Uncertainty assessment of GlobalSoilMap soil available water capacity products: A French case study

Mercedes Roman Dobarco, Hocine Bourennane, Dominique Arrouays, Nicolas P. A. Saby, Isabelle Cousin, Manuel P. Martin

Title: Uncertainty assessment of *GlobalSoilMap* soil available water capacity products: a French case study

Authors:

Mercedes (First name) Román Dobarco (Last name) [a]: mercedes.roman-dobarco@inra.fr

Hocine Bourennane [b] : hocine.bourennane@inra.fr

Dominique Arrouays [a] : dominique.arrouays@inra.fr

Nicolas P. A. Saby [a] : nicolas.saby@inra.fr

Isabelle Cousin [b]: isabelle.cousin@inra.fr

Manuel P. Martin [a]: manuel.martin@inra.fr

Affiliations:

[a] INRA, US1106 Unité InfoSol, 2163 Avenue de la Pomme de Pin, CS 40001 Ardon, F-45075 Orléans Cedex 2, France

[b] INRA, UR0272 Unité de Science du Sol, 2163 Avenue de la Pomme de Pin, CS 40001 Ardon, F-45075 Orléans Cedex 2, France

Corresponding author:

Manuel P. Martin [a]: manuel.martin@inra.fr

Postal address: INRA Centre de recherche d'Orléans, US1106 Unité InfoSol, 2163 avenue de la Pomme de Pin, CS 40001 Ardon, F45075 Orléans CEDEX 2, France

Telephone: +33-238-41-48-21

Uncertainty assessment of *GlobalSoilMap* soil available water capacity products: a French case study.

Abstract

Plant available water capacity (AWC) refers to the maximum amount of water that a soil can store and provide to plant roots. Spatial predictions of AWC through digital soil mapping at high resolution and national extent provide relevant information for upscaling ecological and hydrological models, and assessment of the provision of ecosystem services like water quantity and quality regulation, carbon sequestration, and provision of food and raw materials. However, the spatial predictions of AWC are prone to errors and uncertainties. Moreover, this digital soil mapping process requires using pedotransfer functions (PTFs) due to the lack of sufficient georeferenced measurements of the upper (i.e., soil moisture at field capacity, θ_{FC}) and lower (i.e., soil moisture at permanent wilting point, θ_{PWP}) limits of soil moisture contents defining AWC. This adds an additional source of uncertainty to the final estimates of AWC. The objectives of this study were: 1) to predict AWC for mainland France following the *GlobalSoilMap* (GSM) project specifications on depth intervals and uncertainty, and 2) to quantify the uncertainty of AWC accounting for uncertainty of the soil input variables and the PTFs' coefficients. We first predicted the soil input properties by GSM layer (0–5, 5–15, 15–30, 30–60, 60–100, 100–200 cm), and then applied PTFs for estimating θ_{FC}, θ_{PWP}, and volumetric AWC (cm^3 cm^{-3}). The volume of coarse elements by GSM layer was subtracted before aggregating AWC to estimated soil depth for a maximum of 2 m. The uncertainty of AWC was quantified by first-order Taylor analysis. Independent evaluation indicated that clay had the lowest R^2 (clay R^2 = 0.27, silt R^2 = 0.43 and sand R^2 = 0.46) and RMSE (clay RMSE = 128 g kg^{-1}, silt RMSE = 139 g kg^{-1} and sand RMSE = 172 g kg^{-1}) from the three particle size fractions. However, the model for coarse elements had the worst predictive performance (R^2 = 0.14 and RMSE = 21 %) among all AWC input variables. The performance of the GSM predictions for θ_{FC} and θ_{PWP} had a R^2 of 0.21 and 0.29. When the PTFs were applied to the spatial predictions of sand and clay, the RMSE for θ_{FC} and θ_{PWP} had a relative increase of 25 % and 36 % respectively compared to when they were applied to measured horizon data. Across the majority of mainland France, the main sources of uncertainty of elementary AWC were coarse elements and soil texture, but the

contribution of uncertainty of PTFs' coefficients increased in areas dominated by very sandy and clayey textures. An advantage of the produced maps of θ_{FC}, θ_{PWP} and AWC is that the end users can incorporate associated uncertainties into ecological and agricultural modelling, and decision-making processes involved in soil and water planning.

Keywords: soil available water capacity, digital soil mapping, pedotransfer function, soil moisture at field capacity, soil moisture at permanent wilting point.

1. Introduction[1]

Soil available water capacity (AWC) refers to the maximum amount of water that a soil can store and release to plant roots (Veihmeyer and Hendrickson, 1927), and is a key property for many ecological and hydrological processes. AWC is operationally calculated as the difference between soil moisture at field capacity (θ_{FC}) (i.e., soil moisture remaining in the soil after water has drained by gravitational force) and soil moisture content at permanent wilting point (θ_{PWP}) (i.e., soil water retained so strongly that it is no longer available for plant roots, so plants wither and cannot recover their turgidity) (Silva et al., 2014). AWC is an important variable for agricultural and land use planning, for optimizing irrigation and crop growth of cultivated soils (Tetegan et al., 2015), for assessing soil drought risk (Schwärzel et al., 2009; Poggio et al., 2010; Leenaars et al., 2018), and estimating transport and leaching of pollutants (Marchetti et al. 1997).

Many agricultural and ecological models have AWC, θ_{FC}, or θ_{PWP} as input variables [e.g., STICS (Brisson et al., 1998), CENTURY (Parton et al., 1987), APSIM (O'Leary et al., 2016) SWAT (Arnold et al., 1987; Arnold and Fohrer, 2005)]. Thus, spatially explicit predictions of AWC at high resolution are relevant for upscaling simulation models at regional or national scale, and assessing the provision of some ecosystem services (eg., Dominati et al., 2010) like water quantity and quality regulation, carbon

[1] **Abbreviations**: available water capacity, AWC; digital soil mapping, DSM; *GlobalSoilMap*, GSM; French soil mapping and inventory program dataset, IGCS; French soil monitoring network, RMQS; pedotransfer function, PTF; soil moisture at field capacity, θ_{FC}; soil moisture at permanent wilting point, θ_{PWP}.

sequestration and provision of food, feed, fuel and fiber. Furthermore, uncertainty and scenario analysis should also include the uncertainty of AWC estimates when forecasting carbon sequestration, crop yield and biomass production, and planning efficient water use (Leenaars et al., 2018). Therefore, information of AWC and its spatial variability is important for planning (Poggio et al., 2010), and can help researchers and policy-makers towards the achievement of several United Nations Sustainable Development Goals (e.g., ensuring food security and promoting sustainable agriculture, mitigating climate change, and sustainable water management).

Measuring soil hydraulic properties is time-consuming and requires many human and economic resources. National soil databases rarely contain sufficient georeferenced AWC measurements for applying geostatistical or regression models (Padarian et al., 2014; Viscarra Rossel et al., 2015), and therefore indirect estimates of AWC are calculated at some stage of the digital soil mapping (DSM) process with pedotransfer functions (Poggio et al., 2010; Hong et al., 2013; Ugbaje and Reuter, 2013). Pedotransfer functions (PTFs) are used for translating readily available data (e.g., physical and chemical soil properties) into the data we need (e.g., soil water content) (Bouma, 1989). PTFs estimating soil hydraulic properties often have soil texture class or particle size distribution, bulk density, soil organic carbon, cation exchange capacity, and horizon type among the predictor variables (Wösten et al., 1999; Nemes et al., 2003; Al Majou et al., 2008b; Tóth et al., 2015, Román Dobarco et al. 2019).

AWC predictions with a very high relative error (coefficient of variation) may not be useful for certain applications, e.g. modelling crop yield (Folberth et al., 2016), because the estimates of ecological or agricultural processes produced with unreliable AWC predictions will have consequently a large uncertainty. Hence, to know if the AWC maps can be useful for modelling and decision-making, the AWC maps should provide a measure of the reliability of the predictions and quantified uncertainty (Poggio et al., 2010). Different sources of error are propagated in the process of mapping AWC: measurement errors of the soil profile data, errors due to the PTFs structure and parameters, errors derived from setting the upper and lower limits of AWC in terms of soil water potential, errors derived from the spatial extrapolation, errors of the environmental covariates used for regression modelling (Heuvelink et al., 1989; Carré et al., 2007). Poggio et al. (2010) combined general additive models

(GAM) and geostatistical models for mapping AWC after applying PTFs to individual horizons. They accounted for the uncertainty of the model trend and the local and spatial uncertainty, but did not include the uncertainty due to the PTFs. The uncertainty of soil hydraulic properties due to errors in the PTFs is sometimes small compared to the uncertainty of soil input data (Minasny et al., 1999). Additionally, identifying which input variable (or variables) account for most of the uncertainty of AWC can help to prioritize the input data needed to build DSM products or PTFs that require more improvement.

AWC is included in the soil properties of the *GlobalSoilMap* project (GSM), which aims to produce a digital soil map of the world at 3-arc second resolution providing estimates of uncertainty, following a bottom-up approach (Sanchez et al., 2009; Arrouays et al., 2014). Although AWC is still rarely mapped (Ugbaje and Reuter, 2013), the number of studies on AWC are increasing in the DSM literature from national (Hong et al., 2013, Padarian et al., 2014) to continental extent (Wösten et al., 1999; Ballabio et al., 2016; Tóth et al., 2016). GSM products for AWC are already available for Scotland (Poggio et al., 2010), Nigeria (Ugbaje and Reter, 2013), and Australia (Viscarra Rossel et al., 2015). The objectives of this study were: 1) to predict AWC for mainland France following the GSM specifications, and 2) to quantify the uncertainty of AWC accounting for uncertainty of the soil input variables and the PTFs' coefficients. The incorporation of the uncertainty due to the PTFs' coefficients into the AWC spatial modelling is a novelty relative to previous studies at national extent.

2. Methods

2.1 General framework

AWC is a composite soil property that depends on the difference between the soil moisture at field capacity and at permanent wilting point, on the volume of coarse elements and their ability to store water, and on the total thickness of the soil profile. Under the assumption that the coarse elements are inert and do not contribute to the AWC, the AWC for a unit of soil volume, or elementary AWC, is defined as:

$$\text{elementary AWC } (cm^3 cm^{-3}) = (\theta_{FC} - \theta_{PWP})(1 - R_v) \quad [1]$$

When we consider a soil layer or profile, the total AWC is calculated with the formula:

$$AWC\ (mm) = (\theta_{FC} - \theta_{PWP})(1 - R_v)d \qquad [2]$$

Where θ_{FC} is the volumetric water content at field capacity of the fine fraction (cm^3 cm^{-3}), θ_{PWP} is the volumetric water content at permanent wilting point of the fine fraction (cm^3 cm^{-3}), R_v is the volume fraction of coarse elements, and d is the depth of the soil profile or the thickness of the soil layer considered (mm).

In the DSM literature there are both 1) studies that applied PTFs to horizon or profile data and estimated AWC prior to the spatialization (Vanderlinden et al., 2005; Poggio et al., 2010; Hong et al., 2013), and 2) studies that spatialized the input soil variables first, and then applied the PTFs and equation 2 (Ugbaje and Reuter, 2013). Applying first the PTFs to horizon data or weighed averages of input properties by profile and then interpolating AWC estimates simplifies the DSM process, and can provide better results than spatializing soil properties first and then applying the PTFs (Styc and Lagacherie, 2018). Conversely, the spatial interpolation or spatial modelling of AWC based on environmental-soil relationships should better take place before applying the PTFs because this enables a more efficient use of the spatial distribution characteristics of individual inputs (Heuvelink and Pebesma, 1999), especially for those that are not usually correlated (e.g., soil profile thickness and soil texture). The PTFs' input variables are often correlated in the feature space (i.e., n-dimensional space with all the independent variables) or have some degree of spatial correlation. Thus, their correlation should be considered at spatial interpolation for obtaining plausible estimates of AWC and quantifying its uncertainty more accurately (Heuvelink et al., 2016).

In this study, we first generated maps of the PTFs' soil input properties by each GSM depth interval, taking into account the correlation among variables within each interval but omitting the correlation between different layers. Then we applied suitable PTFs for calculating θ_{FC} and θ_{PWP} by depth interval (Figure 1). Al Majou et al (2008a) found that θ_{FC} measured in situ corresponded best to soil moisture measured at the laboratory at a soil water potential of -10 kPa or pF = 2.0 ($\theta_{2.0}$) for horizons sampled in France, mainly in the Paris basin. We hypothesized that pF = 2.0 represents θ_{FC} across France and θ_{PWP} corresponds to soil moisture at a soil water potential of -1580 kPa or pF = 4.2 ($\theta_{4.2}$). Finally, we summed the AWC spatial predictions of each depth interval to the predicted soil thickness, modelled previously by Lacoste et al. (2016), for a maximum of 2 m:

$$AWC = \sum_{h=1}^{6}(1 - R_h)(\theta_{FC}^{h} - \theta_{PWP}^{h})\, t_h \quad [3]$$

where h = 1,…,6 is each of the GSM depth intervals, R_h is the proportion of soil occupied by coarse elements, θ_{FC}^{h} is the soil moisture at field capacity ($cm^3\ cm^{-3}$) in horizon h, θ_{PWP}^{h} is the soil moisture content at permanent wilting point ($cm^3\ cm^{-3}$), and t_h is the effective thickness (i.e. truncated using soil profile thickness estimates) of the horizon in mm.

2.2 Soil data

2.2.1 Calibration data from the French Soil Mapping and Inventory program

For the DSM model, the calibration data of particle size distribution and coarse elements came from the French soil mapping and inventory program dataset (Inventaire Gestion et Conservation des Sols: IGCS) (Laroche et al., 2014). Data from 81,671 soil profiles and soil cores was extracted from the IGCS dataset. The IGCS observations were originally collected for different studies with the objective of delineating soil-mapping units (Arrouays et al., 2004). Hence, the distribution of the observations was irregular through mainland France (Mulder et al., 2016). Whereas some areas were densely sampled, several areas had very few data or were even practically empty of observations (Figure 2). The horizon data of the profiles was standardized for the six depth intervals specified by the *GlobalSoilMap* project (i.e., 0–5 cm; 5–15 cm; 15–30 cm; 30–60 cm; 60–100 cm; 100–200 cm) (Table 1). For that purpose, we applied equal-area quadratic splines (Bishop et al., 1999) to soil profile data for estimating the average values of input soil properties by depth interval as explained in Mulder et al. (2016).

2.2.2 Evaluation data for soil input properties: French soil monitoring network

The French soil monitoring network (Réseau de Mesures de la Qualité des Sols: RMQS) is based on a systematic random grid of 16 km by 16 km that covers metropolitan France with approximately 2200 sites (Jolivet et al., 2006). Hence, we used data from the first RMQS campaign (2000-2012) as an independent evaluation sample for particle size distribution and coarse elements predictions (Brus et al., 2011). At each RMQS site, a soil pit of approximately 120 cm by 90 cm was dug to the appearance of parent material, and fully described. Samples were collected from each horizon of the soil profile and analyzed at the laboratory for determining the content of sand, silt, and clay using the pipette method

(ISO 13317-2:2001). The soil surveyors estimated visually the content of coarse elements (% volume) on the three faces of the soil pit. In May 2018, the database had data of particle size distribution from 1622 RMQS sites. The particle size distribution of some RMQS sites mostly located in forested areas have not been analyzed in the laboratory yet (Figure 3). Similarly, soil profile data of coarse elements was available for 1662 RMQS sites.

