Meteorological Applications

RMetS Royal Meteorological Society

RESEARCH ARTICLE OPEN ACCESS

Deriving Gridded Soil Moisture Estimates Using Earth Observation Data and a Process Informed Statistical Machine Learning Approach

Rowan Fealy[1] | Kazeem Ishola[2] | Tim McCarthy[3] | Ajay Nair[1] | Rafael de Andrade Moral[4]

[1]ICARUS/Department of Geography, Maynooth University, Maynooth, Co. Kildare, Ireland | [2]ICARUS/National Centre for Geocomputation, Maynooth University, Maynooth, Co. Kildare, Ireland | [3]National Centre for Geocomputation, Maynooth University, Maynooth, Co. Kildare, Ireland | [4]Department of Mathematics and Statistics, Maynooth University, Maynooth, Co. Kildare, Ireland

Correspondence: Rowan Fealy (rowan.fealy@mu.ie)

Received: 7 October 2025 | **Revised:** 28 October 2025 | **Accepted:** 5 December 2025

Keywords: earth observation | energy balance | Ireland | machine learning | soil moisture

ABSTRACT

Soil moisture is classified as an essential climate variable (ECV) and is relevant to understanding hydrological, agricultural and ecological processes. Yet, in spite of its importance, direct observations of soil moisture remain limited globally—those that exist are typically limited in duration and spatial extent. Consequently, alternative approaches for estimating soil moisture have been developed, including water balance ('bucket') models, the use of remotely sensed information and the application of land surface modelling techniques. Spaceborne and land surface modelling based methods offer significant potential for monitoring and modelling soil moisture at a variety of spatial scales; however, their resolution remains relatively coarse for global and continental scale applications. At country scale, land surface models have demonstrated their potential but they require access to computational resources to deliver high resolution products. With the advent of machine- and deep- learning and data fusion techniques, high resolution global and regional soil moisture datasets are increasingly becoming available. Here, we evaluated a statistical machine learning approach to downscale the European Space Agency's (ESA) Climate Change Initiative (CCI) combined passive and active soil moisture product for Ireland using covariates that included both static (e.g., topography) and dynamic (e.g., gridded rainfall and temperature) variables. The model was developed using in situ cosmic ray neutron sensor (CRNS) measurements obtained from a network of sites in the United Kingdom, justified on the basis that the United Kingdom is geographically similar to Ireland in terms of its climate, soil types and land cover management practices. The model was found to perform reasonably well when validated against limited in situ data obtained from available time domain reflectometry (TDR) measurements available from Ireland. The developed model was subsequently used to derive spatial estimates of soil moisture on a 1 km grid across the Republic of Ireland.

1 | Introduction

Understanding the interactions and feedbacks between the land surface and the atmosphere are a foci for international research efforts; much of the recent effort has been directed towards developing a more comprehensive understanding of the role of soil moisture (SM) in modulating climate variability at various spatial and temporal scales (e.g., Dari et al. 2019; Guan et al. 2023; Zignol et al. 2024). Water retained on, or within the land surface as SM—defined as the total amount of water, including water vapour, in an unsaturated soil (AMS 2022)—acts as a limiting control on evaporation and therefore the subsequent partitioning of available energy into sensible and latent heat (Bowen ratio) (e.g., Ishola et al. 2020); and moisture into evaporation

and runoff. The partitioning of energy received at the surface directly influences the development and stability of the atmospheric boundary layer (deep/well mixed; shallow, moist). The importance of SM in limiting evaporation is particularly evident during periods of prolonged high or extreme temperatures over regions where SM is limited (SM limited regime); available energy is preferentially channelled into sensible heat which can act to amplify the initial temperature response (moisture limited regime). This is in contrast to a well-watered site, where available energy is primarily dissipated from the surface through latent heat exchanges (energy limited regime). Depending on the initial perturbation, for example a wet or dry anomaly, SM can also retain a memory of past weather events and therefore can lead to persistence in the climate system (Seneviratne et al. 2010). SM also influences the partitioning of precipitation into runoff and evaporation with consequent impacts for river discharge and in particular, flooding and time of peak flow.

Over vegetated land surfaces, the transfer of soil water to the atmosphere is primarily through evaporation or transpiration of water from leaves through the plant stomata. Plants regulate water (and carbon) transfer through their stomatal apertures; stomata close during times of water stress. When there is a plentiful supply of water and plants are growing, the stomata are open and gases (including water vapour and CO_2) are exchanged in the process of photosynthesis. Consequently, plant CO_2 uptake is directly related to SM. High temperatures, when coupled with reduced SM availability, have a significant impact on ecosystem productivity, primarily reflected in reduced primary production (Ciais et al. 2005; Bastos et al. 2014), with consequent impacts on plant and soil (autotrophic and heterotrophic) respiration and the carbon balance (Reichstein et al. 2013).

Regions can also switch between moisture limited and energy limited regimes over the course of a year or depending on land cover (Seneviratne et al. 2010); for example, Ireland has a maritime temperate climate with ample year-round precipitation, yet frequently experiences seasonal SM deficits in locations where the soils are classified as well-drained (seasonally water limited) (e.g., Ishola et al. 2023). Climate change is likely to result in an increased sensitivity for locations that already experience seasonal SM deficits and result in new areas becoming exposed (Seneviratne et al. 2010). While Ireland does not experience the extremes in temperature associated with continental Europe, it is subject to the occurrence of rainfall extremes, particularly droughts (Noone et al. 2017); the frequency and magnitude of which are projected to increase globally by the end of the century (Seneviratne et al. 2012). These changes, if realised (e.g., Fealy et al. 2018), will have consequences for ecosystems here, as the role of drought stress, rather than heat stress, has been highlighted as the primary factor in limiting ecosystem productivity (De Boeck et al. 2011).

In spite of the recognised importance of SM interactions on climate, and its relevance for understanding hydrological, agricultural and ecological processes, there is a paucity of soil water observations globally. This is due to a variety of reasons; SM often falls between administrative boundaries (meteorology and hydrology), cost and maintenance of instrumentation, heterogeneity of SM makes it difficult to obtain representative measurements (point versus areal), and so forth. Outside of a handful of countries, including the United States of America (USA) and the United Kingdom (UK), few countries have established SM monitoring networks; where data is available, this is often limited in duration and/or extent. In recognition of the importance of SM, many countries, including Ireland, have begun to coordinate and establish monitoring networks (e.g., Irish Soil Moisture Observation Network (ISMON)—Daly et al. 2021). When these new networks are fully established, they will contribute to improving our understanding of SM; however, a gap remains. Globally, efforts to coordinate and standardise SM data have also been established (e.g., International Soil Moisture Network (ISMN)—Dorigo et al. 2021).

In response, a variety of approaches have been developed to acquire, or derive estimates of, SM including from satellite based measurements (e.g., SMAP, AMSR2, SMOS, ASCAT) (e.g., Dorigo et al. 2017), empirical or process based water balance approaches (e.g., SM deficit model, Schulte et al. (2005); simple 'bucket' model based approaches—Raes (2002); Kurnik et al. (2014); soil–vegetation–atmosphere transfer (SVAT) models and the application of land surface modelling (LSM) techniques—for example, Santanello et al. 2016; Ishola et al. 2024) (Table 1) and LSMs employing data assimilation techniques (e.g., Cooper, Blyth, et al. 2021; Pinnington et al. 2021). Water balance based approaches are typical at the catchment scale, where the catchment represents a bounded box with measured inputs (e.g., rainfall) and outputs (e.g., evapotranspiration; streamflow), whereas remote sensing and LSM techniques have been employed to generate global and continental datasets (satellite: surface SM 0–5 cm; LSM: soil profile over soil layers), typically at relatively coarse spatial and/or temporal resolution (Table 1). While spaceborne based methods offer good potential for monitoring (e.g., drying/wetting), significant uncertainties remain with regards to retrieval algorithms and in locations with dense vegetation cover and organic soils (Dorigo et al. 2017). At present, they are also typically limited to a daily temporal resolution. LSM's provide an opportunity to simulate SM at a high temporal resolution (sub-daily) through the soil column; however, accuracy has been found to be dependent on the quality of the boundary or forcing meteorological datasets and physical descriptors of the landscape employed (e.g., soil hydraulic parameters etc., Ishola et al. 2024). The assimilation of remotely sensed SM estimates into a LSM has been identified as one potential way to overcome the limitations in both approaches and increase the skill in land data assimilation estimates of SM (e.g., Liu et al. 2011; Albergel et al. 2017; Pinnington et al. 2021). Recently, Batchu et al. (2023) demonstrated that a deep learning fusion model could reliably generate estimates of SM at ~300 m resolution, using both satellite (e.g., Sentinel 1/2; Soil Moisture Active Passive (SMAP)) and land surface model (e.g., GLDAS) data combined with other high resolution input fields; however, their focus was on developing spatial estimates of SM. Zhang et al. (2023) employing a machine learning approach and similar data types, generated SM estimates at 1 km and daily resolution over the period 2000–2020. At the regional or national scales, a number of high resolution SM products have also been generated using dynamical land surface model based approaches, whereby the smaller domain (relative to a global domain—GLDAS) frees up computational cost enabling outputs in the order of 1 km or less (e.g., de la Martinez Torre et al. 2018; Fealy et al. 2024; Ishola et al. 2024). Where networks of sufficient density exist,

TABLE 1 | Selection of global soil moisture products (direct retrieval, merged, modelled) covering the period to the present day [not exhaustive] (See Karthikeyan, Pan, Wanders, Nagesh Kumar, and Wood 2017; Karthikeyan, Pan, Wanders, and Nagesh Kumar 2017; Peng et al. 2021 and Liu et al. 2020 for a more detailed listing and review of soil moisture products).

Institution	Product	Technique	Spatial/temporal resolution
NASA	SMAP (Soil Moisture Active Passive)	Satellite Direct retrieval	3, 9, 36 km/1–2 Days
NASA	SMAP (Soil Moisture Active Passive) and Sentinel 1	Satellite Direct and merged	1–3 km/1–2 Days
BEC	SMOS (Soil Moisture and Ocean Salinity)	Satellite Direct retrieval	25 km/Daily
EUMETSAT	ASCAT (Advanced Scatterometer)	Satellite Direct retrieval	12.5, 25, 50 km/1–2 Days
NASA	AMSR2 (Advanced Microwave Scanning Radiometer2)	Satellite Direct retrieval	25 km/Daily
ESA	Climate Change Initiative (CCI) Soil Moisture Product	Merged active and passive sensors	~25 km (0.250)/Daily
NOAA	SMOPS (Soil Moisture Operational Products System)	Derived from multi-satellite and sensors	~25 km (0.250)/6 Hourly
NASA	GLDAS (Global Land Data Assimilation System)	Data assimilation with Land Surface Model	~25 km (0.250)/3 Hourly
Sungmin and Orth (2021)	SoMo.ml	Machine Learning (ML) trained on in situ data, scaled using ERA5	~25 km (0.250)/Daily
Han et al. (2023)	GSSM1 km	Machine Learning (ML) trained on 2443 global in situ sites	1 km/Daily
Zhang et al. (2023)	Derived using GLASS, ERA5-reanalysis and auxiliary data	Ensemble machine Learning (XGBoost)	1 km/Daily
Batchu et al. (2023)	Derived using Sentinel-1/2, SMAP, GLDAS and auxiliary data	Deep Learning Data Fusion Model	320 m

in situ measurements have been employed to generate high resolution spatial estimates (e.g., Zhao et al. 2020).

Here, we sought to evaluate an approach to link available and existing large volume earth observation data with available in situ data, including both CRNS and point sensors, to derive gridded spatial estimates of SM for Ireland. A number of challenges exist in this regard. Ireland does not currently have an in situ network for monitoring SM, akin to the comprehensive meteorological or hydrological networks that exist here and elsewhere; SM measurements that do exist have typically been obtained as part of short term (~ < 3–5 years) eddy covariance measurements from specific land cover types (e.g., grass, arable) (e.g., Peichl et al. 2012; Kiely et al. 2018). Consequently, SM measurements that are available are limited in extent, representation (e.g., land cover, soil type, etc.) and duration. Consequently, we obtained in situ data from the UK, which has an established in situ SM monitoring network based on CRNS, experiences a similar climate (predominantly temperate/maritime temperate—Köppen climate zone Cfb) and has soils and land management practices that are comparable to Ireland. Additionally, the availability of suitable optical and thermal satellite-based information is limited due to cloud coverage here, restricting the type of study (e.g., time limited spatial evaluation over available scenes; reliance on derived products from optical based satellite platforms) or satellite platform (e.g., use of satellite based cloud penetrating radar). However, the availability of merged active and passive microwave satellite data products, for example, ESA Climate Change Initiative (CCI) Soil Moisture Product (Plummer et al. 2017), which are insensitive to cloud cover, provide a potential pathway forward.

2 | Review

Earth observation data, and in particular spaceborne satellite data, can provide an essential contribution to complement in situ monitoring of essential climate variables (ECVs) such as SM (Brocca et al. 2011), due to the large area/swath coverage and overpass times. While such data are not a replacement for direct in situ observations, if suitably evaluated, they are complementary and offer the potential for use in a wide range of applications. A variety of approaches exist in this regard ranging from the use of optical and thermal data to microwave data or the combination of both optical and radar.

In an early application of satellite data to estimate surface SM, Carlson et al. (1994) employed the SVAT model with surface

radiant temperature and normalised difference vegetation index (NDVI) (e.g., Taktikou et al. 2016; Zhang and Zhou 2016), derived from the NOAA-11 satellite over an agricultural watershed in Pennsylvania, USA. Following the removal of cloudy pixels and selection of low relief areas, their analysis was confined to two dates in July (7 and 18 July 1990). They found that the model generated outputs that were 'qualitatively realistic' in terms of the spatial distribution, but highlighted the impact of high fractional vegetation amounts. Their approach, which is referred to as the trapezoid or triangle method (e.g., Price 1990), is based on an interpretation of the relationship between a remotely sensed vegetation index and surface temperature, which when plotted resembles a triangle. The triangle method has been widely applied to estimate SM; however, the approach has a number of recognised limitations (e.g., Carlson 2023).

The Water Cloud Model (WCM), originally developed by Attema and Ulaby (1978), is a widely used semi-empirical model applied to estimate SM over vegetated areas using satellite imagery. It employs vegetation descriptors, such as NDVI or Leaf Area Index (LAI) to account for the impact of vegetation on radar backscatter and assumes that only SM varies in the period of interest; all other parameters such as vegetation water content and soil surface roughness are assumed to be sufficiently time invariant. However, agricultural practices associated with vegetation dynamics, ploughing and rainfall events smoothen soil roughness and impact on the vegetation dielectric constant, which led Sabaghy et al. (2018) to conclude that SM retrieval from radar backscatter remains challenging. Additionally, common to all satellite-based methods, SM retrieval is confined to the first few centimetres of the soil layer.

More recent efforts have focused on the use of the physically based Optical TRApezoid Model (OPTRAM), proposed by Sadeghi et al. (2017), which employs shortwave infrared transformed reflectance and NDVI to estimate SM. The OPTRAM model was developed to address limitations with the 'trapezoid' or triangle model, which requires satellite-based thermal data and calibration for each observation date. OPTRAM has been validated with data from a range of satellite platforms, including Landsat-8, MODIS and Sentinel 2, and for different climate conditions (Babaeian et al. 2019). Babaeian et al. (2019) utilised OPTRAM, but replaced NDVI with the soil adjusted vegetation index (SAVI), to estimate SM at field scale. The outputs, which were compared to Time Domain Reflectometry (TDR) soil probes, were found to largely reproduce the measured values and spatial variation in SM, attributed to the high-resolution imagery employed.

The use of passive (radiometer) and active (radar) microwave-based approaches offers a number of advantages over optical based approaches; the backscatter or brightness temperature is directly related to SM and they are not limited by cloud cover. Algorithms employed for both passive and active microwave sensors, required to convert the signal (e.g., brightness temperature from passive sensors; radar backscatter from active sensors) into a surface SM response, are dependent on the difference in dielectric properties between the soil and water (Engman and Chauhan 1995; Karthikeyan, Pan, Wanders, Nagesh Kumar, and Wood 2017). However, measurements from microwave-based approaches are subject to attenuation by the atmosphere and vegetation, particularly at higher frequencies, and by interference from human microwave sources (Peng et al. 2021). Rainfall events can also make it difficult or impossible to separate the 'soil moisture' signal from noise (Karthikeyan, Pan, Wanders, Nagesh Kumar, and Wood 2017; Karthikeyan, Pan, Wanders, and Nagesh Kumar 2017). While the temporal resolution of both passive and active microwave sensors is typically low, there is a marked difference in the spatial resolution between the sensors. For example, the current spatial resolution of passive sensors employed for estimating SM is in the order of 25 km or greater.

The use of statistical and data driven approaches (e.g., Rodriguez-Fernandez et al. 2015) to combine data from both active (e.g., scatterometer) and passive (e.g., radiometers) satellite sensors with the aim of deriving high spatial and or temporal resolution SM estimates (e.g., Kolassa et al. 2017) has also become more widespread (Peng et al. 2017). Such combined approaches, which include the use of both thermal and optical (e.g., Carlson et al. 1994), passive and active microwave (e.g., Liu et al. 2012) and microwave and optical (e.g., Huang et al. 2020; Tong et al. 2020) sensors, can address a number of the limitations associated with the use of a single platform or sensor type. One such initiative is the European Space Agency (ESA) CCI for SM, which developed and continues to refine and improve algorithms to merge active, passive and combined active and passive microwave sensors to produce a long term harmonised and quality-controlled global daily SM product at 0.25° resolution (Gruber et al. 2019).

