



HAL
open science

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

► To cite this version:

Mercedes Roman Dobarco, Hocine Bourennane, Dominique Arrouays, Nicolas P. A. Saby, Isabelle Cousin, et al.. Uncertainty assessment of GlobalSoilMap soil available water capacity products: A French case study. *Geoderma*, 2019, 344, pp.14-30. <10.1016/j.geoderma.2019.02.036>. <hal-02627552>

HAL Id: hal-02627552

<https://hal.inrae.fr/hal-02627552v1>

Submitted on 22 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

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

3

4 **Authors:**

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

6 Hocine Bourennane^b: hocine.bourennane@inra.fr

7 Dominique Arrouays^a: dominique.arrouays@inra.fr

8 Nicolas P. A. Saby^a: nicolas.saby@inra.fr

9 Isabelle Cousin^b: isabelle.cousin@inra.fr

10 Manuel P. Martin^a: manuel.martin@inra.fr

11

12 **Affiliations:**

13 ^a INRA, US1106 Unité InfoSol, 2163 Avenue de la Pomme de Pin, CS 40001 Ardon, F-45075

14 Orléans Cedex 2, France

15 ^b INRA, UR0272 Unité de Science du Sol, 2163 Avenue de la Pomme de Pin, CS 40001

16 Ardon, F-45075 Orléans Cedex 2, France

17

18 **Corresponding author:**

19 Manuel P. Martin^a: manuel.martin@inra.fr

20 Postal address: INRA Centre de recherche d'Orléans, US1106 Unité InfoSol, 2163 avenue de

21 la Pomme de Pin, CS 40001 Ardon, F45075 Orléans CEDEX 2, France

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

23

24

25

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

28 **Abstract**

29 Plant available water capacity (AWC) refers to the maximum amount of water that a soil can store and
30 provide to plant roots. Spatial predictions of AWC through digital soil mapping at high resolution and
31 national extent provide relevant information for upscaling ecological and hydrological models, and
32 assessment of the provision of ecosystem services like water quantity and quality regulation, carbon
33 sequestration, and provision of food and raw materials. However, the spatial predictions of AWC are
34 prone to errors and uncertainties. Moreover, this digital soil mapping process requires using pedotransfer
35 functions (PTFs) due to the lack of sufficient georeferenced measurements of the upper (i.e., soil
36 moisture at field capacity, θ_{FC}) and lower (i.e., soil moisture at permanent wilting point, θ_{PWP}) limits
37 of soil moisture contents defining AWC. This adds an additional source of uncertainty to the final
38 estimates of AWC. The objectives of this study were: 1) to predict AWC for mainland France following
39 the *GlobalSoilMap* (GSM) project specifications on depth intervals and uncertainty, and 2) to quantify
40 the uncertainty of AWC accounting for uncertainty of the soil input variables and the PTFs' coefficients.
41 We first predicted the soil input properties by GSM layer (0–5, 5–15, 15–30, 30–60, 60–100, 100–200
42 cm), and then applied PTFs for estimating θ_{FC} , θ_{PWP} , and volumetric AWC ($\text{cm}^3 \text{cm}^{-3}$). The volume of
43 coarse elements by GSM layer was subtracted before aggregating AWC to estimated soil depth for a
44 maximum of 2 m. The uncertainty of AWC was quantified by first-order Taylor analysis. Independent
45 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
46 RMSE (clay RMSE = 128 g kg^{-1} , silt RMSE = 139 g kg^{-1} and sand RMSE = 172 g kg^{-1}) from the three
47 particle size fractions. However, the model for coarse elements had the worst predictive performance
48 ($R^2 = 0.14$ and RMSE = 21 %) among all AWC input variables. The performance of the GSM predictions
49 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
50 sand and clay, the RMSE for θ_{FC} and θ_{PWP} had a relative increase of 25 % and 36 % respectively
51 compared to when they were applied to measured horizon data. Across the majority of mainland France,
52 the main sources of uncertainty of elementary AWC were coarse elements and soil texture, but the

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

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

59 **1. Introduction¹**

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

70 Many agricultural and ecological models have AWC, θ_{FC} , or θ_{PWP} as input variables [e.g., STICS
71 (Brisson et al., 1998), CENTURY (Parton et al., 1987), APSIM (O'Leary et al., 2016) SWAT (Arnold
72 et al., 1987; Arnold and Fohrer, 2005)]. Thus, spatially explicit predictions of AWC at high resolution
73 are relevant for upscaling simulation models at regional or national scale, and assessing the provision of
74 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} .

