Received: 7 April 2023 | Revised: 21 August 2023 | Accepted: 22 August 2023

DOI: 10.1111/ejss.13414

RESEARCH ARTICLE

European Journal of Soil Science WILEY

An assessment of Sentinel-1 synthetic aperture radar, geophysical and topographical covariates for estimating topsoil particle-size fractions

Sandra Cristina Deodoro[1] | **Rafael Andrade Moral**[2] | **Reamonn Fealy**[3] | **Tim McCarthy**[4] | **Rowan Fealy**[1]

[1]Irish Climate Analysis and Research Units (ICARUS), Department of Geography, Maynooth University, Maynooth, Ireland

[2]Department of Mathematics & Statistics, Maynooth University, Maynooth, Ireland

[3]Teagasc Agrifood Business and Spatial Analysis Department, Ashtown, Ireland

[4]National Centre for Geocomputation, Maynooth University, Maynooth, Ireland

Correspondence
Sandra Cristina Deodoro, Irish Climate Analysis and Research Units-ICARUS, Laraghbryan House, North Campus, Maynooth University, Maynooth, Co. Kildare, Ireland.
Email: sandra.deodoro.2022@mumail.ie

Funding information
National University of Ireland, Maynooth

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 nongeophysical (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 was 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.

KEYWORDS

clay, compositional, linear regression, log-ratio transformations, random forest, sand, Sentinel 1, silt

1 | 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 utilization of Synthetic Aperture Radar (SAR) for estimating soil moisture or carbon content and employing in situ soil texture measurements to retrieve those soil properties (Babaeian et al., 2019; Petropoulos et al., 2015; Pradipta et al., 2022). Additionally, the investigation of soil texture with SAR remote sensing has largely been constrained to bare soil areas (Gholizadeha et al., 2018; Marzahn & Meyer, 2020; Mondejar & Tongco, 2019; Niang et al., 2014), which greatly limits its application. Similar to geophysical data based on electromagnetic signals, SAR data are 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 et al. (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 whether 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).

Highlights

- Compositional LRM with interaction slightly outperformed RF; without interactions, RF was better.
- Sand and clay exhibited different responses for radar and geophysical data in the models.
- Geophysical data were not good predictors to estimate the content of sand, silt and clay on topsoil.
- SAR-based variables were not able to predict soil particle size alone without the covariates.

This paper is organized as follows. Section 2 presents a brief description of both the theory and relevant works. Section 3 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 and 5 present the results of the modelling approaches and discussion; finally, Section 6 concludes by outlining the relevance of findings.

2 | THEORETICAL BACKGROUND

Soil texture affects the radar backscattering coefficient, as sandy soils contain a higher amount of free water than clay soils (Das & Paul, 2015; Srivastava et al., 2006), 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 & Ulaby, 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 adsorbed 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 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 polarization intensity (i.e., VV, VH in Sentinel-1).

A key challenge for utilizing 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

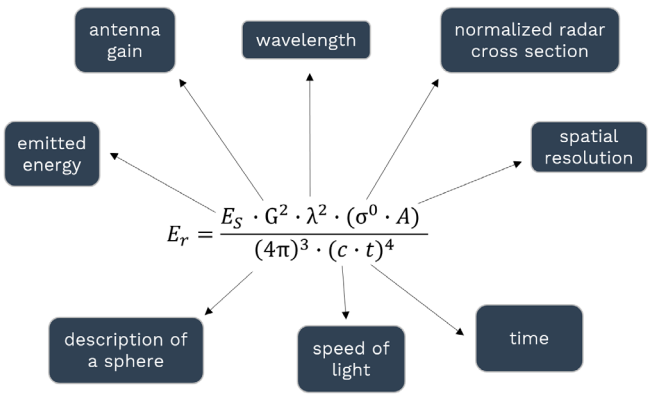

FIGURE 1 Diagram for the radar equation that is used to determine the proportion of energy that is being returned from a target. *Source*: ESA-EO College (2021).

and dielectric constant (including soil and vegetation), which affect the measurement of microwave signals. Cross-polarization is more sensitive to vegetation, as it captures vegetation response (backscattering) better than VV polarization. Barrett et al. (2007), in a study to derive relative changes in soil moisture in the near-surface (0–5 cm) 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 standardization or normalization 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 forest (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 et al. (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, 2005). 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 nonsingular. Compositional data approach has increasingly been employed to estimate soil particle-size fractions in both nonspatial modelling (Chappell et al., 2019; Loosvelt et al., 2013) and explicitly spatial models (Odeh et al., 2003; Wang & Shi, 2017; Wang & Shi, 2018; Zhang & Shi, 2019).

Details about both CLR and ILR transformations as well as the simplicial linear regression model used in this study are found in Supplementary Information—Models (Data S2).

3 | DATA AND METHODS

3.1 | Study region

The study area, representing approximately 24,989 km^2, encompasses a central swath of the Republic of Ireland (Figure 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 in Ireland, accounting for 67% of the national land cover. Within that category, pasture is the main agricultural class (55%) with arable land accounting for a much smaller proportion (4.5%) based on the CORINE 2018 land cover inventory (Supplementary Figure 1).

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 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 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.

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 & O'Sullivan, 2018). In an association between lithology and quaternary sediments