Karthikeyan, Pan, Wanders, and Nagesh Kumar (2017) undertook an assessment of the performance of available remotely sensed SM products with respect to temporal coverage and spatial and temporal performance. They evaluated data from eight passive sensors, including the Scanning Multichannel Microwave Radiometer (SMMR), the Special Sensor Microwave—Imager (SSM/I), the microwave imager of Tropical Rainfall Measuring Mission (TMI), the WindSAT mission, the Advanced Microwave Scanning Radiometer—Earth Observing System (AMSR-E), the Advanced Microwave Scanning Radiometer 2 (AMSR2) mission, the Soil Moisture Ocean Salinity (SMOS) mission and the Soil Moisture Active Passive (SMAP) mission; two active sensors, namely the European remote Sensing (1/2) satellites and the Advanced Scatterometer (ASCAT) from MetOP-A/B; and the combined active and passive data produced by the ESA CCI. Their evaluation was based on an assessment against >1000 in situ monitoring sites, distributed across the contiguous United States (CONUS) region, obtained from the International Soil Moisture network (ISMN). Their findings indicated that while systematic differences were evident between the different data sets, the temporal performance of the ESA CCI SM product was found to be comparable to the other products. Interestingly, higher correlations were found for all products for spatial performance, with respect to the measurement network, than temporal performance. Zhu et al. (2019) found that the ESA CCI combined data product (e.g., active and passive sensors) outperformed both the active-only and passive-only products in their analysis over test sites distributed across China, Spain and Canada. They recommend the combined data product for use in subsequent climatological and hydrological research. These

findings are consistent with Xu et al. (2020) in their analysis of the ESA CCI SM data over the Great Lakes, in the US.

Satellite downscaling techniques (regionalisation), using machine learning, have also been employed (e.g., Mohanty 2013) to estimate sub-grid scale SM and have become more widespread in recent years for estimating SM at global (e.g., Sungmin and Orth 2021; Batchu et al. 2023; Zhang et al. 2023), regional and local scales (e.g., Zappa et al. 2019; Kovačević et al. 2020; Liu et al. 2020) due to the fact that they have been found to perform as well as or better than other techniques (Sabaghy et al. 2018). Conceptually, statistical downscaling approaches have a long history of development and have found widespread application in the environmental sciences, and particularly climate science, where a range of techniques have been developed to downscale coarse resolution information from global climate models to the regional or local scale. Methods, which include those from ordinary linear regression to more complex machine learning algorithms and data fusion techniques, are predicated on the assumption that a relationship can be established between a coarse scale predictor and the regional or local scale phenomenon of interest. For SM estimation, the inclusion of ancillary predictor variables which characterise the local or site-specific geographic context (e.g., elevation, slope, soil texture etc.) has been highlighted as important (e.g., Werbylo and Niemann 2014; Mohanty et al. 2017). The inclusion of modelled data, derived from a land surface (e.g., Batchu et al. 2023) or atmospheric model (e.g., Zhang et al. 2023) has also been found to add value. Ultimately, the selection of a parsimonious suite of predictors or covariates, selected on the basis of both representing physical processes and/or having statistical relevance, are pragmatic choices when downscaling. In the context of downscaling SM from coarser scale estimates of SM, the relationship can be expressed as follows:

$$\text{Soil moisture}_{\text{in situ}} = f(\text{Soil moisture}_{\text{coarse}}, \text{Meteorology}_{\text{high res.}}, \text{Topography}_{\text{high res.}}, \text{Vegetation}_{\text{high res.}}, \text{Soil texture}_{\text{high res.}} \dots) \quad (1)$$

where soil moisture$_{\text{in situ}}$ is the measured response variable of interest; soil moisture$_{\text{coarse}}$ represents the resolution of gridded soil moisture product to be downscaled or regionalised and high res. (or high resolution) is the 'common' resolution of the ancillary covariates which represents the required spatial resolution of the downscaled soil moisture.

While there are numerous statistical methods that can be used for prediction or classification, of these, Random forests (RFs) (Breiman 2001) is one of the more popular and widespread methods employed due to its ability to model interactions and non-linear relationships, while at the same time minimising the potential for overfitting, especially when carried out within a cross-validation framework (Zappa et al. 2019).

Here, we sought to develop a statistical downscaling approach, using machine learning, to downscale the ESA CCI SM product to a higher spatial resolution, using a suite of large scale predictors (Equation 1). Due to the limited SM data available for Ireland, we initially developed and calibrated the model based on data from the UK CRNS network. Following evaluation of the methodological approach, we then employed an equivalent suite of predictors from Ireland to estimate SM here.

3 | Data and Methods

3.1 | Data

3.1.1 | In Situ Data—United Kingdom

Figure 1 illustrates the UK Centre for Ecology and Hydrology (UKCEH) COMOS-UK cosmic ray neutron sensor (CRNS) site network where in situ measurements of SM were available for the period 2013–2019. The daily data, including site metadata, were obtained from the UKCEH Environmental Information Data Centre (EIDC) and contain information on sites, instrumentation data and processing and quality control. Daily data included the date, sitename, CRNS derived volumetric water content (%) and the D86_75M (cm) variable, the effective sensing depth (D86) within the source area footprint (75 m radius), following the method of Köhli et al. (2015). For a detailed description on the data and processing of the CRNS data, see Boorman et al. (2020) and Cooper, Blyth, et al. (2021); Cooper, Bennett, et al. (2021). In addition, metadata for each site was also obtained, outlining the soil type (e.g., mineral, organic, calcareous), altitude, recorded land cover and bulk density (Boorman et al. 2020). Daily aggregated values from the CRNS, obtained from UKCEH, were derived from the daily averaging of sensor counts and not the hourly estimated volumetric water content values (Stanley et al. 2019).

3.1.2 | In Situ Data—Ireland

Only limited in situ data were available at the time the method was developed from four sites in Ireland (Table 2) and were collected as part of the monitoring of surface fluxes associated with eddy covariance measurements. Data was obtained either directly from site owners, as in the case of Johnstown Castle (Teagasc) and the Carlow Grassland (TCD), or through the European Fluxes Database Cluster (http://www.europe-fluxdata.eu/) for Dripsey and Donoughmore (Table 2). Critically, the lack of coherent and/or consistent metadata associated with the limited available in situ data (instrument type; placement and measurement depth; site relocation) means that the subsequent analysis is limited. For example, the in situ (point) measurements at Johnstown Castle appear to be consistent during 2013; however, following a site relocation, the sensors display an offset in readings between both sensors at the measurement site.

3.2 | Gridded Data

Daily 1 km gridded data for maximum and minimum temperature and precipitation (HadUK-Grid v1.1.0.0) were obtained from the UK Meteorological Office and accessible from the CEDA Archive, which forms part of the NERC Environmental Data Service. The daily data, which are available from 1960 to 2021, were interpolated from in situ meteorological observations from across the UK (Hollis et al. 2018). The data are available as CF compliant netCDF files on the British National Grid (EPSG: 27700).

FIGURE 1 | Location of COSMOS-UK sites. *Source:* Adapted from Cooper, Blyth, et al. (2021); Cooper, Bennett, et al. (2021); licenced under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

Daily 1 km gridded data for maximum and minimum temperature and precipitation, covering the Republic of Ireland, were also obtained from the Irish meteorological agency, Met Eireann. The data are provided as comma separated variables (csv) on TM65/Irish National Grid (EPSG: 29902), with the northing and easting associated with each point. The data

TABLE 2 | List of available in situ (point) soil moisture collected as part of various projects (CarboEuropeIP; EPA funded; European Flux Database Cluster data (Papale et al. 2006)).

Site	Owner/source	Land cover	Measurement (sensors × time)	Period	Soil
Johnstown Castle	Teagasc	Grass	2 × 30 min SWC	2013, 2018–2021	Imperfect drainage; Sandy loam
Carlow	TCD	Grass	30 min SWC (%vol)	2003	Moderately drained; medium to loamy textured; grey/ brown podsolic soil
Dripsey[a]	UCC/European Fluxes Database (Kiely et al. 2008)	Grass	2 × 30 min 1 × 30 min 3 × 30 min	2003–2007 2004–2006 2002–2012	Poorly drained/ reclaimed; Organic topsoil
Donoughmore	UCC/European Fluxes Database (Kiely et al. 2008)	Broadleaf forest	30 min SWC (%vol)	2009–2012	Poorly drained; organic soil

Note: The measurement columns indicate the number of sensors, time resolution and depth where known or reported. No meta data were available for soil water content depths and are assumed from published literature.
[a]Reported depths indicated 5 cm (European Flux Database Cluster); Published information indicates measurement depths 5, 10, 25 and 50 cm (Jaksic et al. 2006).

were converted to gridded netCDF format using the ncdf4 (Pierce 2023) and raster (Hijmans 2023) packages in R (R Core Team 2021).

Daily mean temperature was calculated from the available maximum and minimum temperature for both the HadUK-Grid and Met Eireann gridded data. All the gridded meteorological data was subsequently reprojected to WGS84 (EPSG: 4326).

Table 3 provides a description of the available in situ (CRNS) measurements from the UKCEH COSMOS network and candidate predictor variables used in the training dataset for the ML-based approach, covering the UK, Northern Ireland and the Republic of Ireland.

Gridded data on soil textural properties, including bulk density, clay, sand and silt, were obtained from the World Soil Information Service (WoSIS) Soilgrids (Poggio et al. 2021). SoilGrids provides global soil information at 250 m and 1 km resolution at six depth layers. Global maps were derived using machine learning and statistical methods (quantile regression forests) to estimate soil properties based on soil observations from almost 240,000 locations (Poggio et al. 2021). Data from the soil layer representative of the 5–15 cm depth was subsequently employed in the ML model as this layer had the highest frequency of CRNS returns.

3.3 | Satellite Derived Data

Gridded (0.25°) satellite derived soil moisture was obtained from the European Space Agency Climate Change Initiative Soil Moisture (ESA CCI SM) (Preimesberger et al. 2021). Data version v05.2 was originally obtained but subsequently replaced by version v07.1 as the latest release. V07.1 employs an intra-annual bias correction method for harmonisation of sensors with improved temporal and spatial coverage and data is available from 1978 to 2021. Data from the combined active and passive sensors was obtained for the period from 2010 to 2021, as the temporal and spatial coverage prior to this period is relatively poor.

Elevation data was obtained from the EU DEM (v1.0), a hybrid digital elevation model based on SRTM and ASTER GDEM which has been fused using a weighted averaging approach. The data available at a spatial resolution of 500 m was obtained through the Copernicus Land Monitoring Service. The tile locations covering the UK and Ireland were obtained, merged and reprojected to WGS84 (EPSG: 4326). Landscape morphometry indices were derived from the elevation data, including slope, aspect, roughness and index of topographic position (TPI) and wetness (TWI) and general curvature using the raster R package (Hijmans 2023).

MODIS landcover and vegetation indices were obtained from the Application for Extracting and Exploring Analysis Ready Samples (AρρEEARS) service provided by the NASA Earthdata services. The following data were derived from the Terra and Aqua combined Moderate Resolution Imaging Spectroradiometer (MODIS):normalised difference vegetation index (MOD13A1) Version 6.1 (Didan 2021), enhanced vegetation index (MOD13A1) Version 6.1, land cover type (MCD12Q1) Version 6 (Friedl and Sulla-Menashe 2022) and land surface temperature (MOD11A1) (Wan et al. 2021). The vegetation indices were available at 16-day intervals. Following visual inspection of the data, values were subsequently filtered where a pixel quality code of >1 was re-coded to missing. Remaining outliers in the data were removed using the tsoutliers R package (Hyndman 2021) following the method of Chen and Liu (1993). A Savitzky–Golay smoothing filter, using the signal R package (Signal developers 2013), was subsequently applied to the vegetation indices to smooth and temporally interpolate the data to match the temporal resolution of the meteorological and satellite SM data.

TABLE 3 | List of available in situ measurements from the UK CEH COSMOS network and candidate predictor variables for the training dataset covering the UK, Northern Ireland and the Republic of Ireland for the Randon Forest ML based approach.

Name of index	Source	Spatial resolution	Input type	Description
ID	Cosmos UK/CEH		In situ	Identification number of site
Date	Cosmos UK/CEH		In situ	Time range
Location	Cosmos UK/CEH	Point	In situ	Location of the site/area of interest
Latitude	Cosmos UK/CEH	Point	In situ	Latitude of the site/area of interest
Longitude	Cosmos UK/CEH	Point	In situ	Longitude of the site/area of interest
Altitude	Cosmos UK/CEH	Point	In situ	Altitude in meters
Land cover	Cosmos UK/CEH	Areal	In situ	Recorded land cover at site
Soil Moisture	Cosmos UK/CEH	Areal	In situ	Volumetric water content from COSMOS UK locations derived from CRNS
D86_75M	Cosmos UK/CEH	Areal	In situ	Effective depth. D86 at 75 m distance from cosmic-ray sensor; Spatial footprint ~100 s metres.
CCI Soil Moisture	ESA-CCI-Soil moisture	25 km	Satellite	Soil moisture from 25 km resolution ESA-CCI combined
MODIS NDVI	MOD13A1	500 m	Satellite	Normalised Difference Vegetation Index
MODIS EVI	MOD13A1	500 m	Satellite	Enhanced Vegetation Index
MODIS Land Cover Type	MCD12Q1	500 m	Satellite	Land cover type
MODIS Land Surface Temperature	MOD11A1	1 km	Satellite	Land surface temperature
Digital Elevation Model	EU DEM	500 m	Satellite	EU DEM 500 m
Aspect		500 m	Derived	Aspect refers to the compass direction that a hillside or slope faces, value in degrees. 4 neighbour cells.
Slope		500 m	Derived	Slope represents the rate of change of elevation. 4 neighbour cells.
Roughness		500 m	Derived	Roughness is the degree of irregularity of the surface. Calculated as the difference between the maximum and the minimum value of a cell and its 8 surrounding neighbours.
TPI		500 m	Derived	Topographic positional index—difference between the value of a cell and the mean value of its 8 surrounding cells
TRI		500 m	Derived	Topographic ruggedness index—mean of the absolute differences between the value of a cell and the value of its 8 surrounding neighbours.
TWI		500 m	Derived	Topographic wetness
General Curvature		500 m	Derived	General curvature of the landscape

(Continues)

TABLE 3 | (Continued)

Name of index	Source	Spatial resolution	Input type	Description
Bulk Density	SoilGrids	250 m	Soil	Bulk density of the fine earth fraction (5–15 cm depth)
Clay	SoilGrids	250 m	Soil	Proportion of clay particles (< 0.002 mm) in the fine earth fraction (5–15 cm depth)
Sand	SoilGrids	250 m	Soil	Proportion of sand particles (> 0.05 mm) in the fine earth fraction (5–15 cm depth)
Slit	SoilGrids	250 m	Soil	Proportion of silt particles (≥ 0.002 mm and ≤ 0.05 mm) in the fine earth fraction (5–15 cm depth)
Maximum Temperature	HadUK-Grid; Met Eireann	1 km	Weather	Daily maximum air temperature measured between 0900 UTC on day D and 0900 UTC on day D + 1 (°C) interpolated to 1 km grid
Minimum Temperature	HadUK-Grid; Met Eireann	1 km	Weather	Daily minimum air temperature measured between 0900 UTC on day D-1 and 0900 UTC on day D (°C) interpolated to 1 km grid
Precipitation	HadUK-Grid; Met Eireann	1 km	Weather	Daily precipitation (mm) interpolated to 1 km grid

3.4 | Methods

3.4.1 | Machine Learning (ML) Based Approach

An initial examination of a number of statistical (e.g., linear regression) and machine learning (RF; Bayesian Analysis Regression Trees (BART); Chipman et al. 2010) approaches and transformations of the response variable (e.g., logit transformed versus raw values for SM range from 0 to 1) was undertaken. Following this initial evaluation, the RFs method was selected due to its performance relative to the other methods evaluated. RF utilises an ensemble of decision trees to generate predictions or probabilities of observations belonging to a particular category. Predictions generated from decision trees typically change significantly when small changes are made in the dataset, or when new predictors are included. Depending on different decision tree-growing algorithms, decision trees are prone to overfitting the data unless they are pruned adequately. RFs are a way of solving these issues. Since each decision tree within a RF is only allowed to use a random subset of the predictors, overfitting is attenuated, and there is no need for tree pruning (Breiman 2001). Moreover, since the predictions are now an average of the predictions generated by each tree within the RF, the variability is also smaller when changes are made to the data, or when predictors are added/removed. Therefore, RFs are an extension of decision trees that are less prone to overfitting and present less variability. Interpretability is still possible by looking at 'variable importance', which can be calculated as the decrease in explained variability if that particular variable was removed from the analysis (Kuhn 2012). The variables/predictors that yield greater decrease in explained variability are, therefore, the most important in predicting the response variable.

To generate the training dataset, temporally varying (e.g., meteorological data; vegetation indices, land cover) and static (e.g., elevation, soil properties) data were extracted for each COSMOS CRNS site location in the UK for the period from 2013 to 2019. The period from 2013 reflects the start of good/improved in situ (CRNS) station coverage. Prior to training the model, the extracted data were compared to the site information obtained from UKCEH where comparable measurements were available. Site information and locations were also assessed using a visual assessment with Google Maps. A number of issues were identified at this point. For example, one of the sites had incorrect site coordinates, which was subsequently rectified. A more significant methodological issue was identified with regard to the recorded land cover and the land cover extracted from the MODIS Landcover data, specifically where the site recorded land cover did not match the satellite derived information; this is not unexpected due to the differences in scales between the MODIS data and the site level information. Consequently, land cover data were obtained from other sources, including a visual assessment of historical imagery from Google Earth, the Global Landcover (GLC) (2015–2019) and CORINE Land Cover (2018) and the EU Land use and land cover survey (LUCAS) (2013; 2015). While land cover classes were not directly comparable between the different datasets, it was possible to subjectively map the different definitions between the land cover classes. Following this preliminary assessment, the MODIS land cover data were subsequently employed; this was also a pragmatic choice for the subsequent model application to the Irish domain.

TABLE 4 | List of covariates employed in the Random Forest model to estimate soil moisture without (NDVI-) and with NDVI (NDVI+), selected based on the independent training period from 2013 to 2018.

