Assimilation of optical and radar remote sensing data in 3D mapping of soil properties over large areas.

Laura Poggio*¹, Alessandro Gimona¹

* laura.poggio@hutton.ac.uk

¹ The James Hutton Institute - Craigiebuckler, AB158QH, Aberdeen, Scotland (UK)

Abstract

Soil is very important for many land functions. To achieve sustainability it is important to understand how soils vary over space in the landscape. Remote sensing data can be instrumental in mapping and spatial modelling of soil properties, resources and their variability.

The aims of this study were to compare satellite sensors (MODIS, Landsat, Sentinel-1 and Sentinel-2) with varying spatial, temporal and spectral resolutions for Digital Soil mapping (DSM) of a set of soil properties in Scotland, evaluate the potential benefits of adding Sentinel-1 data to DSM models, select the most suited mix of sensors for DSM to map the considered set of soil properties and validate the results of topsoil (2D) and whole profile (3D) models.

The results showed that the use of a mixture of sensors proved more effective to model and map soil properties than single sensors. The use of radar Sentinel-1 data proved useful for all soil properties, improving the prediction capability of models with only optical bands. The use of MODIS time series provided stronger relationships than the use of temporal snapshots. The results showed good validation statistics with a RMSE below 20% of the range for all considered soil properties. The RMSE improved from previous studies including only MODIS sensor and using a coarser prediction grid. The performance of the models was similar to previous studies at regional, national or continental scale. A mix of optical and radar data proved useful to map soil properties along the profile.

The produced maps of soil properties describing both lateral and vertical variability, with associated uncertainty, are important for further modelling and management of soil resources and ecosystem services. Coupled with further information the soil properties maps could be used to assess soil functions and therefore conditions and suitability of soils for a range of purposes.

²⁵ **Key words**

²⁶ Sensors inter-comparison, probabilistic approach, uncertainty assessment, Generalized Additive Models,
²⁷ GlobalSoilMap

1 Introduction

Soil is very important for many land functions, such as food production, capture and storage of water and carbon, and other ecosystem services (Adhikari and Hartemink, 2016). It is therefore important to understand how soil works, how to protect and manage it to achieve sustainability. In order to do so, it is especially important to understand how soils vary over space in the landscape. Remote sensing data can be very important for mapping and spatial modelling of soil properties, resources and their variability (Mulder et al., 2011; Wulf et al., 2014).

The spatial variation in soil has historically been presented as maps of soil classes (i.e. choropleth maps) in which homogeneous regions are partitioned into disjoint polygons. The maps are typically at coarse scales and often based on qualitative assessment. The coverage is not always complete and artifacts are present, due to the different methods used to produce the maps. Normally, the choropleth maps have no associated measures of uncertainty attached to the data, and are not easily updatable (Hartemink et al., 2013). Combining remotely sensed imagery, digital elevation models and soil sample data, a complete coverage can be produced, the accuracy of the estimated soil properties can be improved and the uncertainty more easily assessed and modelled. This can be achieved using digital soil mapping (DSM) methods, i.e. geospatial methods to provide estimates of the continuous variation in soil attributes at fine scales and resolutions that might more adequately represent soil processes (McBratney et al., 2003). DSM allows for the prediction of soil properties or classes using soil information and environmental covariates of soil (Carre et al., 2007).

Most of the recent studies applied a variation of the **scorpan**-kriging approach using a spatial soil prediction function with autocorrelated error (SSPFe) (McBratney et al., 2003): $S = f(s, c, o, r, p, a, n) + e$ where S are soil properties as a function of soil, climate, organisms, relief, parent material, age and spatial position, e refers to the spatially correlated errors. The scorpan model is used for quantitative prediction of soil classes or continuous soil attributes based on empirical observations. It is not trying to explain the factors of soil formation. Remote sensing data are often used as proxies for the **scorpan** factors. Reviews of work on DSM can be found in Minasny and McBratney (2016); Hartemink et al. (2013); Boettinger et al. (2010); Grunwald (2009); Scull et al. (2003); McBratney et al. (2003).

Remote sensing data can be used to obtain qualitative and quantitative information about soil properties and are an essential and very cost effective data source for soil mapping (Agbu et al., 1990; Ben-Dor et al., 2008, 2009; Mulder et al., 2011; Wulf et al., 2014). Remote sensing imagery has been widely used to estimate different soil properties such as organic carbon (e.g. Mulder et al., 2016b; Padarian et al., 2015; Castaldi et al., 2016; de Brogniez et al., 2015; Poggio and Gimona, 2014), texture (e.g. Castaldi et al., 2016; Adhikari et al., 2013) or hydraulic properties (e.g. Padarian et al., 2014; Mohanty, 2013) in

different landscapes, from agricultural to semi-natural. Studies used remote sensing for field scale (e.g. Castaldi et al., 2014; Vaudour et al., 2014) to regional (e.g. Bachofer et al., 2015; Gomez et al., 2012) national (e.g. Mulder et al., 2016b; Adhikari et al., 2013; Poggio and Gimona, 2014), continental (e.g. Ballabio et al., 2016; Viscarra et al., 2015; Hengl et al., 2015; Dewitte et al., 2012; de Brogniez et al., 2015) and global (Hengl et al., 2014; Stockmann et al., 2015) scales. Multispectral and hyperspectral optical sensors were mainly used (e.g. Vaudour et al., 2014; Ben-Dor et al., 2009). However, the spectral range from remote sensing used for soil mapping is not limited to the visible and infrared. Radar, gamma radiometric (e.g. Chaney et al., 2016; Rawlins et al., 2009; Wilford and Minty, 2006; Cook et al., 1996) and radio frequencies (Mendonça-Santos and dos Santos, 2006) can and have also be used. For further examples please refer to the citations above. Remote sensing data have also been used for complex soil evaluation, such as erosion (Schmid et al., 2016) and quality (Obade and Lal, 2013).

The aims of this study were to

1. compare RS sensors with different spatial and temporal resolutions, wavelength ranges and number of bands for DSM of a set of soil properties ;
2. evaluate the potential benefits of adding Sentinel-1 data to DSM models using covariates derived from optical sensors
3. select the most suited mix of sensors for DSM to map the considered set of soil properties and
4. validate the results of topsoil (2D) and whole profile (3D) models.

The selected sensors covered a range of spatial (10 to 1000m), temporal (bi-daily to by-monthly) resolutions and spectral ranges with optical at different wavelengths and radar, in particular: 1. MODIS (Moderate Resolution Imaging Spectro-radiometer) 2. Landsat 3. Sentinel-1 and 4. Sentinel-2 (respectively radar and optical sensor of the European Space Agency Copernicus mission). The soil properties selected were part of the list identified by the *GlobalSoilMap* project (Arrouays et al., 2014) to describe the most important soil characteristics to support ecosystem services delivery. The test area was Scotland, with a high soil variability and landscape diversity.

2 Data and test area

The models were run for the whole of Scotland ($\simeq 78,000\text{km}^2$). The the Shetland islands were excluded because of the high cloud cover.

2.1 Soil information

The Scottish Soils Database contains information and data on soils from locations throughout Scotland. It contains the National Soil Inventory of Scotland (NSIS) profile samples collected on a regular 10 km grid of sampled locations (Lilly et al., 2010) and physical and chemical data from a large number of soil profiles taken to characterize the soil mapping units. The properties considered were measured to a maximum depth of 1 metre. The positional uncertainty of profile locations and the uncertainty of analytical procedures were not taken into account.

The following soil properties were mapped using the available data in the Scottish database:

1. depth of soil (censored at 100cm)
2. pH measured in water
3. organic carbon (percent of content)
4. base saturation (BS, percent): as the percentage saturation of base cations (Ca, Mg, Na and K) as a proportion of the sum of exchangeable cations
5. soil texture: as the relative percentage of sand (> 0.05 mm), silt (0.005-0.05 mm) and clay (< 0.005 mm), with the sum of the components always equals to the unit (one or 100)
6. base saturation (sum of base cations expressed as percentage of the total sum of cations).
7. bulk density (BD, g/cm^3). The measured data were available at a subset of locations (25 % of the total). A statistical model between existing values of BD and pH and loss on ignition organic matter was used to predict the values at the remaining locations (Poggio and Gimona, 2014).

The selected soil properties, excluding bulk density, are primary soil properties derived directly from measured data. BD is a secondary soil property derived with the support of a statistical model from measured parameters, i.e. a pedotransfer function. Table 1 presents a basic summary of the considered properties and the transformations that were used for the modelling. In order to provide validation of the models the data available were split in two sets, training and validation, with a ratio of approximately 3:1. The validation set was randomly sampled. The two resulting data-sets (calibration and validation) have similar spatial distributions.

Table 1: Summary statistics of the soil properties considered in the modelling. All properties but base saturation are listed by the *GlobalSoilMap* specifications (Arrouays et al., 2014). The layers follow the *GlobalSoilMap* specifications with L1=0-5 cm, L2=5-15cm,L3=15-30cm,L4=30-60cm,L5=60-100cm. ALR = additive log ratio.