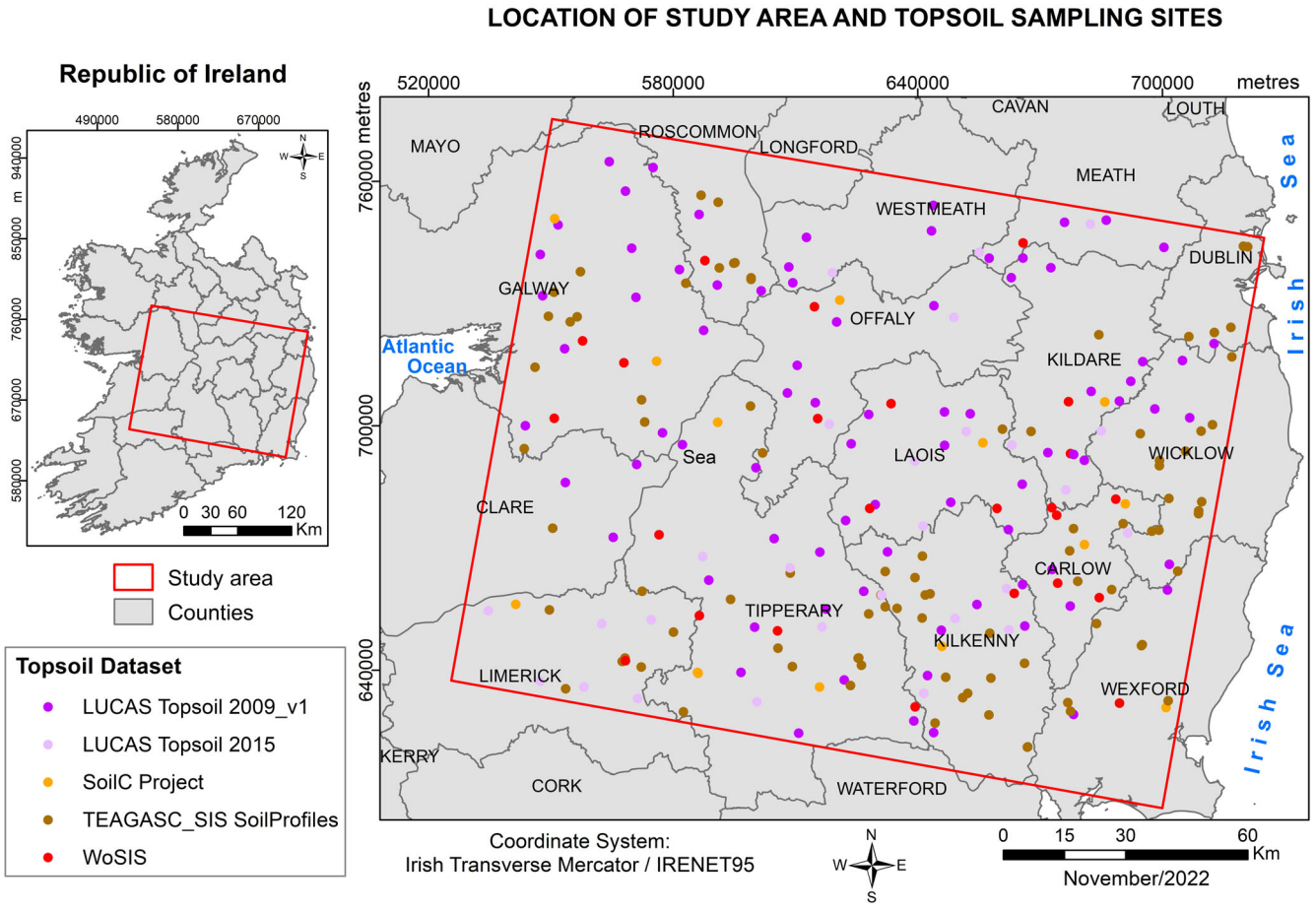

FIGURE 2 Map of the study area with the location of topsoil sampling sites available from different soil surveys and measurement programmes.

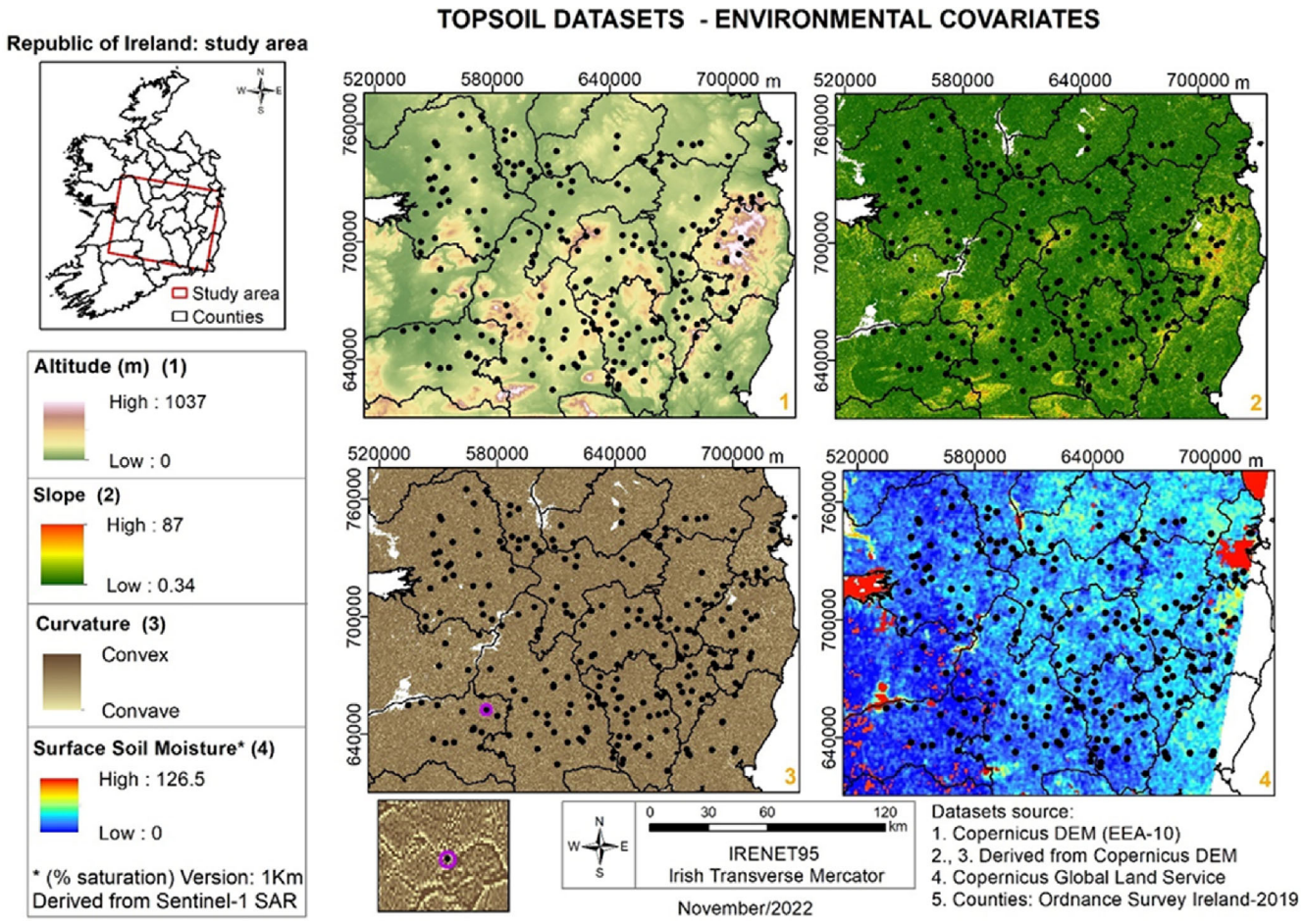

FIGURE 3 Topographic, morphometric and soil moisture for the study domain. Surface soil moisture data were obtained on the same date as that of the Sentinel-1 image (03/04/2021).

(GSI, 2017), the study area is composed of till derived from limestones, shales (Lower Palaeozoic) 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).