Name of index	Description	Variable importance (% of highest reduction in prediction error) NDVI−	Variable importance (% of highest reduction in prediction error) NDVI+
Latitude	Latitude of site/area of interest	4.5	5.65
Longitude	Longitude of site/area of interest	3.8	3.42
Season	Categorical var.—DJF [1], MAM [2], JJA [3] and SON [4]	51.4	49.17
D86_75M [mean]	Site mean effective depth. D86 at 75 m	26.7	22.41
CCI Soil Moisture	Soil moisture from 25 km resolution ESA-CCI combined	100	83.52
NDVI	Normalised Difference Vegetation Index (NDVI)	—	100
MODIS Land Cover	Land cover type	0.3	0.00
Digital Elevation Model	EU DEM 500 m	5.1	5.21
Aspect	Aspect based on 4 neighbour cells.	2.9	3.64
Clay	Proportion of clay particles (5–15 cm depth)	0.0	2.26
Sand	Proportion of sand particles (5–15 cm depth)	2.7	5.68
Mean Temperature	Calculated from the max and min gridded temperature	65.1	54.62
Precipitation	Daily gridded precipitation (mm)	29.4	42.37

TABLE 5 | Pearson's *R* values summarised by MODIS Landcover class for all sites with available data for 2019 (*n*, number of sites available for 2019 = 48).

MODIS landcover class	Pearson's *R*	*n*
1 Evergreen Needleleaf Forest	0.493	1
4 Deciduous Broadleaf Forests	0.825	1
7 Open Shrublands (woody perennials)	0.881	1
9 Savannas (Tree cover 10%–30%)	0.765	3
10 Grasslands	0.886	27
12 Croplands	0.930	15

Additionally, several site locations were excluded from the training data. Due to the spatial resolution of the ESA CCI SM data (0.25°), any CCI grids that overlay a site location less than 12.5 km (half the grid resolution) from the coast were excluded to limit the impact of sea grids on the estimated SM. Other sites were also excluded based on a visual assessment of the measured data. Excluded sites included Redmere, Harwood Forest, Moreton Morrell and Plynlimon which were removed for a number of reasons including poor quality data, including a high number of missing values or recorded high values—for example, Harwood Forest had a mean of 75% VWC, with periods of 100% VWC being recorded over the period of record. This resulted in a total of 36 sites out of 51 available for inclusion in the training data. Screening of time periods was also undertaken to remove the contamination effect of snow cover on the satellite derived SM. A pragmatic approach was taken to exclude dates based on days that recorded a mean temperature less than 0°C and where rainfall had been recorded. Prior screening of the potential covariates was undertaken to obtain a parsimonious suite of predictors for use in the model training process. Screening involved the usual statistical tests of association, principal components analysis (PCA) to identify groups of related predictors and evaluation of relevance from a physical process point of view.

The initial screening identified the effective depth (D86_785M) covariate to be the dominant variable of importance, higher than the ESA CCI SM and meteorological variables. Following investigation of the supporting documentation for the COSMOS-UK data, it was determined that the D86_75M variable is not an independent variable; the calculation of effective depth is a function of the total soil water content and therefore not suitable as a predictor. As the CRNS returns are variable depth measurements and thus require a control, we made the necessary assumption that the effective depth was a characteristic of the instrument

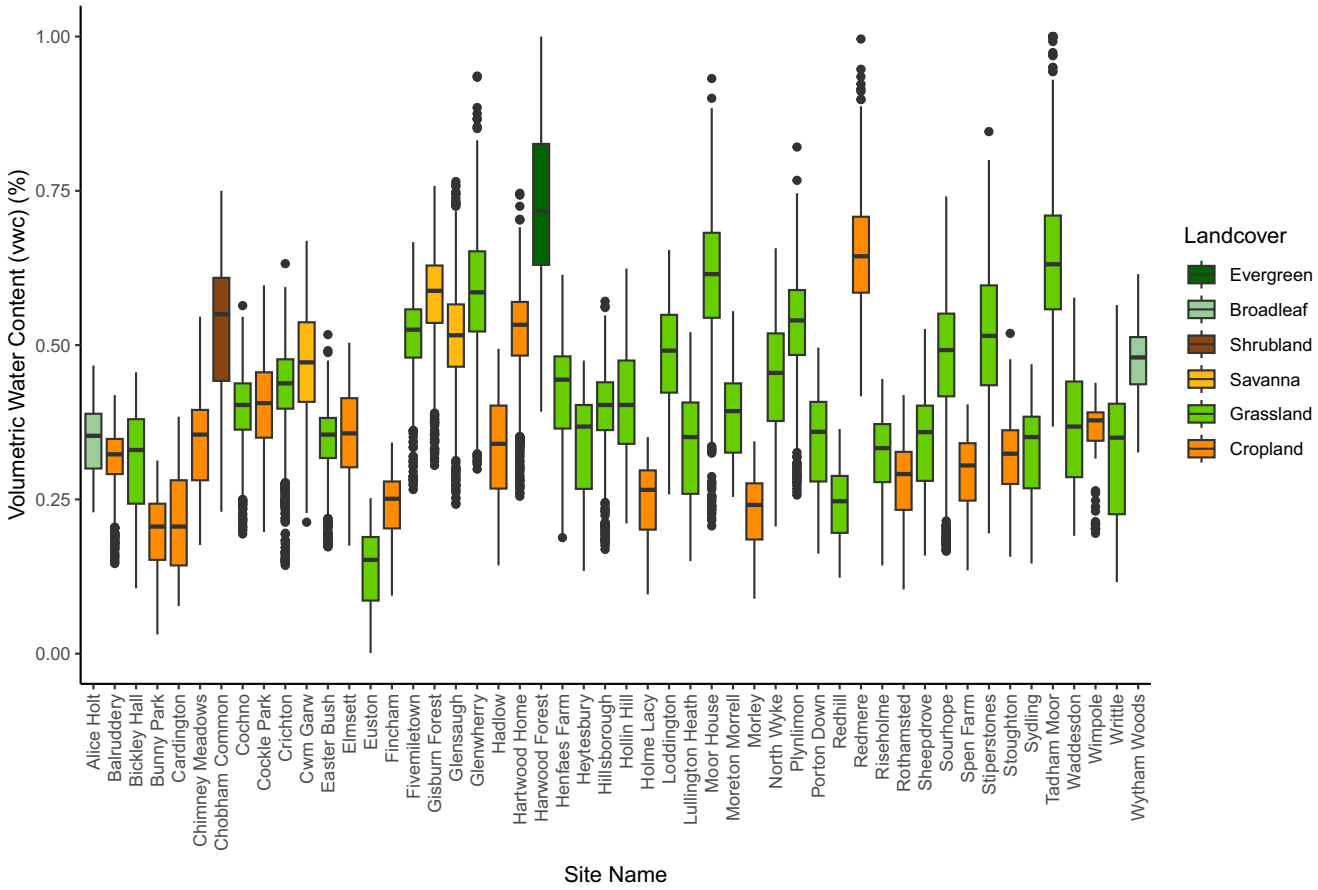

FIGURE 2 | Boxplots of observed volumetric water content (%) from the COSMOS CRNS sites selected for model training (2013–2018) and evaluation (2019) over the period 2013–2019. Sites are colour coded according to their respective MODIS land cover class.

TABLE 6 | Pearson's *R* values summarised by soil data as classified at site ($N = 48$).

Soil type	Pearson's *R*	No. of sites
Mineral soil	0.930	32
Calcareous mineral soil	0.889	9
Organic soil over mineral soil	0.881	1
Organic soil	0.118	6

and site, and that each site had a mean effective depth around which the CRNS returns or counts were obtained.

An extensive assessment of training and evaluation periods and candidate predictors was then undertaken. Initially, the focus of model training was on the period 2013–2018, with the remainder (2019) withheld for model evaluation and selection of the mtry parameter, a tuning parameter for the RFs algorithm, representing the number of predictors randomly selected to grow each tree. The R package CAST (Meyer et al. 2023) was used to evaluate the selection of different space and time folds (e.g., LTO—leave time out; LLO—leave location out). A number of approaches were evaluated including leaving different time (LTO) periods and stations (LLO—one at a time; random selection of sites). On the basis of this evaluation, the inclusion of 2018 in the training period was found to be important, likely due to the fact that it experienced significant SM deficits across the UK and Ireland, associated with the 2018 summer drought which provided important information on the drivers of variation of SM variation. The method was also found to be sensitive to LLO, as particular land cover types had only a small number of representative sites (e.g., Evergreen Needleleaf Forest, Deciduous Broadleaf Forests, Open Shrublands (woody perennials) and Savannas—Table 5). Thus, we opted to employ the LTO method—withholding the year of data from 2019 from the model training. A Feed Forward Selection (FFS) algorithm with LLO validation identified 13 predictors. Values for mtry and the number of trees were also evaluated using a grid search algorithm. Values from 1 to 15 were evaluated for mtry and 500, 100, 1500 for the number of trees. This resulted in the final section of mtry equal to 4 and 500 trees in the RFs.

4 | Results

4.1 | UKCEH Data—Model Training (2013–2018) and Model Validation (2019)

The suite of covariates employed in the final model is outlined in Table 4. While the model trained with and without NDVI did not differ significantly in terms of the model statistics, from a

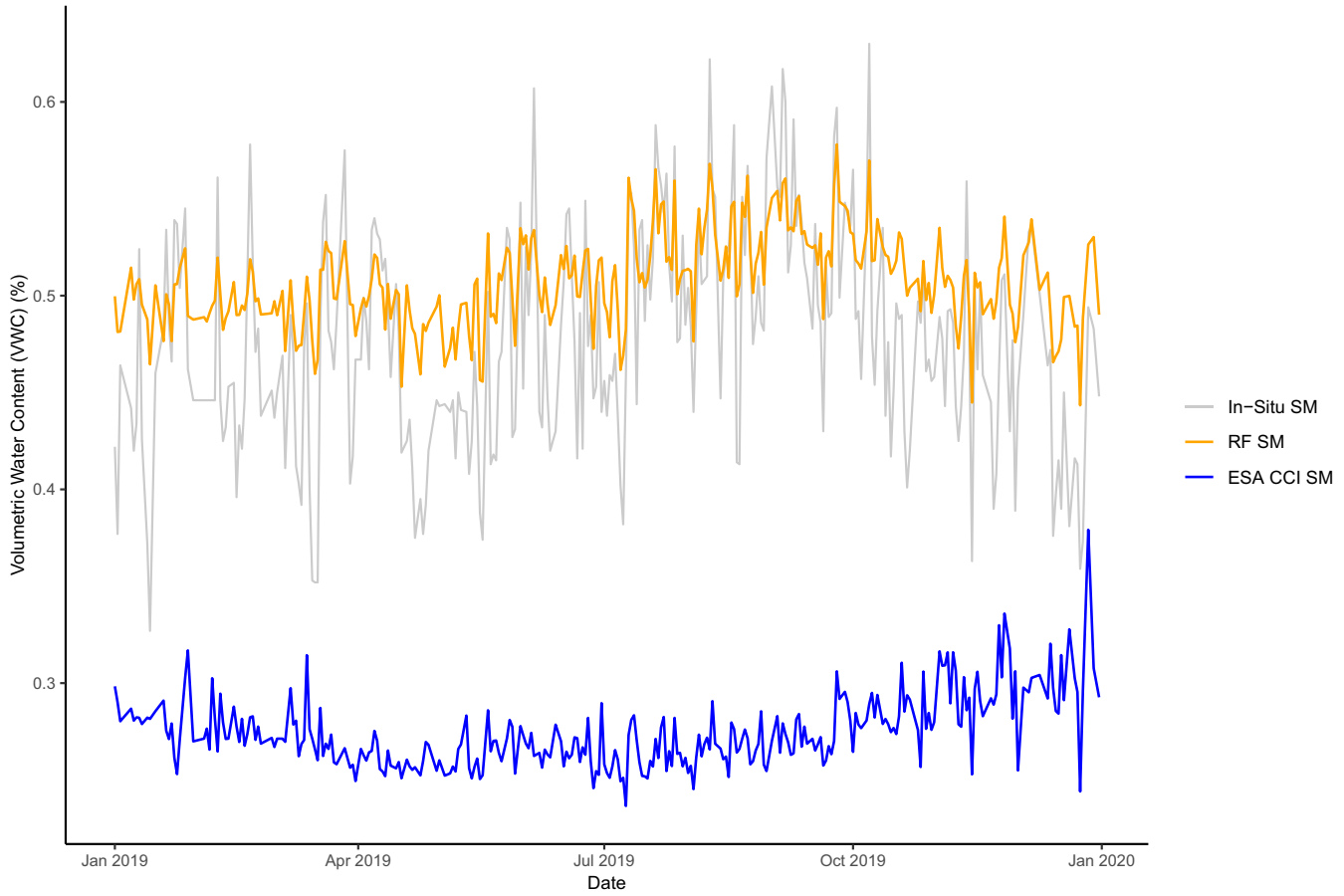

FIGURE 3 | Volumetric water context (%) for Glensaugh (Glens), classified as Savannah land cover and located on organic soil, for the independent evaluation period of 2019 (grey). The Random Forest (RF) estimated VWC are shown in orange and the ESA CCI SM in blue.

physical process point of view NDVI was included in the final model; justified on the basis that NDVI provides a measure of plant health and hence its ability to interact with the atmosphere through transpiration. Based on the independent training period (2013–2018), the model root mean squared error (RMSE) was 0.073 m^3/m^3, the mean absolute error (MAE) was 0.061 m^3/m^3 and the R^2 value was 0.569. As the model output statistics did not include the unbiased RMSE (ubRMSE) metric, originally developed to improve the estimation of the temporal dynamics of SM (Entekhabi et al. 2010), we calculate the ubRMSE and modified Nash-Sutcliff efficiency (mNSE) using the R (R Core Team 2021) package HydroGOF (Zambrano-Bigiarini 2024). We included the mNSE as it is a widely used metric for assessing the goodness of fit in hydrologic and hydroclimatic model assessments. The calculated ubRMSE was 0.079 m^3/m^3 and mNSE was 0.652—findings consistent with the RMSE and R^2 values reported in the model output.

In the absence of significant in situ SM measurements from Ireland, the RF model was initially evaluated against the UK COSMOS CRNS data for 2019, which represents an independent time period withheld from the model training period (2013–2018) (Table 5). At site level, Pearson's R values ranged from a minimum of 0.153 (Glenwherry) to a maximum of 0.969 (Wimpole). In general, the model performed best for the cropland land cover class and poorest for the evergreen forest class (Table 5). The model also appears to perform well over the Broadleaf forest and Open Shrublands land cover type, but this is based on a single site for each of these cover types (Table 5). The model performance over cropland is not unexpected based on the fact that a significant proportion of sites ($n = 15$) are classified as 'arable and horticulture' in the training data. The poorer model performance over the Evergreen Needleleaf site (Harwood Forest—Harwd) is likely associated with the fact that the model was trained on data that did not include an evergreen forest land cover class (Figure 2) as this site was excluded from the training period (2013–2018), as outlined previously.

Table 6 shows the Pearson's R values summarised according to the different soil types as classified at the measurement sites. The model displays the lowest skill in estimating SM over organic soils (Pearson's $R = 0.118$), whereas for mineral and mineral composite soils, the model appears to have good skill. Typically, organic soils are associated with higher volumetric water contents (%) and display a greater range in values compared to the mineral soils; such sites may also have a higher associated measurement uncertainty (e.g., Cooper, Blyth, et al. 2021; Cooper, Bennett, et al. 2021). Figure 3 shows the RF model-estimated volumetric water content for the independent evaluation period (2019) at Glensaugh, characterised by MODIS as Savanna (tree cover 10%–30%, canopy > 2 m) land cover type and on organic soil. Also included is the ESA CCI SM estimated VWC for the co-located grid. The ESA CCI SM values are shown

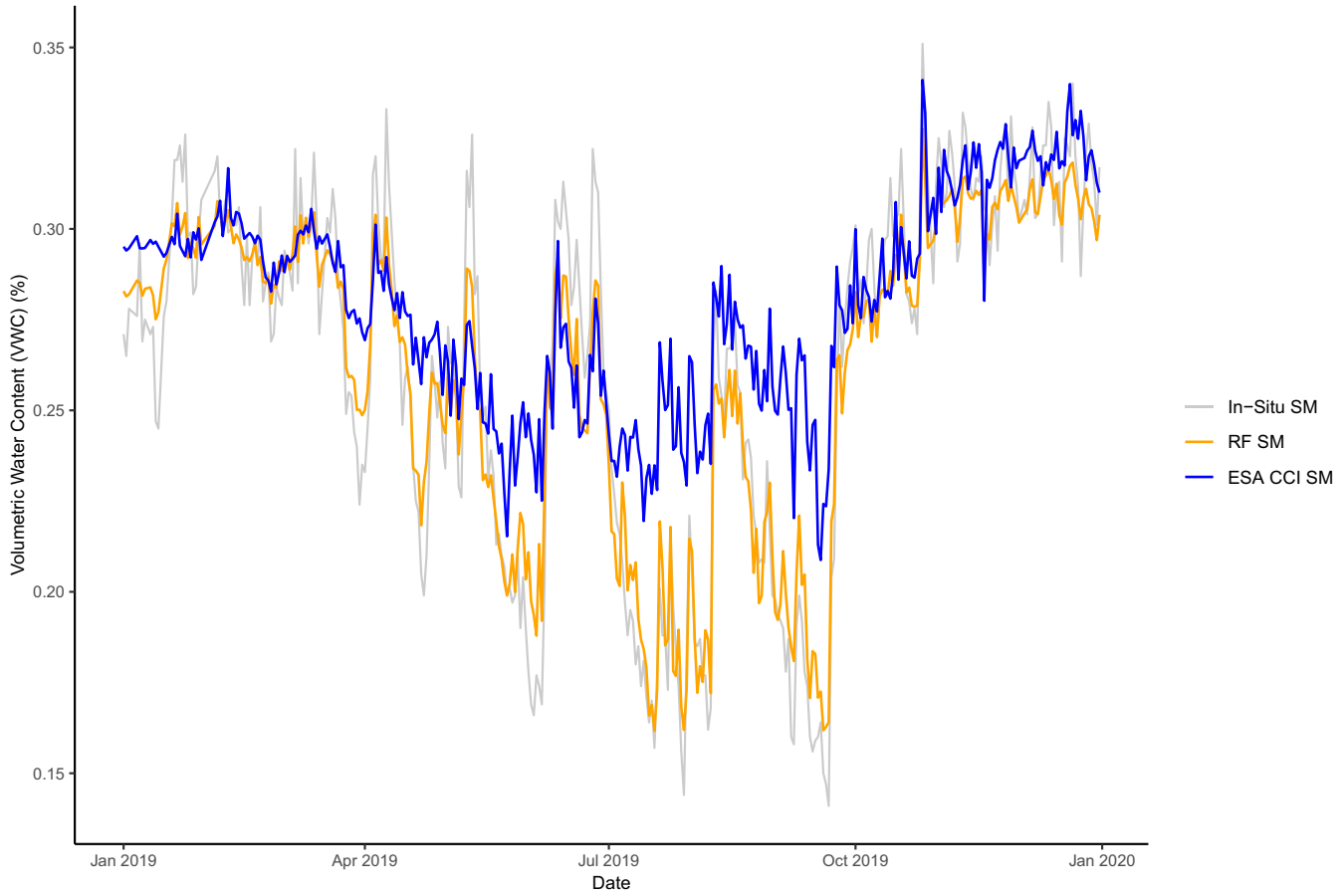

FIGURE 4 | Volumetric water context (%) for Holme Lacy (Hlacy), classified as arable and horticulture on mineral soil, for the independent evaluation period of 2019 (grey). The Random Forest (RF) estimated VWC are shown in orange and the ESA CCI SM in blue.

to significantly underestimate the measured values for this site. While the RF model does not perform well for organic soils, the model estimates lie closer to the mean of the measured values than the ESA CCI data and do appear to capture some of the temporal variability evident at this site (Pearson's $R = 0.615$).

