961	100
202	0.3807938	0.4119145	0.2072917	1	38.079375	41.191451	20.729173	100
203	0.3022041	0.4614617	0.2363341	1	30.220413	46.146174	23.633413	100
204	0.2873503	0.451937	0.2607127	1	28.735033	45.193699	26.071267	100
205	0.6747599	0.213902	0.1113381	1	67.475994	21.390195	11.133811	100
206	0.4123767	0.397104	0.1905192	1	41.237675	39.7104	19.051925	100
207	0.3251884	0.4025736	0.2722379	1	32.518845	40.257363	27.223792	100
208	0.2859534	0.4719437	0.242103	1	28.595335	47.194369	24.210295	100
209	0.6409315	0.2247531	0.1343154	1	64.093153	22.475305	13.431541	100
210	0.700676	0.1945529	0.1047711	1	70.067599	19.455293	10.477108	100
211	0.438394	0.3678911	0.1937149	1	43.8394	36.789112	19.371488	100
212	0.3214038	0.4258198	0.2527764	1	32.140381	42.581982	25.277636	100
213	0.1783974	0.517734	0.3038686	1	17.83974	51.7734	30.38686	100
214	0.4606239	0.3915385	0.1478376	1	46.062385	39.153854	14.783761	100
215	0.4745075	0.3313887	0.1941039	1	47.450747	33.138867	19.410386	100
216	0.3585305	0.4433742	0.1980953	1	35.85305	44.337421	19.809529	100
217	0.2086529	0.4265762	0.364771	1	20.865286	42.657616	36.477098	100

218	0.4086956	0.4221129	0.1691915	1	40.869556	42.21129	16.919154	100
219	0.2961237	0.437453	0.2664234	1	29.612368	43.745296	26.642336	100
220	0.132615	0.5130546	0.3543304	1	13.261503	51.305455	35.433042	100
221	0.3662515	0.4381034	0.1956451	1	36.62515	43.810336	19.564514	100
222	0.2489442	0.4934535	0.2576023	1	24.894422	49.345351	25.760227	100
223	0.5766313	0.2901802	0.1331885	1	57.663129	29.018022	13.318849	100
224	0.5405379	0.3090011	0.1504609	1	54.053792	30.900115	15.046093	100
225	0.2786619	0.4533031	0.2680349	1	27.866195	45.330315	26.803491	100
226	0.1702266	0.5130369	0.3167365	1	17.022658	51.303688	31.673654	100
227	0.174174	0.5003311	0.3254949	1	17.4174	50.033108	32.549493	100
228	0.1923497	0.5319076	0.2757428	1	19.234966	53.190756	27.574278	100
229	0.5064762	0.3214074	0.1721164	1	50.647624	32.140736	17.21164	100
230	0.6070282	0.254305	0.1386668	1	60.702821	25.430503	13.866676	100
231	0.2122133	0.4705284	0.3172583	1	21.221334	47.052841	31.725826	100
232	0.2150966	0.4320126	0.3528908	1	21.509655	43.201263	35.289082	100
233	0.5963074	0.2637506	0.1399421	1	59.630736	26.375058	13.994205	100
234	0.4662842	0.3766767	0.1570391	1	46.628418	37.667668	15.703914	100
235	0.6471353	0.2186438	0.1342209	1	64.713533	21.86438	13.422087	100

Table A4.5 Percentages of land cover categories on predictions from the training data.

Category*	%
Pasture	78.97727
Agriculture (crops; arable land)	10.22727
Forest	6.25
Others	4.545455

* CORINE-Copernicus Land Cover – 2018 (Ireland)

Table A4.6 Percentages of land cover categories on predictions from the test data.

Category*	%
Pasture	64.40678
Agriculture (crops; arable land)	18.64407
Forest	16.94915

* CORINE-Copernicus Land Cover – 2018 (Ireland)

Appendix A – Dataset details

Soil information

Soil data, used to train and validate the models, consisted of 235 soil samples collected by different institutions and programs across Ireland covering both shallow and deep soil layers. The source datasets are as follows:

- Irish Soil Information System (SIS). Modal and Non-Modal profiles (Creamer *et al.*, 2014). Soil Survey carried out by the Irish Agriculture and Food Development Authority (Teagasc);
- LUCAS Topsoil 2009 Survey-v1 (Toth *et al.*, 2013; Orgiazzi *et al.* 2018) and LUCAS Topsoil 2015 Survey (Jones *et al.*, 2020), both soil campaigns were carried out by the European Soil Data Centre (ESDAC) under the programme named Land Use/Land Cover Area Frame Survey (LUCAS);
- Soil Carbon Project 2008. SoilC was a two-year project to estimate soil carbon stocks in representative Irish soils (by field sampling and laboratory soils analysis) and to model soil carbon cycling in Irish grassland. It was carried out by the University College Cork (UCC) and Teagasc. Details on the survey and results can be found in Kiely *et al.* (2009).

- WoSIS Soil Profile Database (June 2022). The World Soil Information Service (WoSIS) is a global soil database that aims to provide the user with standardised and ultimately harmonised soil profile data addressing 45 soil properties (physical and chemical).
- All the dataset encompasses soil particle size fractions as one of the soil properties analysed but only the topmost layer (0-15 cm in depth) was considered in this study as it relates to topsoil.

Microwave remote sensing data

The Sentinel 1 imagery was initially co-registered, multi-looked and speckle-filtered using a Boxcar method at a 7×7 window size, and radiometrically and geometrically calibrated. Calibration was applied to convert the raw images into backscattering coefficients (σ^0), for input in the modelling approach outlined in the Section 3.5. Terrain correction was applied by using the Copernicus Digital Elevation Model for Europe (ESA EEA-10) at 10 m spatial resolution. All datasets were reprojected to the Irish Transverse Mercator projection and IRENET-95 horizontal datum. These procedures were carried out using the Sentinel-1 Toolbox (SNAP-ESA).

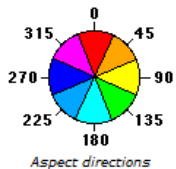
Radar-based data consist of (i) the backscatter coefficients (backscatter intensity) provided by the VV and VH polarisations measured in sigma nought (σ^0) obtained for the in-situ soil sampling locations, georeferenced; (ii) alpha parameter resulting from the dual-pol decomposition, which varies between 0° and 90° and is used to determine the dominant scattering mechanism; (iii) the dual-pol radar vegetation index (RVI, Equation 1) (Nasirzadehdizaji *et al.*, 2019; Gururaj *et al.*, 2019) developed for crop monitoring using Sentinel-1; and (iv) surface soil moisture (SSM) which refers to the relative water content of the top few centimetres of the soil, measured by Sentinel-1 under the Copernicus programme. Likewise the alpha parameter, RVI takes into account the vegetation effect on soil backscattering.

$$RVI = \frac{4\sigma_{VH}^0}{\sigma_{VV}^0 + \sigma_{VH}^0} \quad \text{Equation 1}$$

Environmental data

Topographical data derived from the digital elevation model (DEM) ESA EEA-10 (10 m spatial resolution) was used because of the soil-landscape relationship and an earth-surface characteristic related to soil-surface roughness (micro- and macro-relief of a soil surface), which is a result of variations in soil grains, soil aggregates, soil clods and due to tillage. In this study, the topography covariates consist of altitude, slope, aspect, and curvature of the slope (concave surface or convex surface). They were obtained from the aforementioned DEM by using the surface spatial analyst algorithms available in the ArcGIS® toolbox. ArcGIS is a geographic information system (GIS) software (server and online) to view, edit, manage and analyse geographic data. It was developed and is maintained by Esri. Summary information about the Spatial Analyst ArcGIS® applied in this work for deriving topographical data is presented as follows:

Summary of the procedures for deriving the topographic parameters from the DEM EEA-10

Topographic parameter	Description
Altitude (elevation)	Surface elevation values were extracted directly from the cells (pixels) of the DEM (ESA EEA-10) by using the “extract values to points” algorithm, without additional procedure.
Slope	The slope tool identifies the steepness at each cell of a raster surface. The lower the slope value, the flatter the terrain; the higher the slope value, the steeper the terrain. The output slope raster can be calculated in degrees or percent (percent rise). The percent rise can be better understood when considered as the rise divided by the run, multiplied by 100. When the angle is 45 degrees, the rise is equal to the run, and the percent rise is 100 percent. As the slope angle approaches vertical (90 degrees), the percent rise begins to approach infinity. In this study, slope is measured in degrees.
Curvature	Calculates the curvature of a raster surface, optionally including profile and plan curvature. The primary output is the curvature of the surface on a cell-by-cell basis, as fitted through that cell and its eight surrounding neighbours. Curvature is the second derivative of the surface, or the slope-of-the-slope. Two optional output curvature types are possible: the profile curvature is in the direction of the maximum slope, and the plan curvature is perpendicular to the direction of the maximum slope. In this study, we used the primary output in which a positive curvature indicates the surface is upwardly convex, a negative curvature indicates the surface is upwardly concave, and a value of zero indicates the surface is flat.
Aspect	<p>The Aspect tool identifies the direction the downhill slope faces. The values of each cell in the output raster indicate the compass direction the surface faces at that location. It is measured clockwise in degrees from 0 (due north) to 360 (again due north), coming full circle. Flat areas having no downslope direction are given a value of -1. Broadly, 0=North, 90=East, 180=South, 270=West.</p> 

Source: Spatial Analyst ArcGIS® 10.6 tool help.

Geophysical datasets

The geophysical data were derived from airborne geophysical surveying and provided by the Geological Survey of Ireland (GSI), under the Tellus programme, a national programme to gather geochemical and geophysical data across the island of Ireland. A selection of electromagnetic, radiometric (gamma-ray spectrometer: potassium concentrations (%), equivalent uranium (ppm) and equivalent thorium (ppm)) and magnetic data was obtained from the GSI. The former measures how electrical currents flow through the ground and how it changes due to different types of rock or soil. The radiometric and magnetic data provide information about soil and rocks.

Regarding the magnetic data (merged magnetic field anomaly in nT; 50 m x 50 m cell size), the First Vertical Derivative (1VD) dataset was selected, as it relates to the shallow magnetic responses, reflective of the top most soil layers (Ture *et al.*, 2020).

Electromagnetic data is commonly based on the mapping of apparent resistivity (ohm.m) or apparent conductivity; high resistivity (ρ) means low conductivity and vice-versa. All Tellus electromagnetic maps are published and presented in resistivity units. The electromagnetic resistivity at 25 kHz, reaching up to 50 m from the surface, in the 4F frequency composite image (from all four frequency images 0.9 kHz, 3 kHz, 12 kHz, and 25 kHz) was selected for this research. The rationale for this selection relies on the fact that high frequency means low penetration (see figure below), relevant for topsoil, and the composite image may reveal geological features not immediately apparent in each of the four individual resistivity maps (Ture *et al.*, 2020). The image cell size is 50 m by 50 m (an area on the ground of 2500 m²) containing the average resistivity of

all the points located within that cell. When the soil contains clay, the apparent soil electrical resistivity is influenced by the clay content. The electrical resistivity of clay ranges from ~2 to ~100 ohm.m, whereas the electrical resistivity of sand is usually higher than 1000 ohm. (Besson *et al.*, 2004; Grandjean *et al.*, 2009).

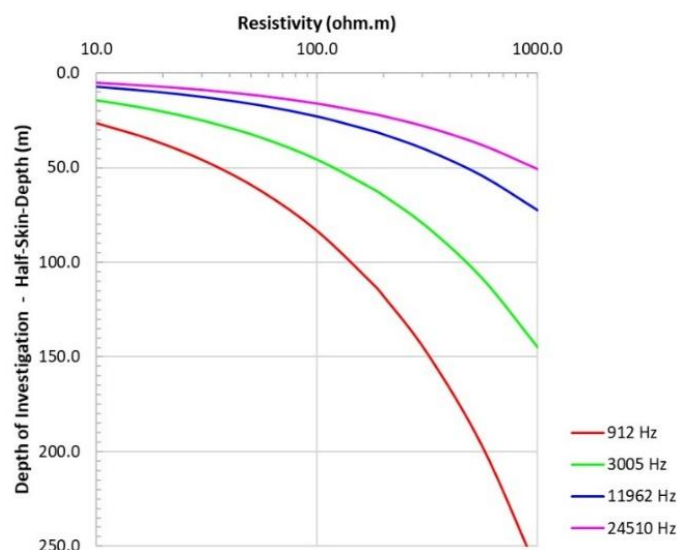


Figure. Approximate depths of investigation for typical Irish rock resistivities, for the four Tellus system frequencies, derived using the half-skin-depth estimator. Source: Ture *et al.* (2020).

REFERENCES

- Besson, A., Cousin, I., Samouëlian, A., Boizard, H. and Richard, G. (2004). Structural heterogeneity of the soil-tilled layer as characterized by 2D electrical resistivity surveying. *Soil and Tillage Research* 79:239-249. DOI:10.1016/j.still.2004.07.012.
- Creamer, R., Simo, I., Reidy, B., Carvalho, J., Fealy, R. M., Hallett, S., Jones, R., Holden, A., Holden, N., Hannam, J., Massey, P.A., Mayr, T., McDonald, E., O'Rourke, S., Sills, P., Truckell, I., Zawadzka, J. and Schulte, R. (2014). Irish Soil Information System. Synthesis Report (2007-S-CD-1-S1). EPA
- Grandjean, G., Cousin, I., Seger, M., Thiesson, J., Lambot, S., Van Wesemael, B., Stevens, A., Samyn, K., Bitri, A., Bernardie, S. (2009). From geophysical parameters to soil characteristics. Report N°BRGM/FP7-DIGISOIL-D2.1, 52 pages. Available at: https://esdac.jrc.ec.europa.eu/public_path/Digisoil-D2.1.pdf (accessed 18 December 2022)
- Jones, A., Fernández-Ugalde, O., Scarpa, S. LUCAS 2015 (2020). Topsoil Survey. Presentation of dataset and results. EUR 30332 EN. Publications Office of the European Union: Luxembourg. 2020, ISBN 978-92-76-21080-1, doi:10.2760/616084, JRC121325
- Kiely, G., Ms. McGoff, N. M., Eaton, J. M., Xu, X., Leahy, P., Carton, O. (2009). SoilC - Measurement and Modelling of Soil Carbon Stocks and Stock Changes in Irish Soils. EPA STRIVE Programme 2001-2007. STRIVE Report. SoilC Final Report – June 12, 2009.
- Nasirzadehdizaji R, Balik Sanli F, Abdikan S, Cakir Z, Sekertekin A, Ustuner M. (2019) Sensitivity Analysis of Multi-Temporal Sentinel-1 SAR Parameters to Crop Height and Canopy Coverage. *Applied Sciences*. 2019; 9(4):655. <https://doi.org/10.3390/app9040655>
- Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A., Fernández-Ugalde, O. (2018). LUCAS Soil, the largest expandable soil dataset for Europe: A review. *European Journal of Soil Science*, 69(1): 140-153, DOI: [10.1111/ejss.12499](https://doi.org/10.1111/ejss.12499)
- Tóth G., Jones A., Montanarella L. (2013) The LUCAS topsoil database and derived information on the regional variability of cropland topsoil properties in the European Union. *Environmental Monitoring and Assessment*, 185 (9), pp. 7409-7425.
- Ture, M., Muller, M., Hodgson, J. (2020). Tellus Interpretation Guide to Airborne Geophysical Data. Report 2020. Geological Survey Ireland. Available at: <https://www.gsi.ie/en-ie/publications/Pages/Tellus-Interpretation-Guide-.aspx> (accessed 2 December 2022)

Appendix A – Models

Simplicial regression model (CoDa) and log-ratio transformation methods

Compositional data analysis (CoDa) is mostly performed in terms of log-transformations and relies on log-ratios between the parts or components of one sample. The theoretical foundations are found in Aitchison (1982). As highlighted by Morais and Thomas-Agnan (2021), in simplicial regression one can use transformations to transport the simplex space S^D into the Euclidean space \mathbb{R}^{D-1} to eliminate the simplex constraints problem.

Three log-transformation methods are typically employed in CoDa analysis, namely additive log-ratio transformation (ALR), centred log-ratio transformation (CLR), and isometric log-ratio transformation (ILR). For the purposes of the work outlined here, we focused on the CLR and ILR methods, as they are symmetric transformations meaning that distances are preserved. The difference between these two methods relies on the fact that in the former, the identity (covariance) matrix is singular whereas, in the ILR transformation, the covariance matrix is non-singular.

In the soil context, the CLR transformation for the soil particle-size fraction can be defined according to the following equations (Equation 1-3):

$$\text{CLR}_{(1)} = \ln \frac{\text{sand}}{\sqrt[3]{\text{sand} \times \text{silt} \times \text{clay}}} \quad (1)$$

$$\text{CLR}_{(2)} = \ln \frac{\text{silt}}{\sqrt[3]{\text{sand} \times \text{silt} \times \text{clay}}} \quad (2)$$

$$\text{CLR}_{(3)} = \ln \frac{\text{clay}}{\sqrt[3]{\text{sand} \times \text{silt} \times \text{clay}}} \quad (3)$$

From Equations 1-3, it can be noted that the geometric mean composed of all compositions of soil is the denominator, and the CLR is a one-to-one transformation. The CLR transformation maps S^D into the subspace U of \mathbb{R}^D , that is, the CLR-transformed vector has D components but belongs to a $D-1$ dimensional subspace. This method is a one-to-one transformation from S^D to U and the inverse transformation $\text{CLR}^{-1} : U \Rightarrow S^D$ is as follows (Equation 4)

$$\text{CLR}^{-1} \mathbf{z} = \mathcal{C}[\exp z_1; \dots; \exp z_D] \quad (4)$$

where \mathcal{C} denotes the closure or constraint operation of compositional data in its *closed* form, that is, as positive vectors whose parts add up to a positive constant value (equal to 100 in this study).

The ILR transformation method is a representation of compositions by orthonormal log-ratio coordinates or orthonormal log-ratio basis (OLR) and it is also a one-to-one transformation from S^D to \mathbb{R}^{D-1} . The identity matrix of dimension $D-1$ is non-singular and plays a decisive role in relating a composition to its OLR-coordinates. Under an orthonormal log-ratio basis, a composition \mathbf{x} in S^D can be expressed as a compositional linear combination of the vector forming the basis. In other words, these $D-1$ independent vectors in S^D are unitary and mutually orthogonal and the compositional operations are reduced to ordinary vector operations when compositions are represented by their olr-coordinates. This is more appropriate for working with

compositional data (CoDa), as the techniques designed for real data hold their coordinates. This means that linear regression models can be built, as such geometric structure dovetails with that of real space \mathbb{R}^{D-1} .