3.2 | Data

3.2.1 | 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 encompass 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 1 and details of each dataset are presented in the Supplementary Information—Datasets (Data S1).

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

TABLE 1 Summary of data used in the modelling approaches.

Type of data	Original dataset	Source	Covariate derived from the original dataset
Topsoil (response variables: sand, silt and clay)	Irish Soil Information System (SIS). Modal and nonmodal profiles (Creamer et al., 2014), which include the legacy survey *An Foras Talúntais-AFT*. Sampling depth varied according to survey purposes, but surficial horizons include 0–5 cm and 5–10 cm.	https://data.gov.ie/dataset/irish-soil-information-system-national-soils-map	-
	LUCAS Topsoil 2009 Survey-v1 (Tóth et al., 2013; Orgiazzi et al., 2018). Sampling depth: 0–20 cm	https://esdac.jrc.ec.europa.eu/content/lucas-2009-topsoil-data	-
	LUCAS Topsoil 2015 Survey (Jones et al., 2020). Sampling depth: 0–20 cm	https://esdac.jrc.ec.europa.eu/content/lucas2015-topsoil-data	-
	Soil Carbon Project 2008 (Kiely et al., 2009). Sampling depth: 0–10 cm, 10–25 cm and 25–50 cm	https://www.ucc.ie/en/hydromet/soilcarbon/	-
	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 polarization: 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. (Gururaj et al., 2019; Nasirzadehdizaji et al., 2019);
Soil moisture (covariate)	Surface soil moisture—SSM calculated from Sentinel 1 for the 3 April 2021 (Copernicus Global Land Service)	European Space Agency https://land.copernicus.eu/global/products/ssm	-
Topography (covariate)	Digital Elevation Model-DEM EEA-10 m (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	

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 differed 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.

3.2.2 | 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 GH, which corresponds to a wavelength of ∼5.55 cm. Sentinel-1 data are available in dual polarization (VV + VH). Both the Single Look Complex (SLC) and Ground Range Detected (GRD) (10 m × 10 m) products were obtained with the backscattering coefficients converted to the same units (σ^0). Sentinel-1 data (VV, VH) were acquired for the 3 April 2021 (Supplementary Figure 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) (Gururaj et al., 2019; Nasirzadehdizaji 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). Likewise the alpha parameter, RVI takes into account the vegetation effect on soil backscattering.

3.2.3 | 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 (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 Supplementary Information—Datasets (Data S1).

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 current flows 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 Supplementary Information—Datasets (Data S1). Table 1 summarizes the datasets used in this study.

3.3 | 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 2 and Figure 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 Supplementary Information—Models (Data S2).

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 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) (Egozcue & Pawlowsky-Glahn, 2006), thus it can be formulated in

TABLE 2 Summary of methods and covariates employed in the two modelling approaches.

Modelling approach (data)	Covariates	Statistical model	Log-ratio transformation
Model 1: Nongeophysical (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

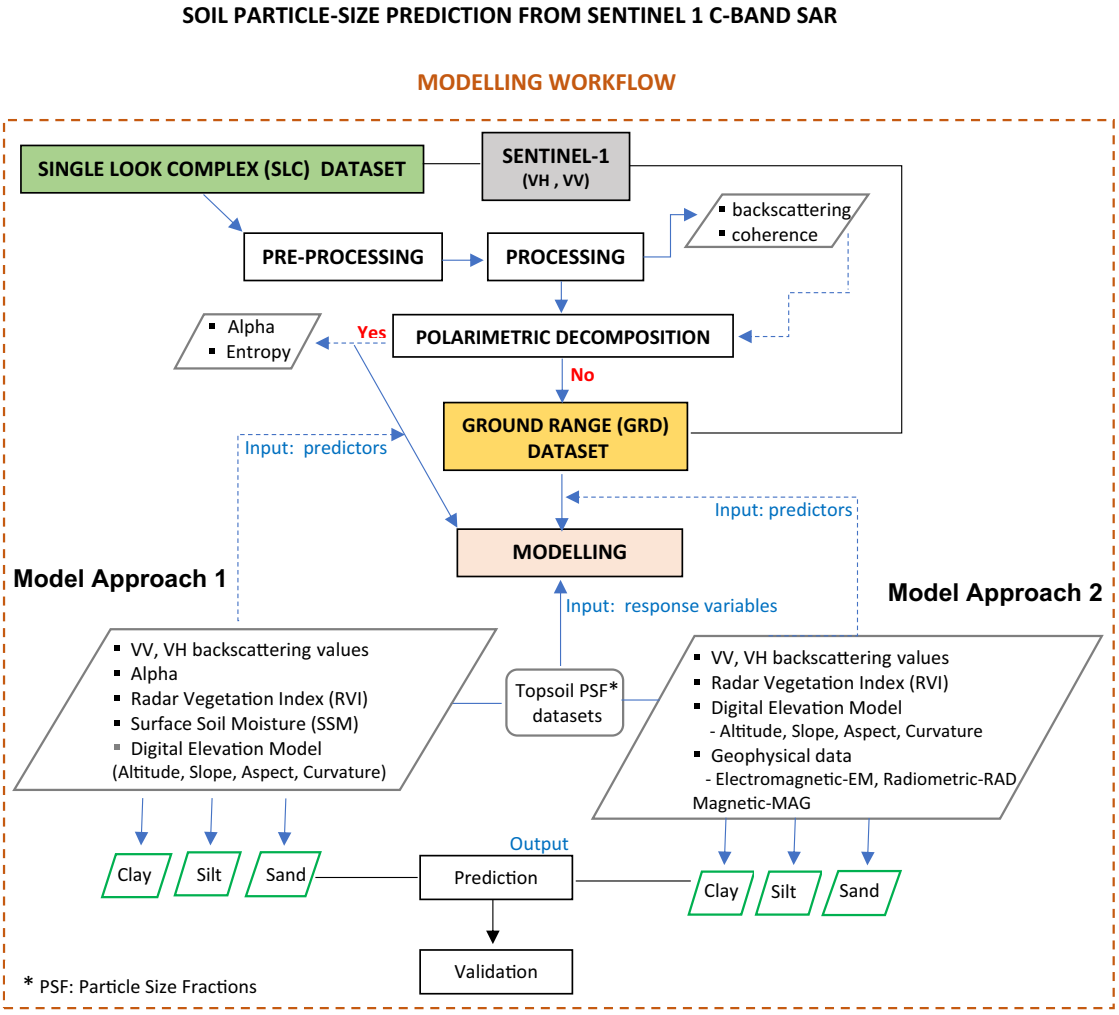

FIGURE 4 Flowchart of the methodology depicting the data used in the models, the processing steps performed for both SLC and GRD products from SAR Sentinel-1, and the modelling approaches (with geophysical and nongeophysical covariates).

terms of orthonormal log-ratio coordinates, following Morais and Thomas-Agnan (2021), and Pawlowsky-Glahn et al. (2015). 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 (Supplementary Tables 1 and 2), we selected the default values available in the R package 'Ranger' (Wright & Ziegler, 2017), which are ntree = 500 and mtry = 3, since different values for ntree and mtry 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 second-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.