2.2.3 Evaluation data for soil hydraulic properties: GEVARNOVIA

The GEVARNOVIA dataset compiled data of physical and chemical properties for 831 horizons collected between 1973 to 2016 by different French institutes (ARVALIS-Institut du végétal, GEVES, INRA, Terres-Inovia) (Cousin et al., 2016), of which 308 horizons came from 108 georeferenced sites. The soil horizons were not sampled following any systematic sampling scheme, and were located mainly in the southwest or northern half of France (Figure 3). The land use was mostly agricultural, with cereals, (wheat, corn, sorghum, oats), sugar beet, and oleaginous crops (rapeseed, sunflower), and some pastures. The parent material varied between loamy materials, calcareous rocks, alluvial deposits, sandy aeolian deposits, and crystalline rocks. This independent evaluation dataset had measurements of particle size distribution, coarse elements, bulk density, and volumetric soil moisture content measured on soil aggregates after equilibrium at -10 kPa ($\theta_{2.0}$) and at -1580 kPa ($\theta_{4.2}$).

2.2.4 Data pretreatment

Particle size distribution constitutes compositional data (i.e., sand, silt, and clay vary between 0 and 1000 g kg^{-1}, and sum up to 1000 g kg^{-1}) that is subject to non-stochastic constraints (Lark and Bishop, 2007). As compositional data, their distributions cannot be drawn from the real space $\mathbb{R}^3$, but from the two-dimensional simplex plane $\mathcal{S}^2$ embedded in this space (Lark and Bishop, 2007). Hence, to avoid negative spurious correlations between the components, and guarantee that their predictions sum up to a constant, the distributions of sand, silt, and clay should not be analyzed independently but based on their ratios (Odeh et al., 2003). Aitchison (1986) proposed the additive log-ratio transformation (*alr*), which is defined as:

$$\boldsymbol{x} = alr(\boldsymbol{z}) = \left(\ln\left(\frac{z_1}{z_D}\right), \ln\left(\frac{z_2}{z_3}\right), \dots, \ln\left(\frac{z_{D-1}}{z_D}\right)\right) \quad [4]$$

where $\boldsymbol{z} = [z_1, z_2, \dots, z_D]^T$ is a composition of D elements, such as $z_i > 0 \; \forall \; i = 1,2, \dots D$ and $\sum_{i=1}^{D} z_i = k$, where k is a constant. The inverse *alr* transformation is defined as:

$$\boldsymbol{z} = \frac{\exp(\boldsymbol{w})}{j^T \exp(\boldsymbol{w})} \, k \qquad [5]$$

where exp(**w**) represents the vector $[\exp(x_1), \exp(x_2), \dots, \exp(x_{D-1}), 1]$ and j is a vector of length D with all elements equal to 1 (Lark and Bishop, 2007). The *alr* transformation is commonly applied for modelling particle size distribution with regression or geostatistical models (Odeh et al., 2003; Lark and Bishop, 2007; Buchanan et al., 2012; Akpa et al., 2014; Ciampalini et al., 2014; Huang et al., 2014; Poggio and Gimona, 2017). We applied the alr function of the rgr R package (Garrett, 2015) to obtain the *alr*-transformed variables:

$$clay_{alr} = \ln\left(\frac{clay}{sand}\right) \qquad [6]$$

$$silt_{alr} = \ln\left(\frac{silt}{sand}\right) \qquad [7]$$

We used sand as the denominator after comparing the evaluation statistics and spatial structure of the model residuals of the three combinations in preliminary tests (results not shown) (Poggio and Gimona, 2017).

2.3 Modelling soil input properties

The DSM process of the soil input properties was based on quantitative relationships between the calibration data and environmental variables related to soil genesis and spatial distribution, as per the *scorpan* framework (McBratney et al., 2003). The *scorpan* model is an extension of the soil genesis model by Jenny (1941), in which the soil system is function of the soil forming factors climate (cl), organisms (o), relief (r), parent material (p), and time (t) (soil = *f(cl,o,r,p,t)*). In addition, the *scorpan* model includes soil (s) and spatial position (n) as factors for predicting the spatial distribution of soil properties (McBratney et al., 2003) plus and error term (ϵ):

$$soil = f(s, c, o, r, p, a, n) + \epsilon \qquad [8]$$

Where c is climate and a is time.

2.3.1 Environmental covariates

We selected 44 covariates describing the *scorpan* factors soil, climate, vegetation, relief and parent material (Table 2). Climatic variables came from the French SAFRAN atmospheric analysis system (Durand et al. 1993). We used a Digital Elevation Model from SRTM (Shuttle Radar Topography Mission) at 90 m (USGS, 2004) to derive primary and secondary relief covariates in ArcGIS (ESRI, Redlands, WA). Soil and parent material were characterized by predominant soil type and parent material by soil mapping unit of the French Soil Geographical Database (Gis Sol, 2011), erosion rates (Cerdan et al., 2010), geophysical gravimetric data, and the Index of River Network Development and Persistence (IRNDP). Vegetation and land use were classified according to Corine Land Cover 2006 data (European Environmental Agency, 2007), ECOCIMAP-II (Faroux et al., 2003), and BD Foret (Institut National de l'Information Géographique et Forestière, 2012). Two vegetation indices derived from remote sensing data were used to describe the photosynthetic capacity of the vegetation cover, the enhanced vegetation index (EVI) and the normalized difference vegetation index (NDVI) (Huete et al., 2002). The MOD13A1 MODIS/Terra Vegetation Indices 16-day composite products at 500 m resolution were retrieved from the online NASA Earthdata Search, courtesy of the NASA EOSDIS Land Processes Distributed Active Archive Center (LP DAAC) (https://earthdata.nasa.gov/) (Didan, 2015). The vegetation indices were collected for the months of January (i.e., minimum vegetation activity) and June (i.e., maximum vegetation activity) for the period 2002-2014. The median of these vegetation indices over the 13 years for each month were used as covariates. All the covariates were projected to the Lambert 93 (EPSG: 2154) associated to the Réseau Géodésique Français 1993 (RGF93), aligned with the SRTM, and resampled to 90 m resolution using nearest neighbor interpolation. Data pre-processing was done in GRASS (GRASS Development Team, 2016), the R software v.3.2.2 (R Core Team, 2015), and the Geospatial Data Abstraction Library v.2.0.1 (GDAL/OGR contributors, 2015).

2.3.2 Soil spatial predictive models

We evaluated the correlation among $clay_{alr}$, $silt_{alr}$, and coarse elements in the feature space as well as the spatial correlation prior to modelling their spatial distribution. In preliminary tests, we also evaluated the spatial correlation and correlation in the feature space of the residuals of the models (Supplemental material S2 p.1). After checking the lack of correlation between coarse elements and the *alr*-variables, the weak spatial structure of the residuals of coarse elements and preliminary mapping exercises of AWC (Román Dobarco et al., 2018) we decided to model separately and differently the *alr*-variables and the coarse elements.

We predicted the *alr*-variables with a regression-cokriging model (Odeh et al., 1994; Hengl et al., 2007). Cubist models for $clay_{alr}$ and $silt_{alr}$ were fitted using the environmental covariates describing *scorpan* factors (Table 2) by GSM depth interval. The Cubist algorithm is a hybridized model that combines tree-based models and linear models. The terminal nodes of the regression tree (leaves) consist on a linear model (Quinlan, 1992). The parameters of the Cubist models were: committees=20, extrapolation = 5, and unbiased=TRUE. We calculated the model residuals at the calibration points, and then fitted a linear model of coregionalisation (LMCR) between the residuals of both variables for each GSM depth interval using the algorithm presented by Goulard and Voltz (1992). The LMCR had two components, a nugget and a spherical variogram. We then interpolated spatially the residuals of $clay_{alr}$ and $silt_{alr}$ by ordinary co-kriging using the closest 10 observations. The final predictions were calculated by adding the kriged residuals to the Cubist predictions and back-transformed to the original scale (equation 5). The kriging variance of the residuals of the *alr*-variables was used afterwards for estimating the prediction uncertainty (see section 2.5).

We modelled coarse elements with quantile regression forests (Meinshausen, 2006) for its ability to provide accurate estimates of uncertainty of predicted soil properties (Vaysse and Lagacherie, 2017). Quantile regression forests is a generalization of random forest models (Breiman, 2001). Random forests is a very popular machine-learning tool for classification or regression that provides an ensemble prediction based on many regression trees. For each regression tree and node, the algorithm incorporates randomness by selecting randomly a subset of features to split on . Quantile regression forest not only provides robust estimates of the conditional mean, but also of the full conditional distribution of the response variable. Whereas random forests keeps the mean value of observations at the nodes, quantile regression forests keeps the values of all observations at the nodes, and can infer estimates for conditional quantiles, prediction intervals, or other statistics from the distribution (Meinshausen, 2006). A detailed description of random forests and quantile random forests can be found in Breiman (2001) and Meinshausen (2006) respectively. We fitted quantile regression forest models for coarse elements by GSM depth, with the settings ntree = 1000 (number of trees), nodesize = 20 (minimum number of observations in terminal nodes), and the default mtry (number of variables randomly sampled as candidates at each split), which in this case was 14. We predicted the mean, the 5th percentile, the 95th percentile, and the standard deviation of coarse elements by GSM depth.

Finally, we mapped $clay_{alr}$, $silt_{alr}$, sand, clay, silt, coarse elements, and their respective standard deviations at 90 m resolution for mainland France. The calculation of the standard deviation of the back-transformed sand, clay, and silt is explained in the Supplemental material S1.

2.4 Functional digital soil mapping of AWC

2.4.1 Pedotransfer functions

The volumetric soil moisture content ($cm^3\ cm^{-3}$) at field capacity or pF = 2.0 ($\theta_{2.0}$) and at permanent wilting point or pF = 4.2 ($\theta_{4.2}$) for the fine fraction were estimated using PTFs developed by Román Dobarco et al. (2019) with the French SOLHYDRO database (Bruand et al., 2003; Al Majou et al., 2008b). These PTFs use the content of clay (%) and sand (%) as predictor variables:

$$\theta_{2.0} = 0.278 + 2.45\ 10^{-3}\ \text{clay} - 1.35\ 10^{-3}\ \text{sand} \quad [9]$$

$$\theta_{4.2} = 0.08 + 4.01\ 10^{-3}\ \text{clay} - 2.93\ 10^{-4}\ \text{sand} \quad [10]$$

The uncertainty of the PTFs' coefficients was calculated by non-parametric bootstrapping (Efron and Tibshirani, 1993). The variance-covariance matrices of the PTFs' coefficients are presented in Table 3.

These PTFs were chosen because: 1) the calibration dataset is representative of a large proportion of sand and clay contents found across France, 2) estimates of uncertainty for the PTFs' coefficients are available, and 3) the propagation of error due to both soil input variables and PTFs' coefficients can be easily calculated with first order Taylor series.

2.5 Uncertainty analysis

We applied a first order Taylor analysis to calculate the variance of θ_{FC} (i.e., $\theta_{2.0}$), θ_{PWP} (i.e., $\theta_{4.2}$), elementary AWC and total AWC estimates. The estimates' variance is considered here as a proxy of prediction uncertainty (Heuvelink et al., 1989). This method relies on the approximation of these estimates (equations 1, 3, 9 and 10 respectively), and of the intermediate variables estimates (i.e. clay, silt, and sand with equation 5). Let Y be an estimate of a given soil property with

$$Y = f(\boldsymbol{z})$$

where *f* is a continuously differentiable function from $\mathbb{R}^n$ into $\mathbb{R}$ and **z** the vector of the *n* input variables of *f*. The approximation of *f* uses a truncated Taylor series centered on the mean values of the *n* input variables $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_n]^T$ (Heuvelink et al., 1989). The variance of *Y=f(z)* is calculated with the formula (Heuvelink et al., 1989):

$$var(Y) \approx \sum_{j=1}^{n}\left\{\sum_{k=1}^{n}\left\{\tau_{jk}\sigma_j\sigma_k \frac{\delta f}{\delta z_j}(\boldsymbol{\mu})\frac{\delta f}{\delta z_k}(\boldsymbol{\mu})\right\}\right\} \quad [11]$$

where z_j and z_k can be soil input variables or PTFs' coefficients, τ_{jk} is the correlation of z_j and z_k, σ_j and σ_k are the standard deviation of the of z_j and z_k, $\frac{\delta f}{\delta z_j}(\boldsymbol{\mu})$ and $\frac{\delta f}{\delta z_k}(\boldsymbol{\mu})$ are the partial derivatives of f(z) around $\boldsymbol{\mu}$. These partial derivatives reflect the model (that is the *f* function) sensitivity to the input variables z_j and z_k. Hence, the variance of a soil property Y (equation 11) can be decomposed in the sum of different terms that consist on the multiplication of the squared model sensitivity by a variance or covariance of different input variables that represent their uncertainty (Dietze, 2017; Heuvelink et al., 1989).

We considered two sources of uncertainty that influence θ_{FC}, θ_{PWP}, elementary AWC, and total AWC estimates: the coefficients (named hereafter β_i, *i* in [0, 2]) of both PTFs, and the soil input properties of the PTFs and elementary AWC computation (namely clay, sand and coarse elements). Because the PTFs' coefficients and the spatial predictions of the input soil properties were determined independently, we assumed that the correlation between their errors was zero (Heuvelink et al., 1989). Hence, the variance estimates for θ_{FC} and θ_{PWP} (equations 9 and 10) can be summarized as the sum of two terms, 1) the product of the sensitivity and uncertainty of the soil input variables, and 2) the product of sensitivity and uncertainty of the PTFs' coefficients:

$$var(\theta) = \sum \tau_{Soil_j Soil_k}\sigma_{Soil_j}\sigma_{Soil_k}\frac{\delta f}{\delta Soil_j}\frac{\delta f}{\delta Soil_k} + \sum \tau_{\beta_j\beta_k}\sigma_{\beta_j}\sigma_{\beta_k}\frac{\delta f}{\delta \beta_j}\frac{\delta f}{\delta \beta_k} \quad [12]$$

The variance of the elementary AWC was similarly decomposed in four groups of sources of uncertainty, given that coarse elements and particle size fractions were uncorrelated, representing: coarse elements, particle size distribution, and the two PTFs. We considered the correlation among the coefficients of each PTF, but omitted the correlation between the coefficients of both PTFs because they were fitted independently. In the case of $clay_{alr}$ and $silt_{alr}$, the error associated to their spatial variation was represented by the cokriging variance. Quantile regression models gave directly the standard deviation of coarse elements predictions (Vaysse and Lagacherie, 2017).

Finally, the variance of total AWC was decomposed into the same four groups. We did not consider the uncertainty of soil profile thickness because that would have required computationally and time demanding Monte Carlo simulations, which would be prohibitive at the desired resolution (e.g., 10 simulations at 2000 pixels required approximately 10 hours with our High Performance Computing facilities, Román Dobarco et al. (2017)). The calculation of the variances by first order Taylor analysis is explained in detail in the Supplemental material S1.

2.6 Evaluation of the functional DSM predictions

Observed horizon data from the RMQS were used for independent evaluation of the predictions of clay, sand, silt, and coarse elements. The measurements of GEVARNOVIA horizons were compared with the predictions of volumetric soil moisture contents of the fine fraction (θ_{FC}, θ_{PWP}). This dataset is not representative of all the pedoclimatic conditions in France and therefore the evaluation statistics may be biased and not suitable for evaluating the whole France DSM approach. However, it is the best data available on θ_{FC} and θ_{PWP} in France at the moment. For each independent observation (*i.e.* an observation of a soil property on a given soil horizon), we calculated the weighted average of the GSM predictions overlapping the horizon attached to the observation. This was done because the sample support differed between the evaluation dataset and the predictions. Hence, a prediction was calculated as $\hat{y} = \sum_{i=1}^{n} \frac{t_i}{t} GSM_i$, where $\hat{y}$ is the estimated value, n is the number of GSM layers overlapping the horizon, GSM_i is the GSM prediction for the layer i, t is the total thickness of the horizon (cm), and t_i is the thickness (cm) of the GSM layer i that overlaps the predicted horizon. The evaluation statistics consisted on the root mean square error (RMSE), coefficient of determination (R^2), Lin's concordance correlation coefficient (Lin, 1989), and the bias, or mean error of prediction. The concordance evaluates both the accuracy and the precision of the prediction, and it is defined as:

$$\rho_C = \frac{2\rho\sigma_{\hat{y}}\sigma_y}{\sigma_{\hat{y}}^2+\sigma_y^2+(\bar{\hat{y}}-\bar{y})^2} \quad [13]$$

where $\bar{\hat{y}}$ and $\bar{y}$ are the means of the predicted and observed values, $\sigma_{\hat{y}}^2$and σ_y^2 their respective variances, and ρ the correlation between predicted and observed values. $\rho_C$can range between -1 and 1, and a value closer to 1 indicates a better fit with the 45° line, or agreement between predictions and observations.