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

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

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

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

108 AWC is included in the soil properties of the *GlobalSoilMap* project (GSM), which aims to produce a
109 digital soil map of the world at 3-arc second resolution providing estimates of uncertainty, following a
110 bottom-up approach (Sanchez et al., 2009; Arrouays et al., 2014). Although AWC is still rarely mapped
111 (Ugbaje and Reuter, 2013), the number of studies on AWC are increasing in the DSM literature from
112 national (Hong et al., 2013, Padarian et al., 2014) to continental extent (Wösten et al., 1999; Ballabio et
113 al., 2016; Tóth et al., 2016). GSM products for AWC are already available for Scotland (Poggio et al.,
114 2010), Nigeria (Ugbaje and Reuter, 2013), and Australia (Viscarra Rossel et al., 2015). The objectives of
115 this study were: 1) to predict AWC for mainland France following the GSM specifications, and 2) to
116 quantify the uncertainty of AWC accounting for uncertainty of the soil input variables and the PTFs'
117 coefficients. The incorporation of the uncertainty due to the PTFs' coefficients into the AWC spatial
118 modelling is a novelty relative to previous studies at national extent.

119 **2. Methods**

120 **2.1 General framework**

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

$$126 \text{ elementary AWC (cm}^3\text{cm}^{-3}\text{)} = (\theta_{FC} - \theta_{PWP})(1 - R_v) \quad [1]$$

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

$$128 \text{ AWC (mm)} = (\theta_{FC} - \theta_{PWP})(1 - R_v)d \quad [2]$$

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

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

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

156
$$AWC = \sum_{h=1}^6 (1 - R_h) (\theta_{FC}^h - \theta_{PWP}^h) t_h \quad [3]$$

157

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

162 **2.2 Soil data**

163 2.2.1 Calibration data from the French Soil Mapping and Inventory program

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

175 2.2.2 Evaluation data for soil input properties: French soil monitoring network

176 The French soil monitoring network (Réseau de Mesures de la Qualité des Sols: RMQS) is based on a
177 systematic random grid of 16 km by 16 km that covers metropolitan France with approximately 2200
178 sites (Jolivet et al., 2006). Hence, we used data from the first RMQS campaign (2000-2012) as an
179 independent evaluation sample for particle size distribution and coarse elements predictions (Brus et al.,
180 2011). At each RMQS site, a soil pit of approximately 120 cm by 90 cm was dug to the appearance of
181 parent material, and fully described. Samples were collected from each horizon of the soil profile and
182 analyzed at the laboratory for determining the content of sand, silt, and clay using the pipette method

183 (ISO 13317-2:2001). The soil surveyors estimated visually the content of coarse elements (% volume)
184 on the three faces of the soil pit. In May 2018, the database had data of particle size distribution from
185 1622 RMQS sites. The particle size distribution of some RMQS sites mostly located in forested areas
186 have not been analyzed in the laboratory yet (Figure 3). Similarly, soil profile data of coarse elements
187 was available for 1662 RMQS sites.

188 2.2.3 Evaluation data for soil hydraulic properties: GEVARNOVIA

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

199 2.2.4 Data pretreatment

200 Particle size distribution constitutes compositional data (i.e., sand, silt, and clay vary between 0 and
201 1000 g kg⁻¹, and sum up to 1000 g kg⁻¹) that is subject to non-stochastic constraints (Lark and Bishop,
202 2007). As compositional data, their distributions cannot be drawn from the real space \mathbb{R}^3 , but from the
203 two-dimensional simplex plane \mathcal{S}^2 embedded in this space (Lark and Bishop, 2007). Hence, to avoid
204 negative spurious correlations between the components, and guarantee that their predictions sum up to
205 a constant, the distributions of sand, silt, and clay should not be analyzed independently but based on
206 their ratios (Odeh et al., 2003). Aitchison (1986) proposed the additive log-ratio transformation (*alr*),
207 which is defined as:

$$208 \quad \mathbf{x} = alr(\mathbf{z}) = \left(\ln\left(\frac{z_1}{z_D}\right), \ln\left(\frac{z_2}{z_3}\right), \dots, \ln\left(\frac{z_{D-1}}{z_D}\right) \right) \quad [4]$$

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

$$211 \quad \mathbf{z} = \frac{\exp(\mathbf{w})}{\mathbf{j}^T \exp(\mathbf{w})} k \quad [5]$$

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

$$218 \quad clay_{alr} = \ln\left(\frac{clay}{sand}\right) \quad [6]$$

$$219 \quad silt_{alr} = \ln\left(\frac{silt}{sand}\right) \quad [7]$$

220 We used sand as the denominator after comparing the evaluation statistics and spatial structure of the
 221 model residuals of the three combinations in preliminary tests (results not shown) (Poggio and Gimona,
 222 2017).