3.3.1 | 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 (1)), mean absolute error (MAE) (Equation (2)) and Nash–Sutcliffe efficiency (NSE) (Nash & Sutcliffe, 1970) (Equation (3)). The latter is a normalized statistical metric that determines the relative magnitude of the residual variance compared with the measured data variance (Nash & 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; −Inf < NSE <0 indicates that the observed mean is a better predictor than the model.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}, \tag{1}$$

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

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \widehat{y}_i|, \tag{2}$$

where y_i, $\widehat{y}_i$, and n represent the same descriptors that those of RMSE.

$$\mathrm{NSE} = 1 - \left[\frac{\sum_{i=1}^{n}\left(Y_i^{obs} - Y_i^{sim}\right)^2}{\sum_{i=1}^{n}\left(Y_i^{obs} - \overline{Y^{obs}}\right)^2}\right], \tag{3}$$

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

4 | RESULTS

4.1 | 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 Supplementary Figures 3 and 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 Supplementary Figure 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) (Supplementary Figure 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 polarization 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^0_{VV}, σ^0_{VH}) and the soil particle-size fractions for either the nontransformed or transformed response variables (Supplementary Figures 6–8).

4.2 | Soil particle-size prediction

4.2.1 | Y-compositional LRM

Modelling approach 1: SAR + topography (morphometry) data

The compositional model for approach 1 (Table 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 (Supplementary Figure 9). The compositional coefficients of the LRM indicated an effect of VH backscattering (σ^0_{VH}) on clay (referred to as inv.ilr. 3 after transformation) (Supplementary Figure 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 3 and Figure 5a–c). Regarding the soil texture classes, the corresponding classes from the soil particle-size fractions predicted (test) are sand (S), loamy sand (LS), sand loam (SL), loam (L), clay loam (CL), silty clay loam (SICL), silty clay (SIC) (Figure 5h).

The inclusion of second-order interaction terms resulted in a disimprovement in the model; resulting in a lower NSE and increase in the RMSE and MAE values (Table 3 and Figure 5d–f). 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 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 (Figure 5g–i). 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, the reference to clay is the textural class rather particle-size fraction.

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 Supplementary Table 3). Conversely, when performing a simple linear regression model, as a univariate approach and with no transformation applied to the data,

TABLE 3 Summary of the evaluation metrics resulted from the modelling approaches validated with LOOCV.

		Model evaluation metrics		
Modelling approach (validation method: LOOCV)	**SPSF**	**RMSE**	**MAE**	**NSE**
Model 1: Y-compositional LRM (ilr transf.) **without** second-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** second-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** second-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** second-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

Note: Model 1: nongeophysical data; Model 2: with geophysical data;
Abbreviation: SPSF, soil particle-size fraction.

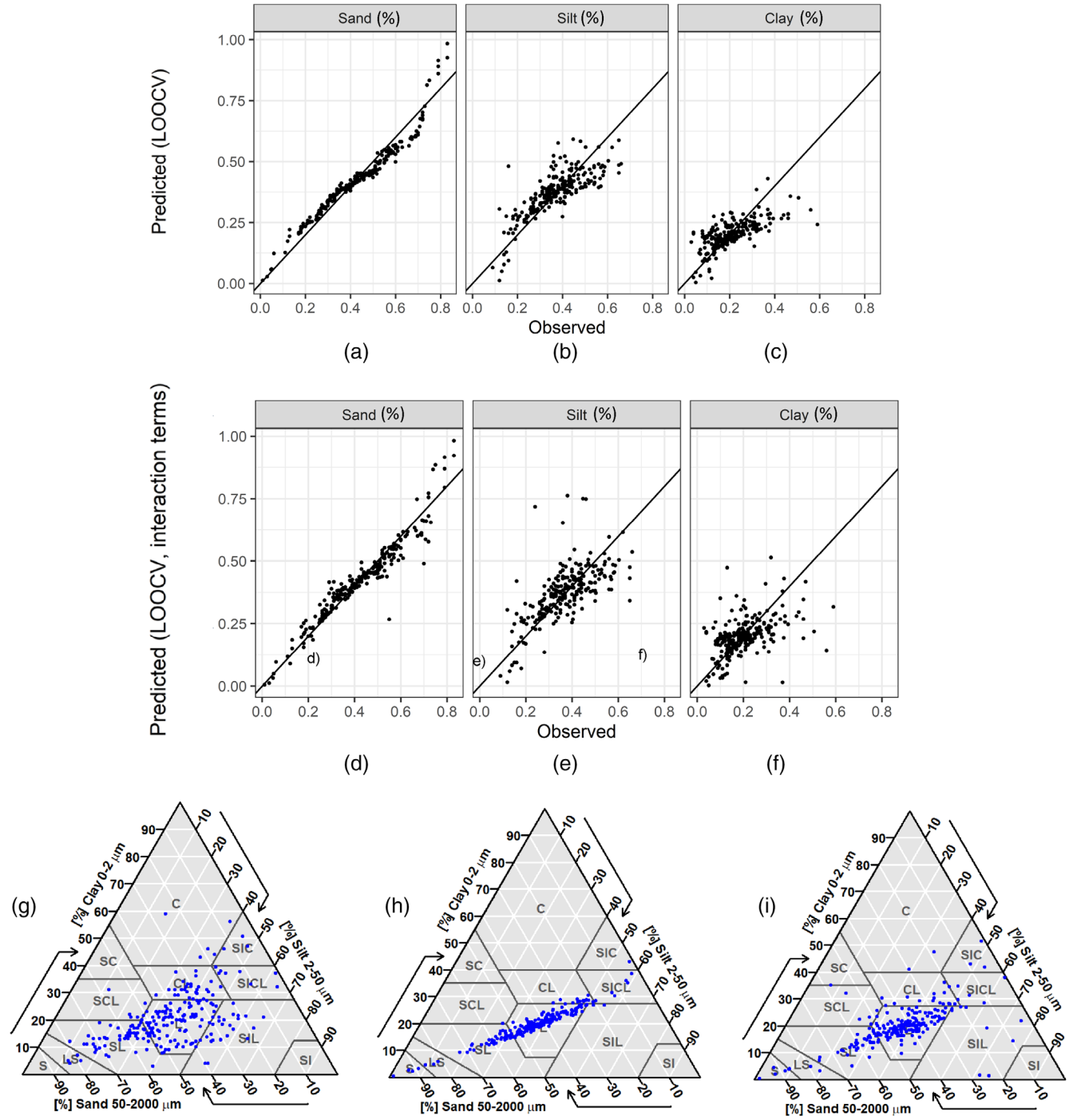

FIGURE 5 Observed versus predicted plots from the Y-compositional linear regression model validated with LOOCV for Modelling approach 1 (SAR + Topography covariates), without second-order interaction terms (Figure 5a–c) and with second-order interaction terms (Figure 5d–f). The responses are measured 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 5g depicts the original soil texture classes ($n = 100\% = 235$) obtained directly from the measured data, while ternary diagrams in the 5h and 5i show classes resulting from predictions of soil particle-size fractions without- and with second-order interaction terms, respectively.