In addition, we assessed the estimation of the prediction uncertainty with the prediction interval coverage probability (PICP) (Shrestha and Solomatine, 2006).

$$PICP = \frac{count\ (LPL_i < y_i < UPL_i)}{n}\ x\ 100 \quad [14]$$

where n is the number of observations in the evaluation dataset, and the numerator the counts that an observation y_i fits within its prediction limits. For a 90 % confidence level, the uncertainty is optimally estimated when the PICP value is close to 90 %. The prediction interval limits for the estimates of the observed horizon data were calculated as $\hat{y}\ \pm 1.64\ \sigma_{\hat{Y}}$ assuming a normal distribution of the estimated variance ($\sigma_{\hat{Y}}^2$) around the mean ($\hat{y}$). The variance of the prediction estimates for the observed horizons was calculated by Taylor series analysis (equation 11), accounting for the global correlation between different GSM layers for the same soil property. For one given soil property, the global correlation coefficients were calculated with all the pixel values for each pair of GSM maps representing the GSM layers.

3. Results

3.1 Spatial structure of model residuals

The linear model of coregionalization parameters are reported in the table 4. The range varied among depths between 160 km at 15–30 cm depth and 252 km for the layer 60–100 cm. Supplemental material S2 p.1a shows the estimates of the cross- and autovariograms of the regression model residuals. The regression residuals of the alr-transformed variables were correlated both spatially (Supplemental material S2 p.1a) and in the feature space (Supplemental material S2 p.1.b) across all depths. The coarse elements residuals were not correlated or had a weak correlation (r < 0.2) with either of the alr-transformed variables in any depth, neither untransformed (Supplemental material S2 p.1.b), or after log transformation (data not shown). The empirical variograms of coarse elements residuals showed some spatial correlation. However, as previous maps produced by regression-kriging of log-transformed coarse elements were unsatisfactory, we decided to exclude the spatial correlation from the model (Román Dobarco et al., 2018). All variograms appear somewhat erratic (Supplemental material S2 p.1.a), which is likely due to the presence of some clusters of points in the dataset (Marchant et al, 2013).

3.2 Independent evaluation

The evaluation statistics for the back-transformed clay, silt, and sand predictions for RMQS horizons indicated that clay had lower R^2 and concordance coefficient than silt and sand (Table 4). The RMSE increased following the trend clay < silt < sand (Table 5). The DSM predictions tended to underestimate clay content, as indicated by a bias of -15 g kg^{-1}. On the other hand, the bias of silt was 19 g kg^{-1}, and the predictions for sand had the smallest bias, of - 3 g kg^{-1}. Small contents of clay, silt, and sand were overestimated while high contents were often largely underestimated (Figure 4). However, the prediction error for the particle size fractions did not show any pattern related to the average depth of the RMQS horizons. Overall, many predictions were dispersed in the scatterplots and fell far from the 1:1 line (Figure 4) and predictions exhibited a RMSE up to 172 g kg^{-1} (Table 5). In comparison with the particle size fractions the predictions for coarse elements had the lowest R^2 and concordance coefficient and a RMSE of 21% (Table 5). The quantile regression forests model strongly underestimated relatively stony soils (> 25 %) and overestimated small contents of coarse elements (Figure 4.d). The prediction error of RMQS horizons with small coarse elements content (< 20 %) increased to some extent with the average horizon depth (Figure 4.d).

The PICP suggested that the uncertainty associated to coarse elements and clay predictions was underestimated (76% and 83 % respectively), but it was close to the expected value of 90 % for silt (86 %) and it was nearly perfect for sand, with a PICP of 90 % (Table 5).

The performance of the DSM predictions for soil moisture at field capacity and soil moisture at permanent wilting point had a R^2 of 0.21 and 0.29, and concordance coefficients of 0.37 and 0.47

respectively. The RMSE and the bias were greater for θ_{FC} than for θ_{PWP} (Table 6). The PICP indicated a large underestimation of the prediction uncertainty, with PICP = 71 % for θ_{FC} and PICP = 77 % for θ_{PWP}. The predictive performance of the PTFs with measured sand and clay was also better for θ_{PWP} than for θ_{FC}, and in both cases, the PICP was smaller than the optimal 90 % (Table 6). The underestimation of higher soil moisture contents at both water potentials increased for the DSM predictions in comparison to applying the PTFs with measured clay and sand (Figure 5).

The DSM predictions underestimated the soil moisture content at both potentials for fine and very fine soil texture classes whereas the DSM predictions tended to overestimate the soil moisture contents of coarse textured soils (Figure 6.a and 6.b). The prediction error by texture class was more or less homogeneous among classes when the PTF was applied directly on measured clay and sand data for θ_{FC} (Figure 6.c). The PTF overestimated θ_{PWP} for very fine texture but the prediction error was smaller for the other texture classes (Figure 6.d).

3.2 Spatial distribution of AWC

The soil AWC to a maximum depth of 2 m had higher estimated values in northern and southwestern France, and along the Rhone river valley (north-south axis in eastern France) (Figure 7). The uncertainty associated to the total AWC followed a similar pattern in its spatial distribution, with higher standard deviation in the north, southwest, and sparse areas in the centre (Sologne) and east (Rhone valley) of France (Figure 7). The higher AWC corresponded mainly to deeper soils or moderately deep soils with silty textures. The relative error (coefficient of variation) was greater than 20 % in most part of France and was greater than 30 % in some areas in the west, in the south along the Mediterranean coast, and in the east (Figure 7).

3.3 Contribution of different sources of uncertainty to the variance of soil moisture at field capacity

We present the results of the decomposition of variance associated to spatial predictions of soil moisture at field capacity for the layer 15-30 cm as an example. The results for soil moisture at permanent wilting point and for the other depths are similar and we provide additional figures in the Supplemental material

S2. The first term of equation 12, or variance associated to soil input properties, is expanded into its different components, the first related to $clay_{alr}$, the second to $silt_{alr}$ and the third to the interaction term between $clay_{alr}$ and $silt_{alr}$:

$$var(\theta)_{soil} = \sum cov(Soil_j, Soil_k)\frac{\delta f}{\delta Soil_j}\frac{\delta f}{\delta Soil_k} = \sigma^2_{clay_{alr}}(\frac{\delta f}{\delta clay_{alr}})^2 + \sigma^2_{silt_{alr}}\left(\frac{\delta f}{\delta silt_{alr}}\right)^2 +$$

$$2\,\tau_{clay_{alr},silt_{alr}}\sigma_{clay_{alr}}\sigma_{silt_{alr}}\frac{\delta f}{\delta clay_{alr}}\frac{\delta f}{\delta silt_{alr}} \qquad [15]$$

where $var(\theta)_{soil}$ is the variance term of the soil moisture at field capacity associated to soil input variables, $\sigma^2_{clay_{alr}}$ and $\sigma^2_{silt_{alr}}$ are the co-kriging variances of $clay_{alr}$ and $silt_{alr}$, $\frac{\delta f}{\delta clay_{alr}}$ and $\frac{\delta f}{\delta silt_{alr}}$ are respectively the sensitivities of the PTF (equation 9) to $clay_{alr}$ and $silt_{alr}$, and $\tau_{clay_{alr},silt_{alr}}$ is the is the correlation between $clay_{alr}$ and $silt_{alr}$.

The sensitivity of elementary soil moisture at field capacity ($cm^3\ cm^{-3}$) to $clay_{alr}$ had higher values in areas with higher predicted values for $clay_{alr}$ (Figure 8 and Figure 9.a). The sensitivity for $silt_{alr}$ had negative values where predictions for $silt_{alr}$ were positive, and positive values where $silt_{alr}$ predictions were negative (Figure 8 and Figure 9.b). The terms resulting from the multiplication of $clay_{alr}$ sensitivity and variance, and the interaction term had higher absolute values in similar regions, but counteracted each other because they had different signs. The total variance (Figure 9.f) was higher in zones with greater co-kriged residual variance (Figure 8.c and 8.d), and where the $clay_{alr}$ term was greater, in absolute value, than the interaction term (Figure 9).

The second term of equation 12 corresponded to the variance of soil moisture related to the PTF coefficients:

$$var(\theta)_{\beta} = \sum cov(\beta_j, \beta_k)\frac{\delta f}{\delta \beta_j}\frac{\delta f}{\delta \beta_k} = \sigma^2_{\beta_0}\left(\frac{\delta f}{\delta \beta_0}\right)^2 + \sigma^2_{\beta_1}\left(\frac{\delta f}{\delta \beta_1}\right)^2 + \sigma^2_{\beta_2}\left(\frac{\delta f}{\delta \beta_2}\right)^2 +$$

$$2\,cov(\beta_0,\beta_1)\frac{\delta f}{\delta \beta_0}\frac{\delta f}{\delta \beta_1} + 2\,cov(\beta_0,\beta_2)\frac{\delta f}{\delta \beta_0}\frac{\delta f}{\delta \beta_2} + 2\,cov(\beta_1,\beta_2)\frac{\delta f}{\delta \beta_1}\frac{\delta f}{\delta \beta_2} \qquad [16]$$

where $var(\theta)_{PTF}$ is the variance term of the soil moisture at field capacity associated to PTF coefficients, $\sigma^2_{\beta_0}$, $\sigma^2_{\beta_1}$, $\sigma^2_{\beta_2}$, $cov(\beta_0,\beta_1)$, $cov(\beta_0,\beta_2)$, $cov(\beta_1,\beta_2)$ are the elements of the variance-covariance matrix for the PTF (in this example at field capacity, Table 3), and $\frac{\delta f}{\delta \beta_i}$ $with\ i = 0, 1, 2$ are the sensitivities of the PTF (equation 9) to the PTF coefficients.

In the case of the PTFs, the variance and covariance of the different terms are spatially constant (Table 3). Therefore, the spatial distribution of the uncertainty (and its different terms) depends on the sensitivity of the PTF to its coefficients ($\frac{\delta f}{\delta \beta_j}\frac{\delta f}{\delta \beta_k}$). We should clarify that $\frac{\delta f}{\delta \beta_0} = 1$, $\frac{\delta f}{\delta \beta_1}$ equals the clay predictions (%), and $\frac{\delta f}{\delta \beta_2}$ the sand predictions (%). Therefore, the sensitivity of both PTFs to the clay and sand coefficients is higher in areas where clay and sand contents are respectively higher (Figure 10). Estimates of clay are higher in the northeast of France, and some areas of southwest. The predictions for sand are higher in the centre (Massif Central) and other areas with predominance of sandy textures (Les Vosges in the northeast, the Sologne in the centre, and les Landes in the southwest). The covariance between the intercept and clay and sand ($cov(\beta_0, \beta_1)$ and $cov(\beta_0, \beta_2)$) is negative for both PTFs (Table 3), which resulted in negative values in their interaction terms (Figure 10.c and 10.d). The components of clay and sand coefficients and their interaction were positive (Figure 10.e, 10.f, and 10.g). The uncertainty due to the PTFs' coefficients had higher values in those areas where the sensitivity to the sand coefficient was greatest (Figure 10.b and Figure 10.h). The maximum value of the variance associated to the PTFs' coefficients is three times smaller than the maximum value of the variance associated to soil input variables (0.00012 vs 0.004).

3.4 Contribution of different sources of uncertainty to the variance of elementary AWC

The decomposition of the variance of elementary AWC for the six GSM layers into different components showed how, in terms of magnitude, the variance associated to the coarse elements and particle size fractions were the most important over mainland France (Supplementary material S2). At the same time, the terms of AWC variance associated to both PTFs' coefficients had small values across the whole area. The sensitivity of the elementary AWC to PTFs' coefficients was greater than the sensitivities to coarse elements, and particle size fractions (Supplemental material S2), but this was compensated by smaller values of the variance-covariance of the PTFs' coefficients (Table 3), resulting in small values of the terms related to the PTFs' coefficients on the variance in the Taylor series analysis.

Conversely, smaller sensitivities to coarse elements, $clay_{alr}$ and $silt_{alr}$ but higher variances (Figure 8) resulted in higher variance associated to these terms. Noteworthy, the variance term that corresponded to the interaction between both alr-variables had a negative sign and reduced the final variance (Supplemental material S2).

Across the six GSM layers, the standard deviation of elementary AWC (cm^3 cm^{-3}) was higher near mountainous regions in the south and southeast, and some areas in the west and east of France (Supplemental material S2 p.5). The standard deviation of elementary AWC increased with soil depth (Supplemental material S2 p.6). The main contribution (%) to the variance in these areas corresponded to the term associated to the coarse elements (red areas in Supplemental material S2 p.6). In regions with lower variance of elementary AWC and mostly low elementary AWC, the main sources of uncertainty for the top soil layers were the PTFs (southwest and centre) (blue in Supplemental material S2 p.6) and soil texture (north) (green in Supplemental material S2 p.6). The contribution (%) of coarse elements to the variance in the southwest and centre increased for the 30–60 cm, 60–100 cm, and 100–200 cm while the contribution (%) of the PTFs decreased (Supplemental material S2 p.7). Conversely, the contribution (%) of coarse elements to the variance of elementary AWC in the north decreased with depth.

4. Discussion

4.1 Previous estimates of AWC in metropolitan France

This study presented the first map of AWC for metropolitan France that provides uncertainty estimates following *GlobalSoilMap* specifications (Arrouays et al., 2014). An advantage of the produced maps of θ_{FC}, θ_{PWP} and AWC is that the potential end users can incorporate estimated uncertainties into ecological and agricultural modelling and perform uncertainty propagation analysis or sensitivity analysis. It is important to stress that since the role of soil thickness was not included in the uncertainty analysis for total AWC, the calculated variance should be considered a conservative uncertainty estimate. We have improved the precision of the estimates of AWC for mainland France compared to a previous map produced by Al Majou et al., (2008b). Al Majou et al., (2008b) applied their PTFs with information on horizon type, horizon thickness, texture, and bulk density provided by the descriptions of soil typological unit (STU) from the 1:1 000 000 Soil Geographical Database of France (King et al., 1995). In these previous studies, the AWC by soil mapping unit was calculated based on the proportion of STU present in each soil mapping unit (Al Majou et al., 2008b). The predictions of this study are provided at a specific resolution (90 m) whereas Al Majou et al. (2008b) provided AWC values by polygons. The spatial patterns in both maps are similar overall, although our estimates are smaller than the AWC predicted by Al Majou et al. (2008b) in the southeast and northeast France, but higher in the north.

Piedallu et al. (2011) also predicted AWC for metropolitan France by first estimating the total AWC of 120,902 soil profiles from the French National Forest Inventory with class-PTFs (i.e., average values of AWC are assigned by classes defined by texture and horizon type) developed by Al Majou et al. (2008b), and then extrapolating spatially with ordinary kriging. The general pattern of AWC was similar between the map by Piedallu et al. (2011) and this study. Piedallu et al. (2001) used texture class estimated by the surveyors in the field, which may enlarge the measurement error compared to laboratory analysis of particle size distribution. However, the high density of observations across metropolitan France may compensate partly the error in the input texture data resulting in a good description of the spatial pattern of AWC. The approach followed by Piedallu et al. (2011) makes the maps suitable for forested areas, since the soil profiles were located in forests and the greater distance between the locations in agricultural land to the soil profiles involves larger uncertainty at these predicted locations. Conversely, the data on particle size fractions distribution and coarse elements used to calibrate our models were not distributed homogeneously across mainland France (Figure 2). In some areas, such as in the south near the Massif Central and in the southwestern coast, the lower density of observations resulted in higher uncertainty of $clay_{alr}$ and $silt_{alr}$ (Figure 8) that propagated into the uncertainty of θ_{FC} and θ_{PWP} (Figure 9). However, particle size fraction was not the main source of uncertainty for elementary AWC in these areas (Figure 11).