For the ILR transformation applied to soil texture, the OLR-coordinates of the soil compositions relative to the orthonormal basis of S^D have to be calculated in terms of partitions, which are presented in a sign matrix (Table 1). Partition is a selection of compositions to be included in the bases. It is a hierarchical grouping of parts of the original compositional vector, starting with the whole composition as a group and ending with each part in a single group (Comas-Cufí & Thió-Henestrosa, 2011). One way to define them is by the sequential binary partition (SBP) (Pawlowsky-Glahn *et al.*, 2015; Egozcue & Pawlowsky-Glahn, 2006). SBP consists of dividing the composition into two groups of parts which are indicated by +1 and -1 (numerator and denominator, respectively) whereing the zeros correspond to parts not included in the partition. With this procedure, data is transformed from the simplex (raw data) to real space (orthonormal coordinates-OLR) with the isometric logratio transformation (ILR) defined by a sequential binary partition.

Table 1. Partition $D = 3^*$

i^{**}	Sand	Silt	Clay	ILR transformation***
1	+1	+1	-1	$x_1^* = \sqrt{\frac{2}{3}} \ln \frac{(sand \cdot silt)^{1/2}}{clay}$
2	+1	-1	0	$x_2^* = \sqrt{\frac{1}{2}} \ln \frac{sand}{silt}$

* Three-dimensional compositional data (sand, silt, clay) or 3-part composition. Default partition built in the CoDaPack software (Comas-Cufí & Thió-Henestrosa, 2011). ** i refers to the i -th OLR-coordinate. *** In OLR-coordinates (balance coordinates).

The inverse transformation $olr^{-1} : \mathbb{R}^D \Rightarrow S^D$ is obtained according to Equation 5:

$$olr^{-1} v = \mathcal{C}(\exp(v \cdot \phi)) \quad (5)$$

where v is the olr x vector, ϕ is the matrix of $(D - 1) \times D$ -matrix and \mathcal{C} denotes the closure or constraint operation of compositional data in its *closed* form.

Thus, applying the log-ratio (linear) transformations, it is possible to use standard unconstrained multivariate analysis on the transformed data because of (i) the one-to-one nature of this transformation, (ii) transferring any inferences back to the simplex, and (iii) the components of the composition. The vector space structure of S^D allows the concepts of linear dependence and independence to be used (i.e. compositional-linear dependence).

In traditional Linear Regression Models (LRM), the most popular fitting method is the least-squares deviation criterion, or ordinary least-squares regression (OLS), which minimizes the sum of squared errors. The (simple linear) regression model is shown in Equation 6,

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad (6)$$

where ε_i are independent and identically distributed as $\text{Normal}(0, \sigma^2)$. Equation 7 presents the estimation of the coefficients $\beta_0, \beta_1, \dots, \beta_r$ of a linear surface into S^D , in CoDa analysis.

$$\hat{x}(t) = \beta_0 \oplus (t_1 \odot \beta_1) \oplus \dots \oplus (t_r \odot \beta_r) = \bigoplus_{j=0}^r (t_j \odot \beta_j), \quad (7)$$

where $t = [t_0, t_1, \dots, t_r]$ are real covariates and are identified as the parameters of the linear surface and $\hat{x}(t)$ is the expected value of the CoDa-response variable. Perturbation (\oplus) and Power (\odot) are the fundamental and the subsidiary operations in the simplex, respectively. They are used in statistics notations such as in linear models equations written specifically for CoDa – the simplicial regression. The former is analogous to translation in real coordinate space or Euclidian space (\mathbb{R}^D) and is used to find a difference between compositions or change from one to another. The latter is the simplicial operation analogous to scalar multiplication in \mathbb{R}^D .

The compositional coefficients of the model, $\beta_j \in S^D$ are to be estimated from the data. Because this model is presented as a least-squares problem in the simplex, it could be formulated in terms of orthonormal log-ratio coordinates (olr). Morais and Thomas-Agnan (2021) synthesised the model formulations in coordinate space wherein the parameters a^* , b^* or B^* are estimated by maximum likelihood in coordinate space for the classical regression (Table 2).

Table 2. Specifications of the compositional models and notations.

Space	Y-compositional model
S^D	$Y_t = a \oplus X_t \odot b \oplus \varepsilon_t$
\mathbb{R}^{D-1}	$Y_t^* = a^* + b^* X_t + \varepsilon_t^*$

$$Y_t, a, b, \varepsilon_t \in S^D_Y, X_t \in \mathbb{R}; Y_t^*, a^*, b^*, \varepsilon_t^* \in \mathbb{R}^{D_Y-1}$$

Source: Adapted from Morais and Thomas-Agnan (2021).

Elastic-net Regression Model ***(non-CoDa response variables, univariate responses)***

The elastic-net penalty (Zou & Hastie, 2005) is a compromise between the ridge-regression penalty ($\alpha = 0$) and the lasso penalty ($\alpha = 1$). This penalty is particularly useful in situation where there are many correlated predictor variables. Ridge regression is known to shrink the coefficients of correlated predictors towards each other; Lasso, on the other hand, is somewhat indifferent to very correlated predictors, and will tend to pick one and ignore the rest (Friedman *et al*, 2010). The Lasso expects many coefficients to be close to zero, and a small subset to be larger and non-zero (Friedman *et al*, 2010).

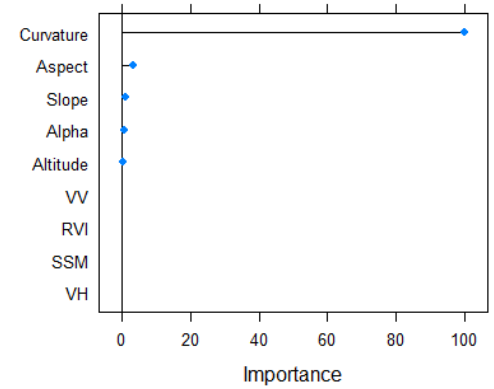
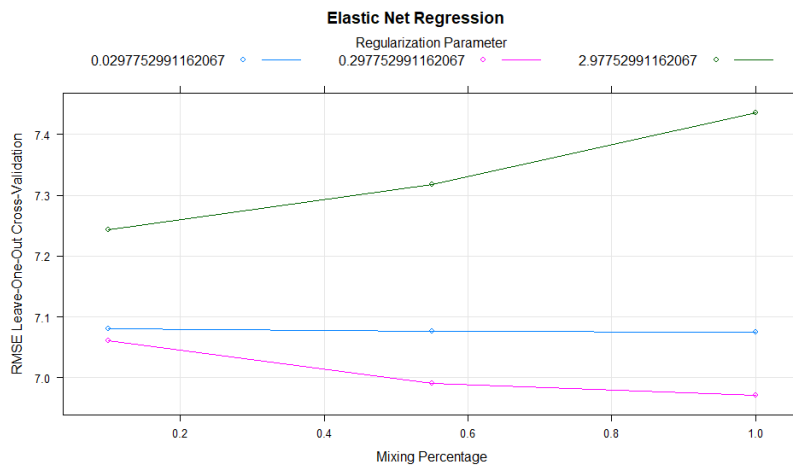
Table: Evaluation metrics for the compositional Random Forest models validated with LOOCV. Approach 1: SAR + Topography covariates and Approach 2: SAR + Topography + Geophysical covariates

Soil PSF*	Evaluation Metrics – Approach 1		
	RMSE	MAE	NSE**
Sand	7.141	5.63	0.81
Silt	7.896	6.113	0.53
Clay	7.309	5.341	0.42
Soil PSF*	Evaluation Metrics – Approach 2		
	RMSE	MAE	NSE**
Sand	14.89	11.65	0.14
Silt	11.22	8.83	0.018
Clay	9.20	6.84	0.11

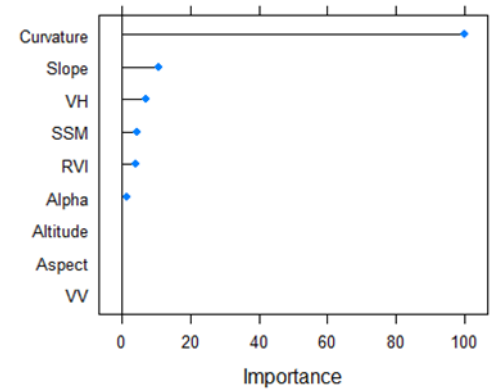
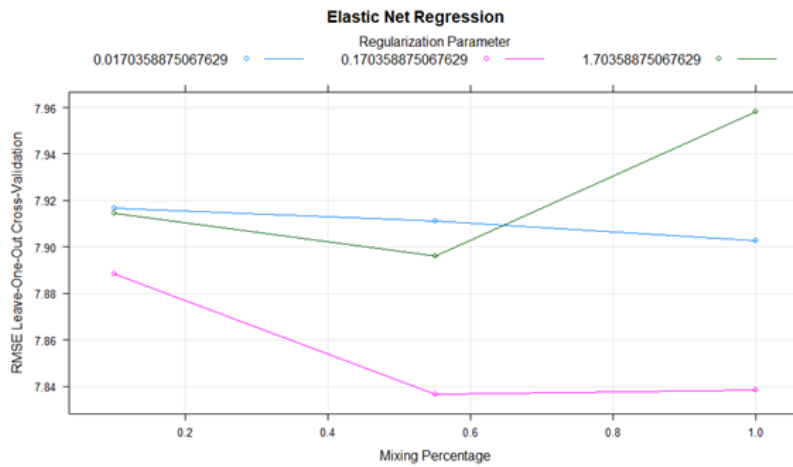
*PSF = particle size fraction; **NSE = Nash-Sutcliffe model Efficiency coefficient (1 : 1 line)

➤ Without geophysical data (Model approach 1)

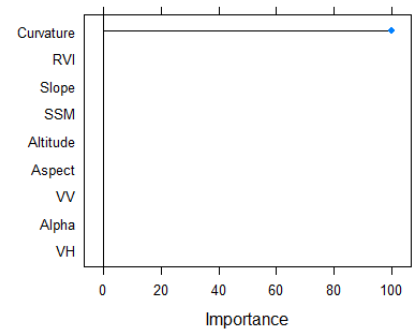
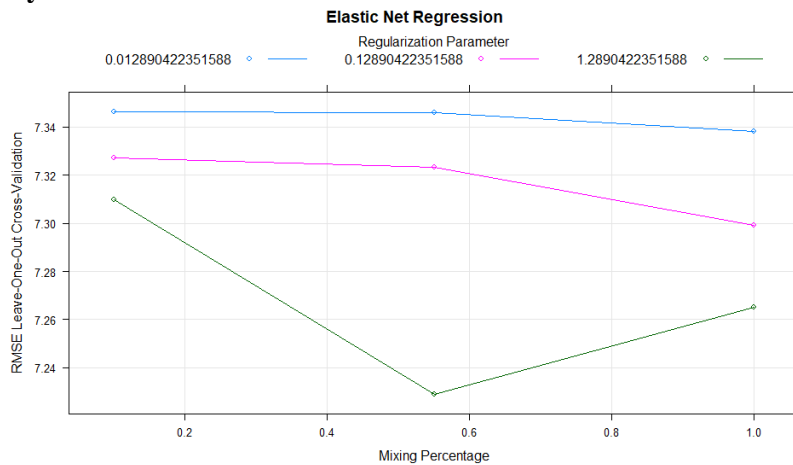
Sand



Silt



Clay



Model – SAND

Train control: method = "LOOCV", number = 10, verboseIter = TRUE

Aggregating results:

Selecting tuning parameters

Fitting alpha = 1, lambda = 0.298 on full training set (234 samples), en method: glmnet

Resampling: LOOCV (9 fold)

* RMSE was used to select the optimal model using the smallest value. The final values used for the model were alpha = 1 and lambda = 0.297753

RMSE score: 7.435291 (training)

Model – SILT

Train control: method = "LOOCV", number = 10, verboseIter = TRUE

Aggregating results:

Selecting tuning parameters

Fitting $\alpha = 0.55$, $\lambda = 0.17$ on full training set (234 samples), en method: glmnet

Resampling: LOOCV (9 fold)

* RMSE was used to select the optimal model using the smallest value. The final values used for the model were $\alpha = 1$ and $\lambda = 0.1703589$

RMSE score: 7.958112 (training)

Model – CLAY

Train control: method = "LOOCV", number = 10, verboseIter = TRUE

Aggregating results:

Selecting tuning parameters

Fitting $\alpha = 0.55$, $\lambda = 1.29$ on full training set (234 samples), en method: glmnet

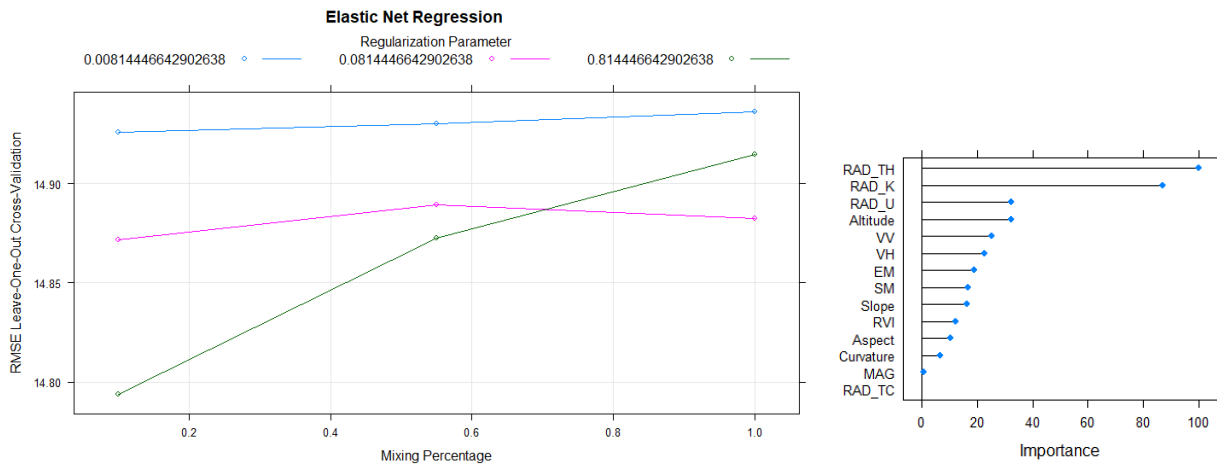
Resampling: LOOCV (9 fold)

* RMSE was used to select the optimal model using the smallest value. The final values used for the model were $\alpha = 0.55$ and $\lambda = 1.289042$

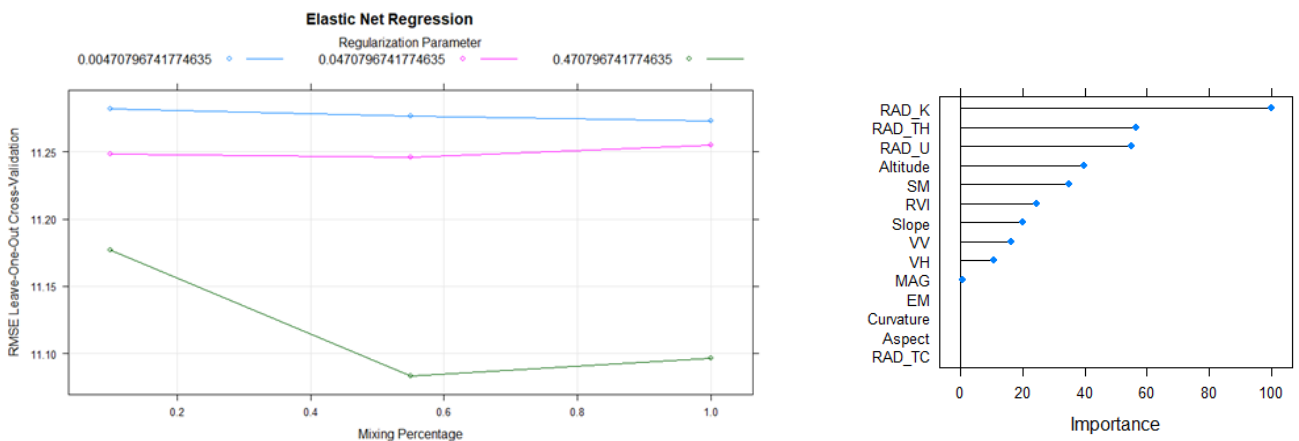
RMSE score: 7.265178 (training)

➤ With geophysical data (Model approach 2)

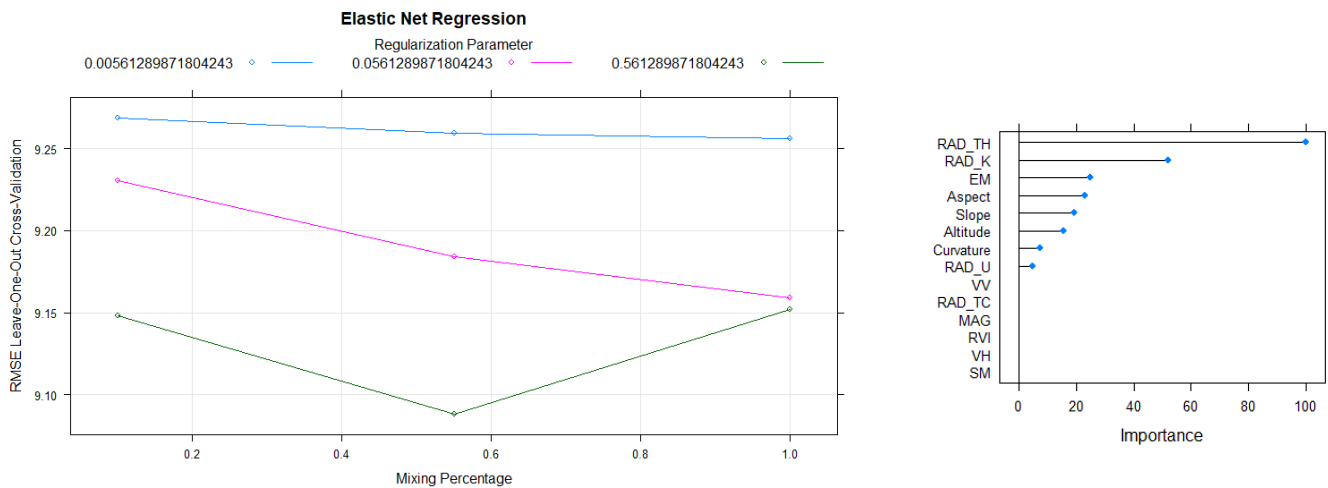
Sand



Silt



Clay



Model – SAND

Train control: method = "LOOCV", number = 10, verboseIter = TRUE

Aggregating results:

Selecting tuning parameters

Fitting alpha = 0.1, lambda = 0.814 on full training set (234 samples), en method: glmnet

Resampling: LOOCV (9 fold)

RMSE was used to select the optimal model using the smallest value. The final values used for the model were alpha = 0.1 and lambda = 0.8144466.