the soil particle-size fractions predicted are not constrained to sum to 100% (Supplementary Figure 11). Moreover, some negative estimates were obtained from both a random proportional split (e.g., 75% and 25%) and LOOCV (Supplementary Figures 11 and 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 (Supplementary Figure 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 3 and Figure 6a–c). The radiometric covariates (thorium and uranium) and altitude were the most statistically significant predictors. An effect was observed for VV on the sand response (Supplementary Figure 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 6h). Note that LS and SIL were obtained with only one sample.

When second-order interaction terms were considered in the model, the agreement between the observed (actual) and the predicted data decreased (Figures 6d–f) as indicated by the evaluation metrics (Table 3). Additionally, more than five soil texture classes were obtained (Figure 6i). Analysis of variance (ANOVA) for the models fitted with and without interactions indicated that the former is statistically significant (p-value = 0.04).

4.2.2 | 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 3 and Figure 7a–c). Similar to the LRM model, the most important variable was curvature (concavity and convexity), according to the impurity method, which is a measure of the variance of the responses for regression in Random Forest (Ranger package) (Supplementary Figure 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 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 Supplementary Table 4).

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

The inclusion of the geophysical covariates resulted in a model with NSE of close to zero (Table 3) indicating that the model simulations have the same explanatory power as the mean of the observations (Figure 8a–c). Loam was the main soil textural class (Figure 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 is 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 (Supplementary Figure 16).

5 | DISCUSSION

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

We obtained different goodness-of-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-polarization (VH) measurements are significantly lower than the co-polarization (VV) (Sano et al., 2020; Ulaby et al., 1978; Ulaby et al., 1979), 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 polarization 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 (Dobson & Ulaby, 1981; Jackson & Schmugge, 1989). The best prediction was obtained for the sand, compared with clay, fraction 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

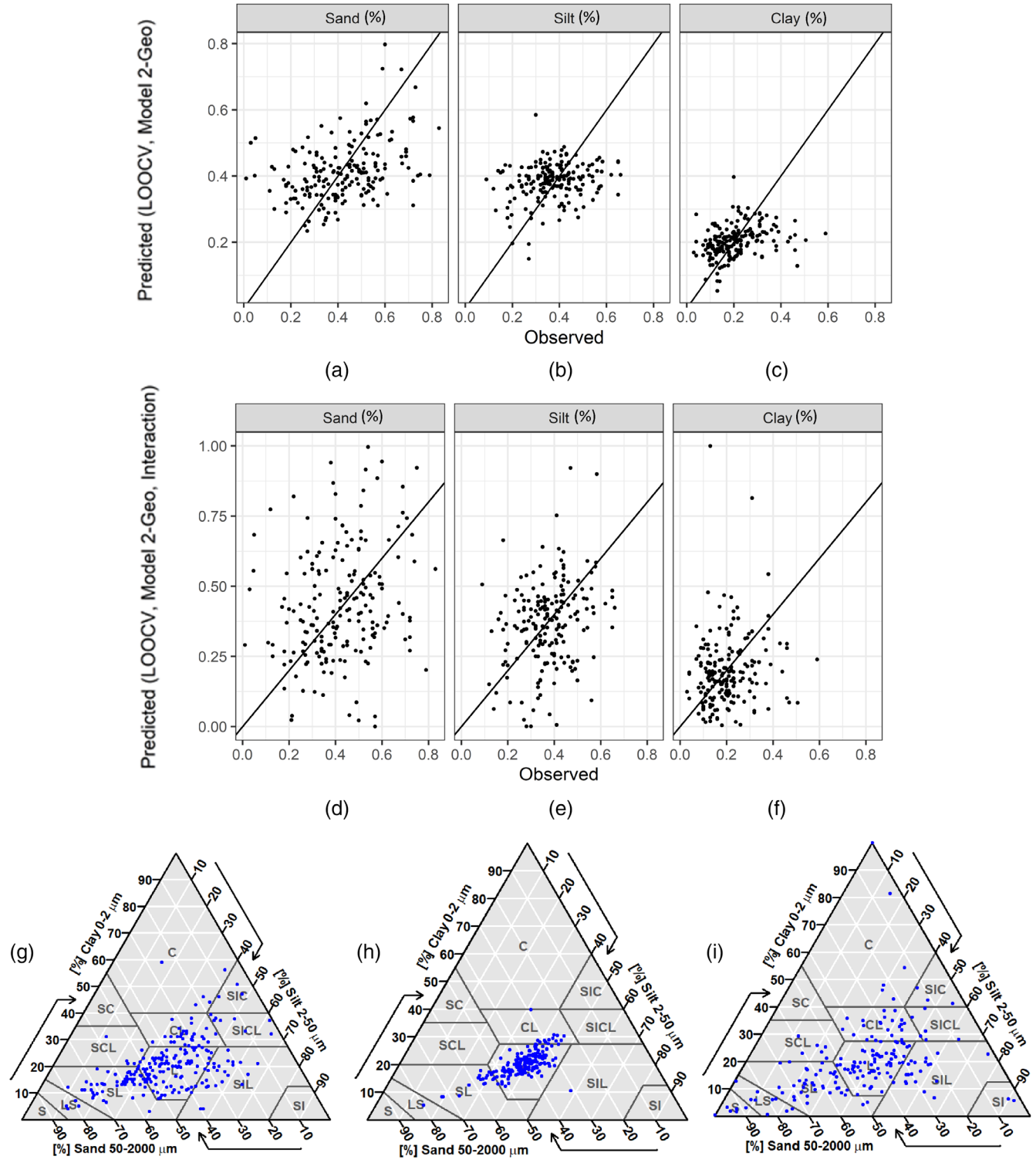

FIGURE 6 Observed versus predicted plots resulting from the Y-compositional linear regression model validated with LOOCV for Modelling approach 2 (SAR + Topography + Geophysical covariates), without second-order interaction terms (Figure 6a–c) and with second-order interaction terms (Figure 6d–f). The 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 (6g) depicts the soil texture classes obtained directly from the measured data, while ternary diagrams in (6h) and (6i) show classes resulting from predictions of soil particle-size fractions without- and with second-order interaction terms, respectively.