Figure 4 shows the RF model estimated volumetric water content for 2019 at Holme Lacy, characterised as cropland land cover overlying mineral soils (Pearson's R 0.937). Also included is the ESA CCI SM estimated VWC for the co-located grid. For this site, the ESA CCI SM values provide a closer match to the measured values, compared to Glensaugh (Figure 3). Extensive areas of similar land cover are likely to be better represented in the ESA CCI SM compared to more heterogeneous landscapes, where the return will be an integrated response from surfaces with different moisture retention properties. While the model generally performed well for cropland sites on mineral soils, a low Pearson's R value (Pearson's R 0.344) was associated with a site that had these characteristics—Balruddery. Balruddery is located close to the mouth of a large river/estuary on the River Tay and was not included in the model training due to its proximity to the coast; however, the ESA CCI SM values lie close to the in situ (CRNS) values at this site suggesting that the ESA values were not overly impacted by the returns from the river/sea surface. Based on the RF model, the modelled soil water content is overestimated particularly during the winter months, likely due to either too much rain or insufficient evapotranspiration in the model. As Balruddery lies on the east coast of Scotland (Figure 1), in the rain shadow of the Scottish highlands, it is possible that the interaction between latitude, longitude and rainfall and temperature in the RF model resulted in the overestimation, particularly evident during the winter and autumn months, at this location.

In general, the model appears capable of estimating SM across the majority of UK sites (Figure 5) with different land cover (Figure 6) and on different soil types, based on the independent evaluation period. Overall, the model was found to perform better for crop and grass land cover types and for mineral soils. The model also appears capable of estimating SM in organic soils, but less skilfully than for mineral soils.

4.2 | Evaluation With Available In Situ Data From Ireland

Following the evaluation of the RF model with the UK COSMOS CRNS data for 2019, the trained model was subsequently employed in conjunction with an equivalent suite of covariates (Table 4) obtained for the Republic of Ireland. Figures 7 and 8 show the comparison between the in situ (point) measurements at Johnstown Castle for 2013 (Figure 7; Table 7) and 2018, 2019 and 2021 (Figure 8; Table 7), the RF model estimates, calculated at a depth of 7.5 cm (the mean effective sensing depth covariate was set equal to 7.5 cm in the RF model), and the ESA CCI SM. The Pearson's R values between the measured and RF estimated SM for 2013 are 0.74 and 0.82 for in situ sensors 1 and

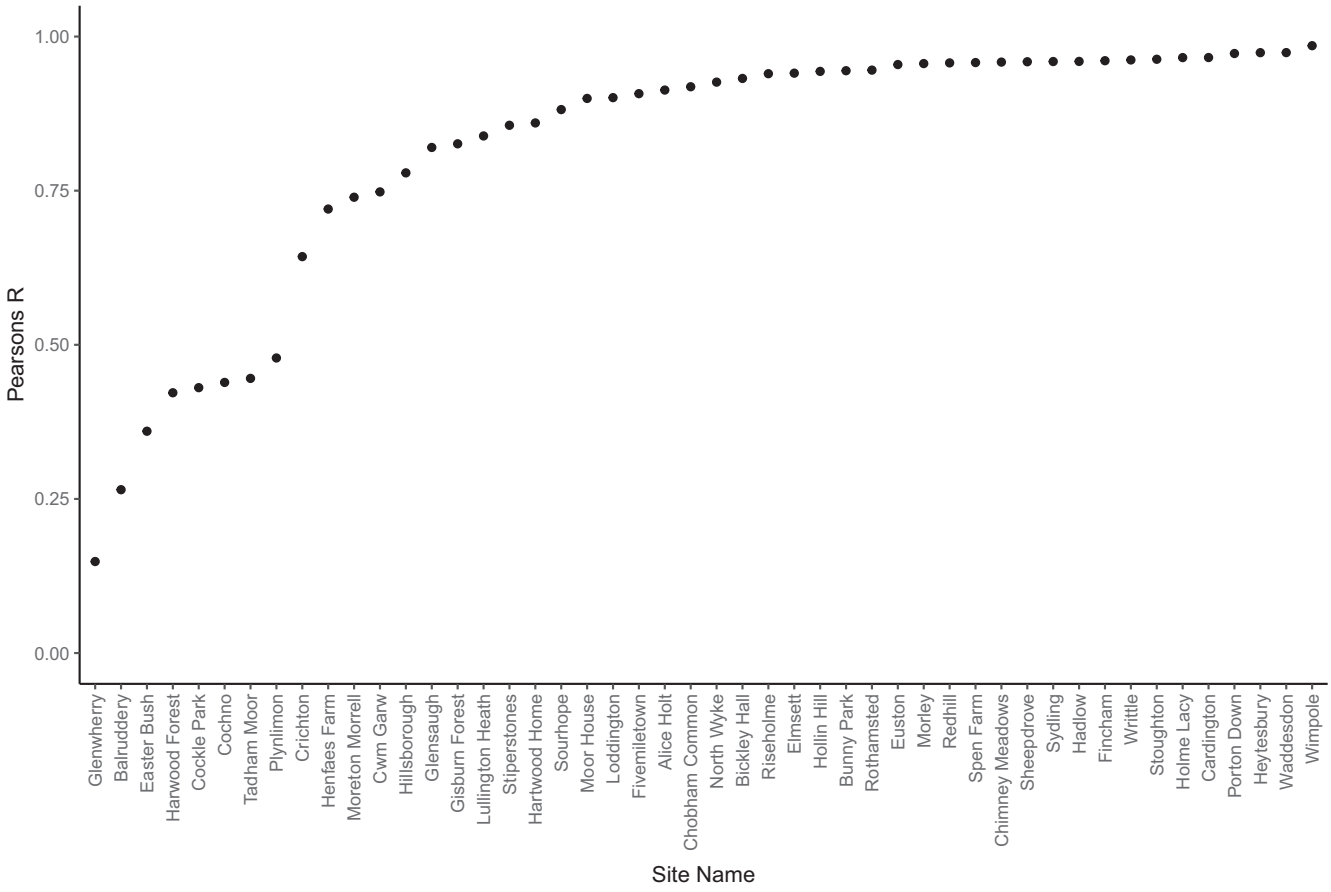

FIGURE 5 | Pearson's *R* values for each of the COSMOS CRNS sites and RF estimated soil moisture for the independent evaluation period (2019).

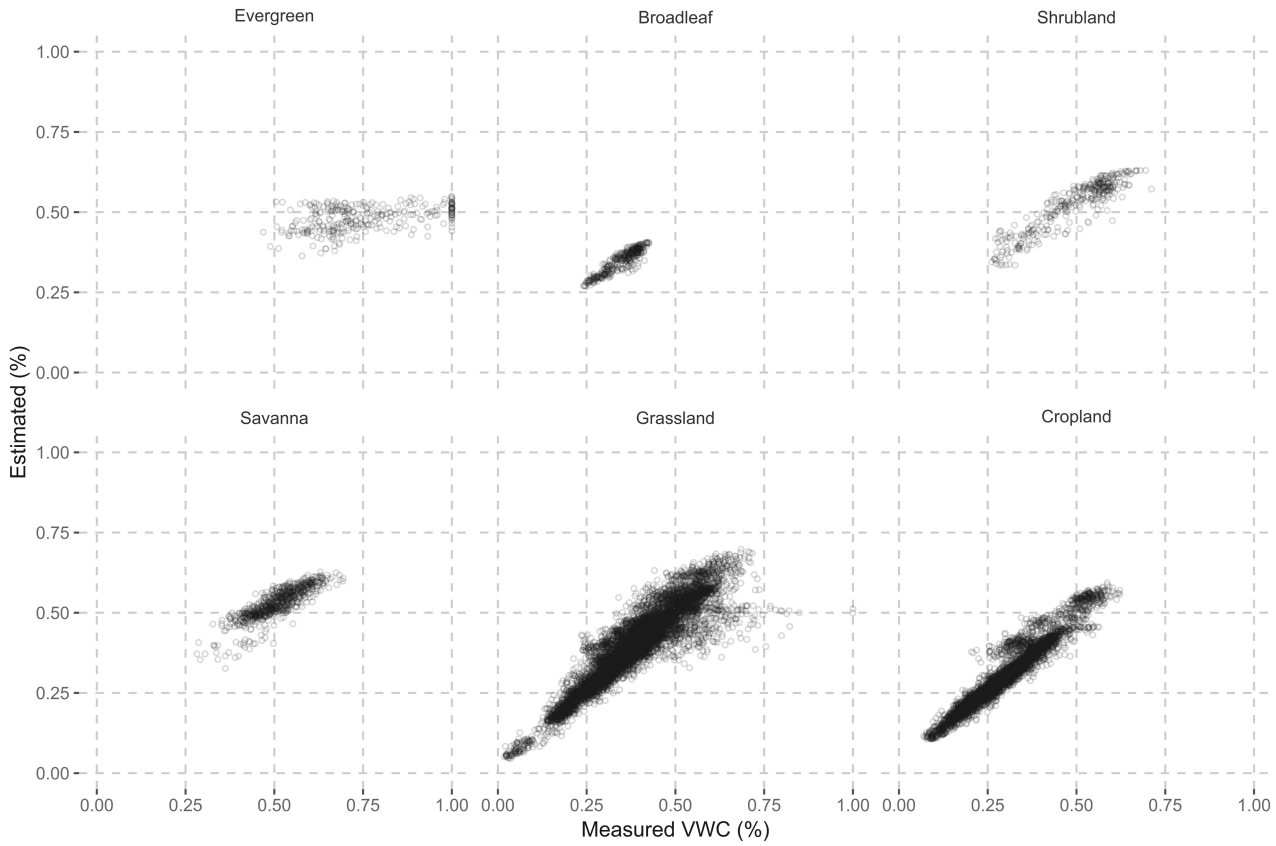

FIGURE 6 | Scatterplots of measured VWC (%) and RF model estimated VWC (%) by MODIS land cover type for the independent evaluation period (2019).

2 respectively. While in situ sensor 2 displays a higher Pearson's *R* value, the mean absolute error (MAE) and root mean square error (RMSE) are higher for this sensor (MAE 0.2, RMSE 0.21) compared to in situ sensor 1 (MAE 0.13, RMSE 0.14). These findings are consistent for the 2018 to 2021 period, with a Pearson's *R* value of 0.7 (MAE 0.05, RMSE 0.07) and 0.8 (MAE 0.18, RMSE 0.18) for in situ sensors 1 and 2, respectively.

Figure 9 shows a comparison between the in situ (point) measurements at Dripsey for the period 2002–2012 (Table 2), estimated RF SM at 11 cm and the respective ESA CCI SM grid. The estimated SM at 11 cm was derived as the mean of the 7.5 and 15 cm model estimated depth layers. This layer was selected on the basis of the reduced MAE and RMSE (Table 6) and absence of clear sensor depth information for this site. While the estimated SM appears to adequately capture the seasonality evident in the measurements, it underestimates the extent of the summer drying, particularly marked in a number of years.

At Johnstown Castle, the RF model overestimates SM with respect to both sensors, whereas the ESA CCI SM is shown to underestimate SM at the site (Figure 10). At Drispey, characterised by grass on organic soil, the mean of the RF estimated values lies much closer to the mean of the in situ (point) measured values. Again, the ESA CCI SM is shown to underestimate the measured values at this site. Similar results are found for Carlow; however, at this site, while the mean values from the ESA CCI SM product are lower, the values lie within the spread of the measured values. At Donoughmore, a site located close to Dripsey and characterised as a broadleaf forest on organic soil, both the RF estimated and ESA CCI SM underestimate the measured values, with the ESA CCI SM displaying a greater under estimation.

4.3 | Comparison With Existing High Resolution Soil Moisture Data Sets—GSSM1 Km

Han et al. (2023) employed a physics-informed machine learning (ML) approach, trained on 2443 global in situ sites, to generate daily 1 km volumetric SM, referred to as the Global Surface Soil Moisture (GSSM1 km) product. The GSSM1 km data provides surface SM representative of the 0–5 cm soil depth layer (Han et al. 2023). As the temporal and spatial resolution of this data is comparable to the outputs generated here, we include a comparison between the RF and GSSM model outputs for two sites, Johnstown Castle and Dripsey, based on data availability; both sites are independent of both the RF and GSSM model training. Figure 11 shows the available in situ (point) data from Johnstown Castle and the RF (7.5 cm) and GSSM1 km modelled outputs for the 1 km grid containing this site; Figure 12 shows the same, but for Dripsey. While the GSSM1 km data are representative of surface conditions, there are clear differences, particularly in the mean of the series, between the measured and GSSM1 km modelled values. Recognising that these differences may partly be accounted for by differences between the measured and GSSM1 km

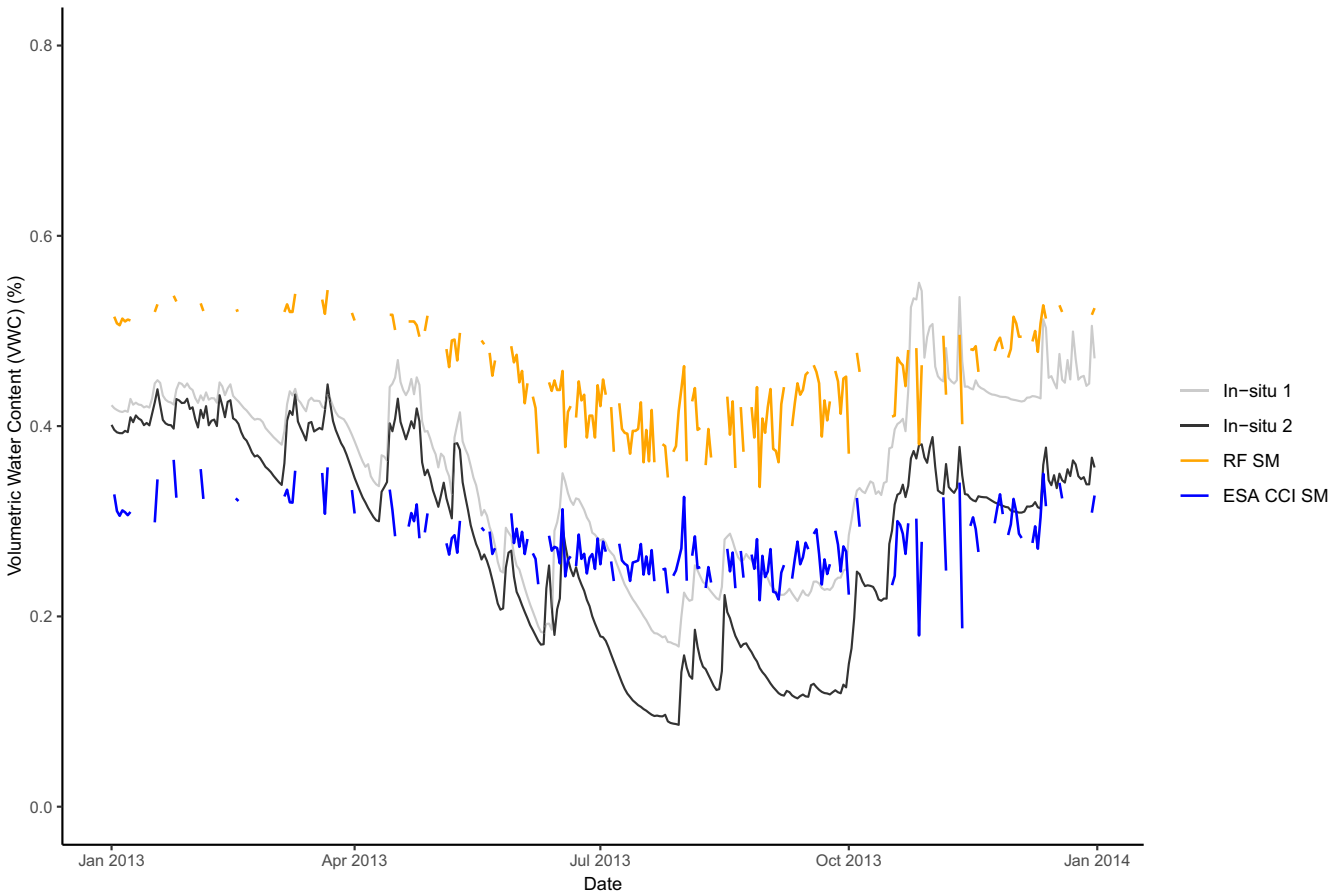

FIGURE 7 | In situ soil moisture from Johnstown Castle (5 cm), RF model estimated soil moisture (7.5 cm), and ESA CCI SM for 2013.

modelled depths, seasonal wetting and drying of the upper soil layers, particularly during wet and dry years, would be expected to diminish differences in SM in these layers.