RMSE score: 15.38857 (training)

Model – SILT

Train control: method = "LOOCV", number = 10, verboseIter = TRUE

Aggregating results:

Selecting tuning parameters

Fitting alpha = 0.55, lambda = 0.471 on full training set (234 samples), en method: glmnet

Resampling: LOOCV (9 fold)

* RMSE was used to select the optimal model using the smallest value. The final values used for the model were alpha = 0.55 and lambda = 0.4707967

RMSE score: 11.09616 (training)

Model – CLAY

Train control: method = "LOOCV", number = 10, verboseIter = TRUE

Aggregating results:

Selecting tuning parameters

Fitting alpha = 0.55, lambda = 0.561 on full training set (234 samples), en method: glmnet

Resampling: LOOCV (9 fold)

* RMSE was used to select the optimal model using the smallest value. The final values used for the model were alpha = 0.55 and lambda = 0.5612899

RMSE score: 9.151947 (training)

Glmnet regression model

(CoDa response variables, multivariate responses)

Glmnet fits generalized linear regression models via penalized maximum likelihood. The regularization path is computed for the lasso or elastic net penalty at a grid of values (on the log scale) for the regularization parameter lambda. It can also fit multi-response linear regression, which is the case of this study. Lasso tend to pick one predictor and ignore the rest, and expects many coefficients to be close to zero, and a small subset to be larger and nonzero.

➤ Model fitted (with geophysical data): ilr.1

Labels of the Lasso's penalties curves

Curve	Predictor	Curve	Predictor
1	VV	8	Radiometric-Thorium (RAD_TH)
2	VH	9	Radiometric-Uranium (RAD_U)
3	RVI	10	Altitude
4	Surface soil moisture (SSM)	11	Slope
5	Electromagnetic (EM)	12	Aspect
6	Magnetic (MAG)	13	Curvature
7	Radiometric-Potassium (RAD_K)	-	-

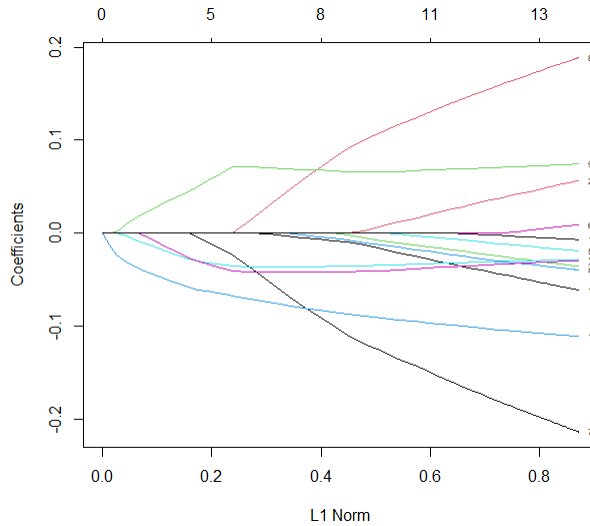


Figure: Lasso's penalties coefficients
ilr.1 model

➤ **Model fitted (with geophysical data): ilr.2**

Labels of the Lasso's penalties curves

Curve	Predictor	Curve	Predictor
1	VV	8	Radiometric-Thorium (RAD_TH)
2	VH	9	Radiometric-Uranium (RAD_U)
3	RVI	10	Altitude
4	Surface soil moisture (SSM)	11	Slope
5	Electromagnetic (EM)	12	Aspect
6	Magnetic (MAG)	13	Curvature
7	Radiometric-Potassium (RAD_K)	-	-

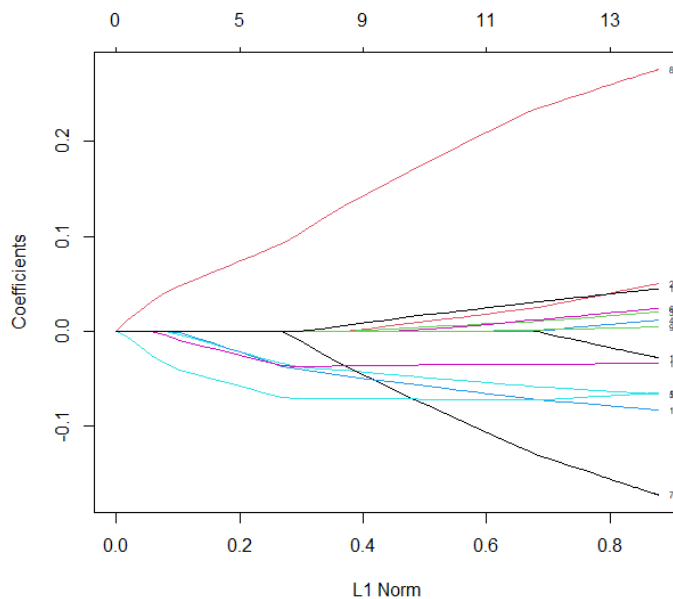


Figure: Lasso's penalties coefficients
ilr.2 model

REFERENCES

- Aitchison, J. (1982). The statistical analysis of compositional data (with discussion). *Journal of the Royal Statistical Society Series B* 44, no. 2, 139-177.
- Comas-Cufí, M., Thió-Henestrosa, S. (2011). CoDaPack 2.0: a stand-alone, multi-platform compositional software. In: Egozcue JJ, Tolosana-Delgado R, Ortego MI, eds. *CoDaWork'11: 4th International Workshop on Compositional Data Analysis*. Sant Feliu de Guíxols; 2011.
- Egozcue, J.J. & Pawlowsky-Glahn, V. (2006). Compositional data in the geosciences. *Geological Society, London, Special Publications: Simplicial geometry for compositional data*, 264, 145-159, Available at: <https://www.lyellcollection.org/doi/abs/10.1144/gsl.sp.2006.264.01.11> (accessed 06 February 2023)
- Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(1), 1–22. <https://doi.org/10.18637/jss.v033.i01>
- Morais & Thomas-Agnan (2021). Impact of Covariates in Compositional Models and Simplicial Derivatives. *Austrian Journal of Statistics*. Vol. 50, 1-15. <https://doi.org/10.17713/ajs.v50i2.1069>
- Pawlowsky-Glahn, V., Egozcue, J. J. & Tolosana-Delgado, R. (2015). *Modeling and Analysis of Compositional Data*, Wiley, 2015, 272 p.
- Zou, H. & Hastie, T (2005). Regularization and variable selection via the elastic net. *J. Royal. Stat. Soc. B.*, 67(2):301–320. Available at: <https://www.jstor.org/stable/3647580> (accessed 22 March 2023)

Appendix B

This supplementary material refers to Chapter 5 (DOI 10.1016/j.jag.103742) and is available at <https://doi.org/10.1016/j.jag.2024.103742>

Appendix B – Figures

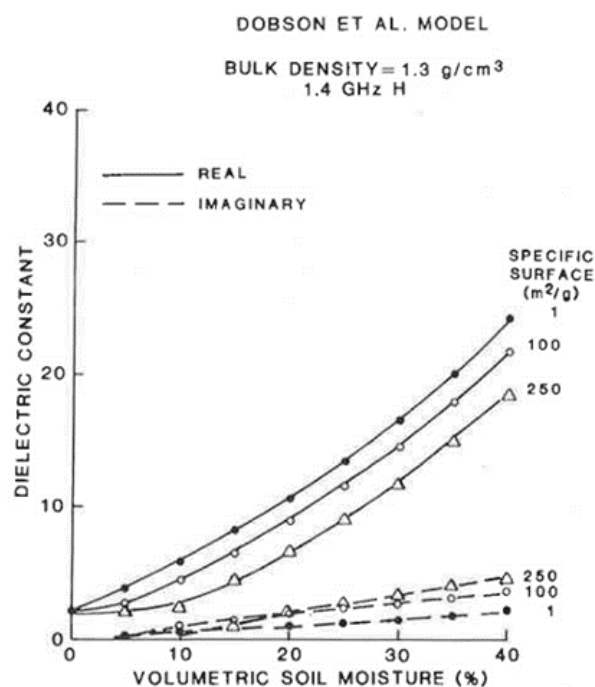


Figure B5.1 Effects of soil specific surface area on the relationship between dielectric constant and volumetric soil moisture. The figure depicts the sensitivity of dielectric constant with change in specific surface area at varied moisture conditions, at a microwave frequency of 1.4 GHz. The figure also indicates that for the same volumetric soil moisture, the real dielectric constant of a sand will be larger than that of a clay because there is more free water in the sand mixture (Jackson, T.J., 1987). The real dielectric constant is a measure of the polarisation of the material or degree of polarisation). The imaginary dielectric constant is a measure of the dielectric losses or energy absorbed. Figure's source: Jackson, T.J. (1987).

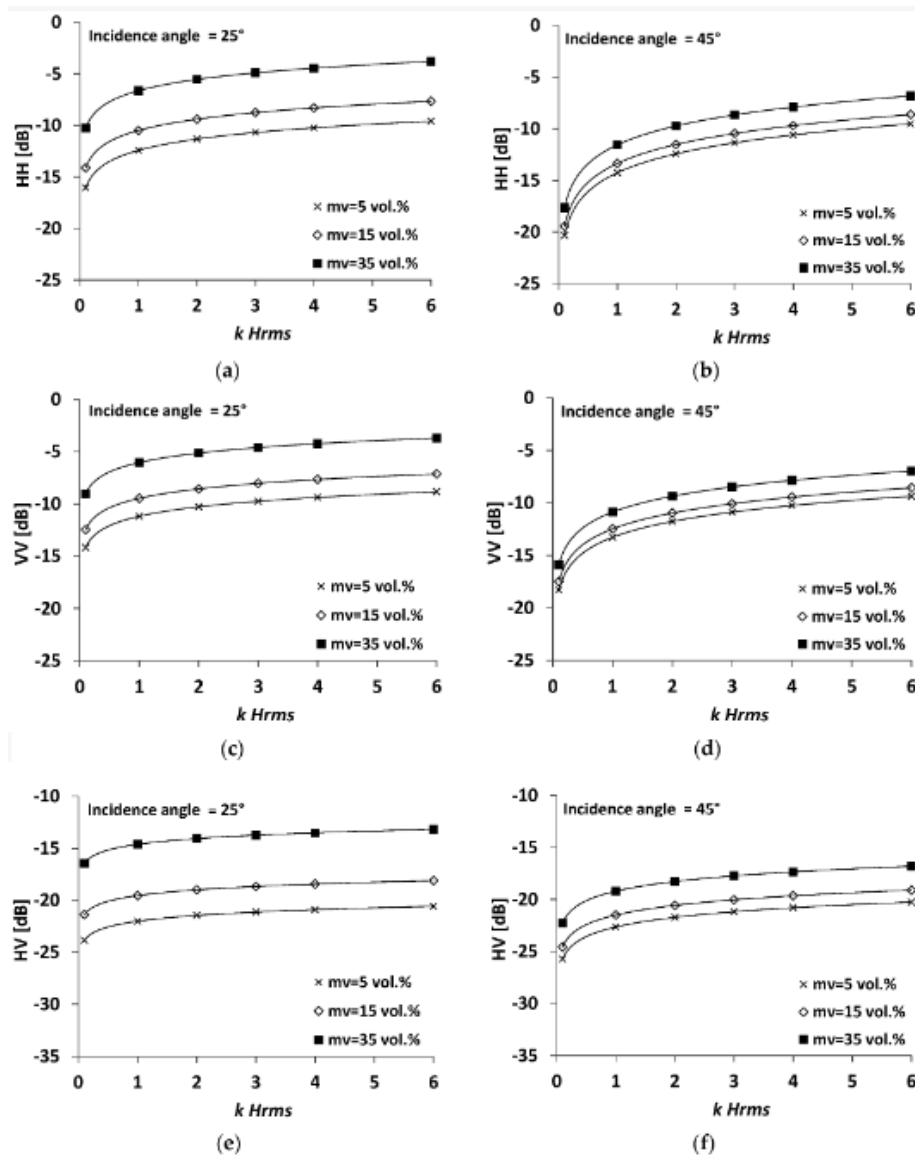


Figure B5.2 Behaviour of incidence angle (θ), surface roughness (k Hrms) and soil moisture (mv) in HH (a and b), VV (c and d), and HV (e and f) polarisations. Source: Baghdadi *et al.* (2016).

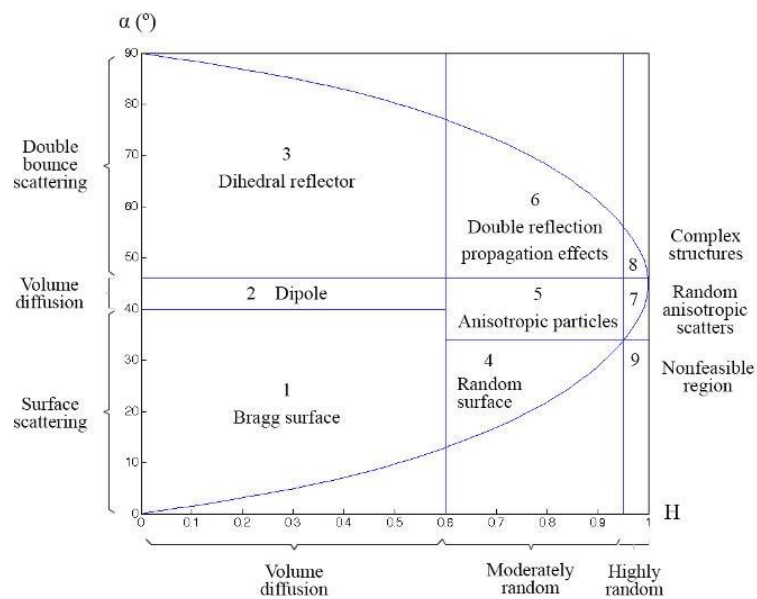
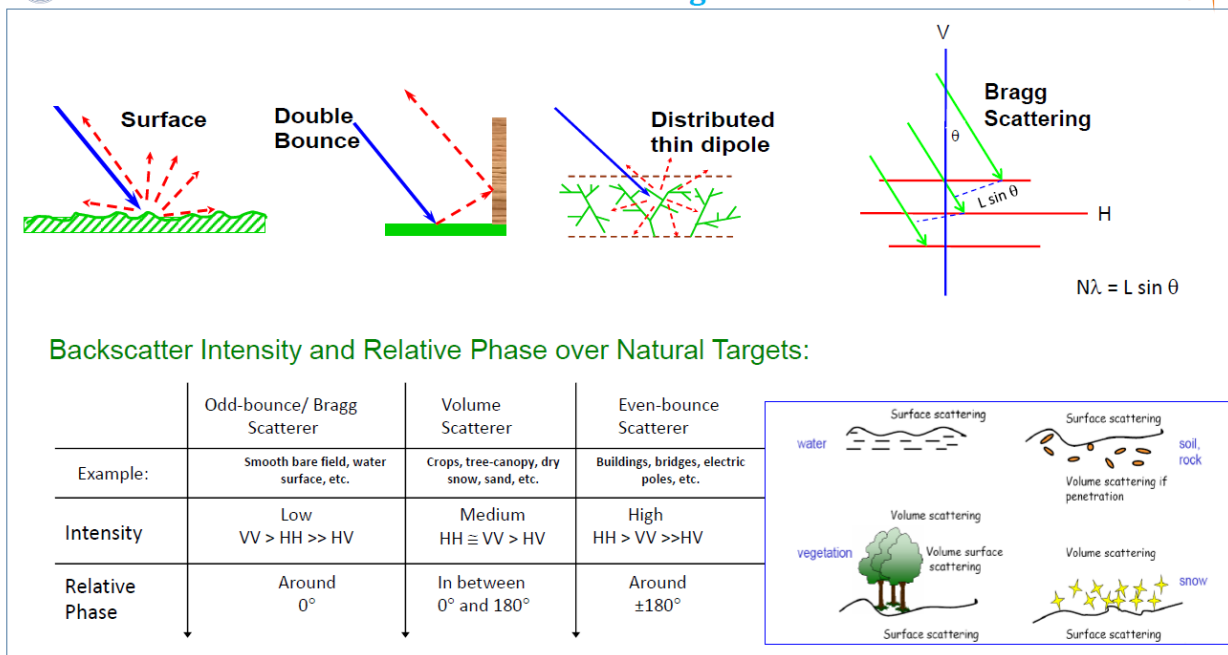
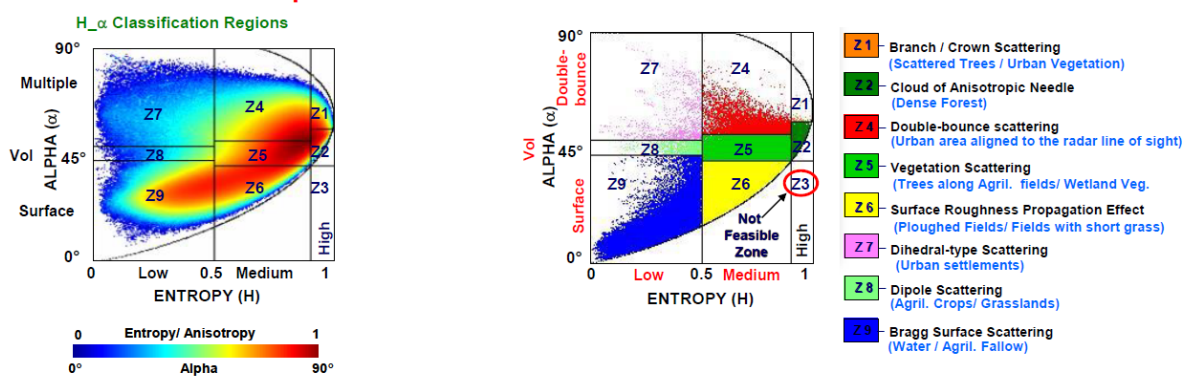