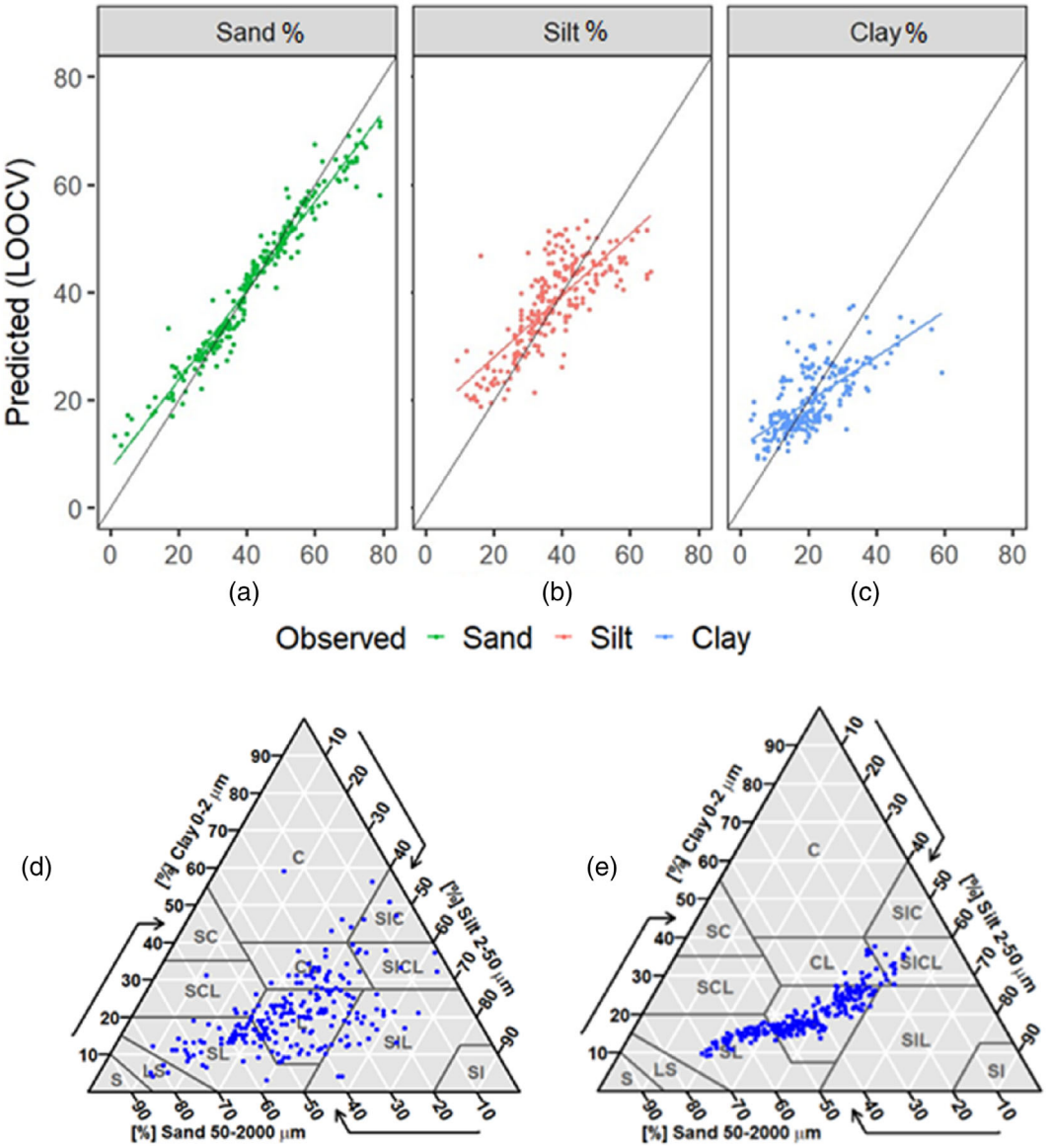

FIGURE 7 Comparative plots of observed and predicted samples (Figure 7a–c) obtained from the Random Forest Modelling with CLR transformation applied to the response variables and validated with LOOCV (**Approach 1: SAR + Topography** covariates). The responses are measured in % content and the black 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 7d and the classes resulting from predictions of soil particle-size fractions are exhibited in 7e.

Mississippi river (riverside and landside). On the landside, the correlation coefficients found with HV polarization were −0.16 and 0.16 for clay and sand, respectively; and 0.16 and −0.25 with VV polarization. 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 1).