4.4 | Spatial Soil Moisture Estimates—1 km

Figure 13 shows the seasonal mean volumetric water content (%) for the RF model, at the model estimated depth of 7.5 cm, and the ESA CCI SM for the period 2010–2021, noting that the ESA CCI SM values represent a shallower layer of soil, typically surface (0–5 cm), reflective of the sensor type. For the model depth layer shown, there is a clear seasonality to SM in Ireland, with the winter months (DJF) displaying the highest VWC ($0.53\,m^3\,m^{-3}$) and summer (JJA) the lowest ($0.44\,m^3\,m^{-3}$), when averaged across the domain. During the summer months, drier regions along the east and south coast are evident. These locations are often subject to seasonal SM deficits due to a combination of low precipitation, higher atmospheric water demand associated with higher summer temperature and freer draining soils.

Both the RF model and ESA CCI SM display a similar geographical and seasonal pattern to SM; however, the RF model estimates higher absolute values for most of the country across all seasons, with the exception of mountainous areas along the south-west and west coast. Whether these spatial differences are attributable to the difference in soil layer depths between the RF model and the ESA CCI SM values or due to the higher resolution associated with the RF model in areas of higher/undulating relief needs to be further examined. These differences, with higher ESA CCI SM values (negative bias), are most marked in the south-west of the island—an area that includes the MacGillycuddy Reeks, a mountainous region with a number of peaks over 1000 m—but with an average elevation of just over ~100 m. In contrast, the RF model estimates higher values (positive bias) at elevation over the Wicklow Mountains on the east coast of the country—which does not display the same differences in height as the south-west and has an average elevation of ~300 m.

5 | Discussion

The RF model, which was trained on data from the UK COSMOS network, demonstrated as good as or closer agreement with the point sensor data than the ESA CCI SM over the independent evaluation period (2019); similar findings applied for the limited number of sites available from Ireland. This indicates the model is not likely to be over-parameterised and at the same time it is reasonably robust. While the model was shown to have lower performance over organic soils for the evaluation period with the UK COSMOS data, the RF model was found to perform reasonably at both Dripsey and Donoughmore, sites with poor or impeded drainage on organic soil (Table 2). With the exception of Carlow, the RF model displays better performance in terms of the covariance with the available measured SM (point) at all other sites compared to the ESA CCI SM product (Table 7); while at three of the sites, the RF model resulted in improved estimates of the mean volumetric water content when compared to the measured in situ (point) values (Figure 10). This is not unexpected based on the resolution of the ESA CCI SM, which is at ~25 km, and therefore reflective of areally integrated processes (including land use, soil etc.) across a large sensor footprint. The findings for the model evaluation with the in situ (point) data from Ireland also need to be caveated with the fact that we are comparing an areally integrated, variable depth, measurement technique, CRNS, with in situ fixed depth point measurements. While CRNS data, and specifically the D86 variable, is calibrated against fixed depth time domain transmissometry (TDT) instruments that are co-located with the CRNS instruments, differences in measured SM values between the different sensors are not unexpected. For example, differences exist not just in the measurement techniques and associated algorithms to convert the measured signal into a SM response, but also in the respective sensor footprints associated with each sensor type (e.g., CRNS—variable spatial footprint in the order of 100 s metres; in situ point—fixed at point location of sensor). Due to these differences in scales, the various techniques are likely to reflect different processes affecting SM (e.g., biomass—Cooper, Blyth, et al. 2021; Cooper, Bennett, et al. 2021). In spite of these differences, the RF model would appear to provide a useful method to downscale coarse resolution SM estimates. In the comparison with the GSSM1 km data product at two sites, Johnstown castle and Dripsey, the RF model estimates values that lie closer to the measured site values (point). The findings highlight the added value in downscaling the ESA CCI SM product to sites of interest and that the downscaled values are capable of estimating SM conditions at the point scale; it also highlights the potential advantage of employing more regionally specific information in downscaling SM estimates.

A key benefit of employing a model-based approach is that it is capable of generating long timeseries of estimated VWC, where the respective covariates are available (e.g., Table 3). Figure 14 shows a timeseries for two locations, Johnstown Castle Co. Wexford and Dripsey Co. Cork, for the period 2022–2022. While the ESA CCI data extends back in time, the temporal and spatial coverage become more patchy, as is evident in the modelled volumetric SM at Johnstown Castle. In spite of this, the model estimated values display both temporal and spatial variations—with marked seasonal and inter-annual variations and differences evident between both sites. To further explore the model estimated values, data was extracted for a selection of years—2011 at Dripsey and 2013 and 2018 at Johnstown Castle, with a specific focus on the summer months when the plant uptake of soil water and atmospheric evaporative demand is at its greatest. Summer 2011 was cooler than average with below average rainfall (Met Eireann, n.d.), while both heat and drought waves were recorded during the summers of 2013 and 2018.

At Dripsey, characterised as a poorly drained site (Table 2), the model estimated VWC values are above the long term climatology (2002–2022) during the summer months of June and July (Figure 15). While recorded rainfall during the summer of 2011 was below average at Cork Airport, a synoptic station located approximately 25 km away, it recorded 52 days of rain during the months of June, July and August; temperatures and sunshine hours also remained below their long term average at

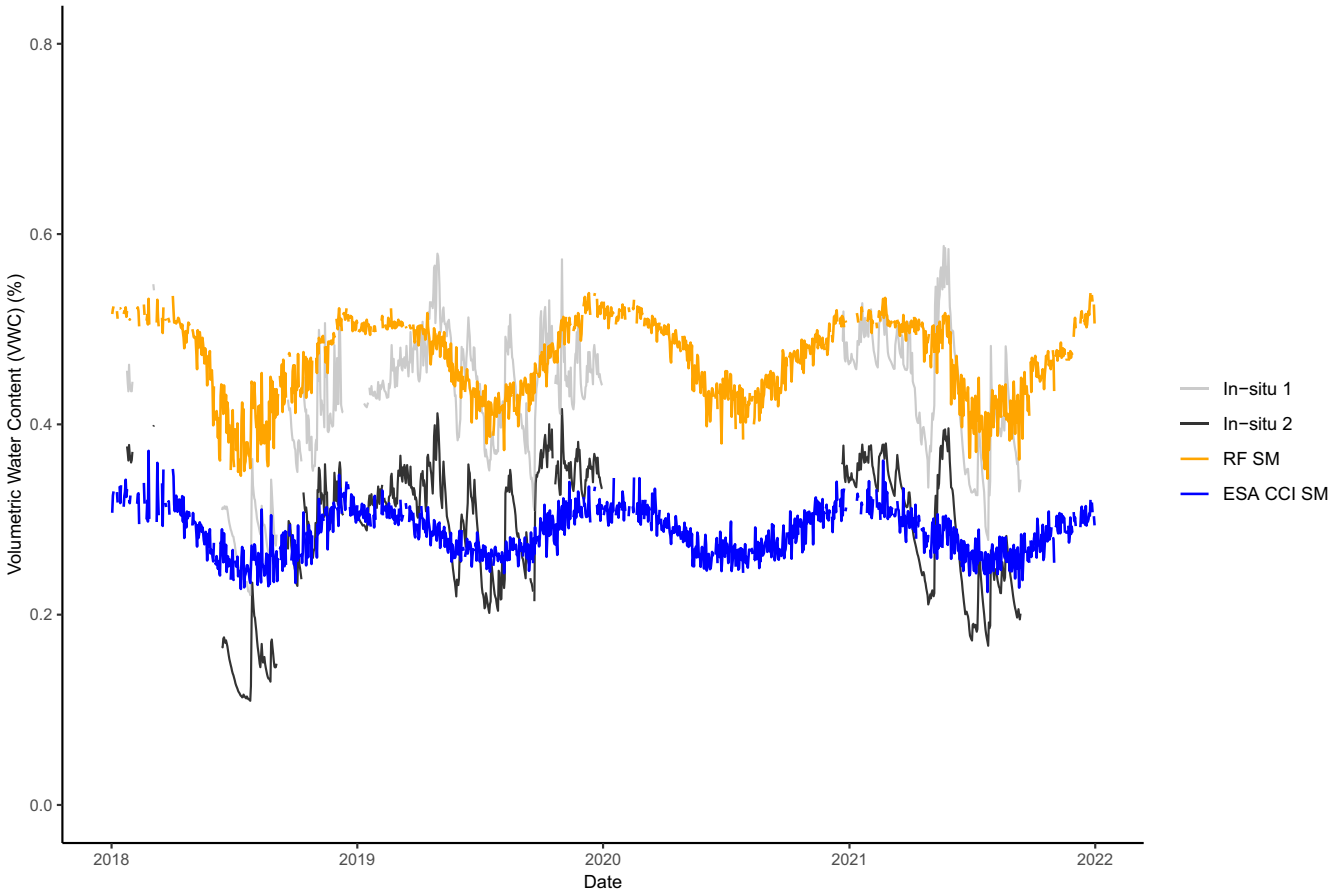

FIGURE 8 | In situ soil moisture from Johnstown Castle (5 cm), RF model estimated soil moisture (7.5 cm), and ESA CCI SM for 2018–2021.

TABLE 7 | Pearson's *R* values between the ESA CCI, RF (depth layer 7.5 cm) and available in situ (point) measurements for available time periods. Model results represent independent sites not included in training the model.

	ESA CCI	RF model		
	Pearson's *R*	**Pearson's *R***	**MAE**	**RMSE**
Johnstown Castle				
2013	0.57; 0.66	0.74; 0.82	0.13; 0.2	0.14; 0.21
2018–2021	0.62; 0.75	0.7; 0.8	0.05; 0.18	0.07; 0.18
Dripsey				
2003–2007	0.54; 0.52	0.66; 0.71	0.1; 0.02	0.13; 0.03
2002–2012[a]	0.51	0.66	0.06 (0.05)	0.07 (0.06)
Carlow	0.66; 0.57	0.55; 0.46	0.09; 0.06	0.11; 0.07
Donoughmore	0.49	0.61	0.05	0.06

Note: The values are shown for the available sensors at each site. Single values reported indicate only one sensor available.
[a]Numbers in brackets for MAE and RMSE for Dripsey are included based on the average model estimated VWC at depth 7.5 and 15 cm.

Cork Airport during the summer of 2011. Newspaper reports on grass growth from the neighbouring county of Kerry indicated that growth was lagging 70 to 80% behind the previous year (Daughton 2011). In spite of the below average rainfall, reduced plant uptake and lower atmospheric vapour demand, coupled with the poorly drained soils, would likely result in above seasonal average VWC which the RF model estimated values indicate.

Figures 16 and 17 show the model estimated values for 2013 and 2018, both dry/drought years at Johnstown Castle, compared to the long term climatology, derived from the period 2002 to 2022. In June 2013, monthly mean temperatures were above the long term average across much of the country, while rainfall was below average except for locations along the north and west of the country (Met Eireann, n.d.). By July, more than 40 stations had recorded heatwave conditions (> 5 days with

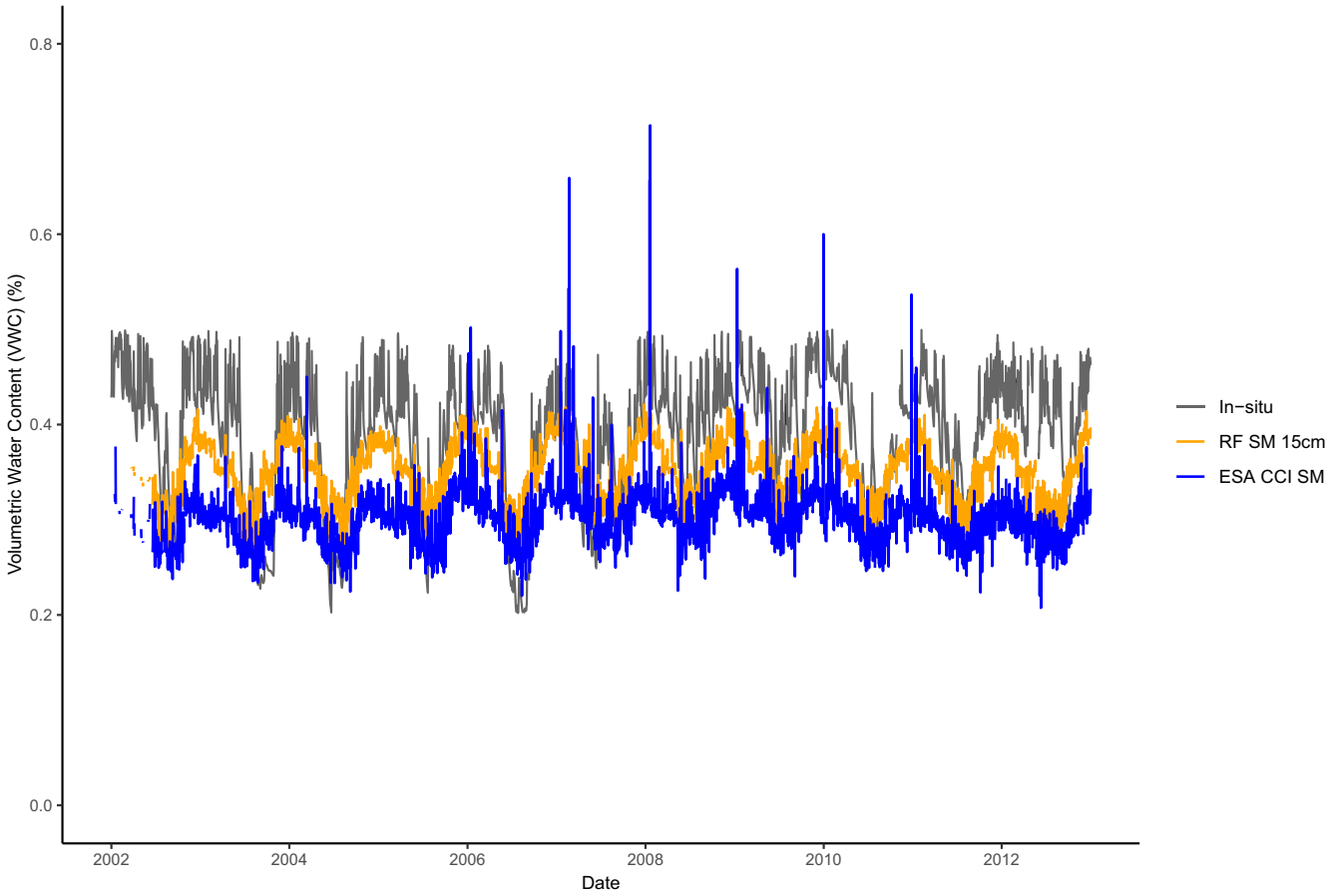

FIGURE 9 | In situ soil moisture from Dripsey, RF model estimated soil moisture (11 cm) and ESA CCI SM for 2002–2012.

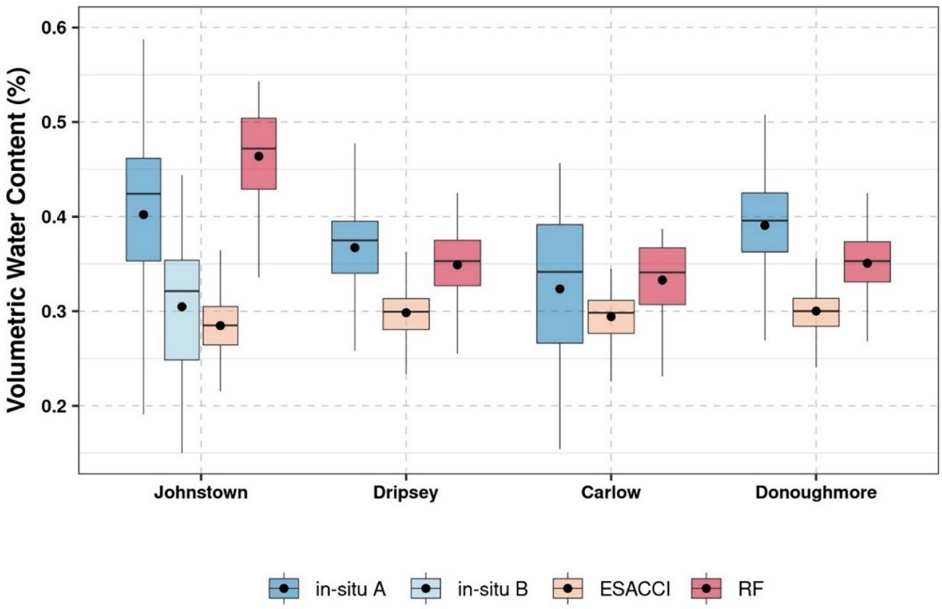

FIGURE 10 | Boxplots of measured in situ (point), ESA CCI SM, and RF model estimated values at the four soil moisture sites located in Ireland for available data at the sites.

temperatures >25°C). During August, temperatures remained at or above average, with variable rainfall; these conditions largely continued into September. Similar to summer 2013, 2018 was characterised by heat and drought wave conditions across Ireland, with drought conditions being recorded from early June to late July. While temperatures remained above average during

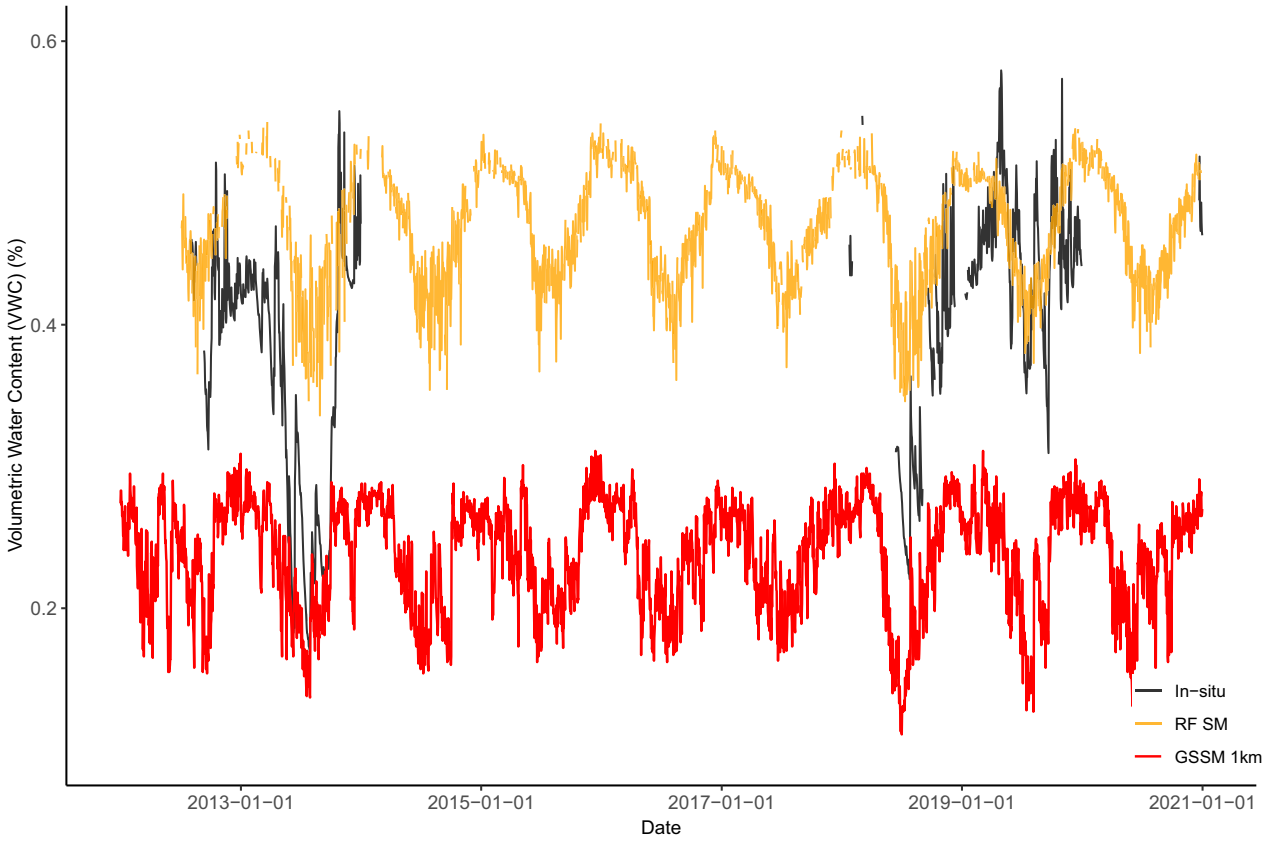

FIGURE 11 | Volumetric water content from the in situ (point) measurements at Johnstown Castle and model outputs from the RF and GSSM1 km models for the period 2013–2021.