Figure B5.3 An example of the H- α classification plane. Source (Olivé, 2015). Available at <https://ddd.uab.cat/record/133501?ln=en>



anup@sac.isro.gov.in

Cloude & Pottier Decomposition



anup@sac.isro.gov.in

Figure B5.4a Type of scattering mechanisms and an example of H-α plane plot resulting from the Cloude-Pottier decomposition with an unsupervised Wishart H-α classification applied to agricultural fields. This figure provides worthwhile information to understand the backscatter intensity for different scattering mechanisms and polarisations as well as the physical properties behind the scattering mechanisms given by the H-α plane plot in that type of land use. Source: Anup Das (n.d. pp 27,70,77). Available at: [https://vedas.sac.gov.in/vedas/downloads/ertd/SAR/L_5_Polarimetric SAR Data Analysis and Applications Dr Anup Das.pdf](https://vedas.sac.gov.in/vedas/downloads/ertd/SAR/L_5_Polarimetric_SAR_Data_Analysis_and_Applications_Dr_Anup_Das.pdf) Accessed 25/01/2023

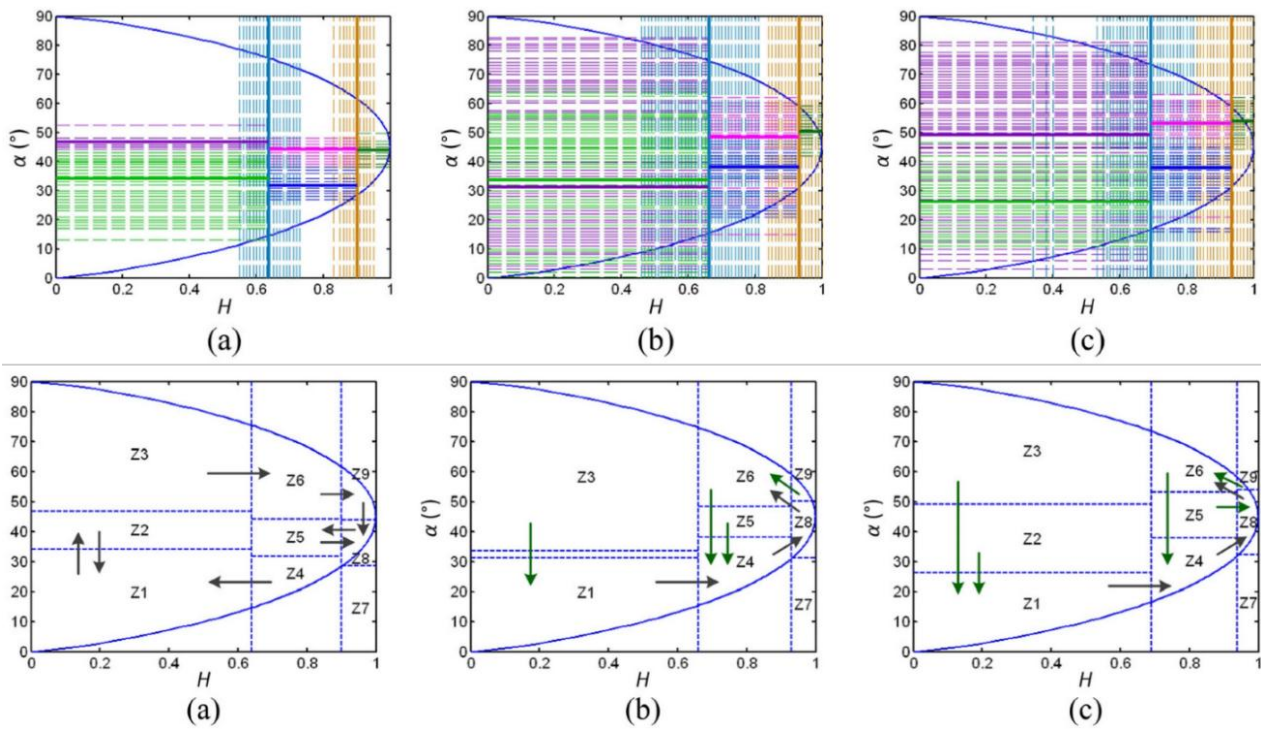


Figure B5.4b The H-alpha plane for dual-pol data exhibiting optimal dividing lines in the dual-polarization H- α plane. (a) HH-VV; (b) HH-HV; (c) HV-VV. (Ji & Wu, 2015). Line 1: divides low and medium entropy zones; Line 2: divides medium and high entropy zones; Line 3: divides Z1 and Z2; Line 4: divides Z2 and Z3; Line 5: divides Z4 and Z5; Line 6: divides Z5 and Z6; Line 7: divides Z8 and Z9. The figure highlight the differences concerning boundary lines and zone thresholds in comparison to Cloude H-alpha plane for full-pol (depicted in Figure B5.4a). Credits and source of the figure: Kefeng Ji & Yonghui Wu (2015), <https://doi.org/10.3390/rs70607447>.

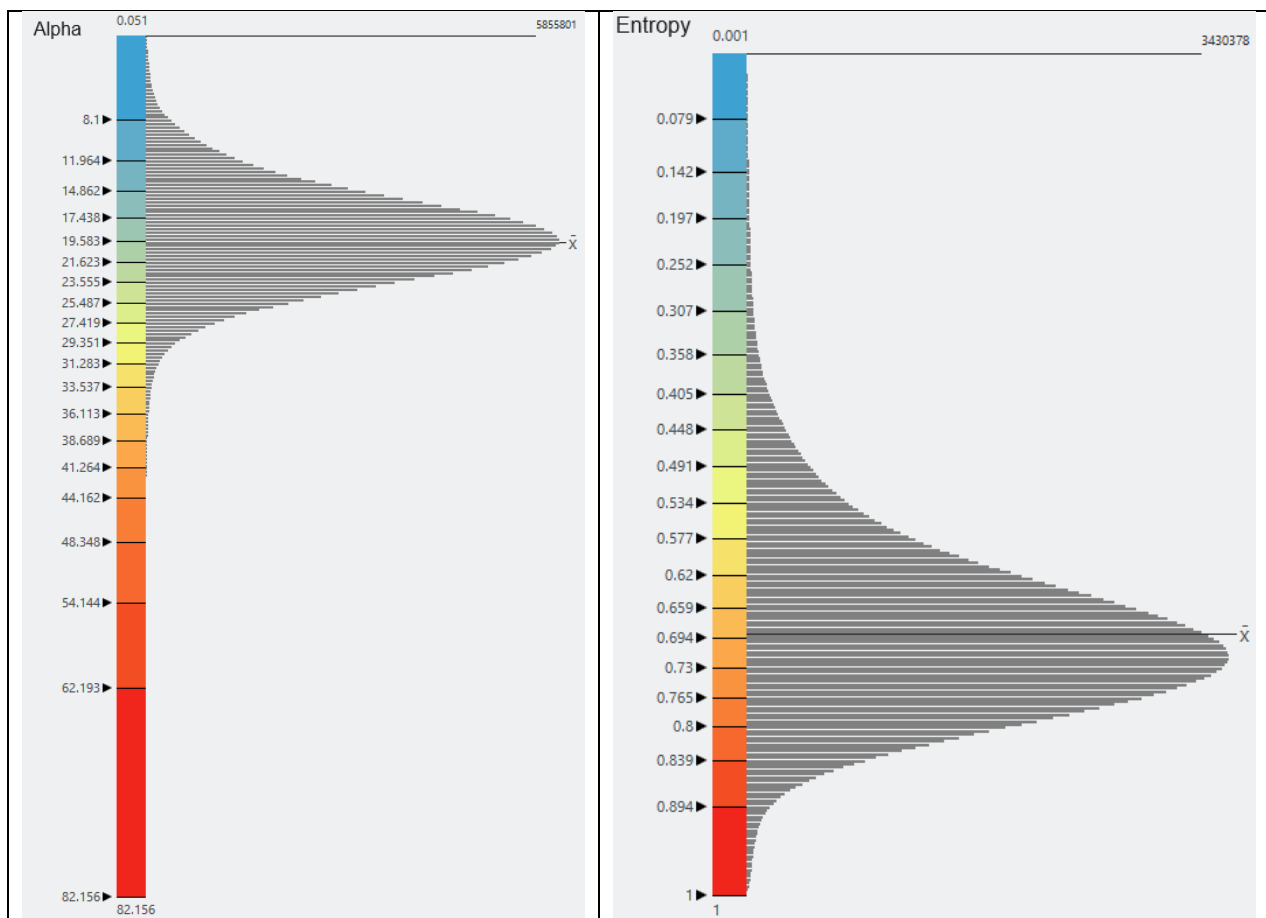


Figure B5.5 Histograms of alpha and entropy values obtained for the whole study area, showing alpha values mostly ranging from 0.051 to approximately 36. Alpha and Entropy parameters reflected the land use/land cover categories within which agriculture is the most occurring land use land cover in Ireland dominated by grass pasture.

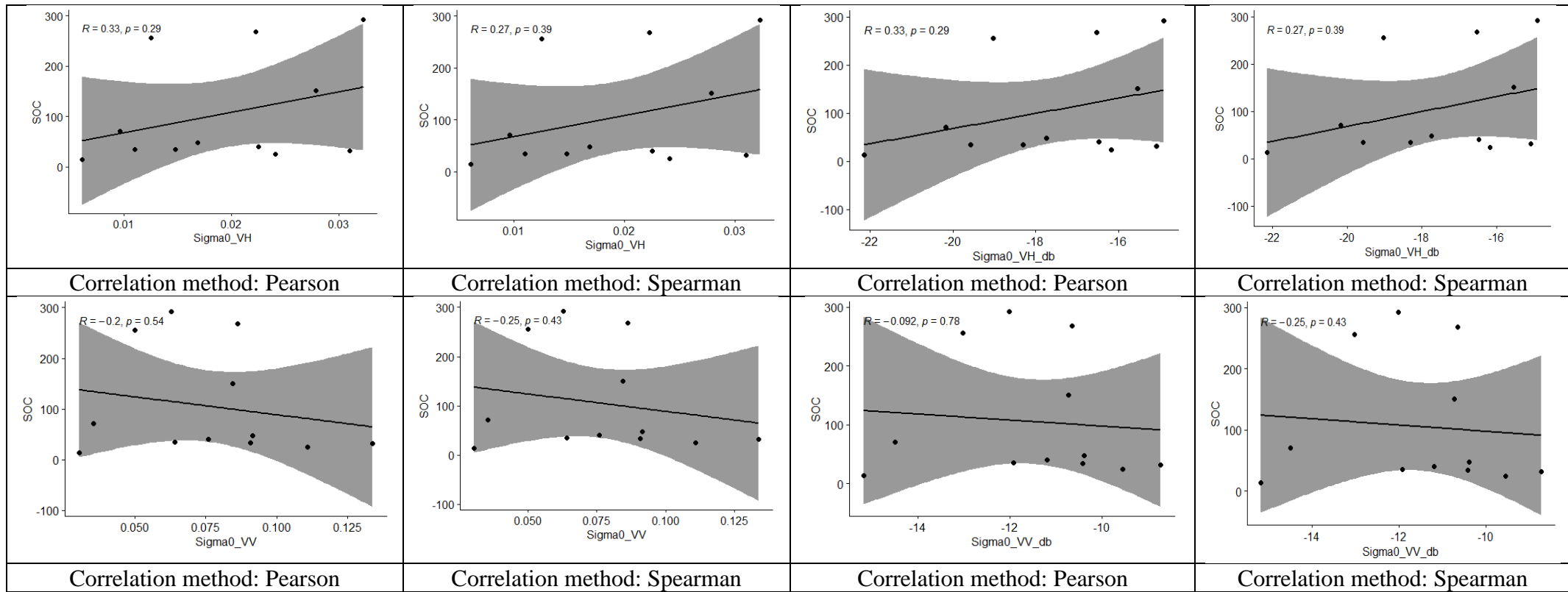


Figure B5.6 Correlation (Pearson and Spearman) plots for SOC and SAR backscattering (sigma null in linear units; sigma null in decibels) for soils samples taken at 0-15 cm. In the y axis, SOC units are g/kg.

Appendix B –Tables

Table B5.1 The dielectric constant for Earth materials

Material	Dielectric Constant
Sand (dry)	3 – 6
Sand (saturated)	20 – 30
Silts	5 – 30
Shales	5 – 15
Clays	5 – 40
Humid soil	30
Cultivated soil	15
Rocky soil	7
Sandy soil (dry)	3
Sandy soil (saturated)	19
Clayey soil (dry)	2
Clayey soil (saturated)	15
Sandstone (saturated)	6
Limestone (dry)	7
Limestone (saturated)	4 – 8
Basalt (saturated)	8
Granite (dry)	5
Granite (saturated)	7

Source: Hubbard *et al.* (1997).

Table B5.2 Summary of the physical scattering characteristics of each of the nine zones in the H- α classification plane. Zone numbers shown in parenthesis refer to corresponding zones in the pol-decomposition scheme depicted in Figure B5.3. Different SAR sensors can yield different arrangements of the zones in the H- α plane but descriptions and interpretations are the same.

Zone	Description
Zone 9(Z1): Low Entropy Surface Scatter	In this zone occur low entropy scattering processes with alpha values less than 42.5° and include bragg surface scattering and specular scattering phenomena. Physical surfaces such as water at L and P-bands, sea ice at L-band, as well as very smooth land surfaces, all fall into this category.
Zone 8(Z2): Low Entropy Dipole Scattering	In this zone occur strongly correlated mechanisms, which have a large imbalance between HH and VV in amplitude. An isolated dipole scatterer would appear, as scattering from vegetation with a strongly correlated orientation of anisotropic scattering elements.
Zone 7(Z3): Low Entropy Multiple Scattering Events	This zone corresponds to low entropy double or even bounce scattering events, such as provided by isolated dielectric and metallic dihedral scatterers. These are characterized by alpha > 47.5°. The lower bound chosen for this zone is dictated by the expected dielectric constant of the dihedrals and by the measurement accuracy of the radar.
Zone 6(Z4): Medium Entropy Surface Scatter	This zone reflects the increase in entropy due to changes in surface roughness and due to canopy propagation effects. In surface scattering theory, the entropy of low-frequency theories like bragg scatter is zero. Likewise, the entropy of high-frequency theories like Geometrical Optics is also zero. Thus as the roughness/correlation length of surface changes, its entropy will increase. Further, a surface cover comprising oblate spheroidal scatterers (leaves or discs for example) will generate an entropy between 0.6 and 0.7.
Zone 5(Z5): Medium Entropy Vegetation Scattering	There is moderate entropy in this zone but with a dominant dipole-type scattering mechanism. The increased entropy is due to a central statistical distribution of orientation angle. Such a zone would include scattering from vegetated surfaces with anisotropic scatterers and a moderate correlation of scatterer orientations.

Zone 4(Z6): Medium Entropy Multiple Scattering	This zone accounts for dihedral scattering with moderate entropy. This occurs for example in forestry applications, where double bounce mechanisms occur at P and L bands following propagation through a canopy. The effect of the canopy is to increase the entropy of the scattering process. A second important process in this category is in urban areas, where dense packing of localized scattering centres can generate moderate entropy with low-order multiple scattering dominant.
Zone 3(Z9): High Entropy Surface Scatter	This class is not part of the feasible region in H- α space i.e., we cannot distinguish surface scattering with entropy $H > 0.9$. This is a direct consequence of our increasing inability to classify scattering types with increasing entropy. It is included to reinforce the idea that increasing entropy does limit our ability to use polarimetric behaviour to classify targets. Thus, radar polarimetry will be most successfully applied to low entropy problems.
Zone 2(Z7): High Entropy Vegetation Scattering	High entropy volume scattering arises when $\alpha = 45^\circ$ and $H = 0.95$. Scattering from forest canopies lies in this region, as does the scattering from some types of vegetated surfaces with random highly anisotropic scattering elements. The extreme behaviour in this class is random noise (i.e., no polarization dependence).
Zone 1(Z8): High Entropy Multiple Scattering	In the region $H > 0.9$, it is still possible to distinguish double bounce mechanisms in a high entropy environment. Again, such mechanisms can be observed in forestry applications or in scattering from vegetation which has a well-developed branch and crown structure.

Source: Cloude & Pottier (1997).

Table B5.3 Approximate Microwave Measurement Depths for Bare Soil and Different Land Surface Covers

Land surface cover	X-band (2.5 – 3.75 cm) (8 – 12 GHz)	C-band (3.75 – 7.5 cm) (4 – 8 GHz)	L-band (15 – 30 cm) (1 – 2 GHz)	P-band (30 – 100 cm) (0.3 – 1 GHz)
Bare soil	~1.25–1.87 cm	~1.87–3.75 cm	~7.5–15 cm	~15–50 cm
Agriculture and pasture	~0.5–0.75 cm	~0.75–1.5 cm	~3–6 cm	~6–20 cm
Forest	~0.25–0.37 cm	~0.37–0.75 cm	~1.5–3 cm	~3–10 cm

Source: Babaeian *et al.* (2019).

Table B5.4 Correlation (Pearson and Spearman) for SOC and SAR backscattering for soils samples taken at 0-15 cm and $\leq 0-6$ cm.