223 **2.3 Modelling soil input properties**

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

$$231 \quad soil = f(s, c, o, r, p, a, n) + \epsilon \quad [8]$$

232 Where c is climate and a is time.

233 2.3.1 Environmental covariates

234 We selected 44 covariates describing the *scorpan* factors soil, climate, vegetation, relief and parent
235 material (Table 2). Climatic variables came from the French SAFRAN atmospheric analysis system
236 (Durand et al. 1993). We used a Digital Elevation Model from SRTM (Shuttle Radar Topography
237 Mission) at 90 m (USGS, 2004) to derive primary and secondary relief covariates in ArcGIS (ESRI,
238 Redlands, WA). Soil and parent material were characterized by predominant soil type and parent
239 material by soil mapping unit of the French Soil Geographical Database (Gis Sol, 2011), erosion rates
240 (Cerdan et al., 2010), geophysical gravimetric data, and the Index of River Network Development and
241 Persistence (IRNDP). Vegetation and land use were classified according to Corine Land Cover 2006
242 data (European Environmental Agency, 2007), ECOCIMAP-II (Faroux et al., 2003), and BD Forêt
243 (Institut National de l'Information Géographique et Forestière, 2012). Two vegetation indices derived
244 from remote sensing data were used to describe the photosynthetic capacity of the vegetation cover, the
245 enhanced vegetation index (EVI) and the normalized difference vegetation index (NDVI) (Huete et al.,
246 2002). The MOD13A1 MODIS/Terra Vegetation Indices 16-day composite products at 500 m resolution
247 were retrieved from the online NASA Earthdata Search, courtesy of the NASA EOSDIS Land Processes
248 Distributed Active Archive Center (LP DAAC) (<https://earthdata.nasa.gov/>) (Didan, 2015). The
249 vegetation indices were collected for the months of January (i.e., minimum vegetation activity) and June
250 (i.e., maximum vegetation activity) for the period 2002-2014. The median of these vegetation indices
251 over the 13 years for each month were used as covariates. All the covariates were projected to the
252 Lambert 93 (EPSG: 2154) associated to the Réseau Géodésique Français 1993 (RGF93), aligned with
253 the SRTM, and resampled to 90 m resolution using nearest neighbor interpolation. Data pre-processing
254 was done in GRASS (GRASS Development Team, 2016), the R software v.3.2.2 (R Core Team, 2015),
255 and the Geospatial Data Abstraction Library v.2.0.1 (GDAL/OGR contributors, 2015).

256 2.3.2 Soil spatial predictive models

257 We evaluated the correlation among clay_{alr} , silt_{alr} , and coarse elements in the feature space as well as the
258 spatial correlation prior to modelling their spatial distribution. In preliminary tests, we also evaluated
259 the spatial correlation and correlation in the feature space of the residuals of the models (Supplemental
260 material S2 p.1). After checking the lack of correlation between coarse elements and the *alr*-variables,

261 the weak spatial structure of the residuals of coarse elements and preliminary mapping exercises of
262 AWC (Román Dobarco et al., 2018) we decided to model separately and differently the *alr*-variables
263 and the coarse elements.

264 We predicted the *alr*-variables with a regression-cokriging model (Odeh et al., 1994; Hengl et al., 2007).
265 Cubist models for clay_{alr} and silt_{alr} were fitted using the environmental covariates describing *scorpan*
266 factors (Table 2) by GSM depth interval. The Cubist algorithm is a hybridized model that combines
267 tree-based models and linear models. The terminal nodes of the regression tree (leaves) consist on a
268 linear model (Quinlan, 1992). The parameters of the Cubist models were: committees=20, extrapolation
269 = 5, and unbiased=TRUE. We calculated the model residuals at the calibration points, and then fitted a
270 linear model of coregionalisation (LMCR) between the residuals of both variables for each GSM depth
271 interval using the algorithm presented by Goulard and Voltz (1992). The LMCR had two components,
272 a nugget and a spherical variogram. We then interpolated spatially the residuals of clay_{alr} and silt_{alr} by
273 ordinary co-kriging using the closest 10 observations. The final predictions were calculated by adding
274 the kriged residuals to the Cubist predictions and back-transformed to the original scale (equation 5).
275 The kriging variance of the residuals of the *alr*-variables was used afterwards for estimating the
276 prediction uncertainty (see section 2.5).

277 We modelled coarse elements with quantile regression forests (Meinshausen, 2006) for its ability to
278 provide accurate estimates of uncertainty of predicted soil properties (Vaysse and Lagacherie, 2017).
279 Quantile regression forests is a generalization of random forest models (Breiman, 2001). Random forests
280 is a very popular machine-learning tool for classification or regression that provides an ensemble
281 prediction based on many regression trees. For each regression tree and node, the algorithm incorporates
282 randomness by selecting randomly a subset of features to split on. Quantile regression forest not only
283 provides robust estimates of the conditional mean, but also of the full conditional distribution of the
284 response variable. Whereas random forests keeps the mean value of observations at the nodes, quantile
285 regression forests keeps the values of all observations at the nodes, and can infer estimates for
286 conditional quantiles, prediction intervals, or other statistics from the distribution (Meinshausen, 2006).
287 A detailed description of random forests and quantile random forests can be found in Breiman (2001)

288 and Meinshausen (2006) respectively. We fitted quantile regression forest models for coarse elements
289 by GSM depth, with the settings ntree = 1000 (number of trees), nodesize = 20 (minimum number of
290 observations in terminal nodes), and the default mtry (number of variables randomly sampled as
291 candidates at each split), which in this case was 14. We predicted the mean, the 5th percentile, the 95th
292 percentile, and the standard deviation of coarse elements by GSM depth.