	Transformation	N	Minimum	Median	Maximum	Layer
Bulk density (g/cm^3)	None	664	0.02	0.22	1.24	L1
		3881	0.04	0.59	1.30	L2
		1889	0.05	0.64	1.27	L3
		2906	0.00	0.74	1.37	L4
		594	0.01	0.76	1.35	L5
C (%)	Logit	665	0.24	30.50	63.57	L1
		3881	0.11	4.78	64.39	L2
		1889	0.11	3.18	62.29	L3
		2906	0.00	1.25	62.66	L4
		595	0.00	0.56	60.70	L5
Base saturation (%)	Logit	651	1.50	20.70	99.10	L1
		3804	0.20	43.20	99.90	L2
		1834	0.00	15.80	98.30	L3
		2833	0.00	42.60	99.80	L4
		576	0.00	34.50	97.70	L5
pH (-)	None	665	3.19	4.33	8.32	L1
		3881	2.98	5.41	8.70	L2
		1889	3.30	5.02	8.76	L3
		2906	3.15	5.69	8.96	L4
		594	3.31	5.46	8.65	L5
Sand (%)	ALR	168	29.29	84.11	100.00	L1
		2906	8.00	70.00	100.00	L2
		1577	3.96	72.92	100.00	L3
		2701	0.00	72.45	100.00	L4
		526	4.00	72.08	100.00	L5
Silt (%)	ALR	168	0.00	9.39	34.00	L1
		2906	0.00	17.71	58.00	L2
		1577	0.00	14.58	58.76	L3
		2701	0.00	16.00	58.42	L4
		526	0.00	17.00	71.00	L5
Clay (%)	ALR	168	0.00	5.32	45.45	L1
		2906	0.00	10.31	49.46	L2
		1577	0.00	11.22	71.13	L3
		2701	0.00	10.00	73.40	L4
		526	0.00	9.00	63.00	L5
Depth (cm, ceiled at 100)	None	3963	38.68	89.94	100.00	-

2.2 Covariates

The covariates included are freely and globally available and were selected to describe, directly or indirectly, the most important **scorpan** factors, namely topography, vegetation, climate and geographical position.

2.2.1 Remote sensing

MODIS

A set of indices was derived from the Terra Moderate Resolution Imaging Spectro-radiometer (MODIS) 8 and 16 day composite products. The data were acquired from the NASA ftp website (<ftp://e4ftl01u.ecs.nasa.gov/MOLT/>) for 12 years between 2000 and 2012. The indices were selected for their capability to differentiate spectral responses from different bare soils, vegetation cover and mixed situations:

- 1) Enhanced Vegetation Index (EVI; Huete et al., 2002)

$$EVI = G \times \frac{(NIR - RED)}{(NIR + C1 \times RED - C2 \times Blue + L)} \quad (1)$$

where NIR (Near InfraRed), red and blue are atmospherically-corrected surface reflectances, L is the canopy background adjustment that addresses non-linear, differential NIR and red radiant transfer through a canopy, C1, C2 are the coefficients of the aerosol resistance term, which uses the blue band to correct for aerosol influences in the red band and G is the gain factor. The coefficients adopted in the MODIS-EVI algorithm are: L=1, C1 = 6, C2 = 7.5, and G = 2.5.

- 2) the Normalised Difference Water Index (NDWI; Gao, 1996)

$$NDWI = \frac{NIR - SWIR}{NIR + SWIR} \quad (2)$$

NDWI was calculated with NIR and SWIR (Short Wave InfraRed) band: SWIR = 2130 (Gu et al., 2008)

- 3) Leaf Area Index (LAI; Knyazikhin et al., 1999) is defined as the one-sided green leaf area per unit ground area in broadleaf canopies and as half the total needle surface area per unit ground area in coniferous canopies.

- 4) Land Surface Temperature (LST; Wan, 1999)

- 5) primary productivity (Running et al., 1999, 2004).

- 6) Phenology information: Phenology is the response of the vegetation to seasonal climatic cycles in irradiance, temperature and rainfall. Therefore phenology constitutes an essential land surface pa-

parameter in atmospheric and climate models. Using fitted functions reduces the uncertainties and leads to more stable measures. Timesat (Jönsson and Eklundh, 2004; Eklundh and Jonsson, 2011) is primarily designed to process time-series of vegetation index derived from satellite spectral measurements. The processing method used was based on least-squares fits to the upper envelope of the vegetation index data with local polynomial functions in the fitting, and the method can be classified as an adaptive Savitzky-Golay filter (Savitzky and Golay, 1964). The parameters calculated are:

(a) length of the season (LOS); time from the start to the end of the season.

(b) seasonal amplitude; difference between the maximum value and the base level, given as the average of the left and right minimum values.

7) snow indices: snow indices were derived from the Snow Cover and its quality assurance (QA) products (MOD10A2, v005) of Terra MODIS (Riggs et al., 2006). Initially the downloaded data were processed for cloud removal, filling the missing values and the resulting data on the presence/absence of snow were considered as binary sequence and processed to obtain indices taking into account the temporal pattern of snow presence (Poggio and Gimona, 2015b).

The medians over 12 years (2000-2011) were used as covariates. The single images were restored to fill the cloud gaps using the approach described in Poggio et al. (2012). The method is not limited by the type of noise or degradation of pixels values, latitude, vegetation dynamics and land uses. The method was able to restore data providing very good to adequate responses in series of simulations of missing data. The reconstructed images reproduced spatial patterns and local features of the measured product, even with substantial amounts of missing pixels (i.e. up to 80% of missing values). The spatio-temporal interpolation for cloud filling was performed using only the MODIS images in close-by dates.

Landsat

Landsat provides high spatial resolution images with a minimum revisit cycle of 16 days, that can be greatly extended due to cloud contamination or duty cycle limitations (Ju and Roy, 2008). Cloud free (cloud cover < 10%) scene of Landsat 8 were acquired for July 2013 as standard level 1 terrain corrected (L1T) product. The images were atmospherically corrected and then mosaicked to provide cover for most of Scotland. The temperature for the thermal bands was processed from digital numbers (DN) to Kelvin (Chander et al., 2009). The clouds were detected using the Automated Cloud-Cover Assessment (ACCA) algorithm from Irish (2000) with the constant values for pass filter one from Irish et al. (2006) using Landsat bands numbers already been processed from DN into reflectance and temperature. Cloud pixels were discarded.

Several bands indices were calculated:

178 1) NDVI

179 2) NDWI (see eq. 2; Gao, 1996) using band SWIR between 2100-2300.

180 3) the Soil Moisture (SM):

$$SM = \frac{NIR}{Blue} \quad (3)$$

181 4) the Soil Colour Index (SCI):

$$SCI = 3 * NIR + Red - Green - 3 * Blue \quad (4)$$

182 Sentinel-1

183 The Sentinel-1 (S1) mission provides data from a dual-polarization C-band Synthetic Aperture Radar
184 (SAR) instrument. This includes the S1 Ground Range Detected (GRD) scenes, processed using the
185 Sentinel-1 Toolbox to generate a calibrated, ortho-corrected product. Each scene was pre-processed with
186 Sentinel-1 Toolbox using the following steps: Thermal noise removal, Radiometric calibration Terrain
187 correction using SRTM 30. The final terrain corrected values were converted to decibels via log scaling
188 $log_value = 10 \times log_{10}(value)$ and quantized to 16-bits. The data were pre-processed, prepared, mosaicked
189 and downloaded from Google Earth Engine (Google Earth Engine Team, 2015). The median of the
190 images available between June 2015 and May 2016 was calculated for VV and VH polarisation. No
191 further transformations were applied.

192 Sentinel-2

193 Sentinel-2 (S2) is a wide-swath, high-resolution, multi-spectral imaging mission supporting Copernicus
194 Land Monitoring studies, including the monitoring of vegetation, soil and water cover, as well as observa-
195 tion of inland waterways and coastal areas. Each band represents TOA reflectance scaled by 10000. The
196 data were mosaicked and downloaded from Google Earth Engine (Google Earth Engine Team, 2015).
197 The same indices calculated with Landsat (i.e. NDVI, NDWI, SM and SCI) were calculated with corre-
198 sponding bands from S2. The obtained mosaicked image (June to September 2016) showed some areas
199 with snow and other smaller areas with some remaining clouds. The clouds and snow free areas allowed
200 to fit the models and validated them.

201 2.2.2 Morphology

202 The Digital Elevation Model (DEM) used as a covariate in the fitted models was SRTM (Shuttle Radar
203 Topography Mission), further processed to fill in no-data voids (Jarvis et al., 2006; Rodriguez et al.,
204 2006). SRTM has a spatial resolution of 90m with global coverage. The measures used were elevation,

Figure 1: Analysis approach with sensor inter-comparison, point data assimilation, validation and prediction workflow.

slope as the steepest slope angle, calculated using the D8 method (O’Callaghan and Mark, 1984), and topographic wetness index (TWI (eq. 5); Sorensen et al., 2006):

$$\ln \frac{a}{\tan(b)} \quad (5)$$

The DEM was resampled to two different resolutions: 100×100m and 1×1 km. The medians in each grid cell of 1×1 km and 100×100 m were used.

3 Methods

The summary of the analysis approach is described in figure 1 with sensor inter-comparison, point data assimilation, validation and prediction workflow.

3.1 Sensors inter-comparison

The considered sensors have different resolutions (table 2) and therefore different approaches were necessary to homogenise the resolutions to the cell size selected for the soil properties interpolation (100m). In particular:

1. Upscaling (all bands with finer resolution, i.e. Landsat, S1 and S2): focal median of the pixels values with a window size to match initial and final resolutions.

Table 2: Summary description of the considered RS sensors.

Sensor	Resolution (m)	Overpass (days)	Dates used
MODIS	250 - 1000	Twice daily	8 days composites from 2000 to 2012 (median and separate dates)
Landsat	30 - 100 (thermal)	ca. every 2 weeks	One composite scene in July 2013
Sentinel-1	10	ca every 12 days	Median of one year (June 2015/May 2016)
Sentinel-2	10	ca every 12 days	Median of 2016 summer (June / September 2016)

Table 3: List of covariates by sensors. All covariates were used in the variables tournament, a subset was considered for the pre-defined models (table 4).