5.2 | Y-compositional LRM—ILRTransformation

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,

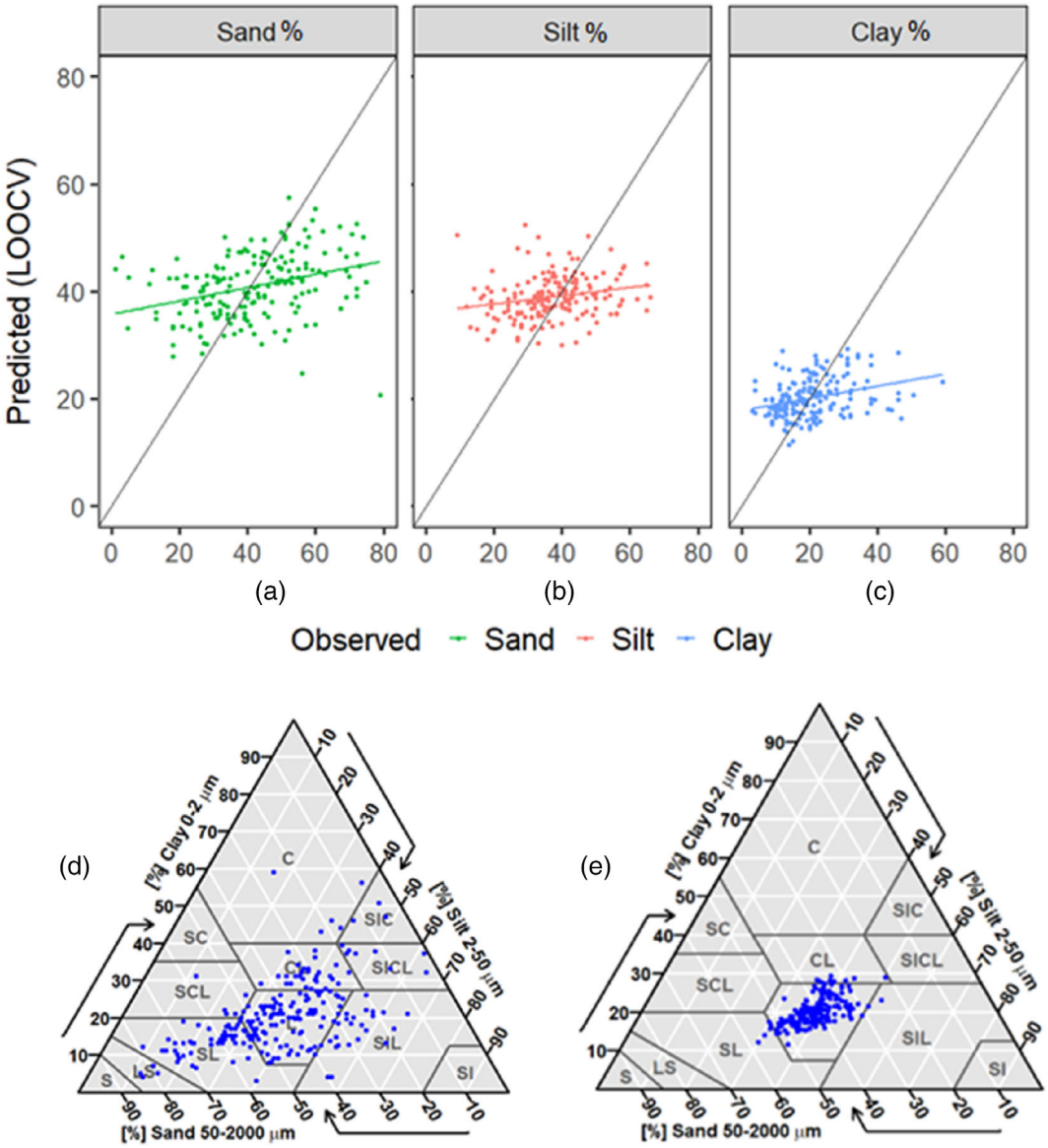

FIGURE 8 Comparative plots of observed and predicted samples (Figure 8a–c) obtained from the Random Forest Modelling with CLR transformation applied to the response variables and validated with LOOCV (**Approach 2: SAR + Topography + Geophysical** covariates). The responses are measured in % content 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 8d, and the classes resulting from predictions of soil particle-size fractions are exhibited in 8e.

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 generalized- and elastic modelling in which any variable with a coefficient of zero was dropped from the model—See Supplementary Information: Models (Data S2). 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.

5.3 | 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 second-order interaction terms. When interactions between covariates were 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 Supplementary Figures 15 and 16, wherein the measure of variable importance was enhanced (e.g., altitude, slope, RVI, SSM) while others are weakened (e.g., curvature).

5.4 | 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. 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^0_{VV}, σ^0_{VH}) 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 in 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 (Supplementary Tables 1 and 2). Moreover, the evaluation strategy did not play a significant role in this case, since data splitting and leave-one-out cross-validation produced similar results. On the contrary, when the linear model was 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 generalized- and elastic modelling under different penalty coefficients—See supplementary information: Models (Data S2). 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 material 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.

5.5 | 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 physical-geographically heterogeneous area. Furthermore, they provide a potential basis and methodological approach to obtain either estimate 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 summarized 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^0_{VV} and σ^0_{VH}) could be used as predictors. The intensity of the backscattering coefficient σ^0 (a normalized measure of the radar return) is a function of the physical and electrical properties of the target, along with the wavelength (λ), polarization and incidence angle (θ) of the radar (Barrett 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 contrary, controlling different land cover conditions in this case could be more of a challenge-based issue than a limitation issue (Supplementary Tables 5 and 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).

6 | 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^0_{VV}, σ^0_{VH}) 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 second-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 (Barrett et al., 2009). This procedure will also follow Jagdhuber (2012:4).

AUTHOR CONTRIBUTIONS

Sandra Cristina Deodoro: Conceptualization; data curation; formal analysis; methodology; visualization; writing – original draft; writing – review and editing. **Rafael Andrade Moral:** Formal analysis; methodology; writing – review and editing; supervision. **Reamonn Fealy:** Investigation; resources; writing – review and editing; supervision. **Tim McCarthy:** Resources; supervision; writing – review and editing. **Rowan Fealy:** Writing – review and editing; supervision; resources; formal analysis.

ACKNOWLEDGEMENTS

We thank Teagasc (Ireland), GSI (Ireland), ESDAC (Europe) and ISRIC Foundation (The Netherlands), for providing or making available the required topsoil data to run the model prediction. We also would like to thank Estevão Batista do Prado (Department of Mathematics and Statistics—Lancaster University-UK) and Antonia Alessandra Lemos dos Santos (Department of Mathematics and Statistics—Hamilton Institute Maynooth University) for their assistance with R-codes. Open access funding provided by IReL.

FUNDING INFORMATION

This work is funded under the scholarship programme 'John and Pat Hume Doctoral Awards Scheme 2020—WISH AWARD', from Maynooth University, Co. Kildare, Ireland.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in Irish Soil Information System (SIS) at http://gis.teagasc.ie/soils/. These data were derived from the following resources available in the public domain: - LUCAS-ESDAC Topsoil 2009 Survey-v1, https://esdac.jrc.ec.europa.eu/content/lucas-2009-topsoil-data - Soil Carbon Project 2008, https://www.ucc.ie/en/hydromet/soilcarbon/ - WoSIS Soil Profile Database, https://www.isric.org/explore/wosis - Sentinel 1, https://scihub.copernicus.eu/ - Surface Soil Moisture, https://land.copernicus.eu/global/products/ssm - Topsoil_Irish Soil Information System (SIS), https://data.gov.ie/dataset/irish-soil-information-system-national-soils-map - LUCAS-ESDAC Topsoil 2015 Survey, https://esdac.jrc.ec.europa.eu/content/lucas2015-topsoil-data - Digital Elevation Model-DEM EEA-10m (Copernicus European DEM, https://spacedata.copernicus.eu/collections/copernicus-digital-elevation-model - Geophysics: Tellus project (Geological Survey Ireland - GSI), https://www.gsi.ie/en-ie/data-and-maps/Pages/Geophysics.aspx

ORCID

Sandra Cristina Deodoro https://orcid.org/0000-0002-6141-7577
Rafael Andrade Moral https://orcid.org/0000-0002-0875-3563
Reamonn Fealy https://orcid.org/0000-0003-4534-1530
Rowan Fealy https://orcid.org/0000-0002-9601-2633

REFERENCES

Aitchison, J. (1982). The statistical analysis of compositional data (with discussion). *Journal of the Royal Statistical Society Series B*, *44*(2), 139–177.