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

296 **2.4 Functional digital soil mapping of AWC**

297 2.4.1 Pedotransfer functions

298 The volumetric soil moisture content (cm³ cm⁻³) at field capacity or pF = 2.0 ($\theta_{2.0}$) and at permanent
299 wilting point or pF = 4.2 ($\theta_{4.2}$) for the fine fraction were estimated using PTFs developed by Román
300 Dobarco et al. (2019) with the French SOLHYDRO database (Bruand et al., 2003; Al Majou et al.,
301 2008b). These PTFs use the content of clay (%) and sand (%) as predictor variables:

$$302 \quad \theta_{2.0} = 0.278 + 2.45 \cdot 10^{-3} \text{ clay} - 1.35 \cdot 10^{-3} \text{ sand} \quad [9]$$

$$303 \quad \theta_{4.2} = 0.08 + 4.01 \cdot 10^{-3} \text{ clay} - 2.93 \cdot 10^{-4} \text{ sand} \quad [10]$$

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

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

310 **2.5 Uncertainty analysis**

311 We applied a first order Taylor analysis to calculate the variance of θ_{FC} (i.e., $\theta_{2.0}$), θ_{PWP} (i.e., $\theta_{4.2}$),
312 elementary AWC and total AWC estimates. The estimates' variance is considered here as a proxy of
313 prediction uncertainty (Heuvelink et al., 1989). This method relies on the approximation of these

314 estimates (equations 1, 3, 9 and 10 respectively), and of the intermediate variables estimates (i.e. clay,
 315 silt, and sand with equation 5). Let Y be an estimate of a given soil property with

$$316 \quad Y = f(\mathbf{z})$$

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

$$321 \quad var(Y) \approx \sum_{j=1}^n \left\{ \sum_{k=1}^n \left\{ \tau_{jk} \sigma_j \sigma_k \frac{\delta f}{\delta z_j}(\boldsymbol{\mu}) \frac{\delta f}{\delta z_k}(\boldsymbol{\mu}) \right\} \right\} \quad [11]$$

322
 323 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
 324 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
 325 $f(\mathbf{z})$ around $\boldsymbol{\mu}$. These partial derivatives reflect the model (that is the f function) sensitivity to the input
 326 variables z_j and z_k . Hence, the variance of a soil property Y (equation 11) can be decomposed in the
 327 sum of different terms that consist on the multiplication of the squared model sensitivity by a variance
 328 or covariance of different input variables that represent their uncertainty (Dietze, 2017; Heuvelink et al.,
 329 1989).

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

$$338 \quad var(\theta) = \sum \tau_{Soil_j Soil_k} \sigma_{Soil_j} \sigma_{Soil_k} \frac{\delta f}{\delta Soil_j} \frac{\delta f}{\delta Soil_k} + \sum \tau_{\beta_j \beta_k} \sigma_{\beta_j} \sigma_{\beta_k} \frac{\delta f}{\delta \beta_j} \frac{\delta f}{\delta \beta_k} \quad [12]$$

339

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

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

353 **2.6 Evaluation of the functional DSM predictions**

354 Observed horizon data from the RMQS were used for independent evaluation of the predictions of clay,
355 sand, silt, and coarse elements. The measurements of GEVARNOVIA horizons were compared with the
356 predictions of volumetric soil moisture contents of the fine fraction (θ_{FC} , θ_{PWP}). This dataset is not
357 representative of all the pedoclimatic conditions in France and therefore the evaluation statistics may be
358 biased and not suitable for evaluating the whole France DSM approach. However, it is the best data
359 available on θ_{FC} and θ_{PWP} in France at the moment. For each independent observation (*i.e.* an observation
360 of a soil property on a given soil horizon), we calculated the weighted average of the GSM predictions
361 overlapping the horizon attached to the observation. This was done because the sample support differed
362 between the evaluation dataset and the predictions. Hence, a prediction was calculated as $\hat{y} =$
363 $\sum_{i=1}^n \frac{t_i}{t} \text{GSM}_i$, where \hat{y} is the estimated value, n is the number of GSM layers overlapping the horizon,
364 GSM_i is the GSM prediction for the layer i , t is the total thickness of the horizon (cm), and t_i is the
365 thickness (cm) of the GSM layer i that overlaps the predicted horizon. The evaluation statistics consisted
366 on the root mean square error (RMSE), coefficient of determination (R^2), Lin's concordance correlation

367 coefficient (Lin, 1989), and the bias, or mean error of prediction. The concordance evaluates both the
368 accuracy and the precision of the prediction, and it is defined as:

$$369 \quad \rho_C = \frac{2\rho\sigma_{\hat{y}}\sigma_y}{\sigma_{\hat{y}}^2 + \sigma_y^2 + (\bar{\hat{y}} - \bar{y})^2} \quad [13]$$

370 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,
371 and ρ the correlation between predicted and observed values. ρ_C can range between -1 and 1, and a value
372 closer to 1 indicates a better fit with the 45° line, or agreement between predictions and observations.