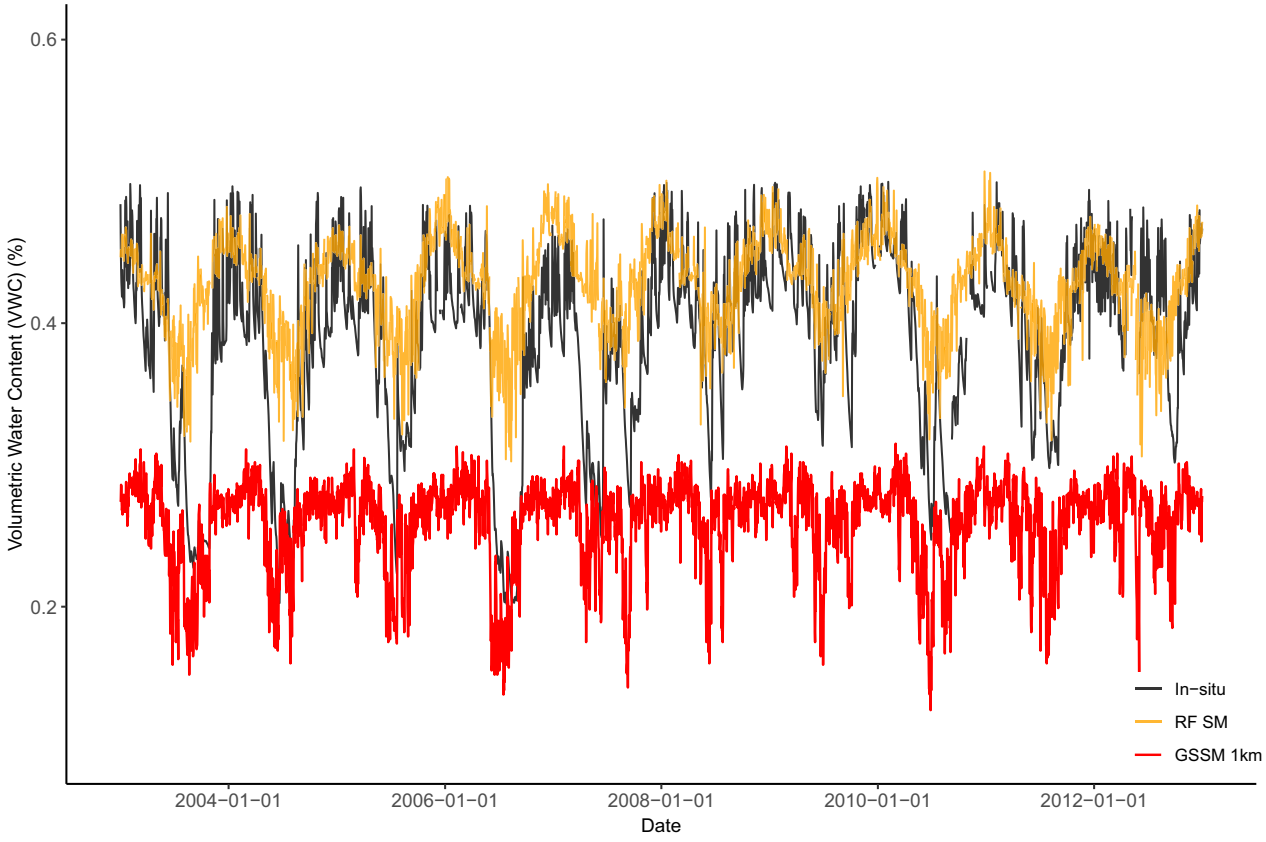

FIGURE 12 | Volumetric water content from the in situ (point) measurements at Dripsey and model outputs from the RF (11 cm) and GSSM1 km models for the period 2003–2012.

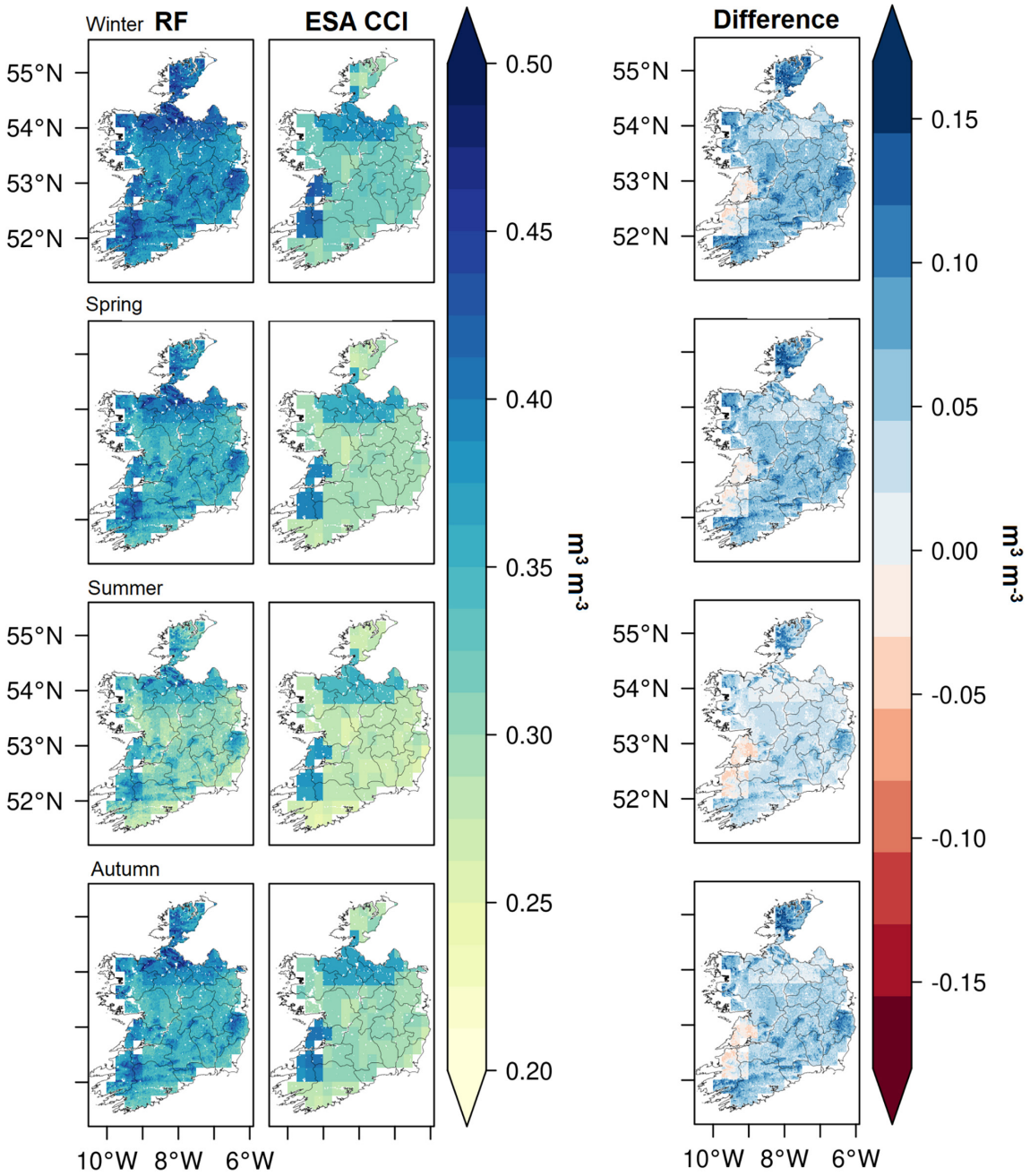

FIGURE 13 | Seasonal mean volumetric water content estimated by the RF model, the ESA CCI SM product and the difference between them (RF model minus ESA CCI SM) for the period 2010–2021.

August, Johnstown Castle recorded 7 wet days, the lowest number of wet days recorded across Ireland (Met Eireann 2018). From Figures 16 and 17, the estimated VWC for both 2013 and 2018 appear to closely track the timing of the driving meteorological conditions during these years (e.g., Ishola et al. 2023).

In addition to hindcasting, the approach outlined here could also be employed in combination with model forecasted meteorology to provide forecasts of SM. Previous work by McDonnell et al. (2018) found that rainfall forecasts, likely to be the limiting factor in forecast length, performed well out to approximately 5 days, which could provide important information for land and water managers. Currently in Ireland, SM deficits are reported which are derived based on rainfall and evaporation and a soil indicator, and represent a relative measure of the amount of water required to bring the soil back to field capacity.

6 | Conclusion

SM has been identified as an essential climate variable (ECV) and plays a central role in understanding and modulating soil–plant–atmosphere interactions. It is also important to measure and monitor SM conditions for agriculture, to support on-farm decision making and assess potential for agricultural drought conditions. SM is also important for many hydrological applications and is of critical importance in understanding the role of extreme climate events. Yet in spite of the recognised

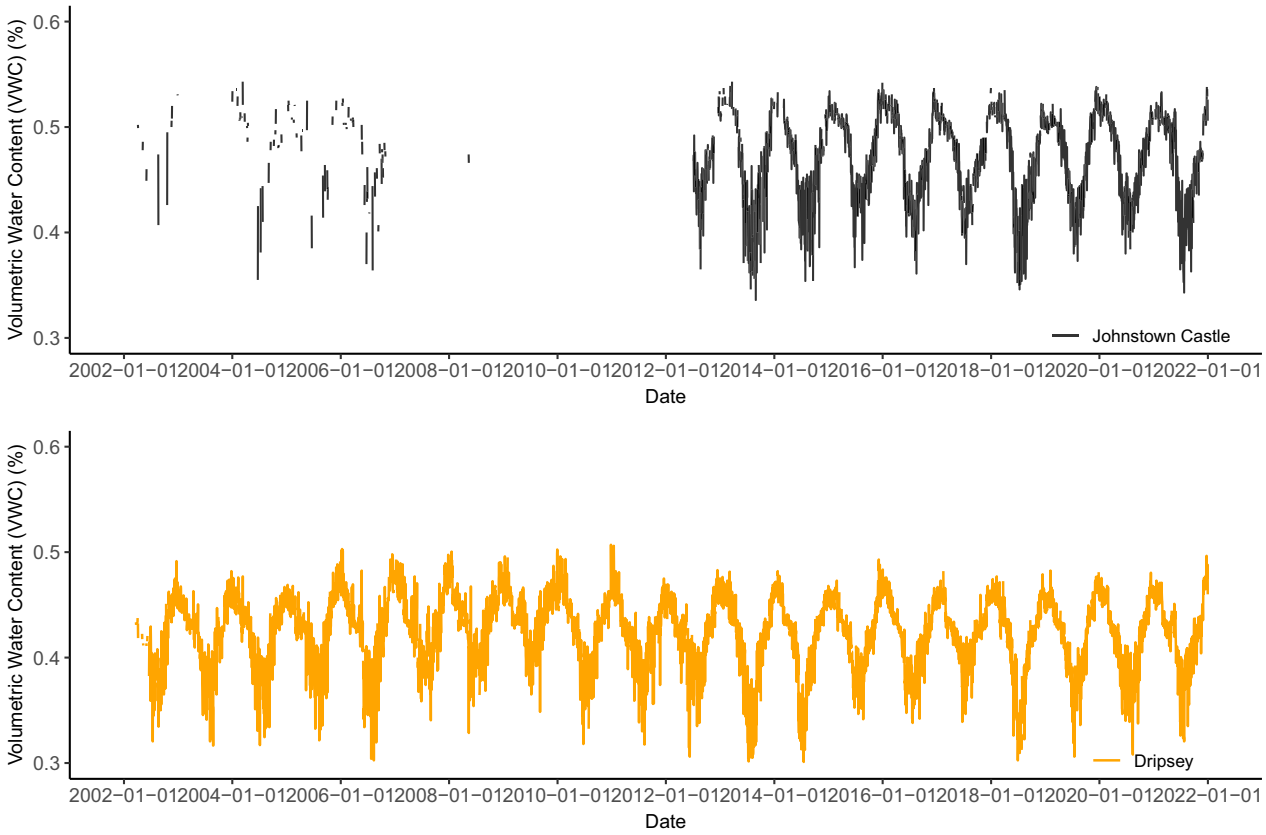

FIGURE 14 | Timeseries of estimated VWC for Johnstown Castle (7.5 cm) and Dripsey (11 cm) based on the Random Forest model for the period 2002–2022. Missing values in the timeseries are due to missing values in the ESA CCI co-variate used in the RF model.

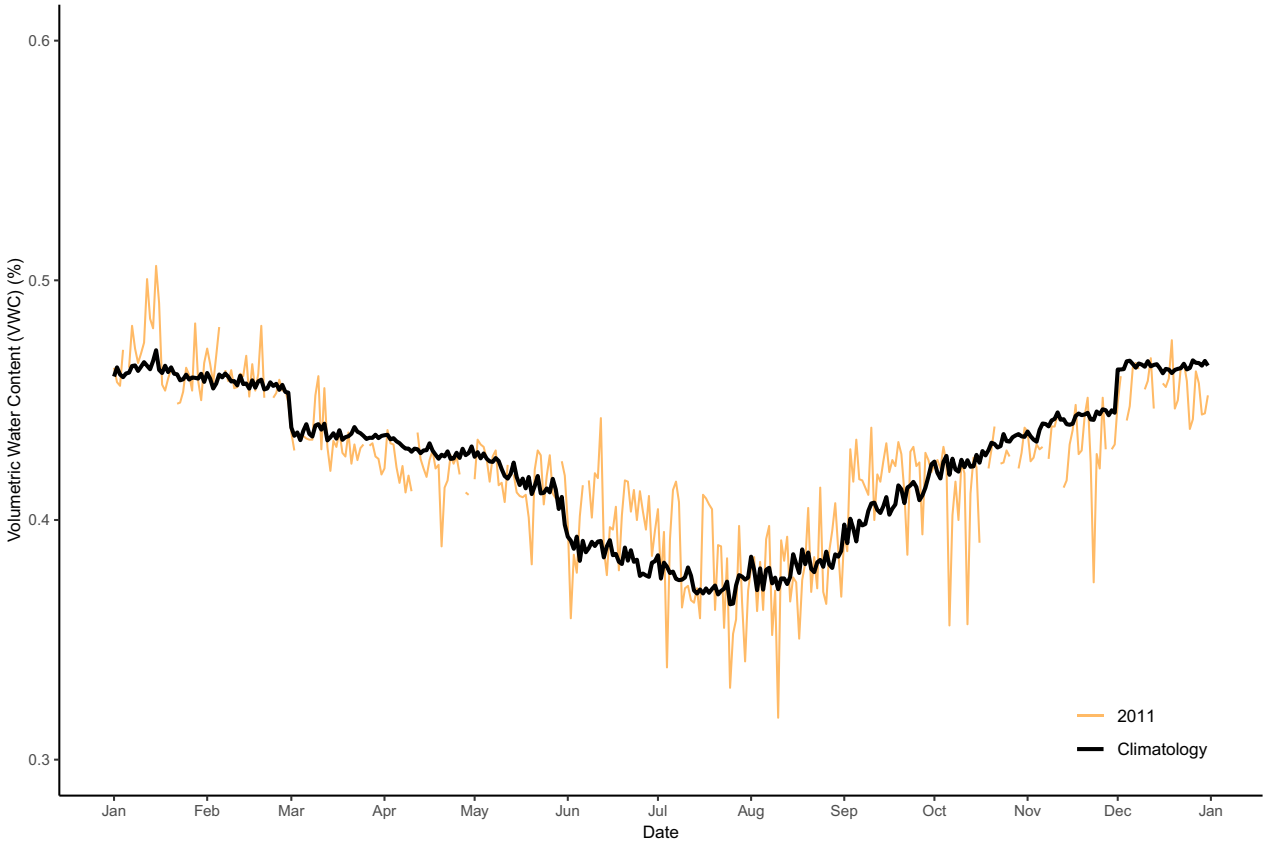

FIGURE 15 | Model estimated VWC at Dripsey for 2011 and the long term climatology derived from the period 2002 to 2022.