Backscattering (0-15cm); n=12	Correlation method: Pearson	Correlation method: Spearman
Sigma0_VH	0.3331781	0.2727273
Sigma0_VH_db	0.3339002	0.2727273
Sigma0_VV	-0.1988311	-0.2517483
Sigma0_VV_db	-0.09185267	-0.2517483
Backscattering (0-3cm, 0-4cm, 0-6cm); n=3	Correlation method: Pearson	Correlation method: Spearman
Sigma0_VH	0.8677421	1
Sigma0_VH_db	0.9177515	-0.5
Sigma0_VV	0.3797253	1
Sigma0_VV_db	0.3720139	-0.5
Backscattering (0-15cm); n=3	Correlation method: Pearson	Correlation method: Spearman
Sigma0_VH	0.5570259	0.5
Sigma0_VH_db	0.6961182	0.5
Sigma0_VV	0.85448	0.5
Sigma0_VV_db	0.9306026	0.5

REFERENCES

- Babaeian, E., Sadeghi, M., Jones, S. B., Montzka, C., Vereecken, H. & Tuller, M. (2019). Ground, proximal, and satellite remote sensing of soil moisture. *Reviews of Geophysics*, 57. <https://doi.org/10.1029/2018RG000618>
- Baghdadi, N., Choker, M., Zribi, M., Hajj, M.E., Paloscia, S., Verhoest, N.E., Lievens, H., Baup, F. and Mattia, F. (2016). A new empirical model for radar scattering from bare soil surfaces. *Remote Sensing*, 8(11), p.920. <https://doi.org/10.3390/rs8110920>
- Hubbard, S. S., J. E. Peterson, Jr., E. L. Majer, P. T. Zawislanski, K. H. Williams, J. Roberts, & F. Wobber (1997). Estimation of permeable pathways and water content using tomographic radar data. *The Leading Edge*, 16, 1623-1628. Available at <http://www.geo.umass.edu/faculty/wclement/dielec.html> (accessed 25 October 2023)
- Jackson, Thomas. J. (1987). Effects of soil properties on microwave dielectric constants. *Transportation Research Board*, 1119, 126-131.
- Ji, K., Wu, Y. (2015). Scattering Mechanism Extraction by a Modified Cloude-Pottier Decomposition for Dual Polarization SAR.. *Remote Sens.* 2015, 7(6), 7447-7470; <https://doi.org/10.3390/rs70607447>

Appendix C

This supplementary material refers to Chapter 6.

Appendix C – Figures

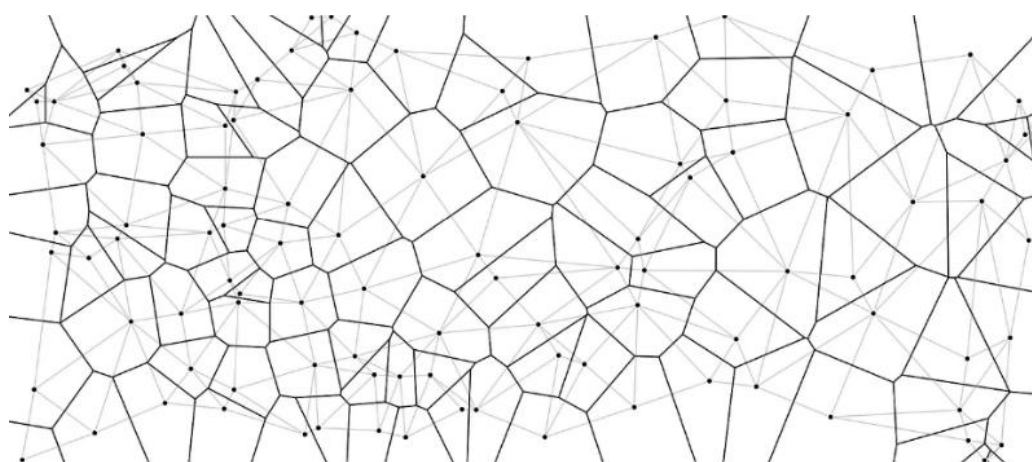


Figure C6.1 Voronoi diagram (black) with Delaunay triangulation (grey). Source: Bellelli (2022). Available at: <https://towardsdatascience.com/the-fascinating-world-of-voronoi-diagrams-da8fc700fa1b>. Accessed 17 Jul 2023.

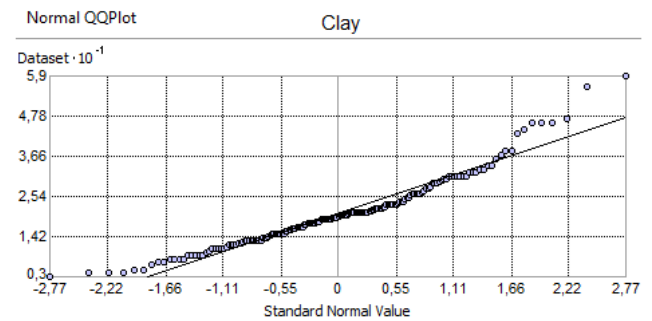
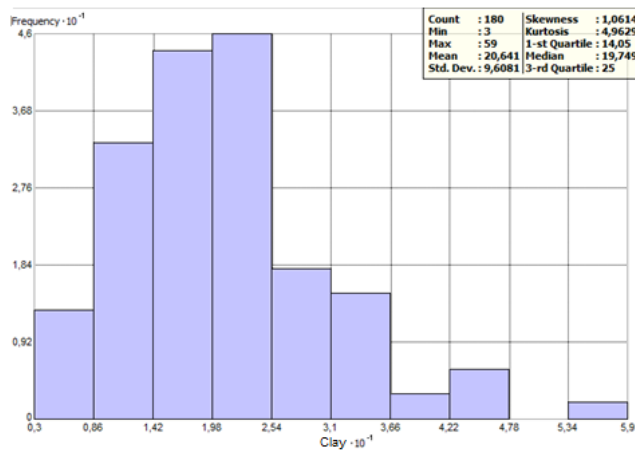
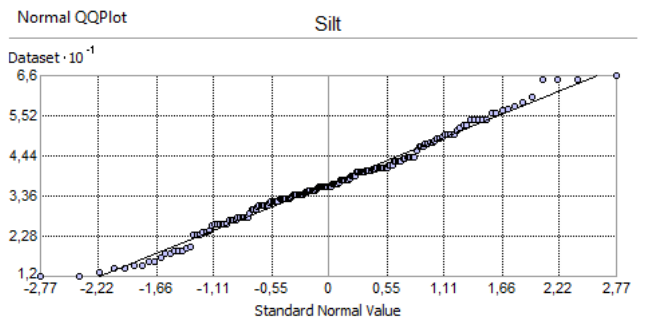
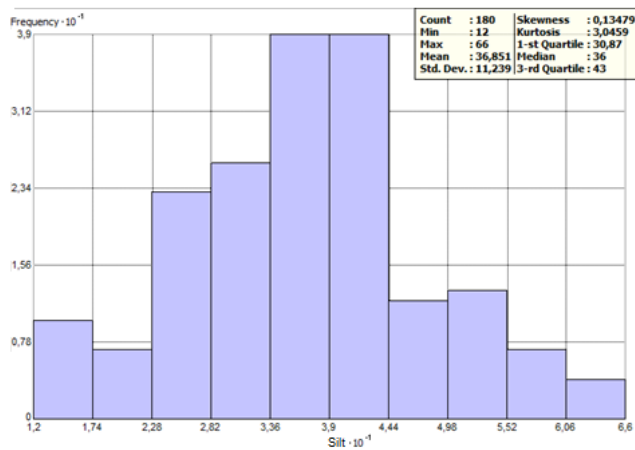
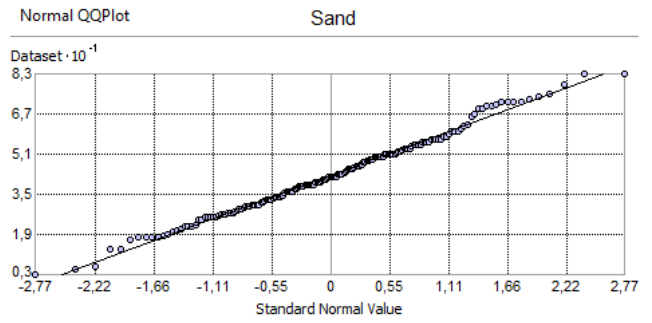
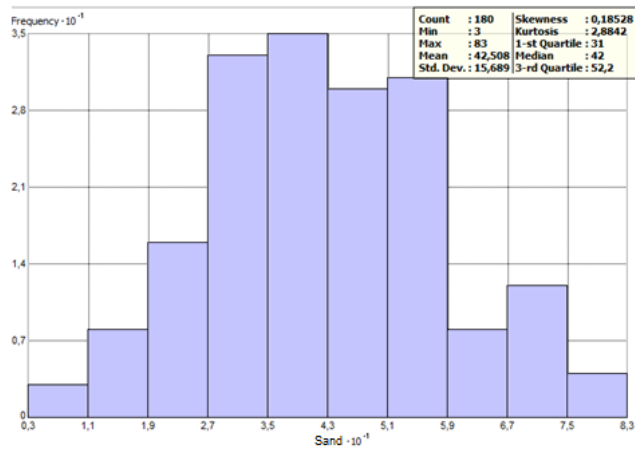


Figure C6.2 Histograms and normal quantile-quantile plots of the response variables

Trend analysis

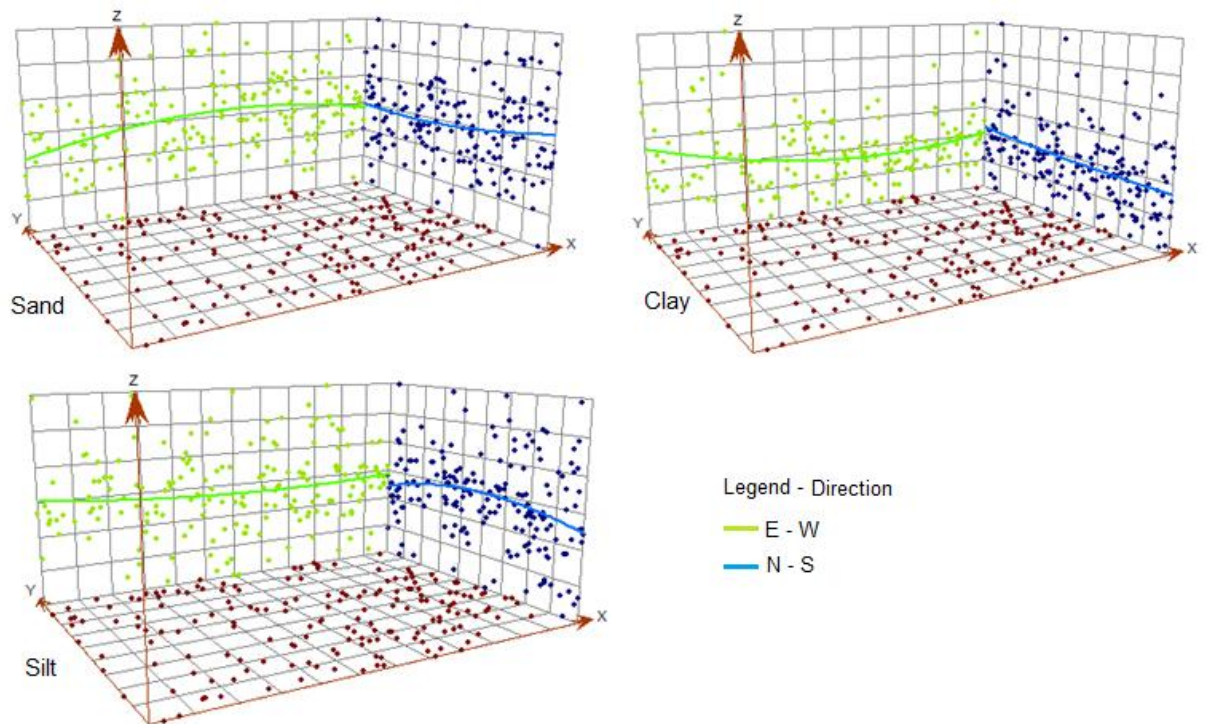


Figure C6.3 Trend analysis plots of response variables (measured data), wherein points are raised and three-dimensionally projected in two directions (north and west). The green line shows a trend in the east-west direction, while the blue line depicts a trend in the north-south direction. Flat curves mean that no trend exists. In the sand plot, curves indicate a slight trend to the south and east, with higher values in these directions since the curves are bent toward them. Conversely, for clay, higher values are towards the west. In the silt plot, higher values are in the centre, however, there is not enough strong influence toward the borders, as the curve is not very sharp. Typically, trends indicate systematic changes in data across an entire study area.

STATIONARITY

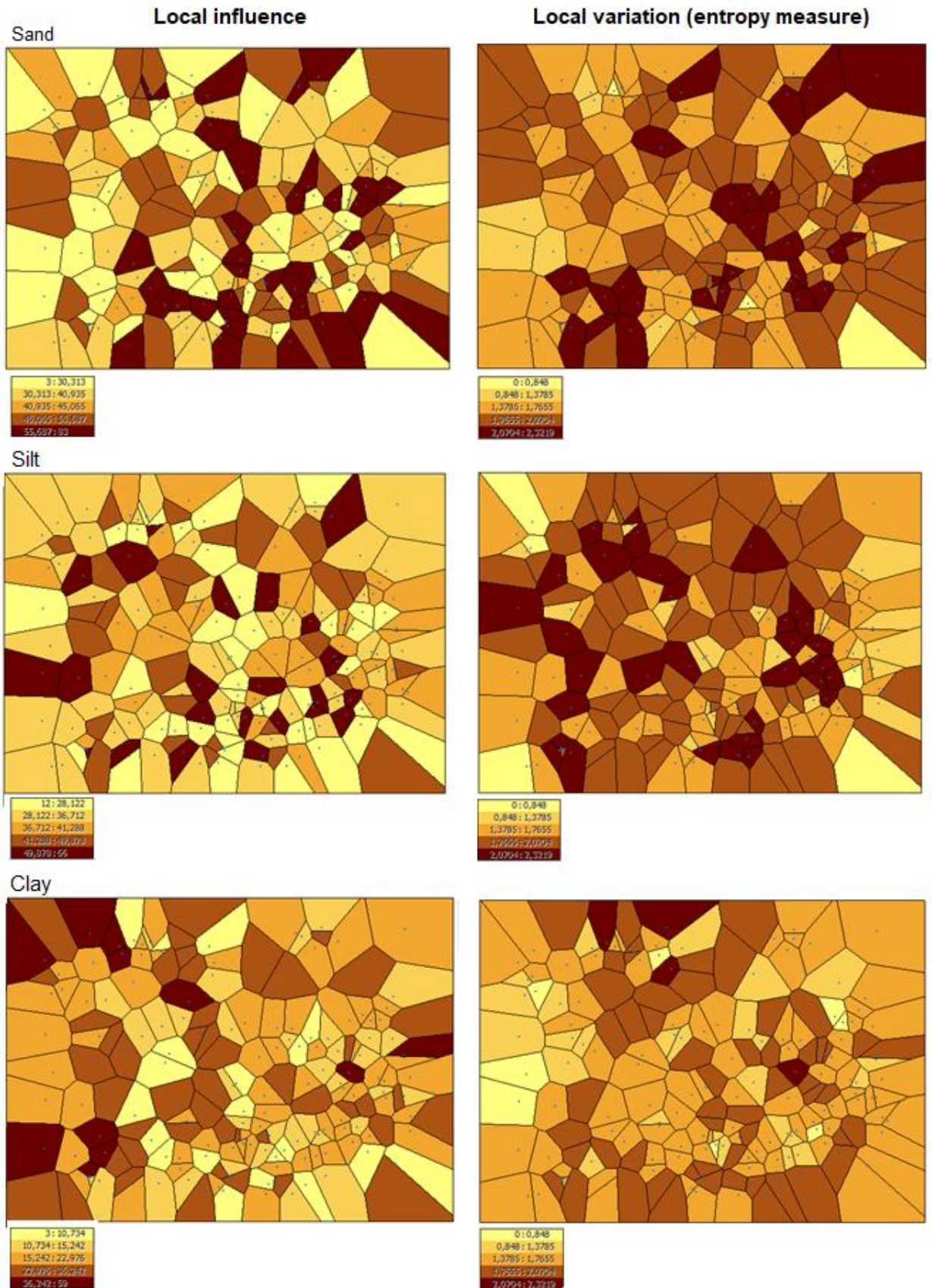


Figure C6.4 Voronoi diagrams depicting stationarity analysis of response variables according to local influence and local variation (entropy).

SPATIAL AUTOCORRELATION BY DISTANCE

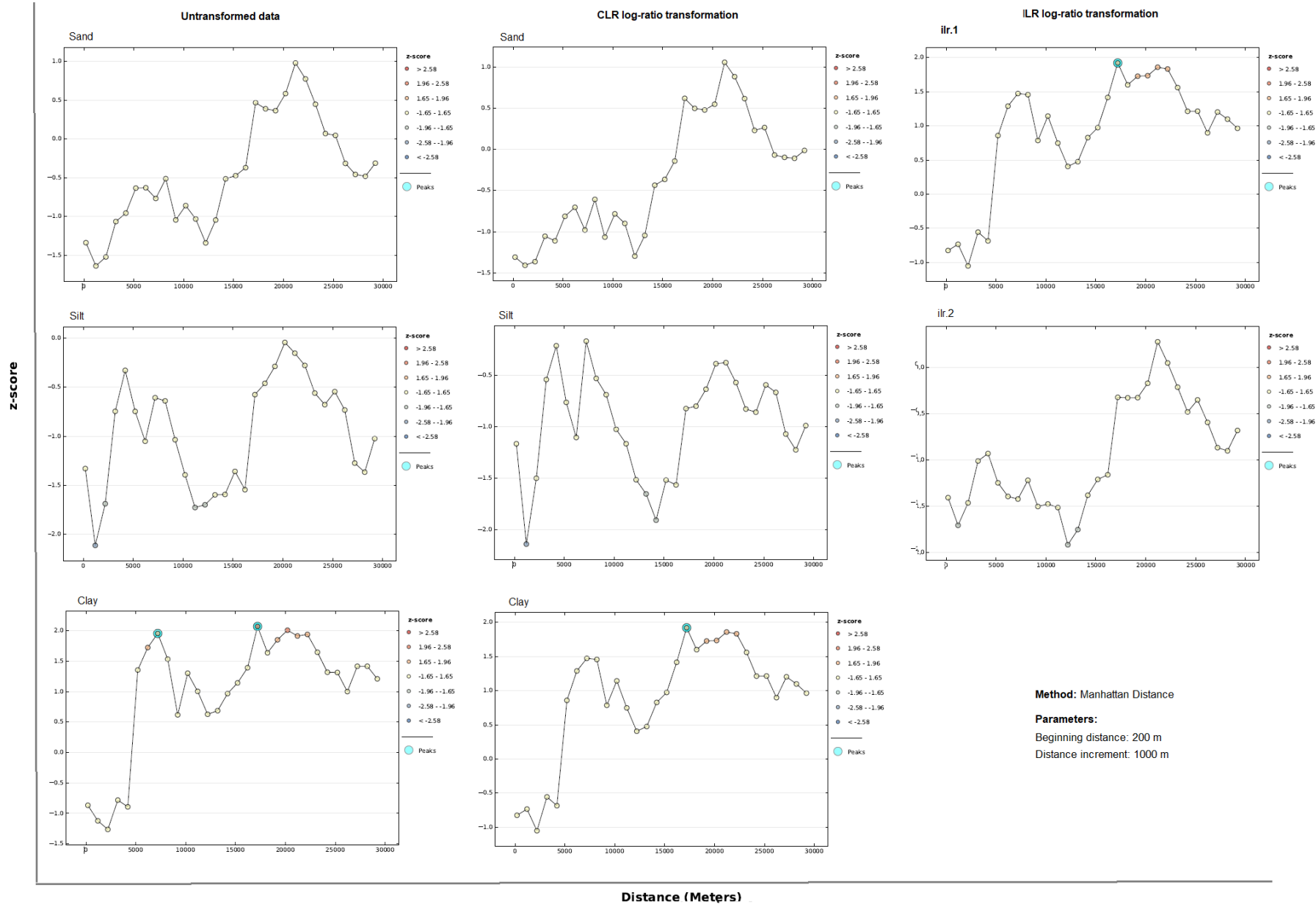


Figure C6.5 Spatial autocorrelation by distance using incremental (regular) distance, measuring the intensity of spatial clustering for each increasing distance. Z-scores determine the intensity of clustering. Peaks reflect distances where the spatial processes for clustering are most pronounced. The colour of each point on the graph corresponds to the statistical significance of the z-score values.