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

$$375 \quad PICP = \frac{\text{count}(LPL_i < y_i < UPL_i)}{n} \times 100 \quad [14]$$

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

385 **3. Results**

386 **3.1 Spatial structure of model residuals**

387 The linear model of coregionalization parameters are reported in the table 4. The range varied among
388 depths between 160 km at 15–30 cm depth and 252 km for the layer 60–100 cm. Supplemental material
389 S2 p.1a shows the estimates of the cross- and autocovariograms of the regression model residuals. The
390 regression residuals of the \ln -transformed variables were correlated both spatially (Supplemental
391 material S2 p.1a) and in the feature space (Supplemental material S2 p.1.b) across all depths. The coarse

392 elements residuals were not correlated or had a weak correlation ($r < 0.2$) with either of the alr-
393 transformed variables in any depth, neither untransformed (Supplemental material S2 p.1.b), or after log
394 transformation (data not shown). The empirical variograms of coarse elements residuals showed some
395 spatial correlation. However, as previous maps produced by regression-kriging of log-transformed
396 coarse elements were unsatisfactory, we decided to exclude the spatial correlation from the model
397 (Román Dobarco et al., 2018). All variograms appear somewhat erratic (Supplemental material S2
398 p.1.a), which is likely due to the presence of some clusters of points in the dataset (Marchant et al, 2013).

399 **3.2 Independent evaluation**

400 The evaluation statistics for the back-transformed clay, silt, and sand predictions for RMQS horizons
401 indicated that clay had lower R^2 and concordance coefficient than silt and sand (Table 4). The RMSE
402 increased following the trend clay < silt < sand (Table 5). The DSM predictions tended to underestimate
403 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
404 the predictions for sand had the smallest bias, of -3 g kg^{-1} . Small contents of clay, silt, and sand were
405 overestimated while high contents were often largely underestimated (Figure 4). However, the
406 prediction error for the particle size fractions did not show any pattern related to the average depth of
407 the RMQS horizons. Overall, many predictions were dispersed in the scatterplots and fell far from the
408 1:1 line (Figure 4) and predictions exhibited a RMSE up to 172 g kg^{-1} (Table 5). In comparison with the
409 particle size fractions the predictions for coarse elements had the lowest R^2 and concordance coefficient
410 and a RMSE of 21% (Table 5). The quantile regression forests model strongly underestimated relatively
411 stony soils ($> 25 \%$) and overestimated small contents of coarse elements (Figure 4.d). The prediction
412 error of RMQS horizons with small coarse elements content ($< 20 \%$) increased to some extent with the
413 average horizon depth (Figure 4.d).

414 The PICP suggested that the uncertainty associated to coarse elements and clay predictions was
415 underestimated (76% and 83 % respectively), but it was close to the expected value of 90 % for silt (86
416 %) and it was nearly perfect for sand, with a PICP of 90 % (Table 5).

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

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

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

431

432 **3.2 Spatial distribution of AWC**

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

441 **3.3 Contribution of different sources of uncertainty to the variance of soil moisture at field 442 capacity**

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

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

$$449 \quad var(\theta)_{soil} = \sum cov(Soil_j, Soil_k) \frac{\delta f}{\delta Soil_j} \frac{\delta f}{\delta Soil_k} = \sigma_{clay_{alr}}^2 \left(\frac{\delta f}{\delta clay_{alr}} \right)^2 + \sigma_{silt_{alr}}^2 \left(\frac{\delta f}{\delta silt_{alr}} \right)^2 +$$

$$450 \quad 2 \tau_{clay_{alr}, silt_{alr}} \sigma_{clay_{alr}} \sigma_{silt_{alr}} \frac{\delta f}{\delta clay_{alr}} \frac{\delta f}{\delta silt_{alr}} \quad [15]$$

451 where $var(\theta)_{soil}$ is the variance term of the soil moisture at field capacity associated to soil input
 452 variables, $\sigma_{clay_{alr}}^2$ and $\sigma_{silt_{alr}}^2$ 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
 453 respectively the sensitivities of the PTF (equation 9) to clay_{alr} and silt_{alr}, and $\tau_{clay_{alr}, silt_{alr}}$ is the is the
 454 correlation between clay_{alr} and silt_{alr}.