Aitchison, J. (2005). A concise guide to compositional data analysis. Retrieved from https://eprints.gla.ac.uk/259608/

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*, *12*(2), 1637–1649. https://doi.org/10.1016/j.asej.2020.09.015

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*, 530–616. https://doi.org/10.1029/2018RG000618

Ballabio, C., Panagos, P., & Montanarella, L. (2016). Mapping topsoil physical properties at European scale using the LUCAS database (2016). *Geoderma*, *261*, 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. ESA Communication Production Office ESTEC. Retrieved from https://earth.esa.int/envisatsymposium/proceedings/sessions/4D1/460228BB.pdf

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

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

Bousbih, S., Zribi, M., Pelletier, C., Gorrab, 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*, 5–32. https://doi.org/10.1023/A:1010933404324

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

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

Cisty, M., Soldanova, V., & Cyprich, F. (2019). Random forest based reclassification of soil texture for hydrological modelling. *Geophysical Research Abstracts*, *21*, EGU2019-14320 Retrieved from https://meetingorganizer.copernicus.org/EGU2019/EGU2019-14320.pdf

Cloude, S. R., & Pottier, E. (1997). An entropy based classification scheme for land applications of Polarimetric SAR. *IEEE Transactions on Geoscience and Remote Sensing*, *35*(1), 68–78.

Comas-Cufí, M., & Thió-Henestrosa, S. (2011). CoDaPack 2.0: A stand-alone, multi-platform compositional software. In J. J. Egozcue, R. Tolosana-Delgado, & M. I. Ortego (Eds.), *CoDaWork'11: 4th international workshop on compositional data analysis*. Sant Feliu de Guíxols.

Creamer, R., & O'Sullivan, L. (2018). *The soils of Ireland. World soils book series*. Springer. https://doi.org/10.1007/978-3-319-71189-8

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 (2007-S-CD-1-S1)*. EPA.

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.

Das, K., & Paul, P. K. (2015). Present status of soil moisture estimation by microwave remote sensing. *Cogent Geoscience*, *1*(1). https://doi.org/10.1080/23312041.2015.1084669

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. https://doi.org/10.1109/TGRS.1981.350328

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*, *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. *The Soil*, *6*, 163–177. https://doi.org/10.5194/soil-6-163-2020

Egozcue, J. J., & Pawlowsky-Glahn, V. (2006). *Compositional data in the geosciences. Geological society, vol. 264* (pp. 145–159). Special Publications: Simplicial geometry for compositional data. https://doi.org/10.1144/gsl.sp.2006.264.01.11

European Space Agency-ESA, Earth Observation College-EO. (2021). Basic principles of radar backscatter. Retrieved from https://eo-college.org/courses/basic-principles-of-radar-backscatter/#learndash-course-content

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*, 137–150. https://doi.org/10.1111/ejss.12862

Filzmoser, P., Hron, K., & Templ, M. (2018). Applied compositional data analysis. With Worked Examples in R https://doi.org/10.1007/978-3-319-96422-5

Geological Survey Ireland. (2017. Scale:1:500,000). Quaternary geological map of Ireland. Retrieved from https://www.gsi.ie/en-ie/data-and-maps/Pages/Quaternary.aspx

Gholizadeha, 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

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 Retrieved from https://esdac.jrc.ec.europa.eu/public_path/Digisoil-D2.1.pdf

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.

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

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. https://doi.org/10.1109/36.20301

Jagdhuber, T. (2012). *Soil parameter retrieval under vegetation CoverUsing SAR polarimetry*. Unpublished PhD thesis. University of Potsdam.

Jones, A., Fernandez-Ugalde, O., Scarpa, S., & LUCAS 2015. (2020, ISBN 978-92-76-21080-1). *Topsoil survey. Presentation of dataset and results. EUR 30332 EN*. Publications Office of the European Union. https://doi.org/10.2760/616084 JRC121325.

Kiely, G., 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. Environmental Protection Agency.

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

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

Met Éireann-The Irish Meteorological Service. (2021). Climate Statement for March 2021. Retrieved from https://www.met.ie/climate-statement-for-march-2021

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

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

Morais & Thomas-Agnan. (2021). Impact of covariates in compositional models and simplicial derivatives. *Austrian Journal of Statistics.*, *50*, 1–15. https://doi.org/10.17713/ajs.v50i2.1069

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. https://doi.org/10.1016/0022-1694(70)90255-6

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*, *9*(4), 655. https://doi.org/10.3390/app9040655

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. https://doi.org/10.2136/sssaj2013.07.0307

Odeh, I. O. A., Todd, A. J., & Triantafilis, J. (2003). Spatial prediction of soil particle-size fractions as compositional data. *Soil Science*, *168*, 501–515. https://doi.org/10.1097/01.ss.0000080335.10341.23

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. https://doi.org/10.1111/ejss.12499

Patton, N. R., Lohse, K. A., Godsey, S. E., Crosby, B. T., & Seyfried, M. S. (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). Wiley, 272 p.

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*, 36–56. https://doi.org/10.1016/j.pce.2015.02.009

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. https://doi.org/10.3390/rs6098190

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

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

Saurette, D. D. (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

Schönbrodt-Stitt, S., Ahmadian, N., Kurtenbach, M., Conrad, C., Romano, N., Bogena, 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. *Frontiers in Water*, *3*, 655837. https://doi.org/10.3389/frwa.2021.655837

Srivastava, H. S., Patel, P., & Navalgund, R. R. (2006). How far SAR has fulfilled its expectation for soil moisture retrieval? In *Microwave Remote Sensing of Atmosphere and Environment-II, AE107* (pp. 1–12). SPIE Digital Library, 64100. https://doi.org/10.1117/12.693946

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.

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), 7409–7425.

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. https://doi.org/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. https://doi.org/10.1109/TGE.1979.294626

Walsh, S. (2012). A summary of climate averages for Ireland 1981-2010. Climatological Note no.14. Met Éireann. Dublin, May 2012. Retrieved from https://www.met.ie/climate-ireland/SummaryClimAvgs.pdf

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*, *324*, 56–66. https://doi.org/10.1016/j.geoderma.2018.03.007

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.

Wolff, C. (2007). Ground penetrating radar. Retrieved from https://www.radartutorial.eu/02.basics/Groundpenetratingradar.en.html#this

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

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

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Deodoro, S. C., Moral, R. A., Fealy, R., McCarthy, T., & Fealy, R. (2023). An assessment of Sentinel-1 synthetic aperture radar, 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