4.2 General approach for mapping AWC

There are multiple modelling trajectories for mapping AWC depending on at which step are applied the PTFs and the spatial extrapolation (Styc and Lagacherie, 2018). We first modelled spatially the AWC input variables, and when possible, jointly, as for particle size distribution, with the aim of capturing their spatial patterns and the relationships between the soil forming factors and the soil properties, thus improving the accuracy of AWC predictions. This approach is similar to that followed by Ugbaje and Reuter (2013) for *GlobalSoilMap* AWC for Nigeria and Tóth at al. (2017) for the European Soil Hydraulic Database (EU-SoilHydroGrids). The methodology chosen for predicting AWC has the advantage that 1) the results can be easily updated when more accurate predictions for the soil input properties or more reliable PTFs are available and 2) some input soil properties (here particle size distribution) come as a side product of AWC estimation.

The predictions of particle size distribution had similar R^2 (Table 5) than previous *GlobalSoilMap* products for France (Mulder et al., 2016), which had R^2 between 0.25 – 0.44 for clay, 0.21 – 0.42 for silt, and 0.19 – 0.33 for sand. The concordance coefficients by Mulder et al. (2016) ranged between 0.34 – 0.53 for clay, 0.37 – 0.61 for silt, and 0.46 – 0.63 for sand, which are comparable to the 0.49, 0.43, and 0.66 of this study (Table 5). This is not surprising because both studies shared the data from the IGCS in the calibration dataset (although Mulder et al. (2016) merged this dataset with the RMQS one), some of the environmental covariates, and applied the cubist algorithm. However, predicting particle size distribution with regression-cokriging allowed us to account for the spatial correlation between particle size fractions of the same GSM layer, hence quantifying the uncertainty of AWC more accurately. On the other hand, the results for coarse elements were less accurate and precise in this study, with R^2 = 0.14 and a concordance coefficient of 0.26 compared to a R^2 between 0.17 – 0.28 and a concordance coefficient between 0.30 – 0.46 by Mulder et al. (2016) for log transformed coarse elements. One possible explanation for this slight performance discrepancy could be, at least for topsoil layers, the better precision of RMQS observations (compared to the IGCS), which was not used for training the models in the present study. Mapping coarse elements is specially challenging. The calibration and evaluation data (volume of coarse elements) was estimated visually by the soil surveyors. More precise methods for measuring the volume of coarse elements are very time consuming and can only be applied in a limited number of sites in national soil monitoring surveys (Jolivet et al., 2018). More generally, we were not able to capture the spatial patterns of distribution of coarse elements, especially in deep layers. Indeed, the input data itself may be partly biased, as coarse elements in surface layers are much more easily estimated than in deeper ones, especially when doing observations by coring. It is also possible that the chosen covariates did not represent well the processes driving the distribution of coarse elements, or these were not accurate enough (e.g., the scale of parent material was 1:1 000 000). Nevertheless, next versions of *GlobalSoilMap* with more accurate predictions on coarse elements can be incorporated in the proposed AWC modelling framework for reducing the prediction error and uncertainty of AWC predictions.

Styc and Lagacherie (2018) compared six possible trajectories for mapping AWC in the Languedoc-Roussillon (France). The modelling approach with best performance consisted in using weighed mean values of the soil input properties involved in the calculation of AWC by profile as training data for DSM and then applying the PTFs, partly because the averaging smooths the variability of soil properties facilitating the spatial modelling (Styc and Lagacherie, 2018). Mapping approaches based on information from modal soil profiles by soil mapping units can produce very suitable maps when the soil maps have sufficient detail (Hong et al., 2013), although they omit the variability within soil mapping units and estimates of uncertainty are often missing. Poggio et al. (2010) calculated the AWC by horizon with a PTF and then mapped AWC combining regression and geostatistics. They considered two sources of uncertainty for the AWC predictions: 1) the uncertainty of the trend described by general additive models, and 2) the uncertainty linked to the spatial extrapolation of the model residuals, omitting uncertainty related to PTFs. They used sequential Gaussian simulations for quantifying the spatial uncertainty, which provides more accurate estimates of the uncertainty than the Taylor series analysis and allows the characterization of the probability distribution of AWC for each pixel, but at high resolution is very computationally demanding.

4.3 Importance of different sources of uncertainty

The contribution of each source of uncertainty to the elementary AWC variance varied spatially. Across the majority of the study area, the main source of uncertainty of elementary AWC was soil input data, either particle size distribution, or coarse elements content. Without excluding the areas where estimated soil depth is shallower than the GSM layer, the terms associated to coarse elements were dominant. However, the contribution of particle size distribution gained importance in deeper GSM layers (60–

100 and 100–200 cm) when excluding areas based on the estimated soil depth by Lacoste et al. (2016). In areas of very sandy and clayey textures the PTFs contributed more to AWC uncertainty. This highlights the importance of developing more reliable PTFs for very coarse and very clayey soils, that although occurring with less frequency, can occupy extensive areas in France (e.g., sandy soils in the Landes of Gascony).

It is also likely that the large prediction error of θ_{FC} and θ_{PWP} estimated with PTFs (Figures 5.b and 5.d) is partly due to the large range in mineral composition within particle size fractions included in the PTFs (i.e., clay and sand) and their influence on AWC. Most of the silty horizons in France (especially in the upper layers), come from loessic aeolian deposits (Arrouays et al., 2011; Bertran et al., 2016). Their size and mineral composition are rather homogeneous. Therefore, their contribution to AWC is mainly determined by the micro-pores formed by stacking particles of nearly equal sizes and similar adsorption properties On the contrary, clay mineralogy in France is very diverse, mainly depending on the parent material from which clay minerals derive and their subsequent evolution by pedogenesis (van Ranst et al., 1995). Similarly, the nature of sands is very diverse in their mineral composition (e.g. pure quartz in the Landes of Gascony, micas and feldspathich sands in the Armorican Massif in Brittany). Moreover, their size and shape are very diverse and may influence their capacity to retain water (Chrétien, 1971).

The capacity of the Taylor analysis for identifying the sources contributing most to the total uncertainty is limited due to the interaction terms and because the variance terms resulting from the product of the model squared sensitivity to input variables and their variance involve variables from different groups (e.g., the variance term of PTFs involves $silt_{alr}$ and $clay_{alr}$ predictions). Hence, it is hard to identify which variable has the variance that we should reduce with the least expense in modelling time or resources required for additional sampling. Therefore, future studies aiming to improve the AWC predictions and to reduce the prediction uncertainty, should first identify the group of soil input variables contributing most in their study area with a global sensitivity analysis. As indicated above, improving the predictions of coarse elements content and particle size fractions, and studying the influence of coarse elements, clay and sand nature on soil moisture content should be among the highest priorities. The effect of $clay_{alr}$

and $silt_{alr}$ interpolation onto θ_{FC} and θ_{PWP} prediction performance (Figure 5.a and 5.c) is another argument for these priorities. A more complete assessment of AWC uncertainty should include the effect of soil depth uncertainty. The latter will likely have an important effect on total AWC uncertainty given its linear relationship with total AWC. The RMSE of soil depth predictions used in this study was 40 cm (Lacoste et al., 2016), suggesting that our estimate of AWC variance largely underestimates the uncertainty. It does not concern the assessment of the uncertainty of elementary AWC (at soil layer level) we provide here.

4.4 Limitations of the produced maps and future directions

The predictive ability of the spatial predictions for θ_{FC} and θ_{PWP} decreased considerably in comparison to applying the PTFs to measured horizon data (Table 6). The change in R^2 for θ_{FC} and θ_{PWP} was of $\Delta R^2 = -0.33$, which in relative terms consisted in a 61 % and 53% reduction compared to the R^2 of the PTFs. The RMSE increased in ΔRMSE = 0.013 $cm^3\ cm^{-3}$ for θ_{FC} and ΔRMSE = 0.015 $cm^3\ cm^{-3}$ for θ_{PWP} that suppose a relative increase of 25 % and 36 % respectively compared to the RMSE of the PTFs. We acknowledge that we could evaluate the spatial predictions of θ_{FC} and θ_{PWP}, which is not always possible for soil hydraulic properties due to the lack of georeferenced observations. However, the evaluation statistics are of limited validity at national extent due to the sample size, distribution and representativeness of pedoclimatic conditions of the evaluation dataset (Table 6 and Figure 3). Ongoing work for France is currently being carried out, for gathering unbiased references of soil water content at different potentials, using the French soil monitoring network (RMQS, Jolivet et al. 2006). The larger prediction error for θ_{FC} and θ_{PWP} spatial predictions were associated to an overestimation of soil moisture for coarse texture, and an underestimation for fine and very fine texture classes (Figure 6). This is likely related to the inability to predict the spatial distribution of clay content in some regions, and in particular, to the underestimation of high clay contents (Table 5 and Figure 4). It is also likely that important drivers such as mineralogical composition of clay and sand are also missing in the PTFs. It is possible that we need more accurate covariates for capturing the processes driving the spatial distribution of clay content and mineralogy assuming the latter could be incorporated into PTFs (e.g., soil geology map, gamma-ray spectrometry). Another source of error is that the soil profiles were not distributed evenly throughout the study area, but were clustered instead. Clustering led to some artifacts during the cokriging step (Padarian et al., 2014; Marchant et al., 2013) despite the benefit of including the correlation between soil particle fractions within a GSM layer. We did not include the spatial correlation among soil properties of multiple layers, which may also have improved the estimates of AWC and its uncertainty (Heuvelink et al., 2016). Angelini et al. (2017) applied structural equation modelling (SEM) to DSM, incorporating pedological knowledge of the interrelations among soil properties and soil processes, and predicting several soil properties at multiple layers simultaneously. Recently, Angelini and Heuvelink (2018) expanded the SEM for soil properties with a geostatistical approach, including the spatial correlation of the model residuals. This methodology could be interesting for mapping AWC, as it would incorporate the interrelations between all soil properties defining AWC (particle size distribution, bulk density, soil organic carbon, soil depth, coarse elements, etc.), within and between soil layers.

We selected PTFs that could incorporate the uncertainty of their coefficients into the AWC predictions and that can be applied to the majority of the study area (Román Dobarco et al., 2019). However, the PTFs by Tetegan et al. (2011) may be more suitable for soils developed from sedimentary rocks, with the additional advantage that they accounted for the capacity of coarse elements for storing water and contributing to AWC. Spatial soil inference systems (Lagacherie and McBratney, 2006) predicting AWC across large areas could apply the most appropriate PTF for each pixel. Another possible limitation is that both the DSM learning dataset (IGCS) and the PTFs deal mostly with cultivated soils. The proposed predictions might not be accurate for other soils such as forest soils, where soil properties other than particle size distribution such as soil organic carbon content and the frequent high amount of coarse elements might shift soils outside of the validity domain of PTFs used here (Román Dobarco et al., 2019).

Another future development is related to the definition of the available water content itself. We set θ_{FC} at pF = 2.0 for metropolitan France based on samples collected mainly in in the Paris basin and southwest of France (Al Majou et al., 2008a). For this set of samples, Bruand et al. (2004) indicated that θ_{FC} at the field corresponded, in the laboratory, to soil moisture content for soil matric potentials between pF = 1.5

and pF = 2.0. Hence, our predictions may underestimate AWC. Conversely, the θ_{FC} is often considered at pF = 2.5 by European PTFs (Toth et al., 2015). The definition of θ_{FC} regarding the soil water potential is another factor that influences the uncertainty of AWC. The ongoing expansion of the database of soil hydraulic properties for France will support the choice of the optimum upper limit of AWC, which may not be the same for different horizon types.

4.5 Conclusions

This study presented spatial predictions of AWC for mainland France following *GlobalSoilMap* specifications to a maximum depth of 2 m. We incorporated two sources of error (spatial estimates of soil input properties and PTFs' coefficients) in the uncertainty analysis carried out with first order Taylor series analysis. The continuous computing and statistical developments will allow improving the quantification of AWC uncertainty with a feasible computing time in future studies, for example with stochastic simulations or Bayesian simulations (Poggio et al., 2016; Beguin et al., 2017; Huang et al., 2017), which would ultimately allow the characterization of the probability distribution of AWC estimates on a pixel base. Overall, this study provides the first estimate of AWC uncertainty, by soil layers or at the whole soil profile level, for mainland France that can be incorporated into ecological and agricultural modelling. The end-users of the AWC maps will be essential for evaluating the usefulness of the maps for assessing the provision of ecosystem services and modelling ecological processes, and to indicate limitations in their exploitation due to the AWC prediction uncertainty.

The reproducible modelling framework allows replacing each component of the AWC calculation (PTFs, soil input properties) when more accurate maps are developed thanks to the selection of covariates that characterize better the processes driving the spatial distribution of soil input properties, the implementation of new regression algorithms, and the acquisition of new soil profile data. Indeed, key aspects for improving AWC estimates are expanding the calibration data on coarse elements (especially for deeper layers), incorporating the nature and soil hydraulic properties of coarse elements into the calculation, improving the estimates of clay and sand and their mineralogy, and improving the prediction of soil depth. According to our initial objectives, a major output of this study is the method we developed to estimate the uncertainty of AWC predictions by taking into account both uncertainties linked to the soil input variables and to the PTFs' coefficients. The second major output is the prediction of AWC and its uncertainty for mainland France according to international specifications, which provides this country a nearly complete set of the mandatory attributes to be predicted according to the GlobalSoilMap initiative.

Acknowledgements

This study was carried in the framework of the Groupement d'intérêt scientifique Sol (Gis Sol) and financed by the French Environment and Energy Management Agency (ADEME), the contract number is 32000753. The GEVARNOVIA dataset was created with contributions from INRA, ARVALIS-Institut du végétal, GEVES, and Terres-Inovia. We would like to thank Alain Bouthier, Marie- Hélène Bernicot, Luc Champolivier, and Aya Labidi for sharing the GEVARNOVIA dataset for the analyses included in this article. We would also like to thank Anne Richer-de-Forges for her help with the calibration dataset for coarse elements, and Line Boulonne and Jean-Philippe Chenu for their help with the RMQS dataset.

References

Achache, J., Debeglia, N., Grandjean, G., Guillen, A., Le Bel, L., Ledru, P., et al., 1997. GEOFRANCE 3D: l'imagerie géologique et géophysique 3D du sous-sol de la France. Mém. Société Géologique Fr. 53–71.

Aitchison, J., 1982. The statistical analysis of compositional data. J. R. Stat. Soc. Ser. B Methodol. 44 (2), 139–177.

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 Sci. Soc. Am. J. 78 (6), 1953–1966. doi:10.2136/sssaj2014.05.0202.

Al Majou, H., Bruand, A., Duval, O. 2008a. Use of in situ volumetric water content at field capacity to improve prediction of soil water retention properties. Can. J. Soil Sci., 88(4), 533–541.

Al Majou, H., Bruand, A., Duval, O., Le Bas, C., Vautier, A., 2008b. Prediction of soil water retention properties after stratification by combining texture, bulk density and the type of horizon. Soil Use Manage. 24(4), 383–391. doi:10.1111/j.1475-2743.2008.00180.x.

Angelini, M.E., Heuvelink, G.B, 2018. Including spatial correlation in structural equation modelling of soil properties. Spat. Stat., 25, 35–51. doi.org/10.1016/j.spasta.2018.04.003.