455 The sensitivity of elementary soil moisture at field capacity (cm³ cm⁻³) to clay_{alr} had higher values in
 456 areas with higher predicted values for clay_{alr} (Figure 8 and Figure 9.a). The sensitivity for silt_{alr} had
 457 negative values where predictions for silt_{alr} were positive, and positive values where silt_{alr} predictions
 458 were negative (Figure 8 and Figure 9.b). The terms resulting from the multiplication of clay_{alr} sensitivity
 459 and variance, and the interaction term had higher absolute values in similar regions, but counteracted
 460 each other because they had different signs. The total variance (Figure 9.f) was higher in zones with
 461 greater co-kringed residual variance (Figure 8.c and 8.d), and where the clay_{alr} term was greater, in
 462 absolute value, than the interaction term (Figure 9).

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

$$465 \quad var(\theta)_{\beta} = \sum cov(\beta_j, \beta_k) \frac{\delta f}{\delta \beta_j} \frac{\delta f}{\delta \beta_k} = \sigma_{\beta_0}^2 \left(\frac{\delta f}{\delta \beta_0} \right)^2 + \sigma_{\beta_1}^2 \left(\frac{\delta f}{\delta \beta_1} \right)^2 + \sigma_{\beta_2}^2 \left(\frac{\delta f}{\delta \beta_2} \right)^2 +$$

$$466 \quad 2 cov(\beta_0, \beta_1) \frac{\delta f}{\delta \beta_0} \frac{\delta f}{\delta \beta_1} + 2 cov(\beta_0, \beta_2) \frac{\delta f}{\delta \beta_0} \frac{\delta f}{\delta \beta_2} + 2 cov(\beta_1, \beta_2) \frac{\delta f}{\delta \beta_1} \frac{\delta f}{\delta \beta_2} \quad [16]$$

467 where $var(\theta)_{PTF}$ is the variance term of the soil moisture at field capacity associated to PTF
 468 coefficients, $\sigma_{\beta_0}^2$, $\sigma_{\beta_1}^2$, $\sigma_{\beta_2}^2$, $cov(\beta_0, \beta_1)$, $cov(\beta_0, \beta_2)$, $cov(\beta_1, \beta_2)$ are the elements of the variance-

469 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
470 the sensitivities of the PTF (equation 9) to the PTF coefficients.

471 In the case of the PTFs, the variance and covariance of the different terms are spatially constant (Table
472 3). Therefore, the spatial distribution of the uncertainty (and its different terms) depends on the
473 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
474 predictions (%), and $\frac{\delta f}{\delta \beta_2}$ the sand predictions (%). Therefore, the sensitivity of both PTFs to the clay and
475 sand coefficients is higher in areas where clay and sand contents are respectively higher (Figure 10).
476 Estimates of clay are higher in the northeast of France, and some areas of southwest. The predictions
477 for sand are higher in the centre (Massif Central) and other areas with predominance of sandy textures
478 (Les Vosges in the northeast, the Sologne in the centre, and les Landes in the southwest). The covariance
479 between the intercept and clay and sand ($cov(\beta_0, \beta_1)$ and $cov(\beta_0, \beta_2)$) is negative for both PTFs (Table
480 3), which resulted in negative values in their interaction terms (Figure 10.c and 10.d). The components
481 of clay and sand coefficients and their interaction were positive (Figure 10.e, 10.f, and 10.g). The
482 uncertainty due to the PTFs' coefficients had higher values in those areas where the sensitivity to the
483 sand coefficient was greatest (Figure 10.b and Figure 10.h). The maximum value of the variance
484 associated to the PTFs' coefficients is three times smaller than the maximum value of the variance
485 associated to soil input variables (0.00012 vs 0.004).

486 **3.4 Contribution of different sources of uncertainty to the variance of elementary AWC**

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

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

499 Across the six GSM layers, the standard deviation of elementary AWC ($\text{cm}^3 \text{ cm}^{-3}$) was higher near
500 mountainous regions in the south and southeast, and some areas in the west and east of France
501 (Supplemental material S2 p.5). The standard deviation of elementary AWC increased with soil depth
502 (Supplemental material S2 p.6). The main contribution (%) to the variance in these areas corresponded
503 to the term associated to the coarse elements (red areas in Supplemental material S2 p.6). In regions
504 with lower variance of elementary AWC and mostly low elementary AWC, the main sources of
505 uncertainty for the top soil layers were the PTFs (southwest and centre) (blue in Supplemental material
506 S2 p.6) and soil texture (north) (green in Supplemental material S2 p.6). The contribution (%) of coarse
507 elements to the variance in the southwest and centre increased for the 30–60 cm, 60–100 cm, and 100–
508 200 cm while the contribution (%) of the PTFs decreased (Supplemental material S2 p.7). Conversely,
509 the contribution (%) of coarse elements to the variance of elementary AWC in the north decreased with
510 depth.