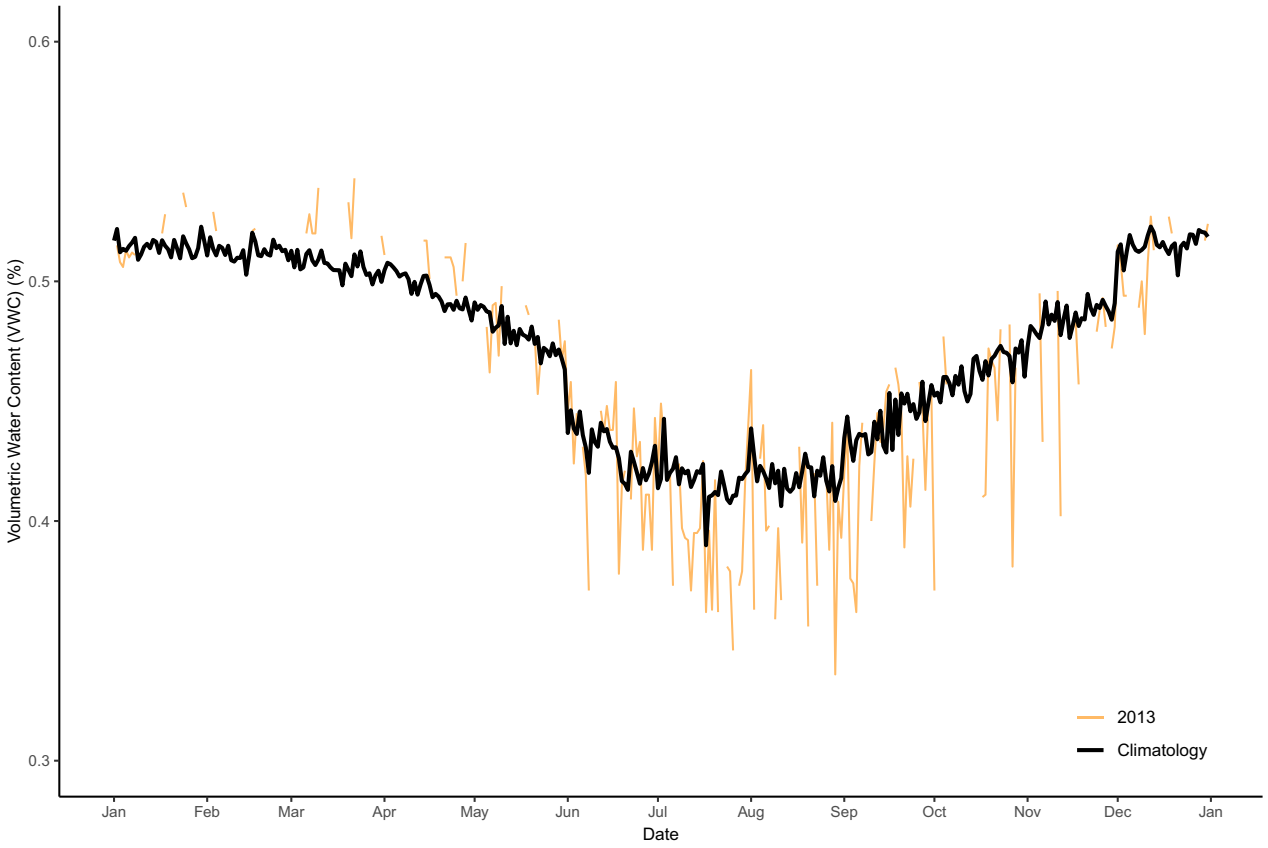

FIGURE 16 | Model estimated VWC at Johnstown Castle for 2013 and the long-term climatology derived from the period 2002 to 2022.

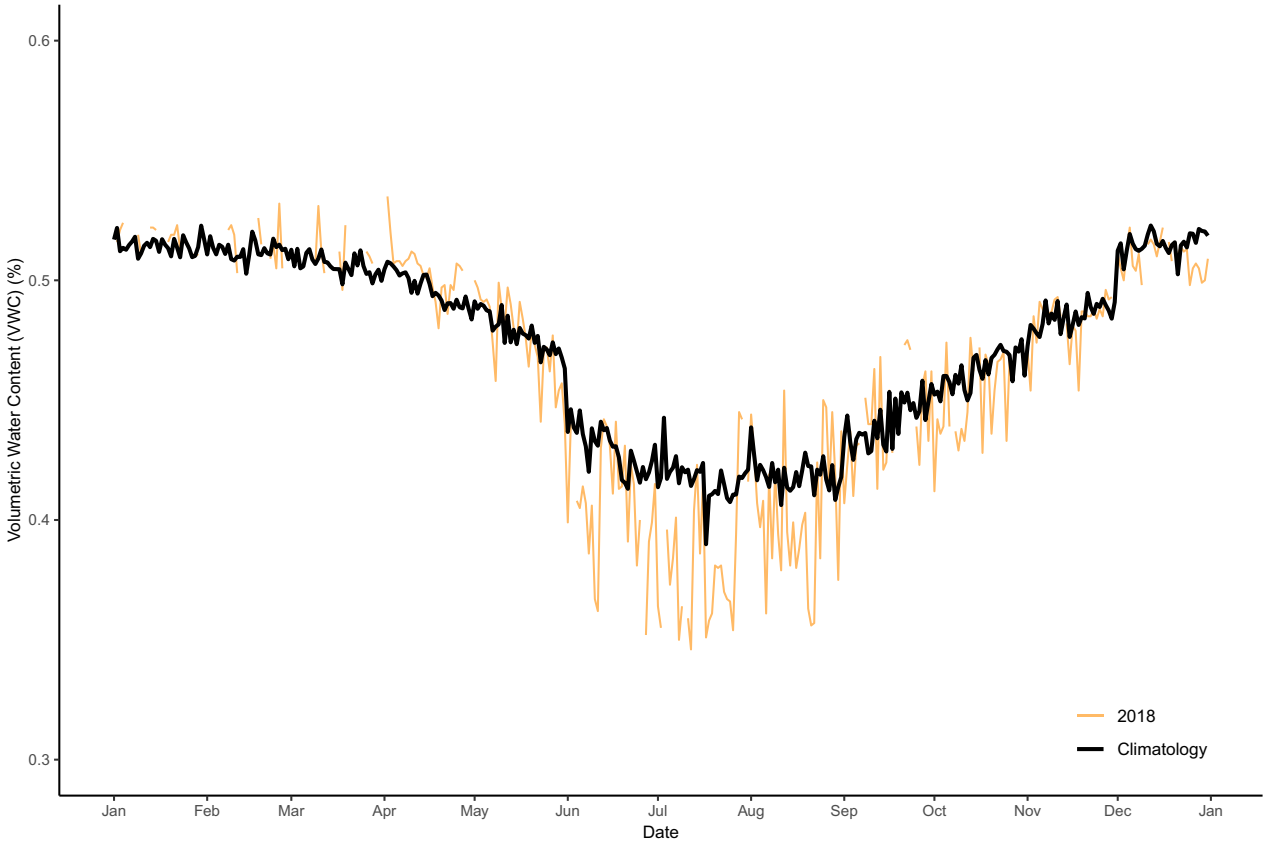

FIGURE 17 | Model estimated VWC at Johnstown Castle for 2018 and the long-term climatology derived from the period 2002 to 2022.

importance of SM, many countries, including Ireland, lack the necessary integrated monitoring infrastructure to measure SM. There is an urgent need to address this from a policy perspective. While remote sensing can support the monitoring of SM, it is not a replacement for in situ measurements.

In response, a number of initiatives have been developed to deploy SM instruments. For example, Terrain-AI, a nationally funded project supported by Science Foundation Ireland (SFI) and Microsoft, deployed a number of time domain reflectometers (TDR) SM probes which are co-located with the existing meteorological or National Agricultural Soil Carbon Observatory (NASCO) flux tower networks funded by the Department of Agriculture, Food and Marine (DAFM). Similarly, University College Dublin (UCD) and Met Éireann, the national meteorological agency, as part of the Joint Working Group on Applied Agricultural Meteorology (AGMET) community and funded by DAFM, have deployed Cosmic-Ray Neutron-Sensor (CRNS) instruments for determining spatial estimates of SM based on the sensor footprint. While each of these deployments is being undertaken as part of separate funded research initiatives, the siting of instruments is being undertaken in a coordinated way between the various projects and institutions to ensure that an optimum network design can be implemented (e.g., ISMON—Daly et al. 2021). The intent is that the sensors deployed will continue beyond the lifetime of the individual funded projects and ultimately contribute to establishing a long term SM monitoring network here.

While the development of national monitoring networks will contribute new measurements to the global monitoring effort, SM can vary significantly over small spatial scales, largely due to its dependence on soil properties, landscape and other factors. Consequently, we need to continue to develop, refine and utilise appropriate remote measurements (e.g., satellite) and modelling (e.g., statistical; simple 'bucket' and land surface models) tools that can fill in the gaps at the spatial and temporal scales necessary to inform agricultural (field scale; managing daily crop water requirements), hydrological (catchment; hours to daily) and meteorological (areal; boundary layer development; energy partitioning) applications.

Following the recommendations of Wei (1995) with regards to the need for a more integrated approach, encompassing the then newly emerging satellite data for monitoring SM, modelling and in situ measurements, Houser et al. (1998) employed a four- dimensional data assimilation technique to integrate data from a passive microwave satellite sensor into a hydrological land surface scheme (TOPMODEL model-based Land–Atmosphere Transfer Scheme—TOPLAT) to estimate SM over an experimental watershed in the USA. The use of data assimilation techniques has been shown to improve the LSMs' estimation of deeper layer SM and surface fluxes. Consequently, the use of data assimilation techniques to ingest single variables or more than one variable into LSMs has become more widespread, particularly over the past 10 years. While a number of challenges remain, with regards to method, pre-processing, computational cost and uncertainty assessment, the benefits of data assimilation (DA) appear to be very promising (e.g., Jun et al. 2021; Yin et al. 2023). Downscaled SM estimates could also provide an additional constraint for DA techniques.

Author Contributions

Rowan Fealy: conceptualization, investigation, funding acquisition, writing – original draft, methodology, validation, visualization, writing – review and editing, software, formal analysis, project administration, data curation, supervision. **Kazeem Ishola:** investigation, methodology, formal analysis, writing – review and editing, visualization. **Tim McCarthy:** validation, writing – review and editing, supervision, data curation, project administration, methodology. **Ajay Nair:** investigation, validation, methodology. **Rafael de Andrade Moral:** investigation, methodology, writing – review and editing, validation, formal analysis.

Acknowledgements

The research detailed here was funded by the Irish Environmental Protection Agency (EPA) Research Programme 2021–2030 (2019-CCRP-MS.66). The EPA Research Programme is a Government of Ireland initiative funded by the Department of the Environment, Climate and Communications. It is administered by the Environmental Protection Agency, which has the statutory function of coordinating and promoting environmental research. A portion of the research was funded under the Terrain-AI project (SFI 20/SPP/3705), supported by the Science Foundation Ireland Strategic Partnership Programme and co-funded by Microsoft. We also wish to thank Gary Lanigan and Matt Saunders for providing access to the measurements from Johnstown Castle and Carlow, respectively. The authors would like to acknowledge the UK Centre for Ecology and Hydrology who own the COSMOS-UK data. COSMOS-UK is supported by the Natural Environment Research Council award number NE/R016429/1 as part of the UK-SCAPE programme delivering National Capability.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Albergel, C., S. Munier, D. J. Leroux, et al. 2017. "Sequential Assimilation of Satellite-Derived Vegetation and Soil Moisture Products Using SURFEX_v8.0: LDAS-Monde Assessment Over the Euro-Mediterranean Area." *Geoscientific Model Development* 10: 3889–3912.

American Meteorological Society (AMS). 2022. "Glossary of Meteorology." https://glossary.ametsoc.org/wiki/Soil_moisture.

Attema, E. P. W., and F. T. Ulaby. 1978. "Vegetation Modeledmodelled as a Water Cloud." *Radio Science* 13, no. 2: 357–364. https://doi.org/10.1029/RS013i002p00357.

Babaeian, E., P. Sidike, M. S. Newcomb, et al. 2019. "A New Optical Remote Sensing Technique for High-Resolution Mapping of Soil Moisture." *Frontiers in Big Data* 2, no. 37: 37. https://doi.org/10.3389/fdata.2019.00037.

Bastos, A., C. M. Gouveia, R. M. Trigo, and S. W. Running. 2014. "Analysing the Spatio-Temporal Impacts of the 2003 and 2010 Extreme Heatwaves on Plant Productivity in Europe." *Biogeosciences* 11: 3421–3435.

Batchu, V., G. Nearing, and V. Gulshan. 2023. "A Deep Learning Data Fusion Model Using Sentinel-1/2, SoilGrids, SMAP, and GLDAS for Soil Moisture Retrieval." *Journal of Hydrometeorology* 24: 1789–1823. https://doi.org/10.1175/JHM-D-22-0118.1.

Boorman, D., J. Alton, V. Antoniou, et al. 2020. "COSMOS-UK User Guide v3.00." http://nora.nerc.ac.uk/id/eprint/528535/1/N528535CR.pdf.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45: 5–32.

Brocca, L., S. Hasenauer, T. Lacava, et al. 2011. "Remote Sensing of Environment Soil Moisture Estimation Through ASCAT and AMSR-E Sensors: An Intercomparison and Validation Study Across Europe." *Remote Sensing of Environment* 115, no. 12: 3390–3408. https://doi.org/10.1016/j.rse.2011.08.003.

Carlson, T. N. 2023. "A Critique of the Triangle Method and a Version Suitable for Estimating Soil Moisture From Satellite Imagery." *Journal of Geography, Environment and Earth Science International* 27, no. 10: 1–16. https://doi.org/10.9734/jgeesi/2023/v27i10713.

Carlson, T. N., R. R. Gillies, and E. M. Perry. 1994. "A Method to Make Use of Thermal Infrared Temperature and NDVI Measurements to Infer Surface Soil Water Content and Fractional Vegetation Cover." *Remote Sensing Reviews* 9: 161–173. https://doi.org/10.1080/02757259409532220.

Chen, C., and L.-M. Liu. 1993. "Joint Estimation of Model Parameters and Outlier Effects in Time Series." *Journal of the American Statistical Association* 88, no. 421: 284. https://doi.org/10.2307/2290724.

Chipman, H. A., E. I. George, and R. E. McCulloch. 2010. "BART: Bayesian Additive Regression Trees." *Annals of Applied Statistics* 4, no. 1: 266–298.

Ciais, P., M. Reichstein, N. Viovy, et al. 2005. "European-Wide Reduction in Primary Productivity Caused by the Heat and Drought in 2003." *Nature* 437: 529–533.

Cooper, E., E. Blyth, H. Cooper, R. Ellis, E. Pinnington, and S. J. Dadson. 2021. "Using Data Assimilation to Optimize Pedotransfer Functions Using Field-Scale In Situ Soil Moisture Observations." *Hydrology and Earth System Sciences* 25: 2445–2458. https://doi.org/10.5194/hess-25-2445-2021.

Cooper, H. M., E. Bennett, J. Blake, et al. 2021. "COSMOS-UK: National Soil Moisture and Hydrometeorology Data for Environmental Science Research." *Earth System Science Data* 13: 1737–1757. https://doi.org/10.5194/essd-13-1737-2021.

Daly, E., K. Finkele, T. Hochstrasser, et al. 2021. The Irish Soil Moisture Observation Network—ISMON. Irish National Hydrology Conference 2021: Proceedings.

Dari, J., R. Morbidelli, C. Saltalippi, C. Massari, and L. Brocca. 2019. "Spatial-Temporal Variability of Soil Moisture: Addressing the Monitoring at the Catchment Scale." *Journal of Hydrology* 570: 436–444.

Daughton, E. 2011. "Grass Growth Below Average in a Summer of Foul Weather, Irish Independent, 17 August." https://www.independent.ie/regionals/kerry/lifestyle/grass-growth-below-average-in-a-summer-of-foul-weather/27410268.html.

De Boeck, H. J., F. E. Dreesen, I. A. Janssens, and I. Nijs. 2011. "Whole-System Responses of Experimental Plant Communities to Climate Extremes Imposed in Different Seasons." *New Phytologist* 189: 806–817.

de la Martinez Torre, A., E. M. Blyth, and E. L. Robinson. 2018. *Water, Carbon and Energy Fluxes Simulation for Great Britain Using the JULES Land Surface Model and the Climate Hydrology and Ecology Research Support System Meteorology Dataset (1961-2015) [CHESS-Land]*. NERC Environmental Information Data Centre. https://doi.org/10.5285/c76096d6-45d4-4a69-a310-4c67f8dcf096.

Didan, K. 2021. "MODIS/Terra Vegetation Indices 16-Day L3 Global 500m SIN Grid V061." Distributed by NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MOD13A1.061.

Dorigo, W., I. Himmelbauer, D. Aberer, et al. 2021. "The International Soil Moisture Network: Serving Earth System Science for Over a Decade." *Hydrology and Earth System Sciences* 25: 5749–5804. https://doi.org/10.5194/hess-25-5749-2021.

Dorigo, W., W. Wagner, C. Albergel, et al. 2017. "ESA CCI Soil Moisture for Improved Earth System Understanding: State-Of-The Art and Future Directions." *Remote Sensing of Environment* 203: 185–215.

Engman, E. T., and N. Chauhan. 1995. "Status of Microwave Soil Moisture Measurements With Remote Sensing." *Remote Sensing of Environment* 51, no. 1: 189–198.