Angelini, M.E., Heuvelink, G.B., Kempen, B., 2017. Multivariate mapping of soil with structural equation modelling. Eur. J. Soil Sci., 68, 575–591. doi:10.1111/ejss.12446.

Arnold, J. G., and J. R. Williams. 1987. Validation of SWRRB: Simulator for water resources in rural basins. J. Water Resour. Plan. Manage. ASCE113(2): 243-256.

Arnold, J. G., and N. Fohrer. 2005. SWAT2000: Current capabilities and research opportunities in applied watershed modeling. Hydrol. Process. 19(3): 563-572.

Arrouays, D., Hardy, R., Schnebelen, N., Le Bas, C., Eimberck, M., Roque, J., et al., 2004. Le programme inventaire gestion et conservation des sols en France. Etude Gest. Sols, 11 (3), 187–197. https://prodinra.inra.fr/record/75841.

Arrouays, D., Grundy, M.G., Hartemink, A.E., Hempel, J.W., Heuvelink, G.B.M., Hong, S.Y., et al., 2014. GlobalSoilMap: toward a fine-resolution global grid of soil properties. Adv. Agron. 125 (125), 93–134. doi.org/10.1016/B978-0-12-800137-0.00003-0.

Arrouays, D., Saby, N.P.A., Thioulouse, J., Jolivet, C., Boulonne, L., Ratié, C., 2011. Large trends in French topsoil characteristics are revealed by spatially constrained multivariate analysis. Geoderma 161, 107–114. doi.org/10.1016/j.geoderma.2010.12.002.

Ballabio, C., Panagos, P., Monatanarella, L., 2016. Mapping topsoil physical properties at European scale using the LUCAS database. Geoderma 261, 110–123. doi.org/10.1016/j.geoderma.2015.07.006.

Beguin, J., Fuglstad, G.A., Mansuy, N., Paré, D., 2017. Predicting soil properties in the Canadian boreal forest with limited data: Comparison of spatial and non-spatial statistical approaches. Geoderma, 306, 195–205. doi.org/10.1016/j.geoderma.2017.06.016.

Bertran, P., Liard, M., Sitzia, L., Tissoux, H., 2016. A map of Pleistocene aeolian deposits in Western Europe, with special emphasis on France. J. Quaternary Sci., 31: 844-856 e2909. doi:10.1002/jqs.2909.

Bishop, T.F.A., McBratney, A.B., Laslett, G.M., 1999. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. Geoderma, 91 (1–2), 27–45. doi.org/10.1016/S0016-7061(99)00003-8.

Bouma, J., 1989. Using soil survey data for quantitative land evaluation. Adv. Soil Sci. 9, 177–213. doi.org/10.1007/978-1-4612-3532-3_4.

Breiman, L., 2001. Random forests. Machine learning, 45(1), 5–32. doi.org/10.1023/A:1010933404324.

Brisson, N., Mary, B., Ripoche, D., Jeuffroy, M. H., Ruget, F., Nicoullaud, B., et al., 1998. STICS: a generic model for the simulation of crops and their water and nitrogen balances. I. Theory and parameterization applied to wheat and corn. Agron., 18(5–6), 311–346.

Bruand, A., Pérez Fernández, P., Duval, O., 2003. Use of class pedotransfer functions based on texture and bulk density of clods to generate water retention curves. Soil Use Manage. 19, 232–242.

Bruand, A., Duval, O., Cousin, I., 2004. Estimation des propriétés de rétention en eau des sols à partir de la base de données SOLHYDRO: Une première proposition combinant le type d'horizon, sa texture et sa densité apparente. Étude et Gestion des Sols, 11 (3), 323 – 332.Brus, D.J., Kempen, B., Heuvelink, G.B.M., 2011. Sampling for validation of digital soil maps. Eur. J. Soil Sci. 62 (3), 394–407. doi:10.1111/j.1365-2389.2011.01364.x.

Buchanan, S., Triantafilis, J., Odeh, I.O.A., Subansinghe, R., 2012. Digital soil mapping of compositional particle-size fractions using proximal and remotely sensed ancillary data. Geophysics 77:WB201–WB211. doi:10.1190/geo2012-0053.1.

Carré, F., McBratney, A.B., Mayr, T., Montanarella, L., 2007. Digital soil assessments: Beyond DSM. Geoderma, 142 (1–2), 69–79. doi.org/10.1016/j.geoderma.2007.08.015.

Cerdan, O., Govers, G., Le Bissonnais, Y., Van Oost, K., Poesen, J., Saby, N.P.A., et al., 2010. Rates and spatial variations of soil erosion in Europe: A study based on erosion plot data. Geomorphol., 122 (1–2), 167–177. doi.org/10.1016/j.geomorph.2010.06.011.

Ciampalini, R., Martin, M.P., Saby, N.P.A., Richer-de-Forges, A.C., Arrouays, D., Nehlig, P., Martelet, G., 2014. Soil texture GlobalSoilMap products for the French region "Centre", in: Arrouays, D., McKenzie, N., Hempel, J., Richer de Forges, A.C., McBratney, A. (Eds.), Globalsoilmap: Basis of the Global Spatial Soil Information System. CRC Press, pp. 121–126.

Coleman, K., Jenkinson, D., Crocker, G.J., Grace, P.R., Klir, J., Körschens, M., et al., 1997. Simulating trends in soil organic carbon in long-term experiments using RothC-26.3. Geoderma, 81(1–2), 29–44. doi.org/10.1016/S0016-7061(97)00079-7.

Cousin, I., Labidi, A., Le Bris, X., Champolivier, L., Bernicot, M.H., Bessard Duparc, P., Bouthier, A., 2016. Evaluating the Available Water Content of soils at large scale to improve the estimation of soil services. 2nd Global Soil Security Conference, 05–06/12/2016.

Chrétien, J., 1971. Essai de caractérisation des sables en tant que squelette minéral du sol. Ann. Agron., 22 (6), 615–654.

Didan, K., 2015. MOD13A1 MODIS/Terra Vegetation Indices 16-Day L3 Global 500m SIN Grid V006. NASA EOSDIS LP DAAC. doi: 10.5067/MODIS/MOD13A1.006.

Dietze, M.C., 2017. Ecological forecasting. Princeton University Press. Princeton, New Jersey, USA.

Dominati, E., Patterson, M., Mackay, A., 2010. A framework for classifying and quantifying the natural capital and ecosystem services of soils. Ecol. Econ., 69(9), 1858-1868. doi.org/10.1016/j.ecolecon.2010.05.002.

Durand, Y., Brun, E., Mérindol, L., Guyomarc'h, G., Lesaffre, B., Martin, E., 1993. A meteorological estimation of relevant parameters for snow models. Ann. Glaciol., 18, 65-71. https://doi.org/10.3189/S0260305500011277.

Durand, Y., Laternser, M., Giraud, G., Etchevers, P., Lesaffre, B., Mérindol, L., 2009: Reanalysis of 44 Yr of Climate in the French Alps (1958–2002): Methodology, Model Validation, Climatology, and Trends for Air Temperature and Precipitation. J. Appl. Meteor. Climatol., 48, 429–449, https://doi.org/10.1175/2008JAMC1808.1.

Efron, B., Tibshirani, R.J., 1993. An introduction to the bootstrap. Monogr. Stat. Appl. Probab. 57. CRC Press, Boca Raton, FL.

European Environmental Agency, 2007. CLC2006 technical guidelines. Technical Report No. 17/2007, European Environmental Agency, Copenhagen, Denmark.

Faroux, S., Kaptué Tchuenté, A.T., Roujean, J.L., Masson, V., Martin, E., Moigne, P.L., 2013. ECOCLIMAP-II/Europe: A twofold database of ecosystems and surface parameters at 1 km resolution based on satellite information for use in land surface, meteorological and climate models. Geosci. Model Dev., 6(2), 563–582. doi.org/10.5194/gmd-6-563-2013.

Folberth, C., Skalský, R., Moltchanova, E., Balkovič, J., Azevedo, L. B., Obersteiner, M., Van Der Velde, M., 2016. Uncertainty in soil data can outweigh climate impact signals in global crop yield simulations. Nat. Commun., 7, 11872. doi.org/10.1038/ncomms11872.

Garrett, R.G., 2015. rgr: applied geochemistry EDA. R Package Version 1.1.10 (http:// CRAN.R-project.org/package=rgr).

GDAL/OGR contributors (2015). GDAL/OGR Geospatial Data Abstraction software Library. Open Source Geospatial Foundation. URL http://gdal.org.

Gis Sol, 2011. L'état des sols de France. Groupement d'intérêt scientifique sur les sols (188 pp.).

Goulard, M., Voltz, M., 1992. Linear coregionalization model: Tools for estimation and choice of cross-variogram matrix. Math. Geol. 24: 269. https://doi.org/10.1007/BF00893750.

GRASS Development Team, 2016. Geographic Resources Analysis Support System (GRASS) Software, Version 7.0. Open Source Geospatial Foundation. Electronic document: http://grass.osgeo.org.

Hengl, T., Heuvelink, G.B.M., Rossiter, D.G., 2007. About regression-kriging: from equations to case studies. Comput. Geosci. 33 (10), 1301–1315. doi.org/10.1016/j.cageo.2007.05.001.

Heuvelink, G.B.M., Burrough, P.A., Stein, A., 1989. Propagation of errors in spatial modelling with GIS Int. J. Geogr. Inf. Syst. 3, 303–322.

Heuvelink, G.B.M., Kros, J., Reinds, G.J., De Vries, W., 2016. Geostatistical prediction and simulation of European soil property maps. Geoderma Reg., 7(2), 201–215. doi.org/10.1016/j.geodrs.2016.04.002.

Heuvelink, G.B.M., Pebesma, W. J., 1999. Spatial aggregation and soil process modelling. Geoderma, 89: 47–65. doi.org/10.1016/S0016-7061(98)00077-9.

Hong, S. Y., Minasny, B., Han, K. H., Kim, Y., Lee, K., 2013. Predicting and mapping soil available water capacity in Korea. PeerJ, 1:e71 https://doi.org/10.7717/peerj.71.

Huang, J., Malone, B.P., Minasny, B., McBratney, A.B., Triantafilis, J., 2017. Evaluating a Bayesian modelling approach (INLA-SPDE) for environmental mapping. Sci. Total Environ., 609, 621–632. doi.org/10.1016/j.scitotenv.2017.07.201.

Huang, J., Subasinghe, R., Triantafilis, J., 2014. Mapping particle-size fractions as a composition using additive log-ratio transformation and ancillary data. Soil Sci. Soc. Am. J., 78(6), 1967–1976. doi:10.2136/sssaj2014.05.0215.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ., 83(1–2), 195–213. DOI: 10.1016/S0034-4257(02)00096-2.

Institut National de l'Information Géographique et Forestière, 2012 Base de Données Forêt. In : BD Forêt ®.

Jolivet, C., Arrouays, D., Boulonne, L., Ratié, C., Saby, N., 2006. Le Réseau de Mesures de la Qualité des Sols de France (RMQS). État d'avancement et premiers résultats. Étude Gest. Sols 13 (3), 149–164.

Jolivet C., Almeida-Falcon J.L., Berché P., Boulonne L., Fontaine M., Gouny L., et al., 2018. Manuel du Réseau de mesures de la qualité des sols. RMQS2: deuxième campagne métropolitaine, 2016–2027, Version 3, INRA, US 1106 InfoSol, Orléans, France.

King, D., Burrill, A., Daroussin, J., Le Bas, C., Tavernier, R., Van Ranst, E., 1995. The EU soil geographical database, in: King, D., Jones, R.J.A. Jones, Thomasson A.J. (Eds), European land information systems for agro-environmental monitoring. Joint Research Centre, Ispra, Italy, pp. 43–60.

Lacoste, M., Mulder, V.L., Richer-de-Forges, A.C., Martin, M.P., Arrouays, D., 2016. Evaluating large-extent spatial modeling approaches: A case study for soil depth for France. Geoderma Reg., 7(2), 137–152. doi.org/10.1016/j.geodrs.2016.02.006.

Lagacherie, P., McBratney, A.B., 2006. Chapter 1 Spatial Soil Information Systems and Spatial Soil Inference Systems: Perspectives for Digital Soil Mapping, in: Lagacherie, P., McBratney, A.B., Voltz, M. (Eds.), Developments in Soil Science, 31. Elsevier, Amsterdam, The Netherlands, pp. 3–22. doi.org/10.1016/S0166-2481(06)31001-X.

Lark, R.M., Bishop, T.F.A., 2007. Cokriging particle size fractions of the soil. Eur. J. Soil Sci. 58 (3), 763–774. doi.org/10.1111/j.1365-2389.2006.00866.x.

Laroche, B., Richer-de-Forges, A., Leménager, S., Arrouays, D., Schnebelen, N., Eimberck, M., et al., 2014. Le programme Inventaire Gestion Conservation des Sols de France : volet Référentiel Régional Pédologique. Étude Gest. Sols, 21 (1), 25–36. https://prodinra.inra.fr/record/282113.

Leenaars, J.G.B., Claessens, L., Heuvelink, G.B.M., Hengl, T., Ruiperez González, M., van Bussel, L.G.J., Guilpart, N., Yang, H., Cassman, K.G., 2018. Mapping rootable depth and root zone plant-available water holding capacity of the soil of sub-Saharan Africa, Geoderma, 324, 18–36. doi.org/10.1016/j.geoderma.2018.02.046.

Lin L.I., 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics 45, 255–268. DOI : 10.2307/2532051.

Marchant, B., Viscarra Rossel, R., Webster, R., 2013. Fluctuations in method-of-moments variograms caused by clustered sampling and their elimination by declustering and residual maximum likelihood estimation. Eur. J. Soil Sci., 64: 401–409. doi:10.1111/ejss.12029.

Marchetti, R., Donatelli, M. and Spallacci, P., 1997. Testing denitrification functions of dynamic crop models. J. Environ. Qual., 26(2), pp. 394–401. doi:10.2134/jeq1997.00472425002600020009x.

Mardhel, V., Gravier, A., 2005. Carte de vulnérabilité intrinsèque simplifiée des eaux souterraines du Bassin Seine-Normandie (Rapport BRGM/RP-54148-FR).

McBratney, A.B., Santos, M.L.M.,Minasny, B., 2003. On digital soil mapping. Geoderma 117 (1–2), 3–52. doi.org/10.1016/S0016-7061(03)00223-4.

Meinshausen, N., 2006. Quantile regression forests. J. Mach. Learn. Res., 7(Jun), 983–999.

Minasny, B., McBratney, A.B., Bristow, K.L., 1999. Comparison of different approaches to the development of pedotransfer functions for water-retention curves. Geoderma 93(3–4), 225–253. doi.org/10.1016/S0016-7061(99)00061-0.

Mulder, V.L., Lacoste, M., Richer-de-Forges, A.C., Arrouays, D., 2016. GlobalSoilMap France: High-resolution spatial modelling the soils of France up to two meter depth. Sci. Total Environ., 73, 1352–1369. doi.org/10.1016/j.scitotenv.2016.07.066.

Nemes, A., Schaap, M.G., Wösten, J.H.M., 2003. Functional evaluation of pedotransfer functions derived from different scales of data collection. Soil Sci. Soc. Am. J. 67(4), 1093–1102. doi:10.2136/sssaj2003.1093.

Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J., 1994. Spatial prediction of soil properties from landform attributes derived from a digital elevation model. Geoderma 63(3–4), 197–214.

Odeh, I.O.A., Todd, A.J., Triantafilis, J., 2003. Spatial prediction of soil particle-size fractions as compositional data. Soil Sci. 168 (7), 501–515. doi: 10.1097/01.ss.0000080335.10341.23.