EBK REGRESSION

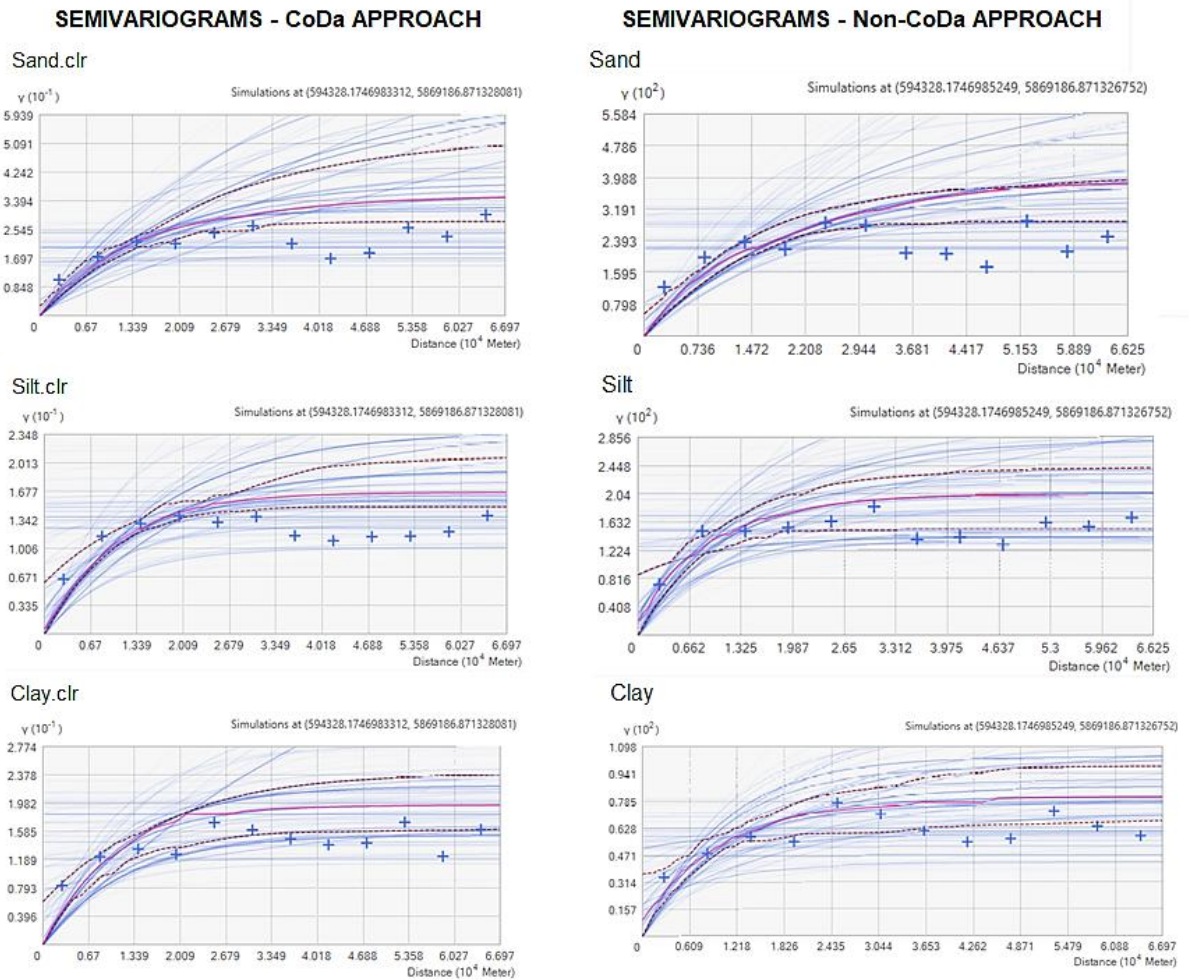
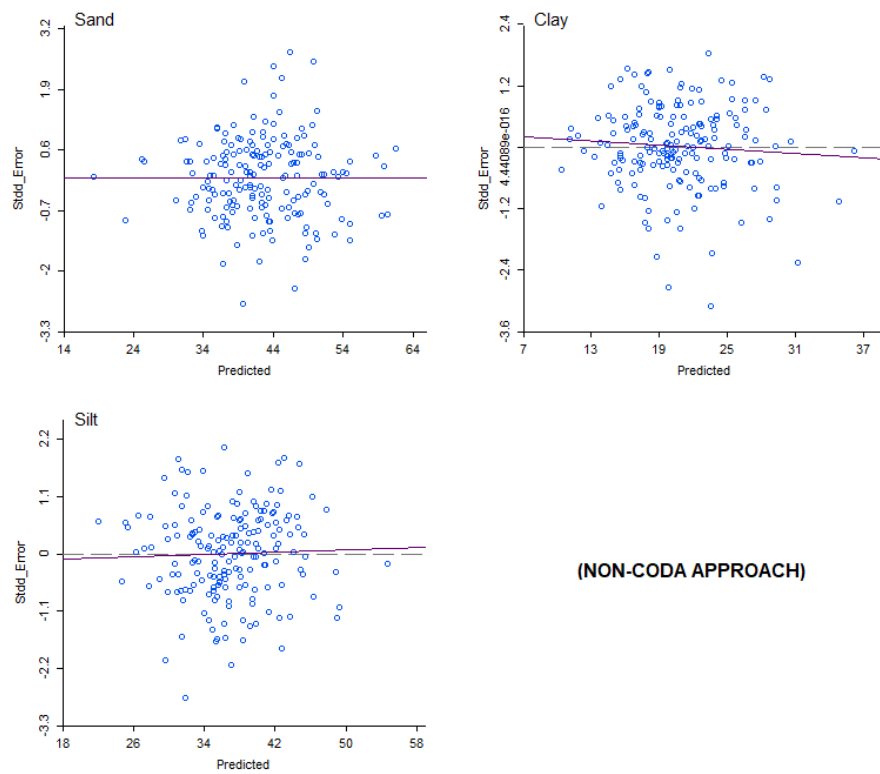


Figure C6.6 Semivariograms fitted for Empirical Bayesian Kriging Regression (EBKR) with compositional approach – CoDa (CLR log-ratio transformation on response variables) and non-CoDa approach (without transformation).

EBK REGRESSION

ERROR DIAGNOSTICS

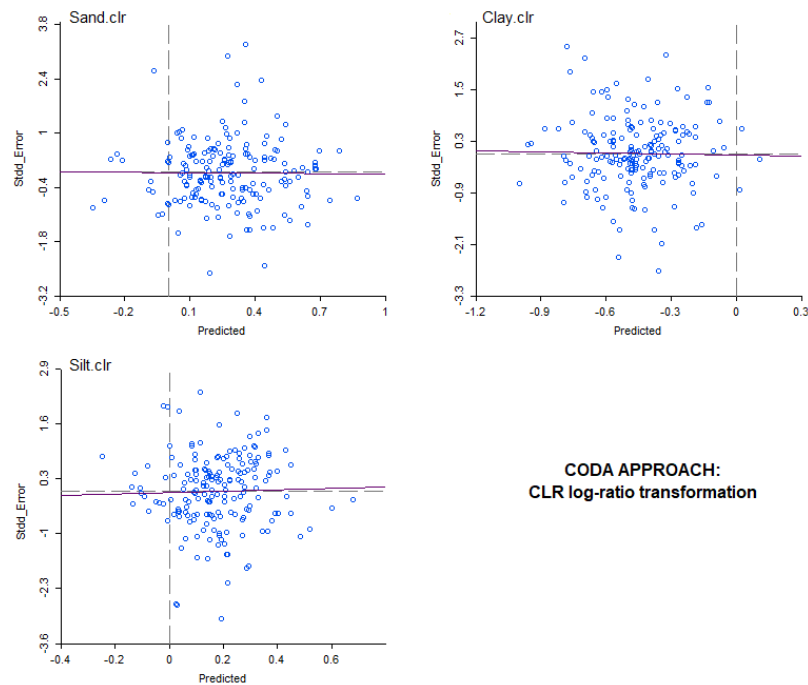


(NON-CODA APPROACH)

Figure C6.7 Error diagnostics plots for the Empirical Bayesian Kriging Regression (EBKR) model fitted without log-ratio transformation applied to response variables. Predicted Vs. Error.

EBK REGRESSION

ERROR DIAGNOSTICS



CODA APPROACH:
CLR log-ratio transformation

Figure C6.8 Error diagnostics plots for the Empirical Bayesian Kriging Regression (EBKR) model fitted with CLR log-ratio transformation applied to response variables. Predicted Vs. Error.

COKRIGING

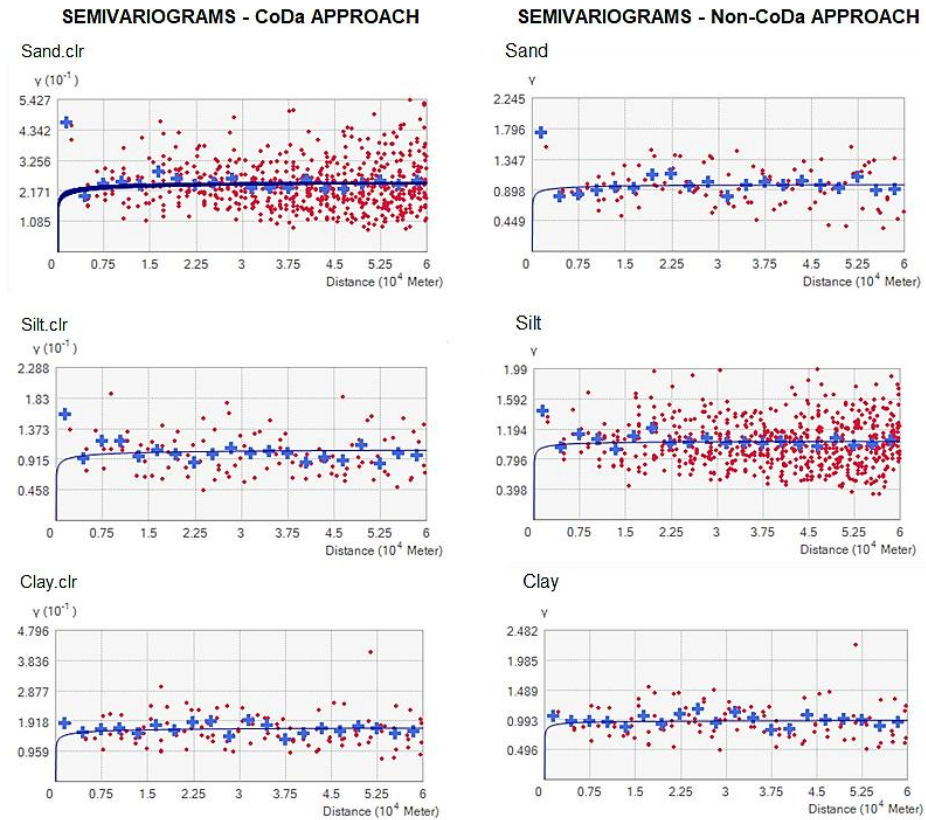


Figure C6.9 Semivariograms fitted for Cokriging (CK) with CoDa approach (CLR log-ratio transformation on response variables) and non-CoDa approach (without transformation). Sand.clr (anisotropic; 199°); Silt.clr (anisotropic; 320°); Clay.clr (anisotropic; 134°); Sand (anisotropic; 199°); Silt (isotropic); Clay (anisotropic; 134°).

COKRIGING

ERROR DIAGNOSTICS

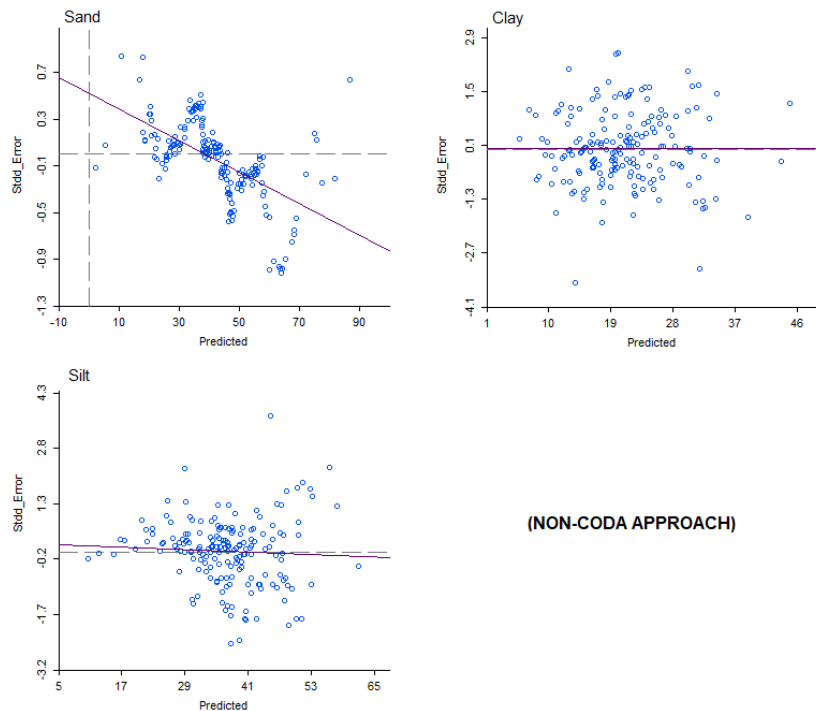


Figure C6.10 Error diagnostics plots for Cokriging with crossvariogram fitted without log-ratio transformation applied to response variables. Predicted Vs. Error. A pattern in the errors for predicted sand is observed in relation to the zero line.

COKRIGING

ERROR DIAGNOSTICS - SAND Error Vs. Predictors

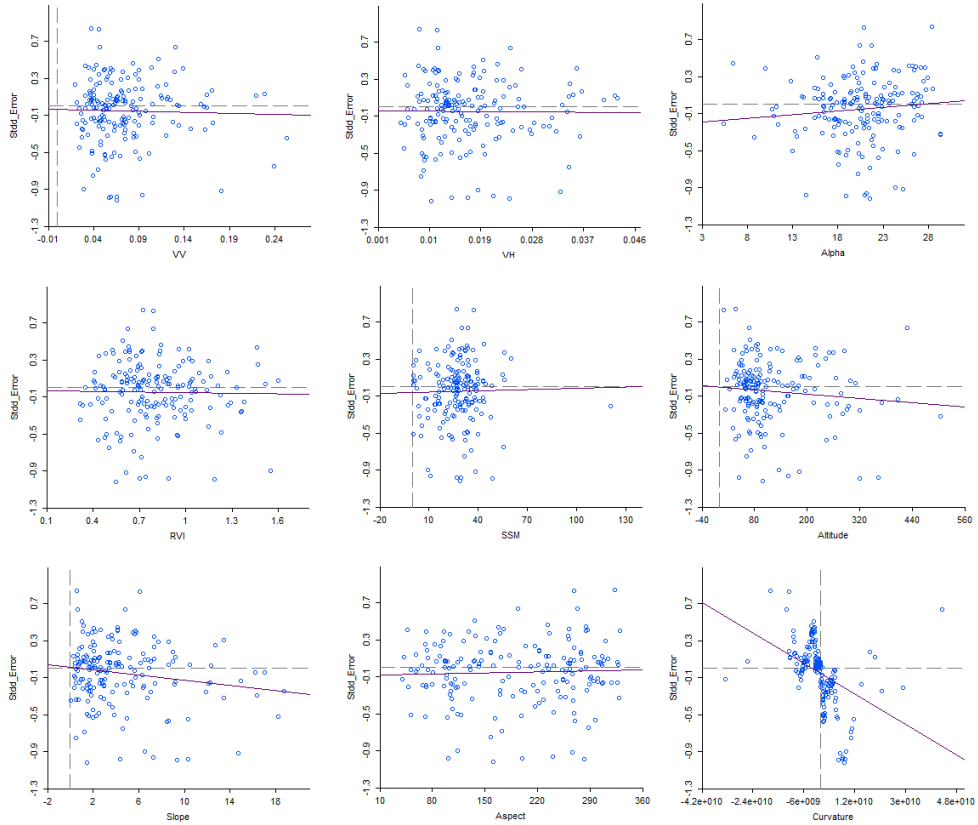
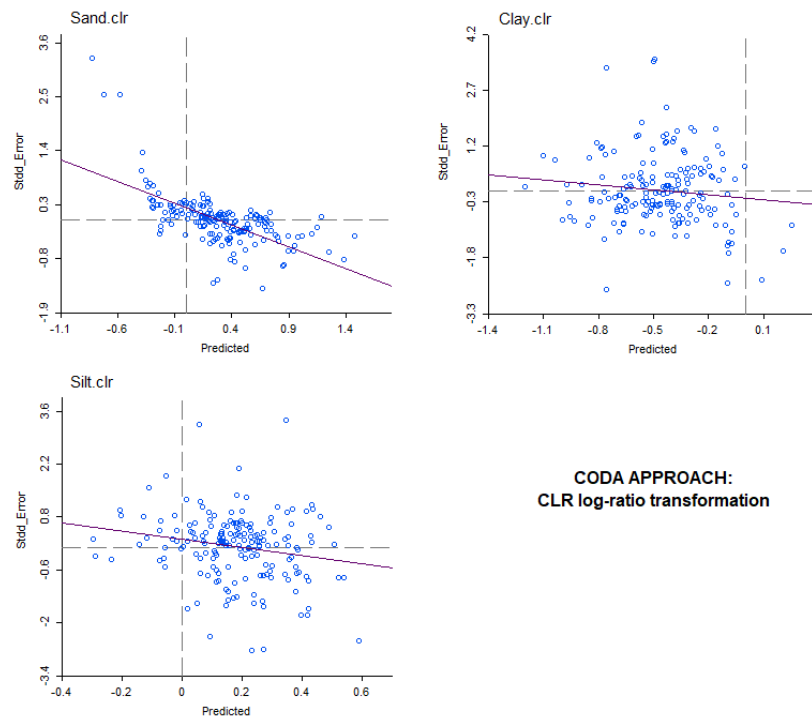


Figure C6.11 Error diagnostics plots of predictors against error for sand estimates resulting from Cokriging fitted without log-ratio transformation applied to response variables. An issue was detected in the curvature plot (bottom right) where points follow the same pattern for sand errors depicted in Figure C6.10. This means curvature has a nonlinear relationship.