Entekhabi, D., E. G. Njoku, P. E. O'Neill, et al. 2010. "The Soil Moisture Active Passive (SMAP) Mission." *Proceedings of the IEEE* 98, no. 5: 704–716. https://doi.org/10.1109/JPROC.2010.2043918.

Fealy, R., C. Bruyere, and C. Duffy. 2018. *Regional Climate Model Simulations for Ireland for the 21st Century*. Environmental Protection Agency.

Fealy, R., T. McCarthy, R. De Andrade Moral, et al. 2024. "SoMoSAT—Soil Moisture Estimates From Satellite-Based Earth Observations." EPA Research Report (2019-CCRP-MS.66), Environmental Protection Agency, Co. Wexford, 1–61.

Friedl, M., and D. Sulla-Menashe. 2022. "MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V061." Distributed by NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MCD12Q1.061.

Gruber, A., T. Scanlon, R. van der Schalie, W. Wagner, and W. Dorigo. 2019. "Evolution of the ESA CCI Soil Moisture Climate Data Records and Their Underlying Merging Methodology." *Earth System Science Data* 11: 717–739. https://doi.org/10.5194/essd-11-717-2019.

Guan, Y., X. Gu, L. J. Slater, J. Li, D. Kong, and X. Zhang. 2023. "Spatio-Temporal Variations in Global Surface Soil Moisture Based on Multiple Datasets: Intercomparison and Climate Drivers." *Journal of Hydrology* 625, no. B: 130095.

Han, Q., Y. Zeng, L. Zhang, et al. 2023. "Global Long Term Daily 1 Km Surface Soil Moisture Dataset With Physics Informed Machine Learning." *Sci Data* 10: 101. https://doi.org/10.1038/s41597-023-02011-7.

Hijmans, R. 2023. "_raster: Geographic Data Analysis and Modeling_." R package Version 3.6–14. https://CRAN.R-project.org/package=raster.

Hollis, D., M. McCarthy, M. Kendon, T. Legg, and I. Simpson. 2018. "HadUK-Grid Gridded and Regional Average Climate Observations for the UK." Centre for Environmental Data Analysis. http://catalogue.ceda.ac.uk/uuid/4dc8450d889a491ebb20e724debe2dfb.

Houser, P. R., W. J. Shuttleworth, J. S. Famiglietti, H. V. Gupta, K. H. Syed, and D. C. Goodrich. 1998. "Integration of Soil Moisture Remote Sensing and Hydrologic Modeling Using Data Assimilation." *Water Resources Research* 34, no. 12: 3405–3420.

Huang, S., J. Ding, B. Liu, et al. 2020. "The Capability of Integrating Optical and Microwave Data for Detecting Soil Moisture in an Oasis Region." *Remote Sensing* 12: 1358. https://doi.org/10.3390/rs12091358.

Hyndman, R. J. 2021. "Detecting Time Series Outliers." https://robjhyndman.com/hyndsight/tsoutliers/.

Ishola, K. A., G. Mills, R. M. Fealy, and R. Fealy. 2023. "A Model Framework to Investigate the Role of Anomalous Land Surface Processes in the Amplification of Summer Drought Across Ireland During 2018." *International Journal of Climatology* 43, no. 1: 480–498.

Ishola, K. A., G. Mills, R. M. Fealy, Ó. Ní Choncubhair, and R. Fealy. 2020. "Improving a Land Surface Scheme for Estimating Sensible and Latent Heat Fluxes Above Grasslands With Contrasting Soil Moisture Zones." *Agricultural and Forest Meteorology* 294: 108151.

Ishola, K., G. Mills, A. Sati, et al. 2024. "Implementation of Global Soil Databases in NOAH-MP Model and the Effects on Simulated Mean and Extreme Soil Hydrothermal Changes." *Hydrology and Earth System Sciences Discussions* 29: 1–39. https://doi.org/10.5194/hess-2023-304.

Jaksic, V., G. Kiely, J. Albertson, et al. 2006. "Net Ecosystem Exchange of Grassland in Contrasting Wet and Dry Years." *Agricultural and Forest Meteorology* 139, no. 334: 323.

Jun, S., J.-H. Park, H.-J. Choi, et al. 2021. "Impact of Soil Moisture Data Assimilation on Analysis and Medium-Range Forecasts in an Operational Global Data Assimilation and Prediction System." *Atmosphere* 12: 1089. https://doi.org/10.3390/atmos12091089.

Karthikeyan, L., M. Pan, N. Wanders, and D. Nagesh Kumar. 2017. "Four Decades of Microwave Satellite Soil Moisture Observations: Part 2. Product Validation and Inter-Satellite Comparisons." *Advances in Water Resources* 109: 236–252. https://doi.org/10.1016/j.advwatres.2017.09.010.

Karthikeyan, L., M. Pan, N. Wanders, D. Nagesh Kumar, and E. F. Wood. 2017. "Four Decades of Microwave Satellite Soil Moisture Observations: Part 1. A Review of Retrieval Algorithms." *Advances in Water Resources* 109: 106–120.

Kiely, G., P. Leahy, C. Lewis, M. Sottocornola, A. Laine, and A.-K. Koehler. 2018. "GHG Fluxes From Terrestrial Ecosystems in Ireland." Research Report No. 227.EPA Research Programme, Wexford. https://www.epa.ie/pubs/reports/research/climate/Research_Report_227.pdf.

Kiely, G., P. Leahy, M. Sottocornola, et al. 2008. *CELTICFLUX: Measurement and Modelling of GHG Fluxes From Grasslands and a Peatland in Ireland*, 118. Report prepared for the Environmental Protection Agency Ireland.

Köhli, M., M. Schrön, M. Zreda, U. Schmidt, P. Dietrich, and S. Zacharias. 2015. "Footprint Characteristics Revised for Field-Scale Soil Moisture Monitoring With Cosmic-Ray Neutrons." *Water Resources Research* 51: 5772–5790.

Kolassa, J., R. H. Reichle, and C. S. Draper. 2017. "Merging Active and Passive Microwave Observations in Soil Moisture Data Assimilation." *Remote Sensing of Environment* 191: 117–130.

Kovačević, J., Ž. Cvijetinović, N. Stančić, N. Brodić, and D. Mihajlović. 2020. "New Downscaling Approach Using ESA CCI SM Products for Obtaining High Resolution Surface Soil Moisture." *Remote Sensing* 12: 1119. https://doi.org/10.3390/rs12071119.

Kuhn, M. 2012. "Variable Importance Using the Caret Package." *Journal of Statistical Software* 6: 1–6.

Kurnik, B., G. Louwagie, M. Erhard, A. Ceglar, and L. Bogataj Kajfež. 2014. "Analysing Seasonal Differences Between a Soil Water Balance Model and In Situ Soil Moisture Measurements at Nine Locations Across Europe." *Environmental Modeling & Assessment* 19, no. 1: 19–34.

Liu, Q., R. H. Reichle, R. Bindlish, et al. 2011. "The Contributions of Precipitation and Soil Moisture Observations to the Skill of Soil Moisture Estimates in a Land Data Assimilation System." *Journal of Hydrometeorology* 12: 750–765.

Liu, Y., W. Jing, Q. Wang, and X. Xia. 2020. "Generating High-Resolution Daily Soil Moisture by Using Spatial Downscaling Techniques: A Comparison of Six Machine Learning Algorithms." *Advances in Water Resources* 141: 103601. https://doi.org/10.1016/j.advwatres.2020.103601.

Liu, Y. Y., W. A. Dorigo, R. M. Parinussa, et al. 2012. "Trend-Preserving Blending of Passive and Active Microwave Soil Moisture Retrievals." *Remote Sensing of Environment* 123: 280–297.

McDonnell, J., K. Lambkin, R. Fealy, D. Hennessy, L. Shalloo, and C. Brophy. 2018. "Verification and Bias Correction of ECMWF Forecasts for Irish Weather Stations to Evaluate Their Potential Usefulness in Grass Growth Modelling." *Meteorological Applications* 25, no. 2: 292–301.

Met Eireann. 2018. "Weather Statement for August 2018." https://www.met.ie/climate/past-weather-statements.

Met Eireann. n.d. "Monthly Weather Bulletin—June 2013, No. 326." https://www.met.ie/climate/past-weather-statements.

Meyer, H., C. Milà, and M. Ludwig. 2023. "_CAST: 'caret' Applications for Spatial-Temporal Models_." R package Version 0.7.1. https://CRAN.R-project.org/package=CAST.

Mohanty, B. P. 2013. "Soil Hydraulic Property Estimation Using Remote Sensing: A Review." *Vadose Zone Journal* 12, no. 4: vzj2013.06.0100.

Mohanty, B. P., M. H. Cosh, V. Lakshmi, and C. Montzka. 2017. "Soil Moisture Remote Sensing: State-Of-The-Science." *Vadose Zone Journal* 16: 1-9 vzj2016.10.0105. https://doi.org/10.2136/vzj2016.10.0105.

Noone, S., C. Broderick, C. Duffy, T. Matthews, R. L. Wilby, and C. Murphy. 2017. "A 250-Year Drought Catalogue for the Island of Ireland (1765–2015)." *International Journal of Climatology* 37: 239–254.

Papale, D., M. Reichstein, M. Aubinet, et al. 2006. "Towards a Standardized Processing of Net Ecosystem Exchange Measured With Eddy Covariance Technique: Algorithms and Uncertainty Estimation." *Biogeosciences* 3: 571–583.

Peichl, M., O. Carton, and G. Kiely. 2012. "Management and Climate Effects on Carbon Dioxides and Energy Exchanges in a Maritime Grasslands." *Agriculture, Ecosystems & Environment* 158: 132–146.

Peng, J., C. Albergel, A. Balenzano, et al. 2021. "A Roadmap for High-Resolution Satellite Soil Moisture Applications - Confronting Product Characteristics With User Requirements." *Remote Sensing of Environment* 252: 112162. https://doi.org/10.1016/j.rse.2020.112162.

Peng, J., A. Loew, O. Merlin, and N. E. C. Verhoest. 2017. "A Review of Spatial Downscaling of Satellite Remotely Sensed Soil Moisture." *Reviews of Geophysics* 55: 341–366. https://doi.org/10.1002/2016RG000543.

Pierce, D. 2023. "_ncdf4: Interface to Unidata netCDF (Version 4 or Earlier) Format Data Files_." R Package Version 1.21. https://CRAN.R-project.org/package=ncdf4.

Pinnington, E., J. Amezcua, E. Cooper, et al. 2021. "Improving Soil Moisture Prediction of a High-Resolution Land Surface Model by Parameterising Pedotransfer Functions Through Assimilation of SMAP Satellite Data." *Hydrology and Earth System Sciences* 25: 1617–1641. https://doi.org/10.5194/hess-25-1617-2021.

Plummer, S., P. Lecomte, and M. Doherty. 2017. "The ESA Climate Change Initiative (CCI): A European Contribution to the Generation of the Global Climate Observing System." *Remote Sensing of Environment* 203: 2–8.

Poggio, L., L. M. de Sousa, N. H. Batjes, et al. 2021. "SoilGrids 2.0: Producing Soil Information for the Globe With Quantified Spatial Uncertainty." *Soil* 7: 217–240.

Preimesberger, W., T. Scanlon, C.-H. Su, A. Gruber, and W. Dorigo. 2021. "Homogenization of Structural Breaks in the Global ESA CCI Soil Moisture Multisatellite Climate Data Record." *IEEE Transactions on Geoscience and Remote Sensing* 59, no. 4: 2845–2862. https://doi.org/10.1109/TGRS.2020.3012896.

Price, J. C. 1990. "Using Spatial Context in Satellite Data to Infer Regional Scale Evapotranspiration." *IEEE Transactions on Geoscience and Remote Sensing* 28: 940–948.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/.

Raes, D. 2002. *BUDGET—A Soil Water and Salt Balance Model. Reference Manual*. K.U.Leuven, Department Land Management.

Reichstein, M., M. Bahn, P. Ciais, et al. 2013. "Climate Extremes and the Carbon Cycle." *Nature* 500, no. 7462: 287–295.

Rodriguez-Fernandez, N. J., F. Aires, P. Richaume, et al. 2015. "Soil Moisture Retrieval Using Neural Networks: Application to SMOS." *IEEE Transactions on Geoscience and Remote Sensing* 53, no. 11: 5991–6007.

Sabaghy, S., J. P. Walker, L. J. Renzullo, and T. J. Jackson. 2018. "Spatially Enhanced Passive Microwave Derived Soil Moisture: Capabilities and Opportunities." *Remote Sensing of Environment* 209: 551–580.

Sadeghi, M., E. Babaeian, M. Tuller, and S. B. Jones. 2017. "The Optical Trapezoid Model: A Novel Approach to Remote Sensing of Soil Moisture

Applied to Sentinel-2 and Landsat-8 Observations." *Remote Sensing of Environment* 198: 52–68. https://doi.org/10.1016/j.rse.2017.05.041.

Santanello, J. A., S. V. Kumar, C. D. Peters-Lidard, and P. M. Lawston. 2016. "Impact of Soil Moisture Assimilation on Land Surface Model Spinup and Coupled Land–Atmosphere Prediction." *Journal of Hydrometeorology* 17: 517–540.

Schulte, R. P. O., J. Diamond, K. Finkele, N. M. Holden, and A. J. Brereton. 2005. "Predicting the Soil Moisture Conditions of Irish Grasslands." *Irish Journal of Agricultural and Food Research* 44: 95–110.

Seneviratne, S. I., T. Corti, E. L. Davin, et al. 2010. "Investigating Soil Moisture–Climate Interactions in a Changing Climate: A Review." *Earth-Science Reviews* 99, no. 3–4: 125–161.

Seneviratne, S., N. Nicholls, D. Easterling, et al. 2012. "Changes in Climate Extremes and Their Impacts on the Natural Physical Environment." In *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*, edited by C. B. Field, V. R. Barros, D. Dokken, et al., 109–230. Cambridge University Press.

Signal Developers. 2013. "signal: Signal Processing." http://r-forge.r-project.org/projects/signal/.

Stanley, S., V. Antoniou, L. A. Ball, et al. 2019. *Daily and Sub-Daily Hydrometeorological and Soil Data (2013-2017) [COSMOS-UK]*. NERC Environmental Information Data Centre. https://doi.org/10.5285/a6012796-291c-4fd6-a7ef-6f6ed0a6cfa5.

Sungmin, O., and R. Orth. 2021. "Global Soil Moisture Data Derived Through Machine Learning Trained With In-Situ Measurements." *Sci Data* 8: 170. https://doi.org/10.1038/s41597-021-00964-1.

Taktikou, E., G. Bourazanis, G. Papaioannou, and P. Kerkides. 2016. "Prediction of Soil Moisture From Remote Sensing Data." *Procedia Engineering* 162: 309–316.

Tong, C., H. Wang, R. Magagi, et al. 2020. "Soil Moisture Retrievals by Combining Passive Microwave and Optical Data." *Remote Sensing* 12: 3173. https://doi.org/10.3390/rs12193173.

Wan, Z., S. Hook, and G. Hulley. 2021. "MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V061." Distributed by NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/MODIS/MOD11A1.061.

Wei, M. Y. 1995. *Soil Moisture: Report of a Workshop Held in Tiburon, California, 25–27 January 1994*. NASA Conference Publication, CP-3319.

Werbylo, K. L., and J. D. Niemann. 2014. "Evaluation of Sampling Techniques to Characterize Topographically-Dependent Variability for Soil Moisture Downscaling." *Journal of Hydrology* 516: 304–316. https://doi.org/10.1016/j.jhydrol.2014.01.030.

Xu, X., B. Shew, S. Zaman, J. Lee, and Y. Zhi. 2020. *Assessment of SMAP and ESA CCI Soil Moisture Over the Great Lakes Basin*, 4590–4593. IGARSS 2020–2020 IEEE International Geoscience and Remote Sensing Symposium. https://doi.org/10.1109/IGARSS39084.2020.9323638.

Yin, J., X. Zhan, M. Barlage, et al. 2023. "Assimilation of Blended Satellite Soil Moisture Data Products to Further Improve Noah-MP Model Skills." *Journal of Hydrology* 621: 129596.

Zambrano-Bigiarini, M. 2024. "hydroGOF: Goodness-of-Fit Functions for Comparison of Simulated and Observed Hydrological Time Series R Package Version 0.6–0.1." https://cran.r-project.org/package=hydroGOF. https://doi.org/10.5281/zenodo.839854.

Zappa, L., M. Forkel, A. Xaver, and W. Dorigo. 2019. "Deriving Field Scale Soil Moisture From Satellite Observations and Ground Measurements in a Hilly Agricultural Region." *Remote Sensing* 11: 2596. https://doi.org/10.3390/rs11222596.

Zhang, D., and G. Zhou. 2016. "Estimation of Soil Moisture From Optical and Thermal Remote Sensing: A Review." *Sensors* 16: 1308.

Zhang, Y., S. Liang, H. Ma, et al. 2023. "Generation of Global 1 Km Daily Soil Moisture Product From 2000 to 2020 Using Ensemble Learning." *Earth System Science Data* 15: 2055–2079. https://doi.org/10.5194/essd-15-2055-2023.

Zhao, C., S. M. Quiring, S. Yuan, D. B. McRoberts, N. Zhang, and Z. Leasor. 2020. "Developing and Evaluating National Soil Moisture Percentile Maps." *Soil Science Society of America Journal* 84: 443–460. https://doi.org/10.1002/saj2.20045.

Zhu, L., H. Wang, C. Tong, W. Liu, and B. Du. 2019. "Evaluation of ESA Active, Passive and Combined Soil Moisture Products Using Upscaled Ground Measurements." *Sensors (Basel)* 19, no. 12: 2718. https://doi.org/10.3390/s19122718.

Zignol, F., W. Lidberg, C. Greiser, J. Larson, R. Hoffrén, and A. M. Ågren. 2024. "Controls on Spatial and Temporal Variability of Soil Moisture Across a Heterogeneous Boreal Forest Landscape." EGUsphere [Preprint]. https://doi.org/10.5194/egusphere-2024-2909.