O'Leary, G.J., Li Liu, D., Ma, Y., Li, F.Y., McCaskill, M., Conyers, M., et al., 2016. Modelling soil organic carbon 1. Performance of APSIM crop and pasture modules against long-term experimental data. Geoderma, 264, 227–237. doi.org/10.1016/j.geoderma.2015.11.004.

Padarian, J., Minasny, B., McBratney, A.B., Dalgliesh, N., 2014. Predicting and mapping the soil available water capacity of Australian wheatbelt. Geoderma Reg., 2–3, 110–118. doi.org/10.1016/j.geodrs.2014.09.005.

Parton, W.J., Schimel, D.S., Cole, C.V., Ojima, D.S., 1987. Analysis of factors controlling soil organic matter levels in Great Plains Grasslands 1. Soil Sci. Soc. Am. J., 51(5), 1173–1179. doi:10.2136/sssaj1987.03615995005100050015x.

Piedallu, C., Gegout, J.C., Bruand, A., Seynave, I., 2011. Mapping soil water holding capacity over large areas to predict potential production of forest stands. Geoderma 160(3–4), 355–366. doi.org/10.1016/j.geoderma.2010.10.004.

Poggio, L., Gimona, A., 2017. 3D mapping of soil texture in Scotland. Geoderma Reg., 9, 5–16. doi.org/10.1016/j.geodrs.2016.11.003.

Poggio, L., Gimona, A., Brown, I., Castellazzi, M., 2010. Soil available water capacity interpolation and spatial uncertainty modelling at multiple geographical extents. Geoderma, 160(2), 175–188. doi.org/10.1016/j.geoderma.2010.09.015.

Poggio, L., Gimona, A., Spezia, L., Brewer, M.J., 2016. Bayesian spatial modelling of soil properties and their uncertainty: The example of soil organic matter in Scotland using R-INLA. Geoderma, 277, 69–82. doi.org/10.1016/j.geoderma.2016.04.026.

Quinlan, J.R, 1992. Learning with continuous classes. Proceedings of the 5th Australian Joint Conference On Artificial Intelligence, pp. 343–348.

Quintana-Seguí, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., et al., 2008. Analysis of near-surface atmospheric variables: validation of the SAFRAN analysis over France. J. Appl. Meteorol. Climatol. 47, 92–107. doi.org/10.1175/2007JAMC1636.1.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Román Dobarco, M., Cousin, I., Le Bas, C., Martin, M.P., 2019. Pedotransfer functions for predicting available water capacity in French soils, their applicability domain and associated uncertainty. Geoderma, 336, 81–95. doi.org/10.1016/j.geoderma.2018.08.022.

Román Dobarco, M., Saby, N.P.A., Bourennane, H., Cousin, I., Arrouays, D., Martin, M.P., 2018. Digital soil mapping and uncertainty propagation of available water capacity for metropolitan France. TERRAENVISION Abstracts Vol. 1, TNV2018-ECO1-396, TERRAenVISION, 1 February 2018, Barcelona, Spain.

Sanchez, P.A., Ahamed, S., Carré, F., Hartemink, A.E., Hempel, J., Huising, J., et al., 2009. Digital soil map of the world. Science, 325(5941), 680–681. doi: 10.1126/science.1175084.

Schwärzel, K., Feger, K.-H., Häntzschel, J., Menzer, A., Spank, U., Clausnitzer, F., Köstner, B., Bernhofer, C., 2009. A novel approach in model-based mapping of soil water conditions at forest sites. For. Ecol. Manage. 258 (10), 2163–2174.

Shrestha, D.L., Solomatine, D.P., 2006. Machine learning approaches for estimation of prediction interval for the model output. Neural Netw. 19, 225–235.

Silva, B.M., Silva É.A., Oliveira, G.C., Ferreira, M.M., Serafim, M.E., 2014. Plant-available soil water capacity: estimation methods and implications. Rev. Bras. Cienc. Solo 38, 464–475.

Styc, Q., Lagacherie, P., 2018. What is the best spatial soil inference system for mapping Available water capacity? A test in Languedoc-Roussillon (France). Geophys. Res. Abstr., 20, EGU2018-6602-1, EGU General Assembly 2018, 9 April 2018, Vienna, Austria.

Tetegan, M., de Forges, A.C.R., Verbeque, B., Nicoullaud, B., Desbourdes, C., Bouthier, A., Arrouays, D., Cousin, I., 2015. The effect of soil stoniness on the estimation of water retention properties of soils: A case study from central France. Catena 129, 95–102.

Tetegan, M., Nicoullaud, B., Baize, D., Bouthier, A., Cousin, I., 2011. The contribution of rock fragments to the available water content of stony soils: Proposition of new pedotransfer functions. Geoderma 165(1), 40–49.

Tóth, B., Weynants, M., Nemes, A., Mako, A., Bilas, G., Tóth, G., 2015. New generation of hydraulic pedotransfer functions for Europe. Eur. J. Soil Sci. 66(1), 226–238. doi:10.1111/ejss.12192.

Tóth, B., Weynants, M., Pásztor, L., Hengl, T. 2017. 3D Soil Hydraulic Database of Europe at 250 m resolution. Hydrological Processes, 31:2662–2666. doi.org/10.1002/hyp.11203.

Ugbaje, S.U., Reuter, H.I., 2013. Functional digital soil mapping for the prediction of available water capacity in Nigeria using legacy data. Vadose Zone J., 12(4). doi:10.2136/vzj2013.07.0140.

USGS, 2004. Shuttle Radar Topography Mission, 1 Arc Second Scene SRTM_u03_n008e004, Unfilled Unfinished 2.0, Global Land Cover Facility. University of Maryland, College Park, Maryland (February 2000).

Vanderlinden, K., Giráldez, J. V., Van Meirvenne, M., 2005. Soil water-holding capacity assessment in terms of the average annual water balance in southern Spain. Vadose Zone J., 4, 317–328.

Van Ranst E., Vanmechelen L., Thomasson A.J., Daroussin J., Hollis J.M., Jones R.J.A., Jamagne M., King D., 1995. Elaboration of an extended knowledge database to interpret the 1:1,000,000 EU soil map for environmental purposes, in: King D., Jones R.J.A., Thomasson A.J. (Eds.). European Land Information Systems for Agro-environmental monitoring, CEC, Luxembourg. 286 pages.

Vaysse, K., Lagacherie, P., 2017. Using quantile regression forest to estimate uncertainty of digital soil mapping products. Geoderma, 291, 55–64. doi.org/10.1016/j.geoderma.2016.12.017.

Veihmeyer, F.J., Hendrickson, A.H. 1927. The relation of soil moisture to cultivation and plant growth. Soil Sci. 3, 498–513.

Viscarra Rossel R.A., Chen C., Grundy M.J., Searle R., Clifford D., Campbell P.H., 2015. The Australian three-dimensional soil grid: Australia's contribution to the GlobalSoilMap project. Soil Res. 53, 845–864. doi.org/10.1071/SR14366.

Wösten, J.H.M., Lilly, A., Nemes, A., Le Bas, C., 1999. Development and use of a database of hydraulic properties of European soils. Geoderma 90(3–4), 169–185. doi.org/10.1016/S0016-7061(98)00132-3.

Figure captions

Figure 1: General framework for modelling the spatial distribution of elementary available water capacity and its uncertainty by *GlobalSoilMap* layer. IGCS: French soil mapping and inventory program dataset; RMQS: French soil monitoring network; GEVARNOVIA: dataset with soil hydraulic measurements; θ_{FC}: soil moisture at field capacity; θ_{PWP}: soil moisture at permanent wilting point; R_v: coarse elements.

Figure 2: Location of soil profiles from the French soil mapping and inventory program dataset (IGCS) used for predicting the spatial distribution of particle size distribution and coarse elements.

Figure 3: Location of evaluation data for the soil input properties (RMQS), and evaluation data for the soil hydraulic properties (GEVARNOVIA).

Figure 4: Scatter plots of observed vs predicted values for RMQS horizons for a) clay (g kg^{-1}), b) silt (g kg^{-1}), c) sand (g kg^{-1}), and d) coarse elements (%). The 1:1 line is indicated in black.

Figure 5: Scatter plots of measured vs predicted values for GEVARNOVIA horizons for: a) soil moisture at field capacity (θ_{FC}) extracted from spatial *GlobalSoilMap* (GSM) predictions, b) soil moisture at field capacity (θ_{FC}) estimated applying the pedotransfer function to measured clay and sand contents, c) soil moisture at permanent wilting point (θ_{PWP}) extracted from spatial GSM predictions, and d) soil moisture at permanent wilting point (θ_{PWP}) estimated applying the pedotransfer function to measured clay and sand contents. Vertical bars represent the prediction intervals.

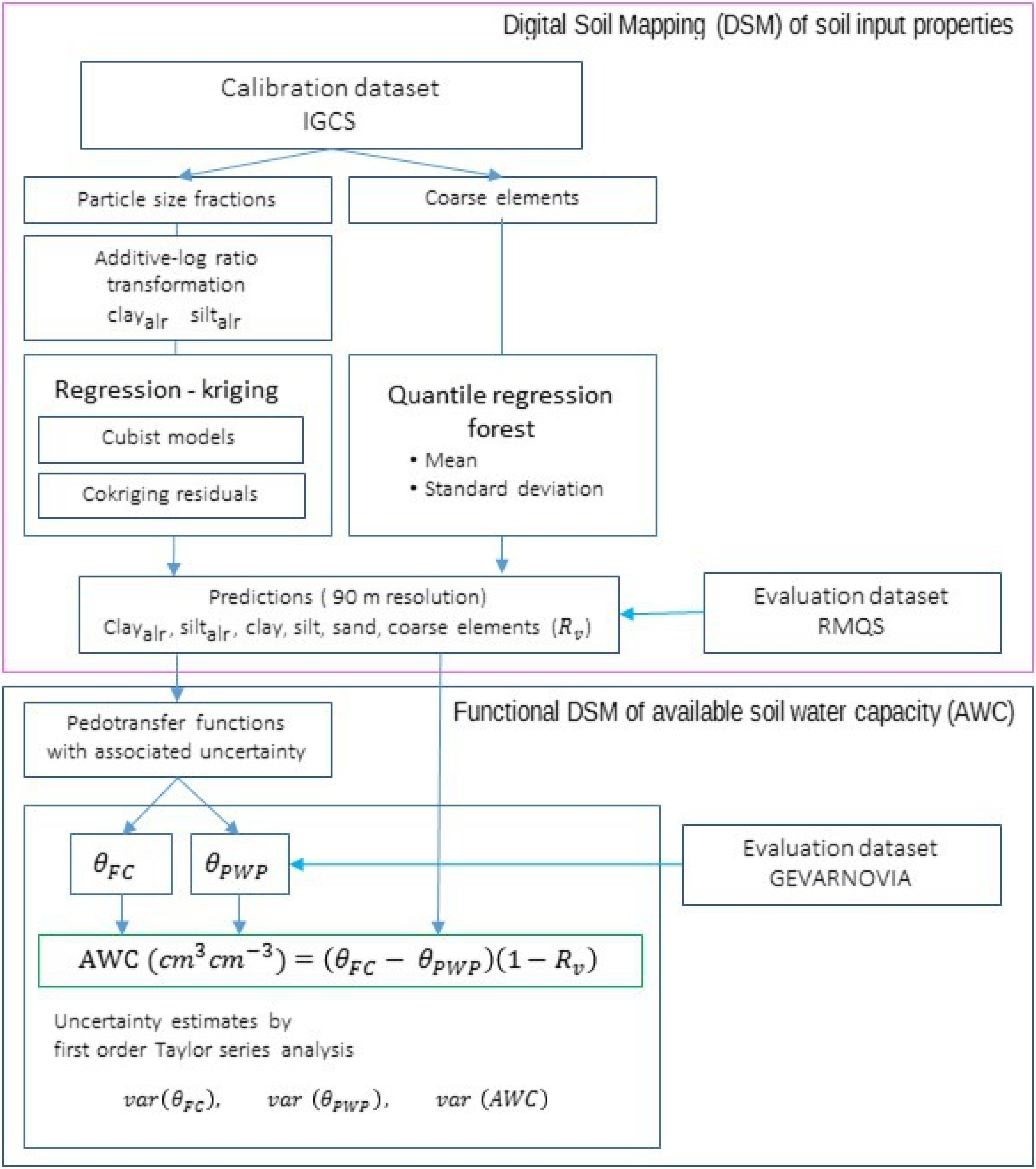
Figure 6: Boxplot of prediction errors (predicted - observed) by texture class: a) soil moisture at field capacity (θ_{FC}) by *GlobalSoilMap* (GSM) estimates, b) soil moisture at field capacity (θ_{FC}) estimated with the pedotransfer function (PTF) on measured clay and sand data, c) soil moisture at permanent wilting point (θ_{PWP}) by *GlobalSoilMap* (GSM) estimates, and d) soil moisture at permanent wilting point (θ_{PWP}) estimated with the pedotransfer function (PTF) on measured clay and sand data. Texture classes: coarse (C), medium (M), medium fine (MF), fine (F), and very fine (VF). The evaluation dataset did not have measurements of θ_{FC} for the very fine texture class.

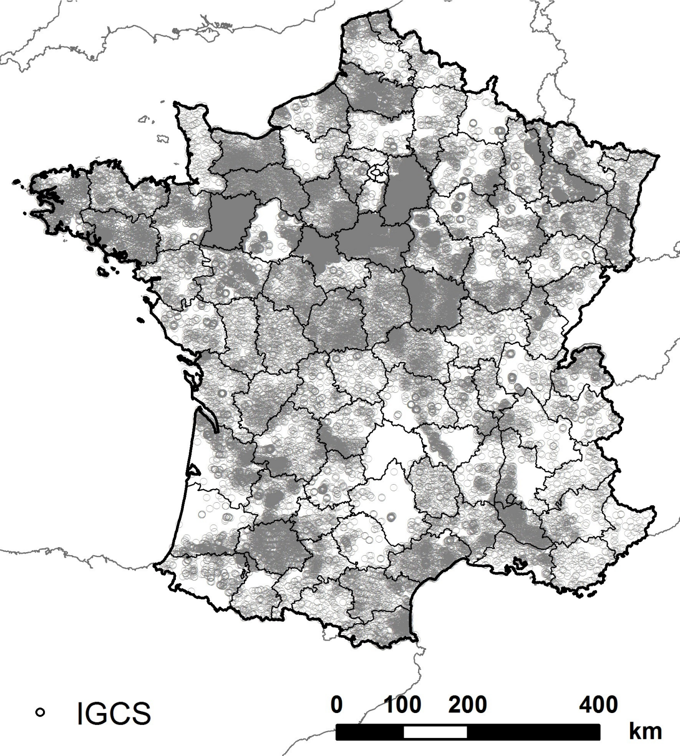
Figure 7: a) Soil thickness (cm), b) total available water capacity (AWC) (mm) to a maximum depth of 2 m, c) standard deviation (SD) of total AWC (mm), and d) relative error of AWC (coefficient of variation, CV).