	Sensor	Predefined model	Variable tournament
Elevation	SRTM	N	Y
Slope		N	Y
TWI		N	Y
EVI	MODIS	Y	Y
Snow index		N	Y
NDWI MODIS		Y	Y
Primary productivity		N	Y
Amplitude		Y	Y
LOS		N	Y
LST day		Y	Y
LST night		Y	Y
NDVI Landsat	Landsat	Y	Y
NDWI Landsat		Y	Y
SCI Landsat		Y	Y
SM Landsat		Y	Y
NDVI S2	Sentinel 2	Y	Y
NDWI S2		Y	Y
SCI S2		Y	Y
SM S2		Y	Y
VV Polarisation	Sentinel 1	Y	Y
VH Polarisation		Y	Y

2. Downscaling (MODIS):

(a) resample: nearest neighbours approach (Burrough and McDonell, 2000).

(b) downscale: geo-statistical approach combining GAMs with kriging using information held in covariates to provide results reproducing more accurately the spatial patterns of the landscape.

It is applicable to a variety of data-sets as it does not rely on predetermined functional relationships (Poggio and Gimona, 2015a; Malone et al., 2012). In this case Northing and Easting were used as covariates.

In order to compare the sensors considered on this study for soil properties, the following analysis approaches were implemented:

1. comparison of pre-defined models. The aim of this step was a rapid comparison of different sensors by evaluating similar sets of indices and bands. Only models using the 2D approach and data for the first 15 cm of soil (topsoil in the following text) were tested;

2. variable tournament. The aim of this step was to select the most appropriate combination of

covariates across all considered (table 3) following a similar approach to what described in Poggio et al. (2013). The final selected combination was used to map the considered soil properties for the test area. Both 2D (topsoil, i.e. 0-15cm) and 3D models (whole profile with 5 layers as defined by the *GlobalSoilMap*) were tested and compared.

The models follow the approach described in section 3.2 derived from Poggio and Gimona (2014).

3.1.1 Pre-defined models

In order to evaluate the performances of the different sensors, sets of covariates were selected to compare:

1. basic indices available for all optical sensors (e.g. EVI/NDVI and NDWI),
2. additional information specific for each sensor (e.g. LAI, phenology),
3. thermal bands (Landsat) and land surface temperature (MODIS).

A subset of all the available covariates (table 3) was selected to have a manageable combinations of models, while exploring a range of different bands and indices.

Models were also run using single dates for MODIS in the 12 years available and the best results were recorded. Finally, models with the MODIS median over a year were considered and the best over the 12 year was reported. In the case of MODIS, products obtained from geostatistical downscaling and resampling were also tested. Table 4 summarises the pre-defined models considered.

In order to quantify the impact of using S1 polarisation bands, models with only optical bands (Landsat, MODIS and S2) and models with a combination of optical and S1 bands were compared for each soil property using the ΔAIC_{S1} (Akaike Information Criterion; Akaike, 1973) defined as

$$\Delta AIC_{S1} = AIC_{optical} - AIC_{optical+S1} \quad (6)$$

where, for each soil property, $AIC_{optical}$ is the AIC of the model using only bands from optical sensors and $AIC_{optical+S1}$ is the AIC of the model integrating optical sensor bands with S1 bands. A higher value for ΔAIC_{S1} indicates a larger improvement while using S1.

3.1.2 Variables tournament and prediction models

In order to model the soil properties at 100 m a combination of indices derived from the sensors described previously was selected. These were used together with the DEM-derived features. An additional smoother of the coordinates was included to consider the spatial structure of the data and to reduce the over-fitting of spatially auto-correlated data (Wood, 2006). The potential 21 covariates were reduced in a variable tournament (Poggio et al., 2013) for each soil property considered. The total pool of covariates was split in three as compromise between the number of covariates, number of point locations available

Table 4: Description of the sets of covariates tested to compare the considered sensors. L = index calculated with Landsat bands, S2 = index calculated with Sentinel-2 bands. U=upscale at 100m; R=resample at 100m; D=downscale at 100m. VV / VH = polarisations of Sentinel-1 (S1). M = MODIS band.

	NDVI	NDWI	Thermal	SM	SCI	VV	VH	EVI	NDWI.2130	Ampl.EVI	LST.day	LST.night	LAI	Change of support	MODIS dates
1	L	L	L											U	NA
2	L	L												U	NA
3	L	L		L	L									U	NA
4								M	M	M				R	median
5								M	M	M				D	median
6								M	M	M	M	M		D	median
7								M	M		M	M	M	R	median
8								M	M					R	median
9								M	M					D	median
10								M	M	M	M	M		R	median
11								M	M	M	M	M		R	best composite in the year
12								M	M	M	M	M		R	best year
13	S2	S2												U	NA
14	S2	S2		S2	S2									U	NA
15						S1	S1							U	NA
16	L	L				S1	S1							U	NA
17	S2	S2				S1	S1							U	NA
18						S1	S1	M	M					U/R	median
19						S1	S1	M	M					U/D	median
20						S1	S1	M	M		M	M		U/R	median
21						S1	S1	M	M		M	M		U/D	median

and the computational load. Batches of seven variables were randomly selected, without replacement, at any given time from the pool of all available covariates and were used to predict the target soil parameter. The method used is GAMs as described in the section 3.2. The resulting significant variables (≤ 0.01 significance level) were stored. The process was repeated for all the covariates in the initial pool. The procedure was then repeated on the reduced set formed by the retained significant variables until a final set of seven or fewer covariates was obtained. The whole variable tournament was run 100 times for each soil property. The covariates in the models with the lowest AIC (plus 2 points) were used in the final model or each soil property. The Δ_{iAIC} threshold was selected based on the analysis of Burnham and Anderson (2013) with Δ_{iAIC} defined as $AIC - AIC_{min}$, where AIC_{min} is the minimum AIC of all models. As a rule of thumb, a $\Delta_{iAIC} \leq 2$ suggests that the models can be considered equivalent (and therefore the choice depend on the number of parameters or other user defined criteria), Δ_{iAIC} values between 2 and 10 indicate strong evidence that the models can be considered different, whereas a $\Delta_{iAIC} > 10$ indicates that there is very strong evidence that the models are different.

The model used for prediction was run with only the variables with significant coefficients selected (≤ 0.01 significance level) from the pool obtained with the variables tournament.

3.2 Soil properties interpolation

An extension of the `scorpan`-kriging approach, i.e. hybrid geostatistical Generalized Additive Models (GAM Wood, 2006), combining GAM with Gaussian simulations (GAM+GS Poggio and Gimona, 2014) was used. The approach features:

- 1) the fitting of a GAM to estimate the trend of the variable, using a 3D smoother with related covariates; and
- 2) Gaussian simulations of GAM residuals as spatial component to account for local details.

The prediction grid had a resolution of 100×100 m for the lateral dimensions and 100cm with 5cm intervals for the vertical dimension. The results were then summarised for the standard *GlobalSoilMap* depths (Arrouays et al., 2014).

The trend was modelled with a 3D spatial approach (Augustin et al., 2009). The model accounts for full 3D-space correlation, exploiting the values of the neighbouring pixels in 3D-space and taking into account the autocorrelated error (Wood, 2006). The trend is modelled using a scale-invariant tensor product smooth of 3D-space dimension, handling the estimation of the 3D space trend and interaction, as described in Augustin et al. (2009). The 3D-space smoother used allows separate smoothing parameters and penalties for the 3D-space dimensions and thus avoids the need to make user specific choices about the relative scaling of 3D-space. The approach does not rely on a regular grid and allows the incorporation

of a wide range of correlation structures (Augustin et al., 2009). The GAM implementation used relied on an internal cross-validation for model fitting (Poggio and Gimona, 2014).

The prediction matrix of the GAM model was used to obtain multiple realisation of the trend, according to the approach suggested by Wood (2006). See also Poggio et al. (2010) for an implementation example. Trend estimations were derived by simulating 100 replicate parameter sets from the posterior distribution of the vector of model parameters.

The GAM produced residuals that were further modelled for spatial correlation (e.g. Poggio et al., 2010) using a Gaussian simulation approach (Journel, 1996; Goovaerts, 1997). A variogram (Cressie, 1993) was fitted for the residuals. Exponential and spherical models (Deutsch and Journel, 1998; Goovaerts, 1997) were tested and the model providing the lowest AIC was retained. Anisotropy was also taken into account and the variograms were fitted accounting for the principal anisotropy axes (Goovaerts, 1997). In the 3D approach one variogram was fit for the whole profile and both lateral and vertical anisotropies were accounted for.

The sum of trend and corresponding residual realisations from the Gaussian simulations was calculated and at each node giving a total of 1000 realisation for each soil properties considered. The realisations thus obtained were summarised with median and upper (quant95) and lower (quant5) percentiles. The variability of the target property values (ΔV) was estimated as the difference between the upper and the lower percentiles:

$$\Delta V = quant95 - quant5 \quad (7)$$

The deviation of the realisations from the median was also expressed as a percentage as a further measure of the spread of credible model outputs:

$$\Delta M = \frac{|\Delta V|}{median} \times 100 \quad (8)$$

A 2D version of the approach describe above was used on a prediction grid at 100 m resolution (see Poggio et al. (2010, 2013); Poggio and Gimona (SUBM) for previous implementations). In particular the 3D smoother was substituted with a 2D smoother of the geographical coordinates and the variograms considered only lateral anisotropy. The 2D approach was used in the pre-defined models, for soil depth and when considering topsoil data (0-15cm).

3.2.1 Soil properties transformation

Table 1 describes the transformation applied to the soil properties for the modelling and mapping. Organic carbon and base saturation were mapped onto the real line by the logit function to provide

(approximately) Normally distributed observations. For these properties a simple back-transformation was used before validation assessment and mapping of the results.

Soil texture was expressed as the relative percentage of sand, silt and clay and the sum of the components always equals 100. Therefore they can be treated as compositional variables and should be transformed using log-ratios (Aitchison, 1986). In particular the Addictive Log-ratio (ALR) transformation was used. ALR is an extension of the logit transformation when the response variable has more than two components. The back-transformation is a numerical approximation following GaussHermite quadrature (Aitchison, 1986). ALR was previously applied to soil texture data (Lark and Bishop, 2007; Akpa et al., 2014; Ballabio et al., 2016; Poggio and Gimona, SUBM) and it was showed (Lark and Bishop, 2007) that ALR transformed variables preserve information on the spatial correlation and maintain the compositional aspect of the variables.