COKRIGING

ERROR DIAGNOSTICS



CODA APPROACH:
CLR log-ratio transformation

Figure C6.12 Error diagnostics plots for Cokriging with crossvariogram fitted with CLR log-ratio transformation applied to response variables. Predicted Vs. Error. A pattern is observed for the errors, in reference to the zero line, for predicted sand.

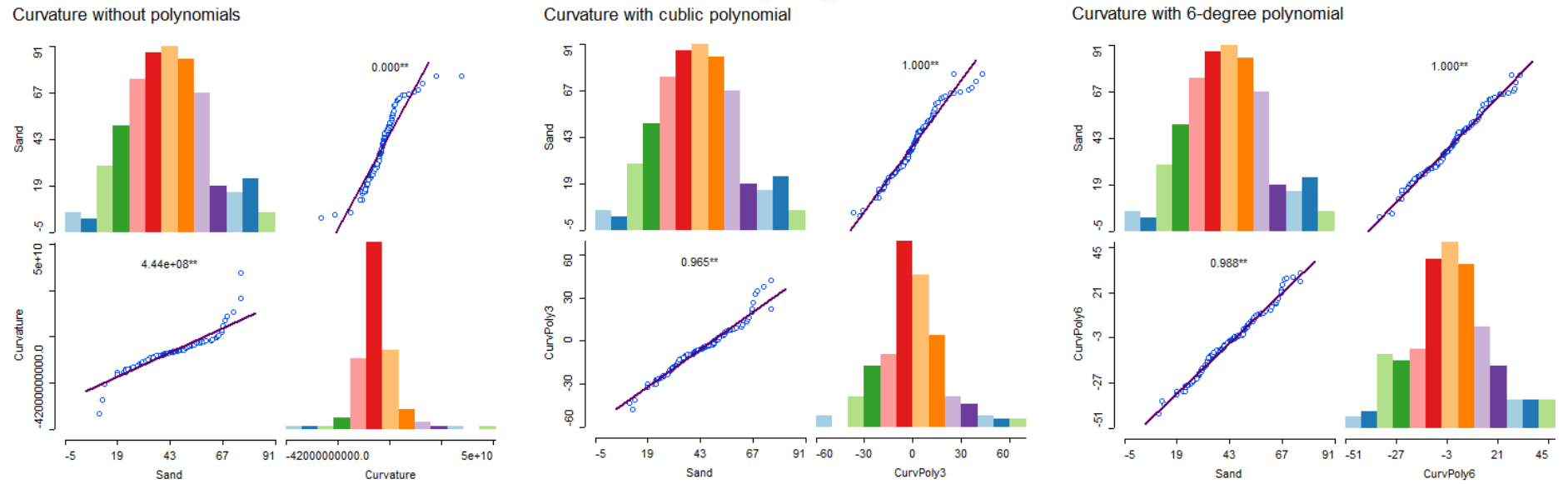


Figure C6.13 Scatter plots matrix for sand and curvature with- and without polynomial functions applied to the covariate (curvature). The 6-degree polynomial appears to be the best solution to deal with deviation from the straight line for the curvature covariate.

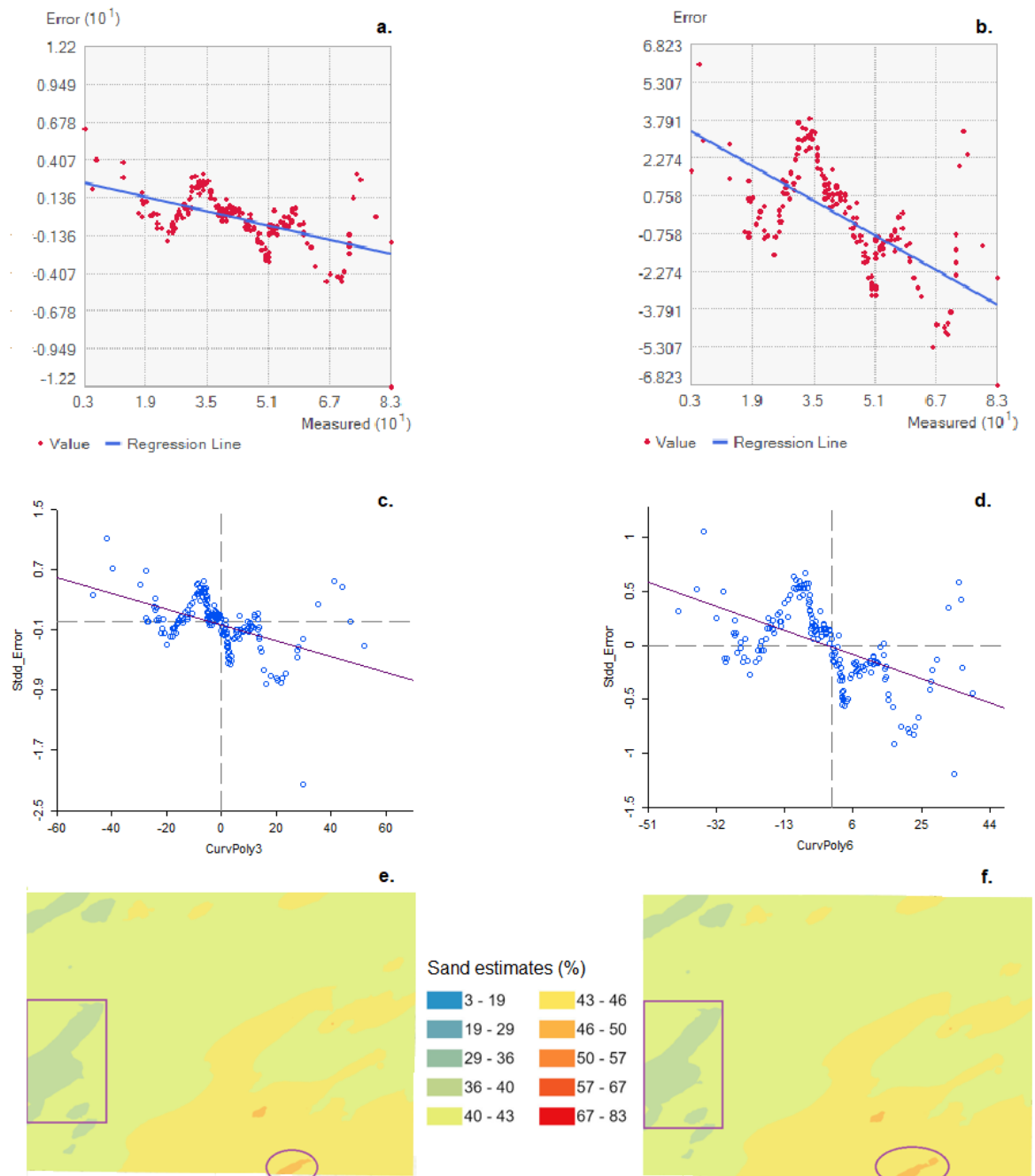
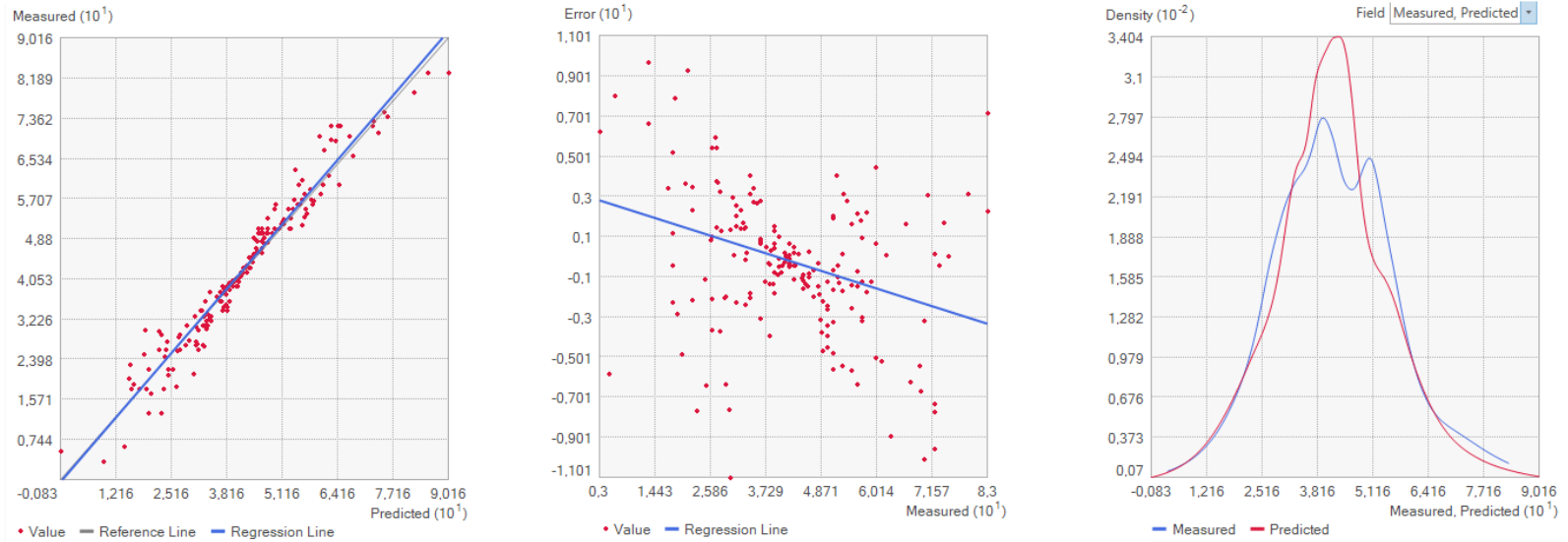


Figure C6.14 Results obtained from the CK model for sand with cubic polynomial and 6-degree polynomial applied to the covariate curvature. Scatter plots for the error and measured data with cubic- and 6-degree polynomials (Figures a and b, respectively); scatter plots for the standardised error and curvature covariate with cubic- and 6-degree polynomials (Figures c and d, respectively); and the respective surface predictions depicted in Figures e (cubic polynomial) and f (6-degree polynomial). Only minor differences were observed as shown by the squares and circles. It is worth noting that the highest values (sand content > 57%) were not estimated, being consistent with the scatter plots.

Cokriging model: Sand (validated with LOOCV)

With 3 most significant interactions

Altitude:Curvature; Altitude:Slope; RVI:Curvature



Error scatterplots for CK-sand against predicted and covariates

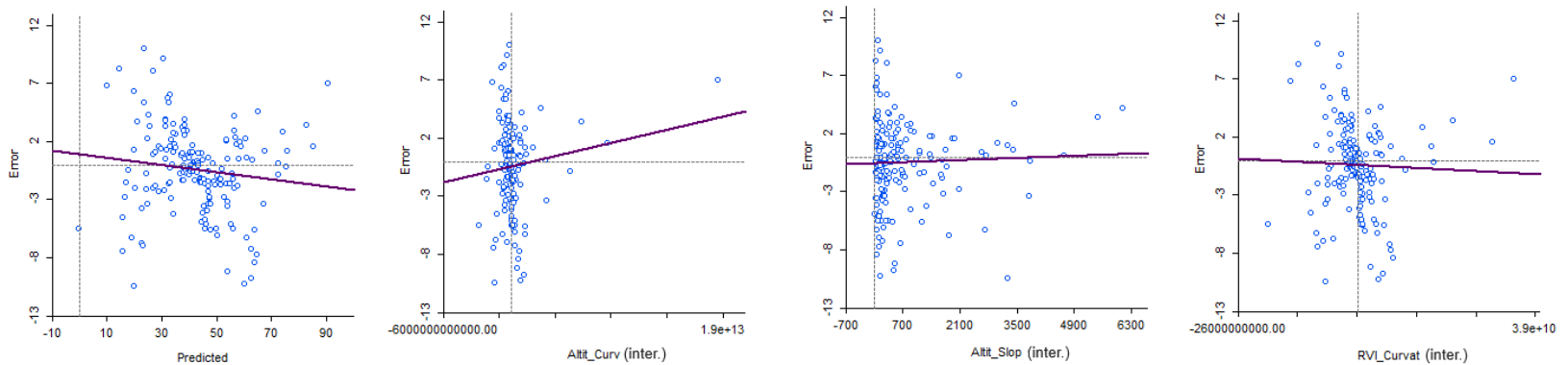


Figure C6.15 Diagnostics plots were obtained from the CK model for sand with interactions terms applied to the covariates before running the model. Notwithstanding still slight overfitting, such a procedure seems to improve the error plots shown in Figures C6.10, C6.11, and C6.14 by accounting for non-linearity.

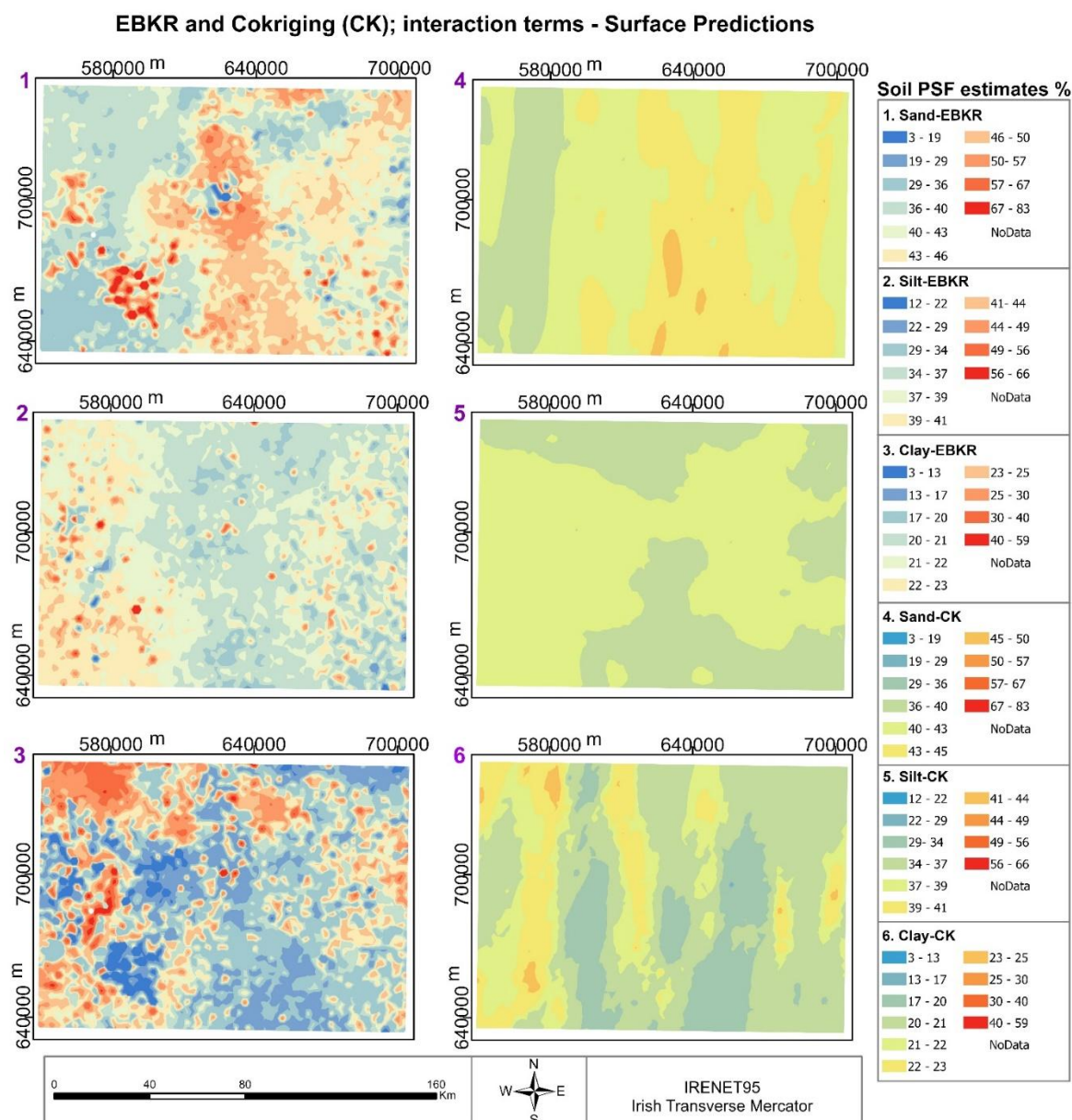


Figure C6.16 A comparison between EBKR and CK models fitted with interaction terms applied to the covariates before running the models. Interaction terms do not appear to improve surface predictions, particularly when spatial direction (anisotropy) is considered in the semivariograms. The highest values estimates (red colours) were stressed with EBKR, especially for sand and clay fractions, account for local variations (See also Figure C6.4).