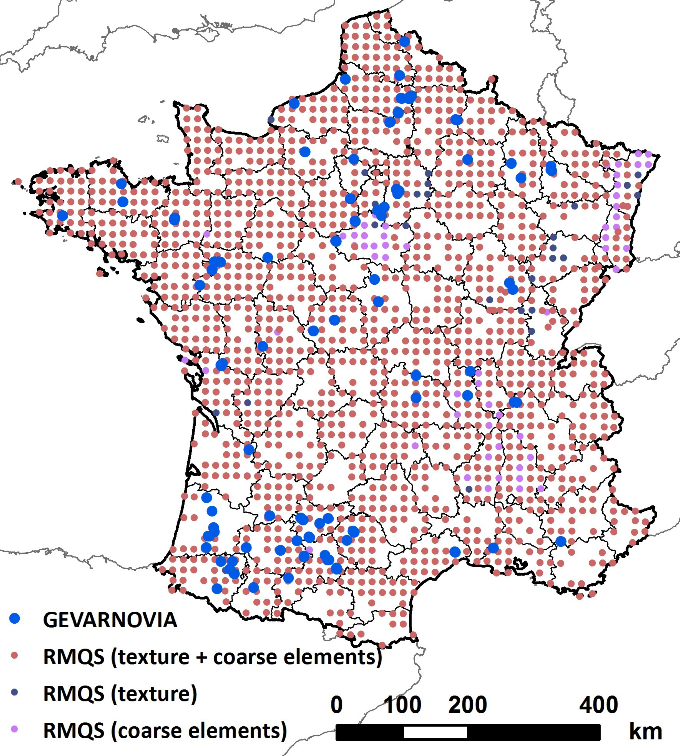
Figure 8: Soil properties used as input for the calculation of AWC and their standard deviation (SD) (15 - 30 cm). a) $Clay_{alr}$, b) $Silt_{alr}$, c) Coarse elements (%), d) $Clay_{alr}$ SD, e) $Silt_{alr}$ SD, and f) Coarse elements SD (%).

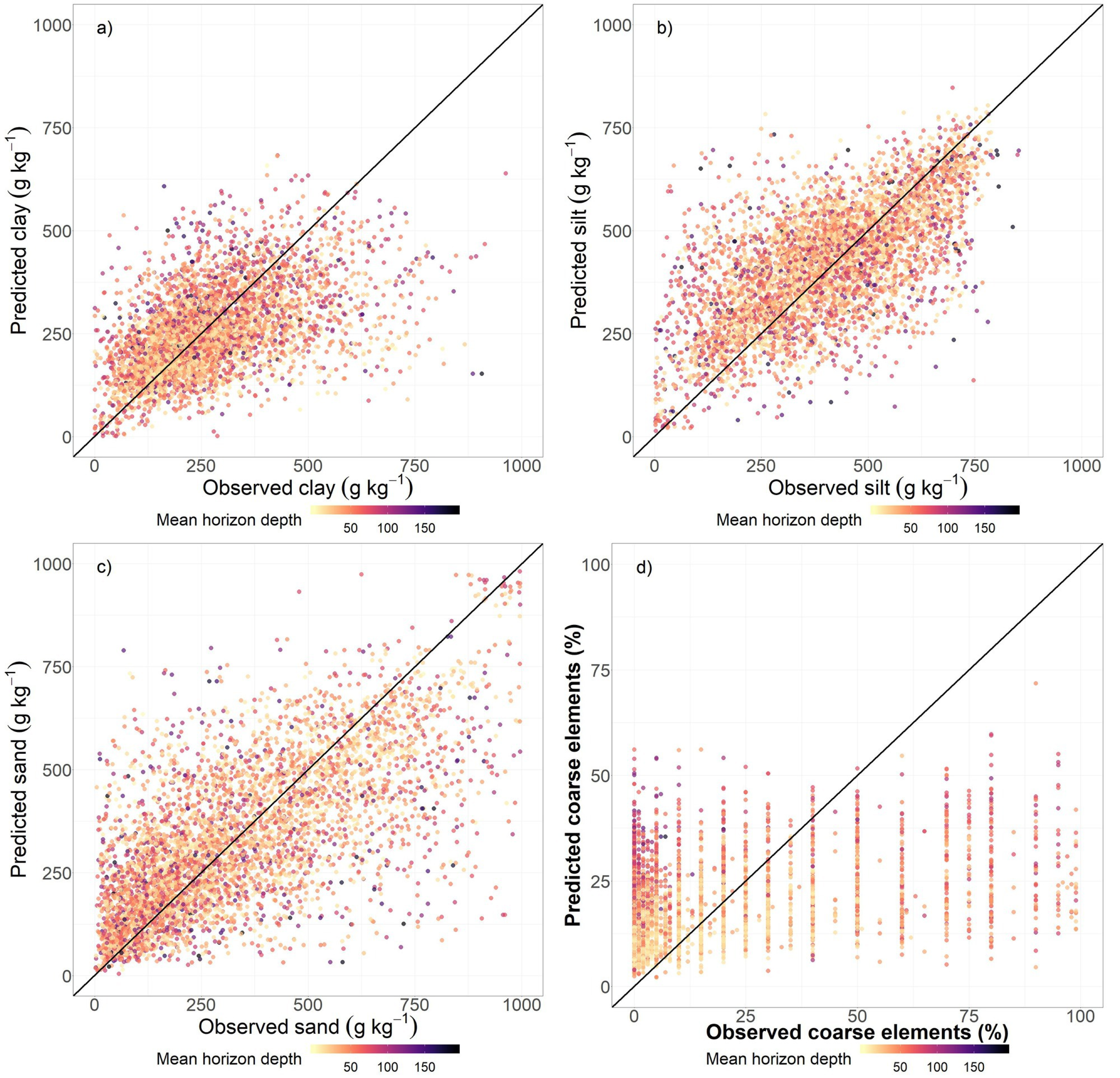
Figure 9: Components of the variance of elementary soil moisture at field capacity ($cm^3\ cm^{-3}$) due to soil input variables for the *GlobalSoilMap* layer 15-30 cm. a) Sensitivity of the PTF to $clay_{alr}$, b) sensitivity of the PTF to $silt_{alr}$, c) uncertainty term associated to $clay_{alr}$ (i.e., the multiplication of the squared sensitivity by the variance), d) uncertainty term associated to $silt_{alr}$, e) uncertainty term associated to the interaction between $clay_{alr}$ and $silt_{alr}$, and f) the total variance due to soil input variables.

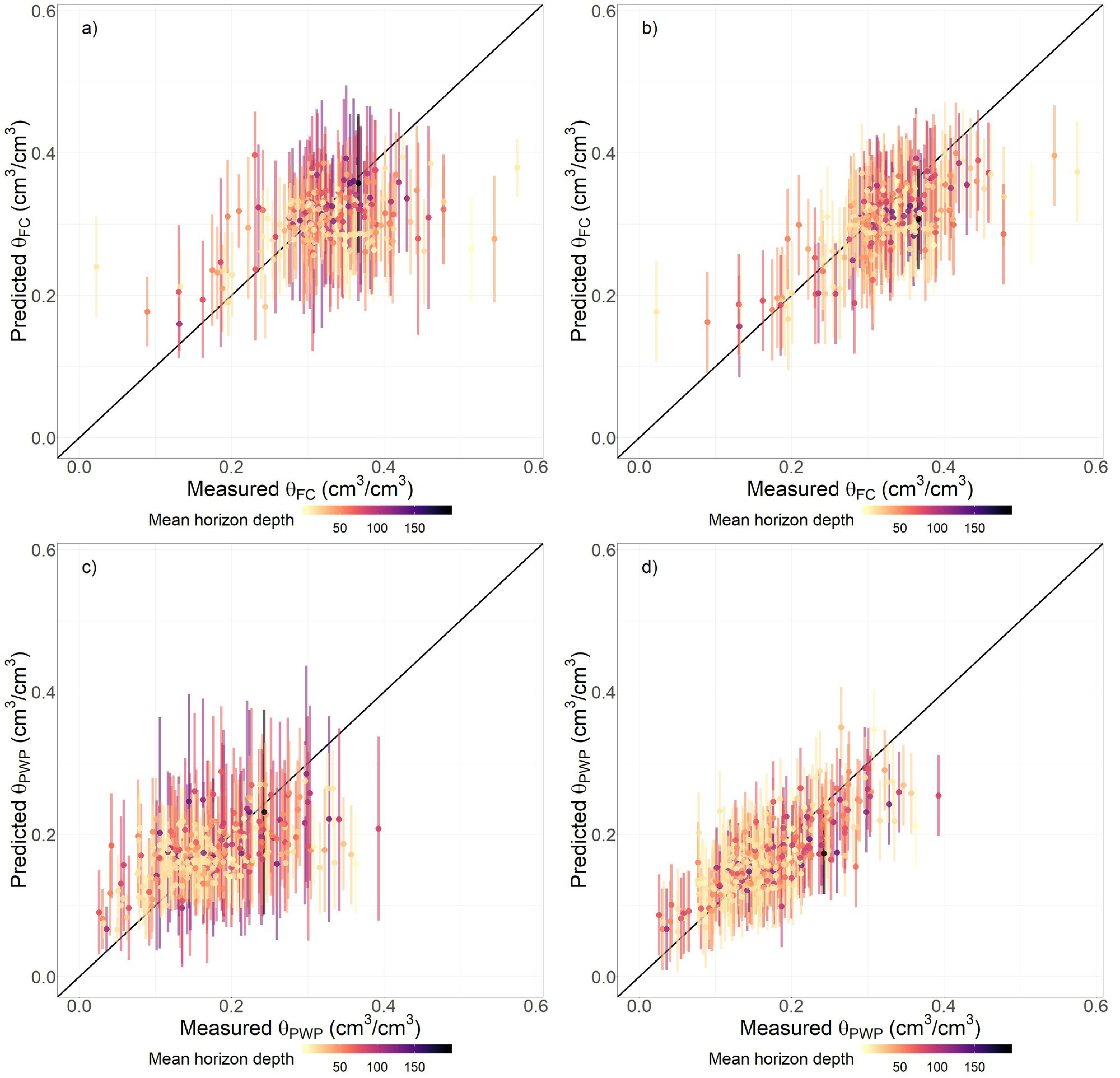
Figure 10: Components of the variance of elementary soil moisture at field capacity ($cm^3\ cm^{-3}$) associated to the PTF's coefficients for the GSM layer 15-30 cm: a) sensitivity of the PTF to clay coefficient (i.e clay predictions %), b) sensitivity of the PTF to sand coefficient (i.e sand predictions %), c) variance term of the interaction between the intercept and clay coefficient, d) variance term of the interaction between the intercept and the sand coefficient, e) variance term of the clay coefficient, f) variance term of the sand coefficient, g) variance term of the interaction between clay and sand, and g) total variance associated to the PTF's coefficients. The sensitivity of the function to the coefficient was 1, and consequently, the term of the variance associated to the intercept was constant ($3.80\ 10^{-5}$ $cm^6\ cm^{-6}$) (not included in the figure).


Digital Soil Mapping (DSM) of soil input properties
Calibration dataset
IGCS
Particle size fractions
Coarse elements
Additive-log ratio transformation
$clay_{alr}$ $silt_{alr}$
Regression - kriging
Cubist models
Cokriging residuals
Quantile regression forest
• Mean
• Standard deviation
Predictions (90 m resolution)
$Clay_{alr}$, $silt_{alr}$, clay, silt, sand, coarse elements (R_v)
Evaluation dataset
RMQS
Functional DSM of available soil water capacity (AWC)
Pedotransfer functions with associated uncertainty
θ_{FC}
θ_{PWP}
Evaluation dataset
GEVARNOVIA
$\text{AWC}\ (cm^3 cm^{-3}) = (\theta_{FC} - \theta_{PWP})(1 - R_v)$
Uncertainty estimates by
first order Taylor series analysis
$var(\theta_{FC})$, $var(\theta_{PWP})$, $var(AWC)$