Given a composition of D elements $\mathbf{z} = [z_1, \dots, z_D]$, such as $z_i > 0 \forall i = 1, \dots, D$ and $\sum_{i=1}^D z_i = 1$, ALR is defined as:

$$\mathbf{x} = ALR(\mathbf{z}) = \left(\ln \frac{z_1}{z_D}, \dots, \ln \frac{z_{D-1}}{z_D} \right) \quad (9)$$

The inverse ALR transformation can be described as:

$$\mathbf{z} = \frac{\exp(\mathbf{w})}{\mathbf{j}^T \exp(\mathbf{w})} \quad (10)$$

where $\exp(\mathbf{w})$ denotes the vector $[\exp(\mathbf{w}_1), \exp(\mathbf{w}_2), \exp(\mathbf{w}_{D-1})]$ and $\mathbf{j}$ is a vector of length D with all elements equal to one (Lark and Bishop, 2007).

Therefore the two ALR components that were interpolated can be defined as:

$$\begin{aligned} ALR1 &= \ln \frac{\text{sand}}{\text{clay}} \\ ALR2 &= \ln \frac{\text{silt}}{\text{clay}} \end{aligned} \quad (11)$$

Sand, silt and clay are the components of soil texture as defined in section 2.1. The approach followed for soil texture is the same as in Poggio and Gimona (SUBM) with different covariates and resolution (100m instead of 250m).

3.3 Validation and summary statistics

All the models were run for 1000 realisations and the validation statistics were calculated for the median of the simulations. The percentage of the validation points contained in the ΔV (eq. 7) interval was calculated. As measure of the spatial variability the spatial structured variance ratio (SSVR) (Vaysse and Lagacherie, 2015) was calculated for the soil properties for the whole profile (3D variogram) and

at certain depths. The SSVR is defined as the complement to 1 of the nugget to sill ratio (Kerry and Oliver, 2008). The values range from zero to one. Values closer to one mean a higher proportion of the data explained by the spatial component. The results of the modelling were summarised for the whole profile and at five depth layers (i.e. 0-5, 5-15, 15-30, 30-60 and 60-100cm) and compared with values from the validation set at corresponding depths. The depths were selected according to the *GlobalSoilMap* specifications (Arrouays et al., 2014). The results were assessed using out-of-sample measures and compared for distribution similarity, spatial structure reproduction, computational load and uncertainty ranges. In particular:

- 1) Root Mean Square Error (RMSE),
- 2) Normalised Root Mean Square Error (nRMSE) using the range of each soil property
- 3) R^2 derived from a linear model between observed and modelled data (R^2LM).

3.4 Software used

The analyses were performed using open source software:

- 1) GRASS GIS (GRASS Development Team, 2014) for data management, preparation and visualisation;
- 2) the R software (R Core Team, 2016) for the statistical modelling. The following packages were used:
 - i) **raster** for data management, preparation and visualisation (Hijmans and van Etten, 2013); ii) **mgcv** for GAM (Wood, 2006); iii) **geoR** (Ribeiro and Diggle, 2001) for fitting the variograms of the residuals;
 - iv) **gstat** (Pebesma, 2004) for kriging; v) **rgdal** for data management (Keitt et al., 2009); vi) **snow** for parallelisation of the computations (Tierney et al., 2015).

The computation of 1000 simulations at 100m for the test area involved the management and processing of large datasets of about seven million pixels for each layer and for each simulation. In order to achieve this the previous developed GAM+GS (see section 3.2) approach had to be modified to run as parallel processes.

4 Results

4.1 Sensor inter-comparison

4.1.1 Pre-defined models

Table 5 presents the three models with the lowest AIC for each soil properties out of the models described in table 4. Models using MODIS data were selected for most of the variables with the models using downscaled data as first choice for six of the seven variables. Models with resampled data were included only in a limited number of cases and never as first choice. The models with downscaled MODIS indices provided better results than simple resampling. This can be explained with the extra-modelling effort. However it is computational intensive and adds a step in the data preparation. Therefore is not suited for all situations and test cases. Models using Landsat and S2 were selected as third best for base saturation and pH, respectively. This could be explained with the use of a snapshot instead of multiple dates that better describes the differences in water soil content and changes in vegetation. Models containing only the simpler indices (e.g. NDVI, NDWI) never appeared as the three lowest AICs. This could explain the increased potential of MODIS where additional information is available or can be derived (e.g. productivity, LAI, phenology). The use of thermal information improved models for both Landsat and MODIS. Sets of covariates containing Land Surface Temperature were in the three lowest AIC models for all considered soil properties.

S1 models were selected for soil texture, for both ALR transformed components. In order to further investigate the impact of S1 on modelling soil properties, further combinations of S1 bands with optical sensors were tested (models 16 to 21 in table 4). The difference in AIC between the model with optical bands without S1 and with it is shown in table 6. For all soil properties S1 reduced the AIC of the model, in some cases by a rather large value (>25). The explained deviance was also increased by 3 to 5 %. The RMSE calculated on the validation set improved for all properties considered.

According to this initial comparison, MODIS data averaged over 12 years provided the models with lowest AIC. However, the differences in the validation results were limited (table 7) with all models providing good RMSE values and reproducing the distribution of the validation set. These results suggest that it is important to have out-of-sample validation, because the effort to maximise in-sample metrics, such as AIC, is not always reflected in improvements in the out-of-sample validation, such as RMSE. It is therefore difficult to suggest the best approach as it depends on data availability, typical cloud cover, application and focus of the model, i.e. to explain relationships or to predict values.

Table 5: Pre-defined models with the three lowest AIC score for each considered soil property for topsoil (0-15 cm) with 2D modelling approach. 1 indicates the model with the lowest AIC. L=Landsat, S1=Sentinel-1, S2=Sentinel-2, M=MODIS, MD=MODIS downscaled, MR=MODIS resampled. For details on the covariates used see table 4.

			pH	BS	Depth	C	BD	ALR1	ALR2
Landsat	L	1							
	L	2							
	L	3		3					
MODIS	MR	4				3	3		
	MD	5				1	1		
	MD	6	1	1			2	1	3
	MR	7			2				
	MR	8							
	MD	9							
	MR	10	2	2	1	2			
	MR	11			3				
Sentinel	S2	13	3						
	S2	14							
	S1	15						2	1
Sentinel1 & optical	S1L	16							
	S1S2	17							
	S1M	18							
	S1M	19							
	S1M	20							
	S1M	21						3	2

Table 6: Δ AIC (see eq. 6) between models with S1 and combination of optical sensors and the same models without S1 bands. $- = \leq 2\Delta$ AIC, $* = 3 - 10\Delta$ AIC, with strong evidence that the models can be considered different, $** = 10 - 15\Delta$ AIC, with very strong evidence that the models can be considered different, $*** = > 15\Delta$ AIC, with extremely strong evidence that the models can be considered different. L=Landsat, S1=Sentinel-1, S2=Sentinel-2, M=MODIS. For details on the covariates used see table 4. Soil data for topsoil (0-15cm) with 2D modelling approach.

		pH	BS	Depth	C	BD	ALR1	ALR2
S1L	16	***	***	***	***	***	*	**
S1S2	17	*	-	*	**	***	**	**
S1M	18	***	***	***	***	***	*	*
S1M	19	***	**	**	***	***	*	*
S1M	20	***	*	***	***	***	***	***
S1M	21	* *	*	***	***	***	***	***

Table 7: RMSE between predicted and validation data for the pre-defined models. Bold numbers are the models with the three lowest AICs (see table 5). L=Landsat, S1=Sentinel-1, S2=Sentinel-2, M=MODIS, MD=MODIS downscaled, MR=MODIS resampled. For details on the covariates used see table 4. Soil data for topsoil (0-15cm) with 2D modelling approach.

			pH	BS	Depth	C	BD	Sand	Silt	Clay
Landsat	L	1	0.76	11.05	7.10	6.86	0.25	11.99	6.32	5.67
	L	2	0.79	11.07	7.10	6.88	0.25	11.93	6.28	5.60
	L	3	0.75	11.03	7.08	6.85	0.24	11.96	6.28	5.63
MODIS	MR	4	0.75	11.03	6.74	6.84	0.24	11.96	6.28	5.63
	MD	5	0.74	11.03	6.76	6.83	0.24	11.96	6.28	5.63
	MD	6	0.73	11.02	6.68	6.84	0.24	11.94	6.27	5.62
	MR	7	0.75	11.05	6.50	6.86	0.25	11.94	6.27	5.62
	MR	8	0.78	11.07	6.80	6.87	0.25	11.92	6.26	5.61
	MD	9	0.77	11.07	6.80	6.87	0.25	11.92	6.27	5.60
	MR	10	0.74	11.01	6.48	6.84	0.24	11.96	6.28	5.63
	MR	11	0.79	11.07	7.04	6.89	0.25	11.95	6.27	5.63
	MR	12	0.76	11.05	6.74	6.85	0.25	11.93	6.27	5.61
Sentinel	S2	13	0.79	10.06	6.83	6.88	0.26	11.99	6.29	5.64
	S2	14	0.81	11.09	7.14	6.91	0.26	11.99	6.29	5.64
	S1	15	0.75	11.05	6.75	6.85	0.25	11.97	6.29	5.63
Sentinel1 & optical	S1L	16	0.78	11.06	6.99	6.88	0.25	11.94	6.27	5.62
	S1S	17	0.76	11.04	6.98	6.85	0.25	11.94	6.27	5.62
	S1M	18	0.79	11.07	7.16	6.89	0.25	11.97	6.29	5.63
	S1M	19	0.78	11.07	7.09	6.87	0.25	11.95	6.28	5.62
	S1M	20	0.74	11.04	6.62	6.84	0.25	11.93	6.27	5.61
	S1M	21	0.75	11.05	6.65	6.86	0.25	11.91	6.26	5.61