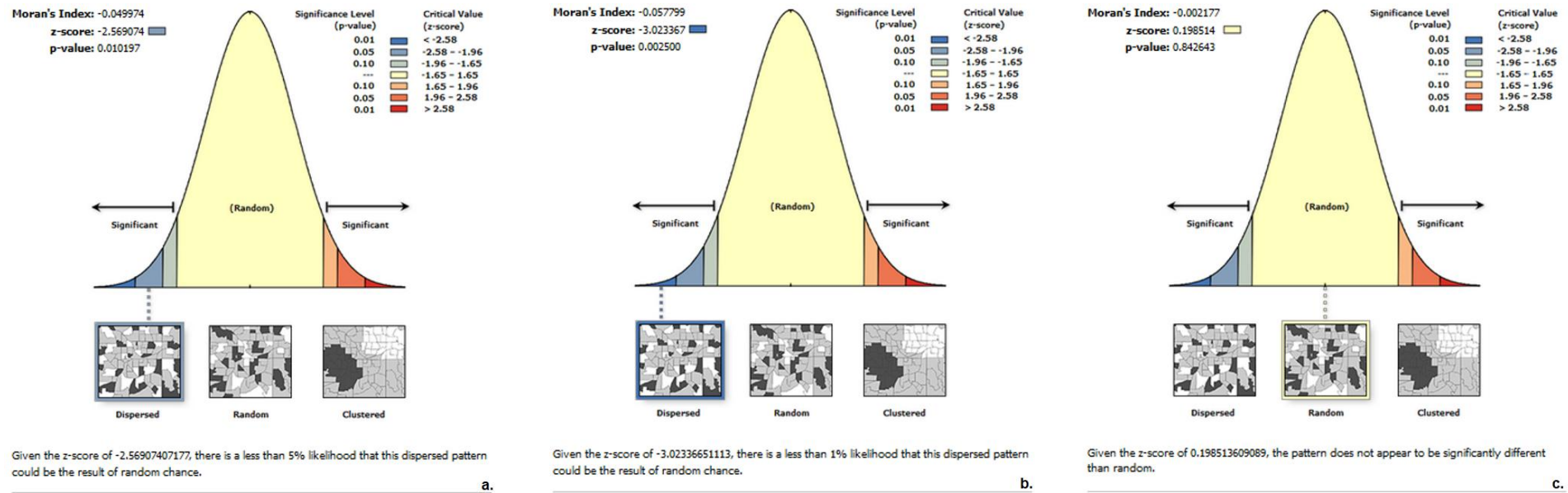


Figure C6.17 The Global Moran Index for the response variables without log-ratio transformation applied to them. Sand: a., Silt: b., Clay: c.

Appendix C – Tables

Table C6.1 The most significant interactions. Linear Regression Model.

Soil PSF	2 nd ORDER INTERACTION TERMS	Pr(> t)
Sand	Altitude * Curvature	1.65e-12
	Altitude * Slope	0.00135
	RVI * Curvature	0.00217
	SSM * Curvature	0.00935
	VV * Curvature	0.04046
Silt	Altitude * Curvature	0.000597
	Slope * Curvature	0.020362
	SSM * Alpha	0.028853
	RVI * Alpha	0.044005
Clay	Altitude * Slope	0.00361
	SSM * Alpha	0.00728
	RVI * Curvature	0.00805
	Slope * Curvature	0.01318
	Slope * Aspect	0.02798
	Alpha * Aspect	0.04470
	RVI * Alpha	0.05163

Appendix C – Models

Simplicial regression model (CoDa) and log-ratio transformation methods⁴

Compositional data analysis (CoDa) is mostly performed in terms of log-transformations and relies on log-ratios between the parts or components of one sample. The theoretical foundations are found in Aitchison (1982). As highlighted by Morais and Thomas-Agnan (2021), in simplicial regression one can use transformations to transport the simplex space S^D into the Euclidean space \mathbb{R}^{D-1} to eliminate the simplex constraints problem.

Three log-transformation methods are typically employed in CoDa analysis, namely additive log-ratio transformation (ALR), centred log-ratio transformation (CLR), and isometric log-ratio transformation (ILR). For the purposes of the work outlined here, we focused on the CLR and ILR methods, as they are symmetric transformations meaning that distances are preserved. The difference between these two methods relies on the fact that in the former, the identity (covariance) matrix is singular whereas, in the ILR transformation, the covariance matrix is non-singular.

In the soil context, the CLR transformation for the soil particle-size fraction can be defined according to the following equations (Equation 1-3):

$$CLR_{(1)} = \ln \frac{\text{sand}}{\sqrt[3]{\text{sand} \times \text{silt} \times \text{clay}}} \quad (1)$$

$$CLR_{(2)} = \ln \frac{\text{silt}}{\sqrt[3]{\text{sand} \times \text{silt} \times \text{clay}}} \quad (2)$$

$$CLR_{(3)} = \ln \frac{\text{clay}}{\sqrt[3]{\text{sand} \times \text{silt} \times \text{clay}}} \quad (3)$$

⁴ This material is an excerpt of the original. Deodoro *et al* (2023). <https://doi.org/10.1111/ejss.13414>

From Equations 1-3, it can be noted that the geometric mean composed of all compositions of soil is the denominator, and the CLR is a one-to-one transformation. The CLR transformation maps S^D into the subspace U of \mathbb{R}^D , that is, the CLR-transformed vector has D components but belongs to a $D-1$ dimensional subspace. This method is a one-to-one transformation from S^D to U and the inverse transformation $CLR^{-1} : U \Rightarrow S^D$ is as follows (Equation 4)

$$CLR^{-1} z = C[\exp z_1; \dots; \exp z_D] \quad (4)$$

where C denotes the closure or constraint operation of compositional data in its *closed* form, that is, as positive vectors whose parts add up to a positive constant value (equal to 100 in this study).

The ILR transformation method is a representation of compositions by orthonormal log-ratio coordinates or orthonormal log-ratio basis (OLR) and it is also a one-to-one transformation from S^D to \mathbb{R}^{D-1} . The identity matrix of dimension $D-1$ is non-singular and plays a decisive role in relating a composition to its OLR-coordinates. Under an orthonormal log-ratio basis, a composition x in S^D can be expressed as a compositional linear combination of the vector forming the basis. In other words, these $D-1$ independent vectors in S^D are unitary and mutually orthogonal and the compositional operations are reduced to ordinary vector operations when compositions are represented by their olr-coordinates. This is more appropriate for working with compositional data (CoDa), as the techniques designed for real data hold their coordinates. This means that linear regression models can be built, as such geometric structure dovetails with that of real space \mathbb{R}^{D-1} .

For the ILR transformation applied to soil texture, the OLR-coordinates of the soil compositions relative to the orthonormal basis of S^D have to be calculated in terms of partitions, which are presented in a sign matrix (Table 1). Partition is a selection of compositions to be included in the bases. It is a hierarchical grouping of parts of the original compositional vector, starting with the whole composition as a group and ending with each part in a single group (Comas-Cufí & Thió-Henestrosa, 2011). One way to define them is by the sequential binary partition (SBP) (Pawlowsky-Glahn *et al.*, 2015; Egozcue & Pawlowsky-Glahn, 2006). SBP consists of dividing the composition into two groups of parts which are indicated by +1 and -1 (numerator and denominator, respectively) where the zeros correspond to parts not included in the partition. With this procedure, data is transformed from the simplex (raw data) to real space (orthonormal coordinates-OLR) with the isometric logratio transformation (ILR) defined by a sequential binary partition.

Table 1. Partition $D = 3^*$

i^{**}	Sand	Silt	Clay	ILR transformation***
1	+1	+1	-1	$x_1^* = \sqrt{\frac{2}{3}} \ln \frac{(\text{sand} \times \text{silt})^{1/2}}{\text{clay}}$
2	+1	-1	0	$x_2^* = \sqrt{\frac{1}{2}} \ln \frac{\text{sand}}{\text{silt}}$

* Three-dimensional compositional data (sand, silt, clay) or 3-part composition. Default partition built in the CoDaPack software (Comas-Cufí & Thió-Henestrosa, 2011). ** i refers to the i-th OLR-coordinate.
*** In OLR-coordinates (balance coordinates).

The inverse transformation $\text{olr}^{-1} : \mathbb{R}^D \Rightarrow S^D$ is obtained according to Equation 5:

$$\text{olr}^{-1} \mathbf{v} = \mathcal{C}(\exp(\mathbf{v} \cdot \Phi)) \quad (5)$$

where \mathbf{v} is the $\text{olr} \mathbf{x}$ vector, Φ is the matrix of $(D - 1) \times D$ -matrix and \mathcal{C} denotes the closure or constraint operation of compositional data in its *closed* form.

Thus, applying the log-ratio (linear) transformations, it is possible to use standard unconstrained multivariate analysis on the transformed data because of (i) the one-to-one nature of this transformation, (ii) transferring any inferences back to the simplex, and (iii) the components of the composition. The vector space structure of S^D allows the concepts of linear dependence and independence to be used (i.e. compositional-linear dependence).

In traditional Linear Regression Models (LRM), the most popular fitting method is the least-squares deviation criterion, or ordinary least-squares regression (OLS), which minimizes the sum of squared errors. The (simple linear) regression model is shown in Equation 6,

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad (6)$$

where ε_i are independent and identically distributed as $\text{Normal}(0, \sigma^2)$. Equation 7 presents the estimation of the coefficients $\beta_0, \beta_1, \dots, \beta_r$ of a linear surface into S^D , in CoDa analysis.

$$\hat{x}(t) = \beta_0 \oplus (t_1 \odot \beta_1) \oplus \dots \oplus (t_r \odot \beta_r) = \bigoplus_{j=0}^r (t_1 \odot \beta_1), \quad (7)$$

where $t = [t_0, t_1, \dots, t_r]$ are real covariates and are identified as the parameters of the linear surface and $\hat{x}(t)$ is the expected value of the CoDa-response variable. Perturbation (\oplus) and Power (\odot) are the fundamental and the subsidiary operations in the simplex, respectively. They are used in statistics notations such as in linear models equations written specifically for CoDa – the simplicial regression. The former is analogous to translation in real coordinate space or Euclidian space (\mathbb{R}^D) and is used to find a difference between compositions or change from one to another. The latter is the simplicial operation analogous to scalar multiplication in \mathbb{R}^D .

The compositional coefficients of the model, $\beta_j \in S^D$ are to be estimated from the data. Because this model is presented as a least-squares problem in the simplex, it could be formulated in terms of orthonormal log-ratio coordinates (olr). Morais and Thomas-Agnan (2021) synthesised the model formulations in coordinate space wherein the parameters a^* , b^* or B^* are estimated by maximum likelihood in coordinate space for the classical regression (Table 2).

Table 2. Specifications of the compositional models and notations.
Source: Adapted from Morais and Thomas-Agnan (2021).

Space	Y-compositional model
S^D	$Y_t = a \oplus X_t \odot b \oplus \varepsilon_t$
\mathbb{R}^{D-1}	$Y_t^* = a^* + b^* X_t + \varepsilon_t^*$
$Y_t, a, b, \varepsilon_t \in S^D_Y, X_t \in \mathbb{R}; Y_t^*, a^*, b^*, \varepsilon_t^* \in \mathbb{R}^{D_Y-1}$	

Appendix D. Declarations of Co-authorship

This form is to accompany the submission of a PhD that contains research reported in published or unpublished work. **Please include one copy of this form for each co-authored work.** This form along with the published work should, under normal circumstances, appear in an Appendix.

Authorship Declaration Form

Publication Details:

Thesis Chapter/Pages	Chapter 4 (71 – 93)
Publication title	<i>An assessment of Sentinel 1 SAR, geophysical and topographical covariates for estimating topsoil particle size fractions.</i>
Publication status	Published
Type of publication	Research article
Publication citation	Deodoro, S.C., Moral, R. de A.; Fealy, R.M.; McCarthy, T., & Fealy, R. (2023). <i>An assessment of Sentinel 1 SAR, geophysical and topographical covariates for estimating topsoil particle size fractions. European Journal of Soil Science</i> , volume 74, issue 5, September–October 2023, e13414, DOI: 10.1111/ejss.13414, https://doi.org/10.1111/ejss.13414

Nature/extent of my contribution to the work detailed above is as follows:



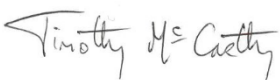
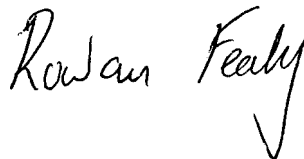
Nature/Extent of Contribution	Lead author
For this paper, I developed its conceptualisation and the research design, proposed a method framework, carried out data- collate and curation, models fitting and formal analysis of the results, and the writing of the paper.	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No

The following co-authors contributed to the work (all contributing co-authors):

Name	Nature of contribution
Rafael de Andrade Moral	Revision of the original text, particularly the statistical analysis; minor edits to the original text; proof-reading
Réamonn Fealy	Revision of the original text, particularly the soil analysis; provision of Teagasc soil legacy data; minor edits to the original text; proof-reading
Tim McCarthy	Revision of the original text; minor edits to the original text; proof-reading
Rowan Fealy	Revision of the original text; edits to the original text; proof-reading

The undersigned hereby certify that the above declaration correctly reflects the nature and extent of the student's and co-author's contribution to this work

	Name	Signature	Date
Student	Sandra Cristina Deodoro	<i>Sandra Cristina Deodoro</i>	09/09/2024

Co-author 1	Rafael de Andrade Moral		24/09/2024
Co-author 2	Réamonn Fealy		22/10/2024
Co-author 3	Tim McCarthy		22/10/2024
Co-author 4	Rowan Fealy		24/09/2024



This form is to accompany the submission of a PhD that contains research reported in published or unpublished work. **Please include one copy of this form for each co-authored work.** This form along with the published work should, under normal circumstances, appear in an Appendix.

Authorship Declaration Form

Publication Details:

Thesis Chapter/Pages	Chapter 5 (94 – 124)
Publication title	<i>Using the surface scattering mechanism from dual-pol SAR data to estimate topsoil particle size fractions.</i>
Publication status	Published
Type of publication	Research article
Publication citation	Deodoro, S.C., Moral, R. de A.; Fealy, R.M.; McCarthy, T., & Fealy, R. (2024). Using the surface scattering mechanism from dual-pol SAR data to estimate topsoil particle size fractions. <i>International Journal of Applied Earth Observation and Geoinformation</i> , 128, e10374, https://doi.org/10.1016/j.jag.2024.103742

Nature/extent of my contribution to the work detailed above is as follows:


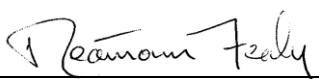
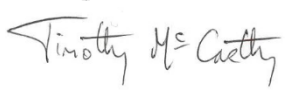
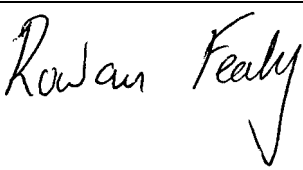
Nature/Extent of Contribution	Lead author
For this paper, I developed its conceptualisation and the research design, proposed an approach to deal with vegetated soil in SAR images by testing an existing method, carried out collation- and data curation, models fitting and formal analysis of the results, and the writing of the paper.	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No

The following co-authors contributed to the work (all contributing co-authors):

Name	Nature of contribution
Rafael de Andrade Moral	Revision of the original text, particularly the statistical analysis; minor edits to the original text; proof-reading
Réamonn Fealy	Revision of the original text, particularly the soil analysis; provision of Teagasc soil legacy data; minor edits to the original text; proof-reading
Tim McCarthy	Revision of the original text; minor edits to the original text; proof-reading
Rowan Fealy	Revision of the original text; edits to the original text; proof-reading

The undersigned hereby certify that the above declaration correctly reflects the nature and extent of the student's and co-author's contribution to this work

	Name	Signature	Date
Student	Sandra Cristina Deodoro	<i>Sandra Cristina Deodoro</i>	09/09/2024

Co-author 1	Rafael de Andrade Moral		24/09/2024
Co-author 2	Réamonn Fealy		22/10/2024
Co-author 3	Tim McCarthy		22/10/2024
Co-author 4	Rowan Fealy		24/09/2024



Book chapter / Chapter 2 (partially)

This form is to accompany the submission of a PhD that contains research reported in published or unpublished work. **Please include one copy of this form for each co-authored work.** This form along with the published work should, under normal circumstances, appear in an Appendix.

Authorship Declaration Form

Publication Details:

Thesis Chapter/Pages	Chapter 2 (24 – 42)
Publication title	<i>The use of remote sensing data and techniques for above ground biomass estimation.</i>
Publication status	Unpublished (under review)
Type of publication	Book chapter
Publication citation	Deodoro, S.C., Fealy, R., & McCarthy, T. (under review). <i>The use of remote sensing data and techniques for above ground biomass estimation.</i> In: Moral, R. de, A. (ed.) <i>Wildlife Monitoring: Integrating Conservation and Innovation in Human-Altered Landscapes.</i> Springer.

Nature/extent of my contribution to the work detailed above is as follows:

Nature/Extent of Contribution	Lead author
For this book chapter, I developed its conceptualisation and wrote the original text.	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No

The following co-authors contributed to the work (all contributing co-authors):

Name	Nature of contribution
Rowan Fealy	Revision of and edits to the original text.
Tim McCarthy	Revision of and edits to the original text.
Rafael de Andrade Moral	Book's editor.

The undersigned hereby certify that the above declaration correctly reflects the nature and extent of the student's and co-author's contribution to this work

	Name	Signature	Date
Student	Sandra Cristina Deodoro	<i>Sandra Cristina Deodoro</i>	09/09/2024
Co-author 1	Rowan Fealy	<i>Rowan Fealy</i>	24/09/2024
Co-author 2	Tim McCarthy	<i>Timothy McCarthy</i>	24/09/2024
Co-author 3	Rafael de Andrade Moral	<i>Rafael de Andrade Moral</i>	24/09/2024