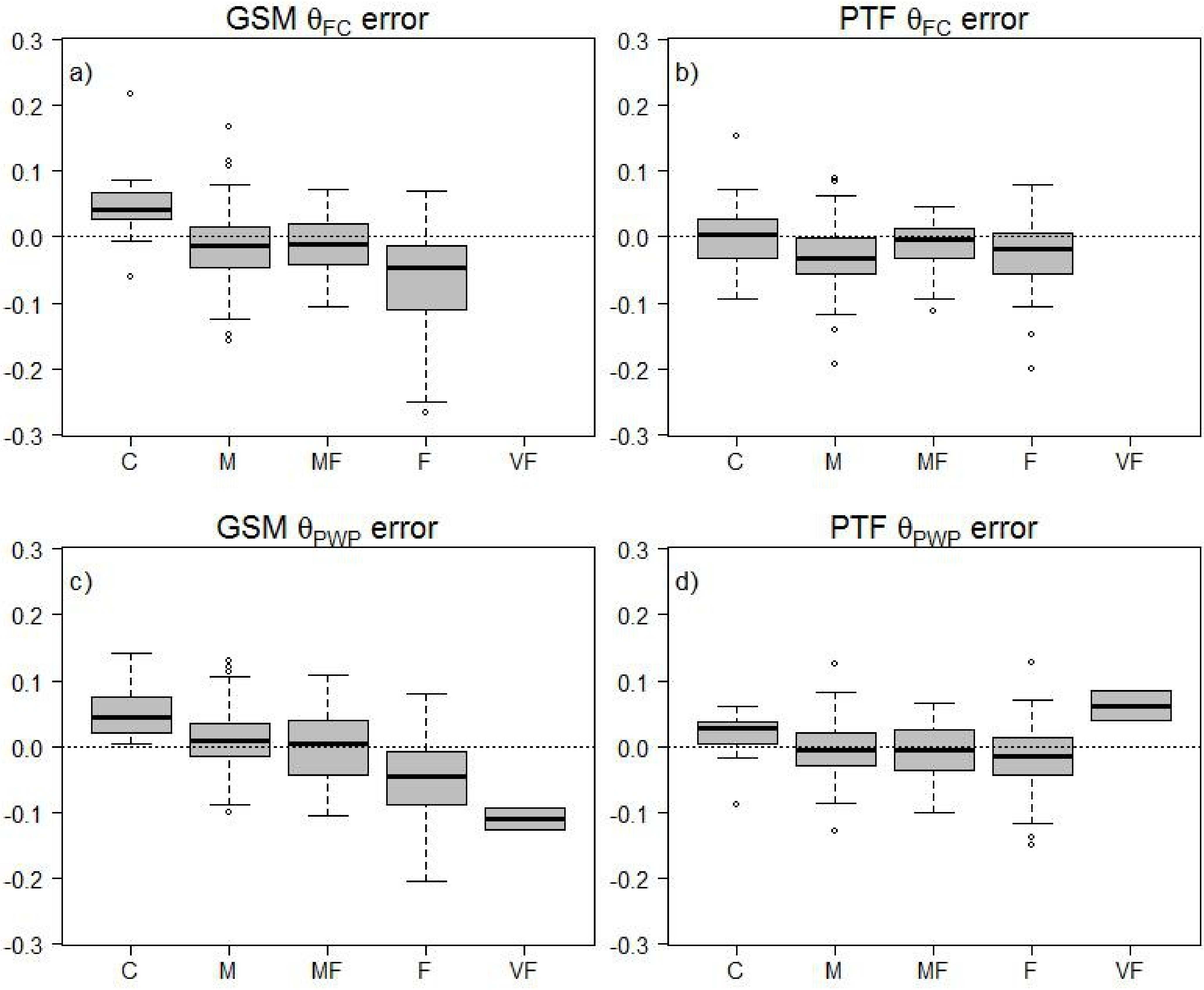

IGCS
0
100
200
400
km

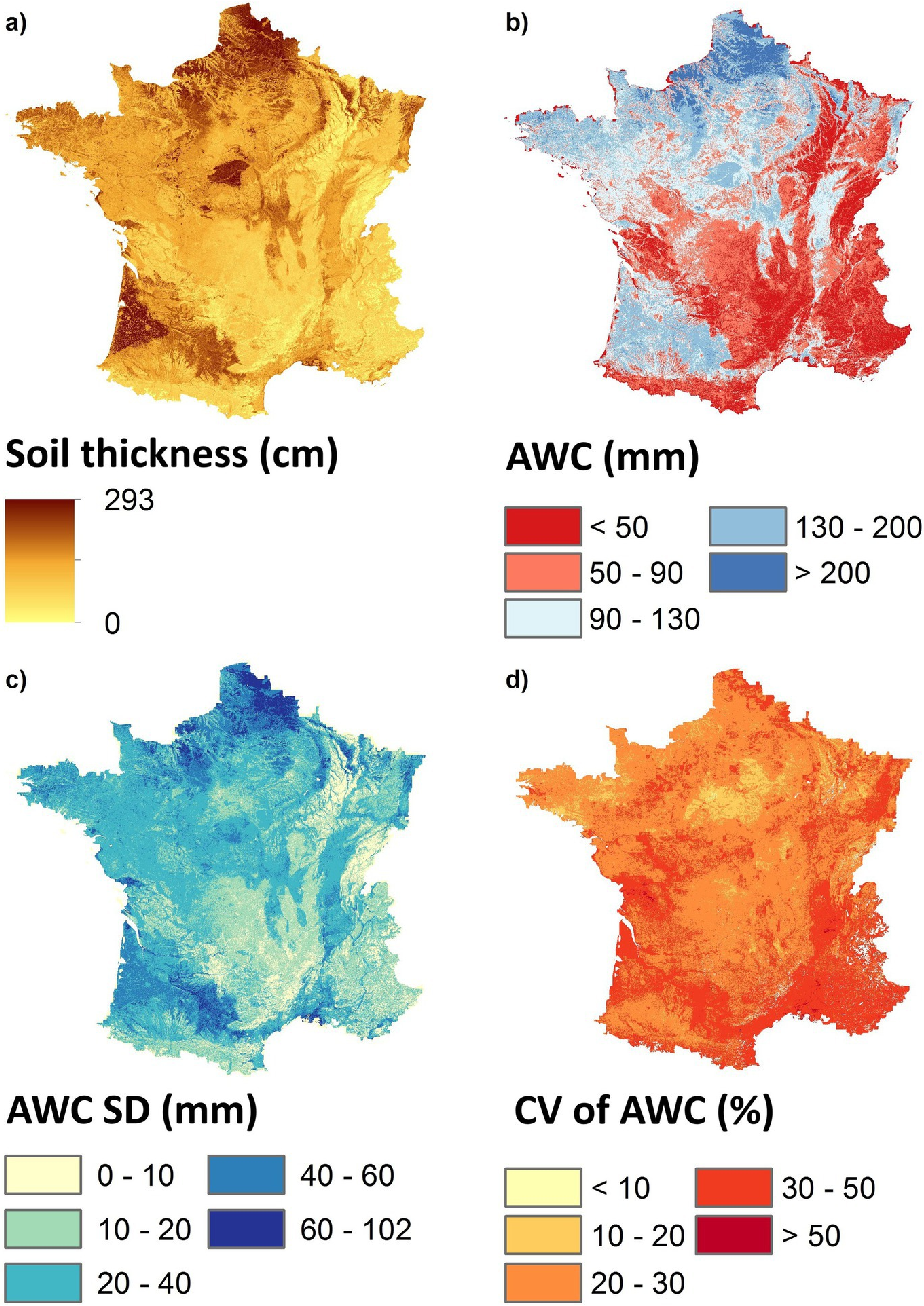

GEVARNOVIA
RMQS (texture + coarse elements)
RMQS (texture)
RMQS (coarse elements)
0
100
200
400
km

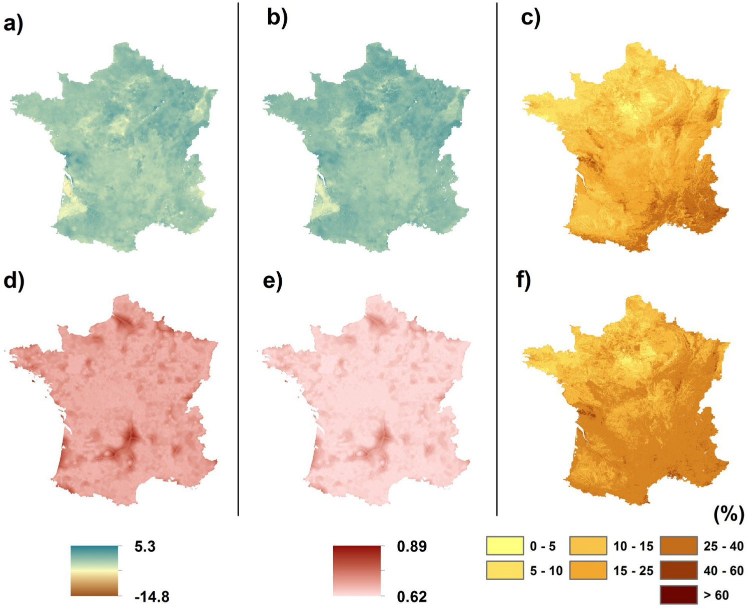

a)
Predicted clay (g kg⁻¹)
Observed clay (g kg⁻¹)
Mean horizon depth
50
100
150
b)
Predicted silt (g kg⁻¹)
Observed silt (g kg⁻¹)
Mean horizon depth
c)
Predicted sand (g kg⁻¹)
Observed sand (g kg⁻¹)
Mean horizon depth
d)
Predicted coarse elements (%)
Observed coarse elements (%)
Mean horizon depth



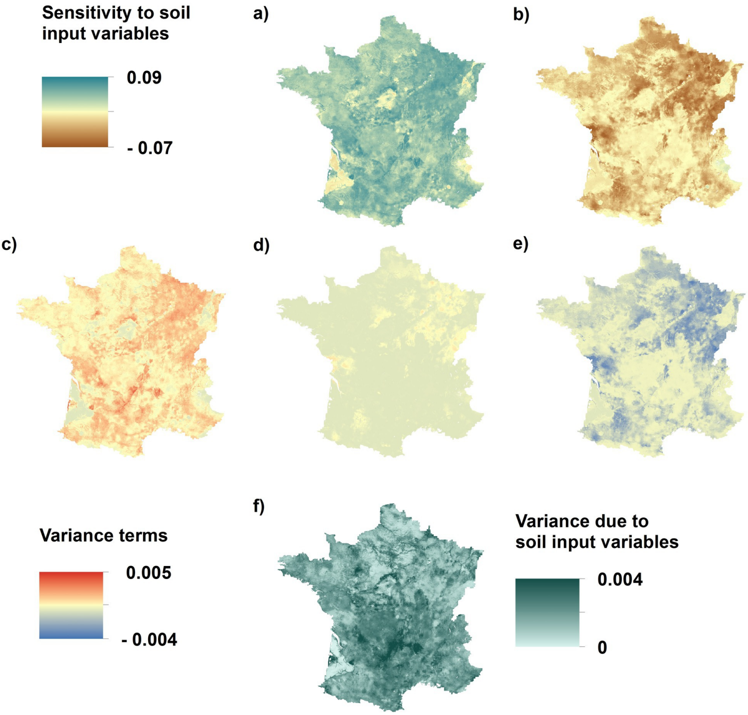
a)
b)
c)
d)
Predicted θFC (cm3/cm3)
Measured θFC (cm3/cm3)
Predicted θPWP (cm3/cm3)
Measured θPWP (cm3/cm3)
Mean horizon depth
50 100 150
0.0 0.2 0.4 0.6

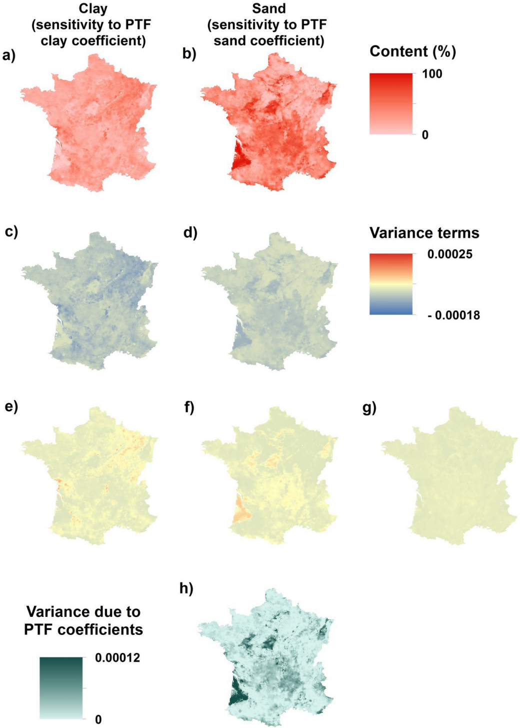

GSM θ_{FC} error
PTF θ_{FC} error
GSM θ_{PWP} error
PTF θ_{PWP} error
a)
b)
c)
d)
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
C
M
MF
F
VF


a)
b)
Soil thickness (cm)
293
0
AWC (mm)
< 50
50 - 90
90 - 130
130 - 200
> 200
c)
d)
AWC SD (mm)
0 - 10
10 - 20
20 - 40
40 - 60
60 - 102
CV of AWC (%)
< 10
10 - 20
20 - 30
30 - 50
> 50



a)
b)
c)
d)
e)
f)
5.3
-14.8
0.89
0.62
(%)
0 - 5
5 - 10
10 - 15
15 - 25
25 - 40
40 - 60
> 60


Sensitivity to soil input variables
0.09
- 0.07
a)
b)
c)
d)
e)
f)
Variance terms
0.005
- 0.004
Variance due to soil input variables
0.004
0


Clay
(sensitivity to PTF
clay coefficient)
Sand
(sensitivity to PTF
sand coefficient)
a)
b)
Content (%)
100
0
c)
d)
Variance terms
0.00025
- 0.00018
e)
f)
g)
h)
Variance due to
PTF coefficients
0.00012
0

Tables

Table 1: Number of observations by *GlobalSoilMap* layer in the calibration dataset for the particle size distribution and coarse elements models.

Depth (cm)	Particle size distribution (N)	Coarse elements (N)
0–5	36381	51966
5–15	35614	53552
15–30	35614	53516
30–60	31687	50500
60–100	25005	47900
100–200	13183	45169

Table 2: Description of the environmental covariates used for fitting regression models for particle size distribution and coarse elements. Soil forming factors: soil (S), climate (C), organisms (O), relief (R), parent material (P). SAFRAN applies an optimal interpolation of observations from meteorological stations (1958-present) and surface analyzes from numerical weather prediction systems at 8 km resolution (Quintana-Seguí et al., 2008; Durand et al., 2009). The IRNDP is a proxy for permeability of the geological material, and it is calculated from the comparison between the observed hydrological network and the theoretical network based on topographic conditions (Mardhel and Gravier, 2005).

Source	Variables	Soil forming factor	Scale/ resolution	Reference
SAFRAN	Mean annual potential evapotranspiration, precipitation, and temperature statistics (minimum, median, mean, maximum)	C	8 km	Quintana-Seguí et al. (2008)
SRTM	Elevation, slope, elevation above channel network, slope height, mid-slope position, multiresolution valley bottom flatness index (Gallant and Dowling, 2003), multiresolution ridgetop flatness index (Gallant and Dowling, 2003), topographic wetness index (Böhner et al., 2002), compound topographic index, curvature, longitudinal curvature, transversal curvature, exposition, heat load index, linear aspect, roughness, surface area ratio, slope position, surface relief ratio	R	90 m	USGS (2004)
	Erosion rate	S, R	1:1000000	Cerdan et al. (2010)
French Soil Geographical Database	Soil type, parent material	S, P	1:1000000	Gis Sol (2011)
	Index of Development and Persistence of Hydrological Network	P	1:50000	Mardhel and Gravier (2005)
Gravimetric data	Gravimetric data: Bouger anomaly, free-air bouguer anomaly, Bouguer gravity anomaly.	R, P	4 km	Achache et al. (1997)
Corine Land Cover 2006	Land use	O	250 m	EEA (2007)
BD Forêt version 1.0	Natural and semi-natural vegetation type	O		IGN (2012)
ECOCLIMAP-II	Land use	O	1 km	Faroux et al. (2003)
MODIS	Enhanced vegetation index: median for January (2002-2014), median for June (2002-2014). Normalized difference vegetation index: median for January (2002-2014), median for June (2002-2014)	O	500 m	Didan (2015)

Table 3: Variance-covariance matrices of PTFs coefficients for estimating soil moisture at field capacity ($\theta_{2.0}$) and at permanent wilting point ($\theta_{4.2}$).

$\theta_{2.0}$

	Intercept	Clay	Sand
Intercept	$3.80\ 10^{-5}$	$-9.93\ 10^{-7}$	$-3.85\ 10^{-7}$
Clay	$-9.93\ 10^{-7}$	$3.17\ 10^{-8}$	$7.05\ 10^{-9}$
Sand	$-3.85\ 10^{-7}$	$7.05\ 10^{-9}$	$9.09\ 10^{-9}$

$\theta_{4.2}$

	Intercept	Clay	Sand
Intercept	$1.84\ 10^{-5}$	$-4.07\ 10^{-7}$	$-1.97\ 10^{-7}$
Clay	$-4.07\ 10^{-7}$	$1.04\ 10^{-8}$	$3.79\ 10^{-9}$
Sand	$-1.97\ 10^{-7}$	$3.79\ 10^{-9}$	$3.76\ 10^{-9}$

Table 4: Fitted parameters for the linear model of coregionalization for the cubist residuals of $clay_{alr}$ and $silt_{alr}$ at each *GlobalSoilMap* depth interval. The covariogram models were spherical.

Depth	Variable	N	Nugget	pSill	Range (m)
0–5	$Silt_{alr}$	36159	0.41	0.45	190098
	$Clay_{alr}$		0.48	0.61	
	$Silt_{alr}$ x $Clay_{alr}$		0.37	0.42	
5–15	$Silt_{alr}$	36108	0.39	0.33	178104
	$Clay_{alr}$		0.45	0.48	
	$Silt_{alr}$ x $Clay_{alr}$		0.34	0.31	
15–30	$Silt_{alr}$	35401	0.35	0.32	160970
	$Clay_{alr}$		0.40	0.50	
	$Silt_{alr}$ x $Clay_{alr}$		0.30	0.32	
30–60	$Silt_{alr}$	31494	0.58	0.48	170776
	$Clay_{alr}$		0.62	0.74	
	$Silt_{alr}$ x $Clay_{alr}$		0.48	0.46	
60–100	$Silt_{alr}$	24849	1.83	0.56	252306
	$Clay_{alr}$		1.50	0.91	
	$Silt_{alr}$ x $Clay_{alr}$		1.31	0.60	
100–200	$Silt_{alr}$	13086	3.03	1.25	167139
	$Clay_{alr}$		2.50	1.42	
	$Silt_{alr}$ x $Clay_{alr}$		1.87	1.16	

Table 5: Independent evaluation statistics for clay, silt, sand, and coarse elements from observed RMQS horizons.

Variable	N	R^2	Concordance	RMSE	bias	PICP (%)
Clay (g kg^{-1})	4970	0.27	0.49	127.7	-15.3	83
Silt (g kg^{-1})	4970	0.43	0.63	138.6	19.3	86
Sand (g kg^{-1})	4970	0.46	0.66	171.8	-2.7	90
Coarse elements (%)	4988	0.14	0.26	21.0	3.3	76

Table 6: Independent evaluation statistics for soil moisture at field capacity (θ_{FC}) and soil moisture at permanent wilting point (θ_{PWP}) measured at the laboratory on horizon samples (GEVARNOVIA dataset). The soil moisture contents estimates were calculated applying pedotransfer functions (PTFs) to measured particle size distribution (PSD) data from horizon samples, or applying the PTFs to weighed averages of *GlobalSoilMap* (GSM) spatial predictions.

Estimate origin	Variable	N	R^2	Concordance	RMSE	bias	PICP (%)
PTFs on measured horizon PSD	θ_{FC} (cm^3cm^{-3})	236	0.54	0.65	0.052	-0.02	84.3
	θ_{PWP} (cm^3cm^{-3})	308	0.62	0.75	0.042	-0.005	85.1
GSM prediction	θ_{FC} (cm^3cm^{-3})	236	0.21	0.37	0.065	-0.02	71.2
	θ_{PWP} (cm^3cm^{-3})	308	0.29	0.47	0.057	-0.0004	76.6