Table 8: Significant covariates selected with the variables tournament and included in the 2D modelling approach for topsoil only data (0-15cm). Significant level of $p < 0.01$. 1=covariate selected, 0=covariate not selected.

	pH	C	BD	BS	ALR1	ALR2	Depth	Sum
Elevation	1	1	1	1	0	0	1	5
Slope	0	1	1	0	0	0	1	3
TWI	0	0	0	0	1	1	0	2
EVI	1	0	1	0	1	0	1	4
Snow index	0	0	0	0	0	0	0	0
NDWI MODIS	1	0	0	0	0	0	0	1
Primary productivity	0	0	0	0	0	0	0	0
Amplitude	1	1	1	1	1	0	1	6
LOS	1	0	0	1	1	0	0	3
LST days	1	1	0	0	0	0	0	2
LST night	1	0	1	0	1	0	0	3
NDVI Landsat	0	0	0	0	0	1	1	1
NDWI Landsat	0	1	1	1	1	1	0	5
SCI Landsat	0	0	1	0	1	1	1	4
SM Landsat	0	1	1	0	0	0	0	2
NDVI S2	0	0	0	0	1	0	0	1
NDWI S2	0	0	0	1	1	1	0	3
SCI S2	0	0	0	0	0	0	0	0
SM S2	1	1	0	0	0	0	0	2
VV Polarisation	0	0	0	0	1	1	0	2
VH Polarisation	0	0	0	0	1	1	0	2
Explained deviance (%)	38.1	38.3	38.5	58.6	50.1	37.2	41.1	—

4.1.2 Variables tournament

Table 8 and table 9 present the significant variables for the 2D and the 3D models from a pool including morphological features, coordinates smoother and indices from the RS sensors. DEM-derived data are important for all considered soil properties. This is similar to what found from previous studies and as expected given the morphology of Scotland. Elevation and mountains are important factor and proxies for soil formation and development. Land surface temperature is also important for mostly of the considered soil properties. Again the long term median of LST is a proxy for climate and morphology. The different vegetation and water indices play also an important role for most of the considered soil properties. Snow derived indices are considered significant only for a reduced number of soil properties, such as bulk density and soil depth.

In general the variables considered are highly correlated among them self and mostly describe the morphological and climate gradients existing in Scotland.

Table 9: Significant covariates selected with the variables tournament and included in the models for 3D data (whole profile censored at 100cm). Significant level of $p < 0.01$. 1=covariate selected, 0=covariate not selected.

	pH	C	BD	BS	ALR1	ALR2	Sum
Elevation	1	1	1	1	0	0	4
Slope	0	0	0	1	1	0	2
TWI	0	0	0	0	1	1	2
EVI	0	1	0	1	0	0	2
Snow index	0	0	0	0	0	0	0
NDWI MODIS	0	0	0	0	0	0	0
Primary productivity	0	0	0	0	0	1	1
Amplitude	1	1	0	0	0	0	2
LOS	1	0	0	1	0	0	2
LST days	1	0	0	0	0	0	1
LST night	1	1	1	0	0	1	4
NDVI Landsat	0	0	0	0	0	0	0
NDWI Landsat	0	0	0	0	1	0	1
SCI Landsat	0	0	0	0	0	0	0
SM Landsat	1	1	0	1	1	0	4
NDVI S2	0	0	0	0	0	0	0
NDWI S2	0	1	0	0	1	0	2
SCI S2	0	0	0	0	0	0	0
SM S2	1	0	0	0	0	0	1
VV Polarisation	0	0	1	0	1	1	3
VH Polarisation	0	1	0	0	1	1	3
Explained deviance (%)	39.9	53.5	47.8	33.9	34.6	21.4	—

4.2 Prediction and uncertainty of selected soil properties

Figure 2 shows the validation results for the 2D and 3D models at 100m. The validation matrices showed good agreement with low RMSE and rather high R^2 . For all considered properties substantial proportions of their variance were spatially structured, with SSVR > 0.5 . The validation results for 100m resolution predictions improved compared to previous tests at 1000m (e.g. Poggio et al., 2013). This could be explained by the covariates being closer to the data support size. This is an indication that lower resolution covariates capture in more details the relationships with soils, thus reducing the impact of the spatial auto-correlation on the results. The validation metrics are generally decreasing with depth, with increasing values for RMSE. However there are exceptions, such as C content, with lower RMSE values at higher depths. This can be explained with the lower C contents at depth.

Figure 3 shows the validation assessment of the distribution of the modelled values at validation points and their spatial structure. For all soil properties considered the reproduction of values shown by qq-plots were rather good for most of the distribution range. The better agreements were found for soil texture, pH and bulk density. The results obtained for soil texture at 100m with the use of a mix of sensors appeared closer to the validation values than when using 250m resolution and only MODIS bands and morphological features (Poggio and Gimona, SUBM). The soil property with larger differences in values distributions was organic carbon, especially for mid range values. This was already shown in Poggio and Gimona (2014); Poggio et al. (2016) and it can be explained with the typical binomial distribution of organic soils in Scotland. The variograms shown in figure 3 indicate a similar spatial trend for the modelled and validation data with a lower sill due to the predicted data being smoother as expected from the modelling process. However there are exceptions such as BS (figure 3f) where the variogram of predicted values is different at short ranges despite the good explained deviance of the models (table 8 for topsoil and table 9 for the whole profile). This highlights the fact that samples are not sufficient to explain the short scale variability of the soil properties, despite the relatively large concentration of soil profiles available in Scotland.

Figure 4 shows the percentage of the validation values that fall within the calculated the interval between 5 and 95th percentiles. The percentages are higher for predictions at 100m than at 1000m. For 100m predictions the values are between 60 and 80% with lowest values for clay. The values are similar across the soil properties considered and they decrease with depth.. This is also shown in figure 3 where the qq-plots and the variograms for the different simulations are shown. The qq-plots show a rather good fit. For most of the soil properties considered the figure showed a smaller spread of the realisation predictions for the lower values. Exceptions were carbon and sand where the spread was larger for lower values. This pattern is linked with the number of samples in the tails of the range as for both carbon

Figure 2: Validation summary statistics and SSVR values for 2D and 3D models at 100m resolution. L1=0-5cm; L2=5-15cm; L3=15-30cm; L4=30-60cm; L5=60-100cm.

(a) pH - values

(b) pH - variogram

(c) C - values

(d) C - variogram

Figure 3: Validation plots: qq-plots between predicted and measured quantiles of the distribution and experimental variograms for the 1000 realisations. The blue line in the variogram plots refers to the validation data, while in the qq-plots represents the 1-to-1 line.

(e) BS - values

(f) BS - variogram

(g) BD - values

(h) BD - variogram

Figure 3: Continued.

(i) Sand - values

(j) Sand - variogram

(k) Silt - values

(l) Silt - variogram

(m) Clay - values

(n) Clay - variogram

Figure 3: Continued.

Figure 4: Proportion of the validation points contained within ΔV (eq. 7). L1=0-5cm; L2=5-15cm; L3=15-30cm; L4=30-60cm; L5=60-100cm.

and sand there is a smaller number of samples in the lower range, while for the other properties there is a smaller number of samples in the higher range.

4.3 Spatial distribution

Figure 5 show the mapping results for the selected properties. Not all the maps for all layers could be presented here. The properties presented are an example of primary (pH) and secondary (BD) soil properties. The results for topsoil only produced with a 2D modelling approach are presented in supplement A. The maps for depth, organic carbon and soil texture (presented in supplement B) have a very similar the spatial pattern of previous results (Poggio and Gimona, 2014, SUBM).

The figures show the median and the uncertainty (eq. 8) for the two selected properties. The spatial distribution shows more acidic soils on the west and on the mountains where the largest areas of peat are found. Higher pH is predicted on the east coast and on plain regions where most of the agricultural good quality land is located (Brown et al., 2011). BD follows a similar pattern with lower values on the west coast and higher values for the agricultural soils.

The analysis of the uncertainty shows that lower variability was modelled on the West coast and mountains area for both properties. The uncertainty for BD is lowest where larger patches of peat are located with larger values for agricultural soils, where absolute values are also higher. The spatial pattern is different when considering the percent of the uncertainty from the median (eq. 8). For the pH the lowest values can be found for the more acidic soils, with highest values for agricultural soils. In case of BD, the highest values of ΔM can be found on the West and mountainous soils with lower values for agricultural soils. The patterns can be explained because of the higher values in the regions where the short scale variability of the considered soil properties is larger. This uncertainty map considered only

the spatial modelling uncertainty and not the uncertainty due to the model used to derived BD values (Poggio et al., 2014; Heuvelink and Burrough, 2002).

Figure 5 shows also the change of the spatial pattern with depth. For both properties there is a decrease of median values from topsoil (L1) to 15, 30 and 60 cm layers in the areas with alpine and organic soils (West coast and mountains). The values for the deeper agricultural soils show a more uniform pattern with depth. The uncertainty maps show a pattern with increasing values of uncertainty for deeper layers. The increase in uncertainty can be explained by the change in relationships between the soil forming factors and the covariates used.

(a) Median: pH

(b) ΔV : pH

Figure 5: Spatial distribution of median and uncertainty, ΔV (eq. 7) for selected soil properties for five depth layers using 3D modelling approach. L1=0-5cm; L2=5-15cm; L3=15-30cm; L4=30-60cm; L5=60-100cm.

(c) Median: BD

(d) ΔV : BD

Figure 5: Continued.