511 **4. Discussion**

512 **4.1 Previous estimates of AWC in metropolitan France**

513 This study presented the first map of AWC for metropolitan France that provides uncertainty estimates
514 following *GlobalSoilMap* specifications (Arrouays et al., 2014). An advantage of the produced maps of
515 θ_{FC} , θ_{PWP} and AWC is that the potential end users can incorporate estimated uncertainties into
516 ecological and agricultural modelling and perform uncertainty propagation analysis or sensitivity
517 analysis. It is important to stress that since the role of soil thickness was not included in the uncertainty
518 analysis for total AWC, the calculated variance should be considered a conservative uncertainty
519 estimate. We have improved the precision of the estimates of AWC for mainland France compared to a
520 previous map produced by Al Majou et al., (2008b). Al Majou et al., (2008b) applied their PTFs with
521 information on horizon type, horizon thickness, texture, and bulk density provided by the descriptions

522 of soil typological unit (STU) from the 1:1 000 000 Soil Geographical Database of France (King et al.,
523 1995). In these previous studies, the AWC by soil mapping unit was calculated based on the proportion
524 of STU present in each soil mapping unit (Al Majou et al., 2008b). The predictions of this study are
525 provided at a specific resolution (90 m) whereas Al Majou et al. (2008b) provided AWC values by
526 polygons. The spatial patterns in both maps are similar overall, although our estimates are smaller than
527 the AWC predicted by Al Majou et al. (2008b) in the southeast and northeast France, but higher in the
528 north.

529 Piedallu et al. (2011) also predicted AWC for metropolitan France by first estimating the total AWC of
530 120,902 soil profiles from the French National Forest Inventory with class-PTFs (i.e., average values of
531 AWC are assigned by classes defined by texture and horizon type) developed by Al Majou et al. (2008b),
532 and then extrapolating spatially with ordinary kriging. The general pattern of AWC was similar between
533 the map by Piedallu et al. (2011) and this study. Piedallu et al. (2001) used texture class estimated by
534 the surveyors in the field, which may enlarge the measurement error compared to laboratory analysis of
535 particle size distribution. However, the high density of observations across metropolitan France may
536 compensate partly the error in the input texture data resulting in a good description of the spatial pattern
537 of AWC. The approach followed by Piedallu et al. (2011) makes the maps suitable for forested areas,
538 since the soil profiles were located in forests and the greater distance between the locations in
539 agricultural land to the soil profiles involves larger uncertainty at these predicted locations. Conversely,
540 the data on particle size fractions distribution and coarse elements used to calibrate our models were not
541 distributed homogeneously across mainland France (Figure 2). In some areas, such as in the south near
542 the Massif Central and in the southwestern coast, the lower density of observations resulted in higher
543 uncertainty of clay_{air} and silt_{air} (Figure 8) that propagated into the uncertainty of θ_{FC} and θ_{PWP} (Figure
544 9). However, particle size fraction was not the main source of uncertainty for elementary AWC in these
545 areas (Figure 11).

546 **4.2 General approach for mapping AWC**

547 There are multiple modelling trajectories for mapping AWC depending on at which step are applied the
548 PTFs and the spatial extrapolation (Styc and Lagacherie, 2018). We first modelled spatially the AWC

549 input variables, and when possible, jointly, as for particle size distribution, with the aim of capturing
550 their spatial patterns and the relationships between the soil forming factors and the soil properties, thus
551 improving the accuracy of AWC predictions. This approach is similar to that followed by Ugbaje and
552 Reuter (2013) for *GlobalSoilMap* AWC for Nigeria and Tóth et al. (2017) for the European Soil
553 Hydraulic Database (EU-SoilHydroGrids). The methodology chosen for predicting AWC has the
554 advantage that 1) the results can be easily updated when more accurate predictions for the soil input
555 properties or more reliable PTFs are available and 2) some input soil properties (here particle size
556 distribution) come as a side product of AWC estimation.

557 The predictions of particle size distribution had similar R^2 (Table 5) than previous *GlobalSoilMap*
558 products for France (Mulder et al., 2016), which had R^2 between 0.25 – 0.44 for clay, 0.21 – 0.42 for
559 silt, and 0.19 – 0.33 for sand. The concordance coefficients by Mulder et al. (2016) ranged between 0.34
560 – 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,
561 and 0.66 of this study (Table 5). This is not surprising because both studies shared the data from the
562 IGCS in the calibration dataset (although Mulder et al. (2016) merged this dataset with the RMQS one),
563 some of the environmental covariates, and applied the cubist algorithm. However, predicting particle
564 size distribution with regression-cokriging allowed us to account for the spatial correlation between
565 particle size fractions of the same GSM layer, hence quantifying the uncertainty of AWC more
566 accurately. On the other hand, the results for coarse elements were less accurate and precise in this study,
567 with $R^2 = 0.14$ and a concordance coefficient of 0.26 compared to a R^2 between 0.17 – 0.28 and a
568 concordance coefficient between 0.30 – 0.46 by Mulder et al. (2016) for log transformed coarse
569 elements. One possible explanation for this slight performance discrepancy could be, at least for topsoil
570 layers, the better precision of RMQS observations (compared to the IGCS), which was not used for
571 training the models in the present study. Mapping coarse elements is specially challenging. The
572 calibration and evaluation data (volume of coarse elements) was estimated visually by the soil surveyors.
573 More precise methods for measuring the volume of coarse elements are very time consuming and can
574 only be applied in a limited number of sites in national soil monitoring surveys (Jolivet et al., 2018).
575 More generally, we were not able to capture the spatial patterns of distribution of coarse elements,
576 especially in deep layers. Indeed, the input data itself may be partly biased, as coarse elements in surface