5 Discussion

Digital soil mapping relies on the use of statistical models to produce digital representations of spatial soil distribution using point soil observations and spatially exhaustive environmental covariates (McBratney et al., 2003; Scull et al., 2003; Florinsky, 2012). DSM requires sufficient and appropriately distributed point soil data within the area being mapped (Carre et al., 2007). The model structure explores only the empirical relationship among environmental conditions and soil properties (Grunwald, 2009). It should be born in mind, that the covariates are only approximations of the true environmental conditions and might serve only as proxies of environmental variables that have a casual relationship with soil properties. However, remote sensing proved effective to help build the empirical relationships in a range of different environmental conditions (see reviews of Mulder et al. (2011) and Wulf et al. (2014) and references therein).

In order to produce soil properties maps with finer cell size, higher resolution covariates are necessary. DEM is available at least at 90m resolution. RS data are available at different resolutions. Quite often used in DSM are Landsat (30m) and MODIS (500 to 1000m). MODIS has a coarser spatial resolution, but higher overpass and higher spectral dimensions. Landsat has a higher spatial resolution, but a lower overpass and it is more difficult to get cloud free scenes (especially in oceanic location with moist temperate climate). S2 data became recently available with an slightly shorter overpass and finer spatial resolution compared to Landsat. However the shorter time series increases the difficulties to select cloud free scenes. S1 has a good overpass and not many problems with cloud cover. Radar sensors were used in DSM especially for physical properties (texture, water content) (Niang et al., 2014; White et al., 2015; Vrieling, 2006; Moran et al., 2004).

This study explored the relationships between soil properties and readily available sensors (Landsat, MODIS, S1 and S2) with different spatial and temporal resolutions, wavelength ranges and number of bands in Scotland, a region with high soil variability at short distances. The high cloud coverage in Scotland (Perry and Hollis, 2005) was an important factor to take into account. Therefore both time series and radar data played an important role in the modelling.

The results showed good validation statistics with a RMSE below 20% of the range for all considered soil properties. The RMSE improved from previous studies including only MODIS sensor and using a coarser prediction grid. The performance of the models was similar to previous studies at regional (e.g. Vaysse and Lagacherie, 2015; Dobarco et al., 2016), national (e.g. Adhikari et al., 2013; Lacoste et al., 2016; Mulder et al., 2016a) or continental scale (Hengl et al., 2015; Ballabio et al., 2016; Viscarra et al., 2015). The use of an integrated 3D procedure helps to reduce the propagation of the uncertainty due to separately fitting a model for depth followed by interpolations (Mulder et al., 2016a; Viscarra et al.,

2015; Hengl et al., 2014). Compared to the use of only machine learning approaches (e.g. Mulder et al., 2016a), the use of a hybrid geostatistical approach included the modelling of the spatial structure of the residuals. Despite the increased computational load, this approach provides a better modelling of the spatial pattern of the data (e.g. Goovaerts, 1997). This study presents a first attempt for Scotland to map the required soil properties following the *GlobalSoilMap* specifications (Arrouays et al., 2014) and shows the importance of RS covariates in the process. Some primary properties (e.g. coarse fragments, cation exchangeable capacity, depth for root) are missing. From the secondary properties available water capacity was mapped following a similar approach (Poggio et al., 2010) but further work is needed to update the resolution of the predictions. Further properties could be derived in particular related to hydrological properties and soil moisture (Tóth et al., 2015; Mohanty, 2013).

Further work will be needed to compare the Scottish DSM products with EU-wide (Ballabio et al., 2016; de Brogniez et al., 2015) and global products (Hengl et al., 2014). Previous work (Mulder et al., 2016b) suggested that a bottom-up approach may be more favourable. However, the variety of methods used, differences in data input, covariates used and processes involved need to be taken into account when making a direct comparison.

Previous studies used mostly one sensor, mainly Landsat or MODIS. However, in this study, the combination of optical and radar sensors gave the best models and most accurate predictions results. The use of radar S1 data proved useful for all soil properties and texture in particular. Previous studies showed the use of radar for DSM (e.g. Niang et al., 2014; Akpa et al., 2014; Pasolli et al., 2011). However, to our knowledge, this is the first test with S1 data for DSM applications over a variety of soil properties. The median of one year data was used and further investigations are needed to assess the results when accounting for the seasonality within S1 data.

Combining optical with radar sensors has a great potential to overcome issues, especially in regions with high cloud cover. Sensors with more frequent overpass, but coarser resolution can provide the seasonality element to sensors with less frequent overpass but more detailed resolution. Sensors with higher spatial resolution can provide information of the shorter range variability. Readily available radar bands can add information without the issue of the cloud cover. The combination of all these information is powerful for further predictions such as DSM.

A large portion of the soil variation is still largely spatially structured and is likely to remain unpredictable unless different covariates (such as Lidar or gamma radiometric (e.g. Stockmann et al., 2015; Beamish, 2014; Cook et al., 1996)) or further ground data become available (Vaysse and Lagacherie, 2015; Samuel-Rosa et al., 2015). However, future hyperspectral sensors may support better relationships with soil properties and improve the results further (Castaldi et al., 2016).

6 Conclusions

A combination of optical and radar remote sensing data has the potential to support the assessment of soil properties and associated soil functions over wide areas. The understanding of spatial variability of soil properties has large implications for a range of land functions such as food and water security and prevention of land degradation, promoting global environment management and monitoring.

This study explored the relationships between soil properties and readily available sensors (Landsat, MODIS, S1 and S2) with varying spatial, temporal and spectral resolution in Scotland. Different soil properties were mapped at 100m resolution for Scotland at five depths using a geostatistical approach taking into account the full 3D spatial relationships between input data. The spatial uncertainty of the process was taken into account with a large number of simulations.

The results showed that the use of a mixture of sensors proved more effective to model and map soil properties than the individual sensors. In particular:

1. the use of S1 improved the prediction capability of models with only optical bands,
2. the use of MODIS time series and their statistics provided stronger relationships than the use of snapshots, and
3. models at finer resolution (i.e. 100m) had lower RMSE results for validation.

Future work may include the use of hyperspectral sensors, seasonality variation of Sentinel-1 and more Sentinel-2 data once available. More soil properties can be modelled, in particular related to soil moisture and water regimes. Particular efforts is needed to focus on uncertainty propagation of both modelling and mapping stages. The produced maps of soil properties describing both lateral and vertical variability, with associated uncertainty, are important for further modelling and management of soil resources and ecosystem services. Coupled with further information the soil properties maps could be used to assess conditions and suitability of soils for a range of purposes.

Acknowledgements

This work was funded by the Scottish Government's Rural and Environment Science and Analytical Services division. Many thanks to the team that sampled and analysed the soils and to the team that set up the database. MODIS data are distributed by the Land Processes Distributed Active Archive Centre (LP DAAC), located at the U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center (lpdaac.usgs.gov). Landsat data were available from the U.S. Geological Survey. Sentinel data were available from the European Space Agency and Copernicus Service. Thanks are due to Dr Luigi Spezia (BioSS) for comments on an early version of this manuscript.

References

- Adhikari, K., Hartemink, A. E., 2016. Linking soils to ecosystem services - A global review. *Geoderma* 262, 101–111.
- Adhikari, K., Kheir, R. B., Greve, M. B., Bocher, P. K., Malone, B. P., Minasny, B., McBratney, A. B., Greve, M. H., 2013. High-Resolution 3-D Mapping of Soil Texture in Denmark. *Soil Science Society of America Journal* 77 (3), 860–876.
- Agbu, P., Fehrenbacher, D., I.J., J., 1990. Soil property relationships with SPOT satellite digital data in east central Illinois. *Soil Science Society of America Journal* 54, 807–812.
- Aitchison, J., 1986. *The statistical analysis of compositional data*. Chapman & Hall, London.
- Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B., Csaki, F. (Eds.), *Second International Symposium on Information Theory*. Akademia Kiado, Budapest, Hungary, pp. 267–281.
- Akpa, S. I. C., Odeh, I. O. A., Bishop, T. F. A., Hartemink, A. E., 2014. Digital Mapping of Soil Particle-Size Fractions for Nigeria. *Soil Science* 78 (6), 1953–1966.
- Arrouays, D., Grundy, M. G., Hartemink, A. E., Hempel, J. W., Heuvelink, G. B., Hong, S. Y., Lagacherie, P., Lelyk, G., McBratney, A. B., McKenzie, N. J., d.L. Mendonca-Santos, M., Minasny, B., Montanarella, L., Odeh, I. O., Sanchez, P. A., Thompson, J. A., Zhang, G.-L., 2014. Chapter three - globalsoilmap: Toward a fine-resolution global grid of soil properties. Vol. 125 of *Advances in Agronomy*. Academic Press, pp. 93 – 134.
- Augustin, N., Musio, M., von Wilpert, K., Kublin, E., Wood, S., Schumacher, M., 2009. Modeling spatiotemporal forest health monitoring data. *Journal of the American Statistical Association* 104 (487), 899–911.
- Bachofer, F., Queneherve, G., Hochschild, V., Maerker, M., 2015. Multisensoral Topsoil Mapping in the Semiarid Lake Manyara Region, Northern Tanzania. *Remote Sensing* 7 (8), 9563–9586.
- Ballabio, C., Panagos, P., Monatanarella, L., 2016. Mapping topsoil physical properties at European scale using the LUCAS database. *Geoderma* 261, 110–123.
- Beamish, D., 2014. Peat Mapping Associations of Airborne Radiometric Survey Data. *Remote Sensing* 6 (1), 521–539.

- Ben-Dor, E., Chabrilat, S., Dematt, J., Taylor, G., Hill, J., Whiting, M., Sommer, S., 2009. Using imaging spectroscopy to study soil properties. *Remote Sensing of Environment* 113, Supplement 1, S38 – S55, imaging Spectroscopy Special Issue.
- Ben-Dor, E., Taylor, R. G., Hill, J., Dematte, J. A. M., Whiting, M. L., Chabrilat, S., Sommer, S., 2008. Imaging spectrometry for soil applications. In: Sparks, DL (Ed.), *Advances in Agronomy*. Vol. 97 of *Advances in Agronomy*. pp. 321+.
- Boettinger, J., Howell, D., Moore, A., Hartemink, A., Kienast-Brown, S. (Eds.), 2010. Digital soil mapping: bridging research, environmental application, and operation. *Progress in Soil Science*. Springer.
- Brown, I., Poggio, L., Gimona, A., Castellazzi, M., 2011. Climate change, drought risk and land capability for agriculture: implications for land-use in scotland. *Regional Environmental Change* 11 (3), 503–518.
- Burnham, K., Anderson, D., 2013. *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach*. Springer.
- Burrough, P. A., McDonell, R., 2000. *Principles of Geographical Information System*. Oxford University Press.
- Carre, F., McBratney, A. B., Minasny, B., 2007. Estimation and potential improvement of the quality of legacy soil samples for digital soil mapping. *Geoderma* 141 (1?2), 1–14.
- Castaldi, F., Casa, R., Castrignano, A., Pascucci, S., Palombo, A., Pignatti, S., 2014. Estimation of soil properties at the field scale from satellite data: a comparison between spatial and non-spatial techniques. *European Journal of Soil Science* 65 (6, SI), 842–851.
- Castaldi, F., Palombo, A., Santini, F., Pascucci, S., Pignatti, S., Casa, R., 2016. Evaluation of the potential of the current and forthcoming multispectral and hyperspectral imagers to estimate soil texture and organic carbon. *Remote Sensing of Environment* 179, 54–65.
- Chander, G., Markham, B. L., Helder, D. L., 2009. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sensing of Environment* 113 (5), 893–903.
- Chaney, N. W., Wood, E. F., McBratney, A. B., Hempel, J. W., Nauman, T. W., Brungard, C. W., Odgers, N. P., 2016. Polaris: A 30-meter probabilistic soil series map of the contiguous united states. *Geoderma* 274, 54 – 67.
- Cook, S., Corner, R., Groves, P., Grealish, G., 1996. Use of airborne gamma radiometric data for soil mapping. *Australian Journal of Soil Research* 34, 183–194.

627 Cressie, N., 1993. Statistics for Spatial Data. Wiley, New York.

628 de Brogniez, D., Ballabio, C., Stevens, A., Jones, R. J. A., Montanarella, L., van Wesemael, B., 2015.
629 A map of the topsoil organic carbon content of europe generated by a generalized additive model.
630 European Journal of Soil Science 66 (1), 121–134.

631 Deutsch, C., Journel, A., 1998. GSLIB: Geostatistical Software Library and User’s Guide, second edition
632 Edition. Oxford University Press, New York.

633 Dewitte, O., Jones, A., Elbelrhiti, H., Horion, S., Montanarella, L., 2012. Satellite remote sensing for
634 soil mapping in Africa: An overview. progress in Physical Geography 36 (4), 514–538.

635 Dobarco, M. R., Orton, T. G., Arrouays, D., Lemerrier, B., Paroissien, J.-B., Walter, C., Saby, N. P.,
636 2016. Prediction of soil texture using descriptive statistics and area-to-point kriging in region centre
637 (france). Geoderma Regional 7 (3), 279 – 292.

638 Eklundh, L., Jonsson, P., 2011. Timesat 3.1 Software Manual. Tech. rep., Lund University, Sweden.

639 Florinsky, I., 2012. Digital terrain analysis in soil science and geology. Academic Press.

640 Gao, B., 1996. NDWI - a normalized difference water index for remote sensing of vegetation liquid water
641 from space. Remote Sensing of Environment 58, 257 – 266.

642 Gomez, C., Lagacherie, P., Coulouma, G., 2012. Regional predictions of eight common soil properties
643 and their spatial structures from hyperspectral Vis-NIR data. Geoderma 189, 176–185.

644 Google Earth Engine Team, 12 2015. Google earth engine: A planetary-scale geo-spatial analysis plat-
645 form. <https://earthengine.google.com>.

646 Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press.

647 GRASS Development Team, 2014. Geographic Resources Analysis Support System (GRASS GIS) Soft-
648 ware, version 6.5.0svn.
649 URL <http://www.grass.osgeo.org>

650 Grunwald, S., 2009. Multi-criteria characterization of recent digital soil mapping and modeling ap-
651 proaches. Geoderma 152 (3-4), 195–207.

652 Gu, Y., Hunt, E., Wardlow, B., Basara, J. B., Brown, J. F., Verdin, J. P., 2008. Evaluation of MODIS
653 NDVI and NDWI for vegetation drought monitoring using Oklahoma Mesonet soil moisture data.
654 Geophysical Research Letters 35 (22), L22401.

655 Hartemink, A. E., Krasilnikov, P., Bockheim, J., 2013. Soil maps of the world. *Geoderma* 207-208, 256
656 – 267.

657 Hengl, T., de Jesus, J. M., MacMillan, R. A., Batjes, N. H., Heuvelink, G. B. M., Ribeiro, E., Samuel-
658 Rosa, A., Kempen, B., Leenaars, J. G. B., Walsh, M. G., Gonzalez, M. R., 2014. SoilGrids1km-Global
659 Soil Information Based on Automated Mapping. *PLoS ONE* 9 (8).

660 Hengl, T., Heuvelink, G. B. M., Kempen, B., Leenaars, J. G. B., Walsh, M. G., Shepherd, K. D., Sila, A.,
661 MacMillan, R. A., Mendes de Jesus, J., Tamene, L., Tondoh, J. E., 06 2015. Mapping soil properties
662 of africa at 250 m resolution: Random forests significantly improve current predictions. *PLoS ONE*
663 10 (6), 1–26.

664 Heuvelink, G. B. M., Burrough, P. A., 2002. Developments in statistical approaches to spatial uncertainty
665 and its propagation. *International Journal of Geographical Information Science* 16 (2), 111–113.

666 Hijmans, R. J., van Etten, J., 2013. raster: Geographic data analysis and modeling. R package version
667 2.1-25.
668 URL <http://CRAN.R-project.org/package=raster>

669 Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., Ferreira, L. G., 2002. Overview of the radio-
670 metric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*
671 83 (1-2), 195 – 213.

672 Irish, R., Barker, J., Goward, S., Arvidson, T., 2006. Characterization of the landsat-7 ETM+ Automated
673 Cloud-Cover Assessment (ACCA) Algorithm. *Photogrammetric Engineering and Remote Sensing* 72,
674 1179–1188.

675 Jarvis, A., Reuter, H., Nelson, A., Guevara, E., 2006. Hole-filled seamless SRTM data V3. Tech. rep.,
676 International Centre for Tropical Agriculture (CIAT).

677 Jönsson, P., Eklundh, L., 2004. TIMESAT - a program for analyzing time-series of satellite sensor data.
678 *Computer and Geosciences* 30, 833–845.

679 Journel, A., 1996. Modelling uncertainty and spatial dependence: Stochastic imaging. *International*
680 *Journal of Geographical Information Systems* 10 (5), 517–522.

681 Ju, J., Roy, D., 2008. The availability of cloud-free Landsat ETM+ data over the conterminous United
682 States and globally. *Remote Sensing of Environment* 112, 1196–1211.

- Keitt, T., Bivand, R., Pebesma, E., Rowlingson, B., 2009. rgdal: Bindings for the Geospatial Data Abstraction Library. R package version 0.6-21.
URL <http://CRAN.R-project.org/package=rgdal>
- Kerry, R., Oliver, M., 2008. Determining nugget:sill ratios of standardized variograms from aerial photographs to kriging sparse soil data. *Precision Agriculture* 9, 33–56.
- Knyazikhin, Y., Glassy, J., Privette, J. L., Tian, Y., Lotsch, A., Zhang, Y., Wang, Y., Morisette, J. T., Votava, P., Myneni, R. B., Nemani, R. R., Running, S. W., 1999. MODIS Leaf Area Index (LAI) and Fraction of Photosynthetically Active Radiation Absorbed by Vegetation (FPAR) Product (MOD15) Algorithm, Theoretical Basis Document. Tech. rep., NASA Goddard Space Flight Center.
- Lacoste, M., Mulder, V., de Forges, A. R., Martin, M., Arrouays, D., 2016. Evaluating large-extent spatial modeling approaches: A case study for soil depth for france. *Geoderma Regional* 7 (2), 137 – 152.
- Lark, R., Bishop, T., 2007. Cokriging particle size fractions of the soil. *European Journal of Soil Science* 58, 763–774.
- Lilly, A., Bell, J., Hudson, G., Nolan, A., Towers, W., 2010. National Soil Inventory of Scotland 1 (NSIS1): site location, sampling and profile description. (1978-1998). Tech. rep., Macaulay Institute.
- Malone, B. P., McBratney, A. B., Minasny, B., Wheeler, I., 2012. A general method for downscaling earth resource information. *Computers & Geosciences* 41, 119 – 125.
- McBratney, A., Santos, M., Minasny, B., 2003. On digital soil mapping. *Geoderma* 117, 3–52.
- Mendonça-Santos, M., dos Santos, H., 2006. Chapter 3 the state of the art of brazilian soil mapping and prospects for digital soil mapping. In: P. Lagacherie, A. M., Voltz, M. (Eds.), *Digital Soil Mapping An Introductory Perspective*. Vol. 31 of *Developments in Soil Science*. Elsevier, pp. 39 – 601.
- Minasny, B., McBratney, A., 2016. Digital soil mapping: A brief history and some lessons. *Geoderma* 264, Part B, 301 – 311, soil mapping, classification, and modelling: history and future directions.
- Mohanty, B. P., 2013. Soil Hydraulic Property Estimation Using Remote Sensing: A Review. *Vadose Zone Journal* 12 (4).
- Moran, M., Peters-Lidard, C., Watts, J., McElroy, S., 2004. Estimating soil moisture at the watershed scale with satellite-based radar and land surface models. *Canadian Journal Of Remote Sensing* 30 (5), 805–826.

712 Mulder, V., de Bruin, S., Schaepman, M., Mayr, T., 2011. The use of remote sensing in soil and terrain
713 mapping - A review. *Geoderma* 62, 1–19.

714 Mulder, V., Lacoste, M., de Forges, A. R., Arrouays, D., 2016a. Globalsoilmap france: High-resolution
715 spatial modelling the soils of france up to two meter depth. *Science of The Total Environment*, –.

716 Mulder, V., Lacoste, M., de Forges, A. R., Martin, M., Arrouays, D., 2016b. National versus global
717 modelling the 3d distribution of soil organic carbon in mainland france. *Geoderma* 263, 16 – 34.

718 Niang, M. A., Nolin, M. C., Jegou, G., Perron, I., 2014. Digital Mapping of Soil Texture Using
719 RADARSAT-2 Polarimetric Synthetic Aperture Radar Data. *Soil Science Society of America Jour-*
720 *nal* 78 (2), 673–684.

721 Obade, V. d. P., Lal, R., 2013. Assessing land cover and soil quality by remote sensing and geographical
722 information systems (GIS). *Catena* 104, 77–92.

723 O’Callaghan, J. F., Mark, D. M., 1984. The extraction of drainage networks from digital elevation data.
724 *Computer Vision Graphics Image Processing* 28, 323–344.

725 Padarian, J., Minasny, B., McBratney, A., 2015. Using google’s cloud-based platform for digital soil
726 mapping. *Computers & Geosciences* 83, 80 – 88.

727 Padarian, J., Minasny, B., McBratney, A., Dalglish, N., 2014. Predicting and mapping the soil available
728 water capacity of australian wheatbelt. *Geoderma Regional* 273, 110 – 118.

729 Pasolli, L., Notarnicola, C., Bruzzone, L., Bertoldi, G., Della Chiesa, S., Niedrist, G., Tappeiner, U.,
730 Zebisch, M., 2011. Polarimetric RADARSAT-2 imagery for soil moisture retrieval in alpine areas.
731 *Canadian Journal Of Remote Sensing* 37 (5, SI), 535–547.

732 Pebesma, E., 2004. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences* 30,
733 683–691.

734 Perry, D., Hollis, D., 2005. The development of a new set of long-term climate averages for the UK.
735 *International Journal of Climatology* 25, 1023–1039.

736 Poggio, L., Gimona, A., 2014. National scale 3D modelling of soil organic carbon stocks with uncertainty
737 propagation - An example from Scotland. *Geoderma* 232-234, 284–299.

738 Poggio, L., Gimona, A., 2015a. Downscaling and correction of regional climate models outputs with a
739 hybrid geostatistical approach. *Spatial Statistics*.

740 Poggio, L., Gimona, A., 2015b. Sequence-based mapping approach to spatio-temporal snow patterns
741 from MODIS time-series applied to Scotland. *International Journal of Applied Earth Observation and*
742 *Geoinformation* 34, 122–135.

743 Poggio, L., Gimona, A., SUBM. Digital soil mapping for soil texture: Scotland as test case. *Geoderma*
744 *Regional*.

745 Poggio, L., Gimona, A., Brewer, M., 2013. Regional scale mapping of soil properties and their uncertainty
746 with a large number of satellite-derived covariates. *Geoderma* 209-210, 1–14.

747 Poggio, L., Gimona, A., Brown, I., 2012. Spatio-temporal MODIS EVI gap filling under cloud cover: an
748 example in Scotland. *ISPRS Journal of Photogrammetry and Remote Sensing* 72, 56–72.

749 Poggio, L., Gimona, A., Brown, I., Castellazzi, M., 2010. Soil available water capacity interpolation and
750 spatial uncertainty modelling at multiple geographical extents. *Geoderma* 160, 175–188.

751 Poggio, L., Gimona, A., Spezia, L., Brewer, M., 2014. Example of Bayesian uncertainty for digital soil
752 mapping. In: *Proceedings of the 6th Global Workshop on Digital Soil Mapping*. Nanjing - China.

753 Poggio, L., Gimona, A., Spezia, L., Brewer, M., 2016. Bayesian spatial modelling of soil properties and
754 their uncertainty: the example of soil organic matter in Scotland using R-INLA. *Geoderma* 277, 69–82.

755 R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for
756 Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
757 URL <http://www.R-project.org/>

758 Rawlins, B. G., Marchant, B. P., Smyth, D., Scheib, C., Lark, R. M., Jordan, C., 2009. Airborne
759 radiometric survey data and a dtm as covariates for regional scale mapping of soil organic carbon
760 across northern ireland. *European Journal of Soil Science* 60 (1), 44–54.

761 Ribeiro, P., Diggle, P., 2001. geoR: a package for geostatistical analysis. *R-NEWS* 1 (2), 14–18.
762 URL <http://CRAN.R-project.org/doc/Rnews/>

763 Riggs, G., Hall, D., Salomonson, V., 2006. MODIS Snow Products User Guide to Collection 5. Tech. rep.

764 Rodriguez, E., Morris, C., Belz, J., Chapin, E., Martin, J., Daffer, W., Hensley, S., 2006. An assessment
765 of the SRTM topographic products. Tech. Rep. JPL D-31639, NASA-Jet Propulsion Laboratory.

766 Running, S., R. R., Nemani, F. A., Heinsch, M., Zhao, M., Reeves, Hashimoto., H., 2004. A continuous
767 satellite-derived measure of global terrestrial primary production. *BioScience* 54 (6), 47–560.

Running, S., Nemani, R., Glassy, J., Thornton, P., 1999. MODIS daily photosynthesis (PSN) and annual net primary production (NPP) product Algorithm Theoretical Basis Document. Tech. rep.

Samuel-Rosa, A., Heuvelink, G., Vasques, G., Anjos, L., 2015. Do more detailed environmental covariates deliver more accurate soil maps? *Geoderma* 243?244, 214 – 227.

Savitzky, A., Golay, M., 1964. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry* 36 (8), 1627–1639.

Schmid, T., Rodriguez-Rastrero, M., Escribano, P., Palacios-Orueta, A., Ben-Dor, E., Plaza, A., Milewski, R., Huesca, M., Bracken, A., Cicuendez, V., Pelayo, M., Chabrillat, S., 2016. Characterization of Soil Erosion Indicators Using Hyperspectral Data From a Mediterranean Rainfed Cultivated Region. *IEEE Journal Of Selected Topics In Applied Earth Observations And Remote Sensing* 9 (2), 845–860.

Scull, P., Chadwick, O., McArthur, D., 2003. Predictive soil mapping: A review. *Progress in Physical Geography* 27 (2), 7.

Sorensen, R., Zinko, U., Seibert, J., 2006. On the calculation of the topographic wetness index: evaluation of different methods based on field observations. *Hydrology and Earth System Sciences* 10, 101–112.

Stockmann, U., Malone, B. P., McBratney, A. B., Minasny, B., 2015. Landscape-scale exploratory radiometric mapping using proximal soil sensing. *Geoderma* 239, 115–129.

Stockmann, U., Padarian, J., McBratney, A., Minasny, B., de Brogniez, D., Montanarella, L., Hong, S. Y., Rawlins, B. G., Field, D. J., 2015. Global soil organic carbon assessment. *Global Food Security* 6, 9 – 16.

Tierney, L., Rossini, A. J., Li, N., Sevcikova, H., 2015. snow: Simple Network of Workstations. R package version 0.4-1.
URL <https://CRAN.R-project.org/package=snow>

Tóth, B., Weynants, M., Nemes, A., Mak, A., Bilas, G., Tth, G., 2015. New generation of hydraulic pedotransfer functions for europe. *European Journal of Soil Science* 66 (1), 226–238.

Vaudour, E., Gilliot, J. M., Bel, L., Brechet, L., Hamiache, J., Hadjar, D., Lemonnier, Y., 2014. Uncertainty of soil reflectance retrieval from SPOT and RapidEye multispectral satellite images using a per-pixel bootstrapped empirical line atmospheric correction over an agricultural region. *International Journal of Earth Observation and Geoinformation* 26, 217–234.

797 Vaysse, K., Lagacherie, P., 2015. Evaluating digital soil mapping approaches for mapping globalsoilmap
798 soil properties from legacy data in Languedoc-Roussillon (France). *Geoderma regional* 4, 20–30.

799 Viscarra, R. A., Chen, C., Grundy, M. J., Searle, R., Clifford, D., Campbell, P. H., 2015. The Australian
800 three-dimensional soil grid: Australia’s contribution to the GlobalSoilMap project. *Soil Research* 53 (8),
801 845–864.

802 Vrieling, A., 2006. Satellite remote sensing for water erosion assessment: A review. *Catena* 65 (1), 2–18.

803 Wan, Z., 1999. MODIS Land-Surface Temperature Algorithm Theoretical Basis Document (LST ATBD).
804 Tech. rep., NASA.

805 White, L., Brisco, B., Dabboor, M., Schmitt, A., Pratt, A., 2015. A Collection of SAR Methodologies
806 for Monitoring Wetlands. *Remote Sensing* 7 (6), 7615–7645.

807 Wilford, J., Minty, B., 2006. Chapter 16 the use of airborne gamma-ray imagery for mapping soils and
808 understanding landscape processes. In: P. Lagacherie, A. M., Voltz, M. (Eds.), *Digital Soil MappingAn
809 Introductory Perspective*. Vol. 31 of *Developments in Soil Science*. Elsevier, pp. 207 – 610.

810 Wood, S., 2006. *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC Press.

811 Wulf, H., Mulder, T., Schaepman, M., Keller, A., Jrg, P., 2014. Remote sensing of soils. Tech. Rep.
812 00.0338.PZ/435-0501, University of Zurich.