577 layers are much more easily estimated than in deeper ones, especially when doing observations by
578 coring. It is also possible that the chosen covariates did not represent well the processes driving the
579 distribution of coarse elements, or these were not accurate enough (e.g., the scale of parent material was
580 1:1 000 000). Nevertheless, next versions of *GlobalSoilMap* with more accurate predictions on coarse
581 elements can be incorporated in the proposed AWC modelling framework for reducing the prediction
582 error and uncertainty of AWC predictions.

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

598 **4.3 Importance of different sources of uncertainty**

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

604 100 and 100–200 cm) when excluding areas based on the estimated soil depth by Lacoste et al. (2016).
605 In areas of very sandy and clayey textures the PTFs contributed more to AWC uncertainty. This
606 highlights the importance of developing more reliable PTFs for very coarse and very clayey soils, that
607 although occurring with less frequency, can occupy extensive areas in France (e.g., sandy soils in the
608 Landes of Gascony).

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

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

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

638 **4.4 Limitations of the produced maps and future directions**

639 The predictive ability of the spatial predictions for θ_{FC} and θ_{PWP} decreased considerably in comparison
640 to applying the PTFs to measured horizon data (Table 6). The change in R^2 for θ_{FC} and θ_{PWP} was of
641 $\Delta R^2 = -0.33$, which in relative terms consisted in a 61 % and 53% reduction compared to the R^2 of the
642 PTFs. The RMSE increased in $\Delta \text{RMSE} = 0.013 \text{ cm}^3 \text{ cm}^{-3}$ for θ_{FC} and $\Delta \text{RMSE} = 0.015 \text{ cm}^3 \text{ cm}^{-3}$ for
643 θ_{PWP} that suppose a relative increase of 25 % and 36 % respectively compared to the RMSE of the
644 PTFs. We acknowledge that we could evaluate the spatial predictions of θ_{FC} and θ_{PWP} , which is not
645 always possible for soil hydraulic properties due to the lack of georeferenced observations. However,
646 the evaluation statistics are of limited validity at national extent due to the sample size, distribution and
647 representativeness of pedoclimatic conditions of the evaluation dataset (Table 6 and Figure 3). Ongoing
648 work for France is currently being carried out, for gathering unbiased references of soil water content at
649 different potentials, using the French soil monitoring network (RMQS, Jolivet et al. 2006). The larger
650 prediction error for θ_{FC} and θ_{PWP} spatial predictions were associated to an overestimation of soil
651 moisture for coarse texture, and an underestimation for fine and very fine texture classes (Figure 6). This
652 is likely related to the inability to predict the spatial distribution of clay content in some regions, and in
653 particular, to the underestimation of high clay contents (Table 5 and Figure 4). It is also likely that
654 important drivers such as mineralogical composition of clay and sand are also missing in the PTFs. It is
655 possible that we need more accurate covariates for capturing the processes driving the spatial distribution
656 of clay content and mineralogy assuming the latter could be incorporated into PTFs (e.g., soil geology
657 map, gamma-ray spectrometry). Another source of error is that the soil profiles were not distributed

658 evenly throughout the study area, but were clustered instead. Clustering led to some artifacts during the
659 cokriging step (Padarian et al., 2014; Marchant et al., 2013) despite the benefit of including the
660 correlation between soil particle fractions within a GSM layer. We did not include the spatial correlation
661 among soil properties of multiple layers, which may also have improved the estimates of AWC and its
662 uncertainty (Heuvelink et al., 2016). Angelini et al. (2017) applied structural equation modelling (SEM)
663 to DSM, incorporating pedological knowledge of the interrelations among soil properties and soil
664 processes, and predicting several soil properties at multiple layers simultaneously. Recently, Angelini
665 and Heuvelink (2018) expanded the SEM for soil properties with a geostatistical approach, including
666 the spatial correlation of the model residuals. This methodology could be interesting for mapping AWC,
667 as it would incorporate the interrelations between all soil properties defining AWC (particle size
668 distribution, bulk density, soil organic carbon, soil depth, coarse elements, etc.), within and between soil
669 layers.

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

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

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

690 **4.5 Conclusions**

691 This study presented spatial predictions of AWC for mainland France following *GlobalSoilMap*
692 specifications to a maximum depth of 2 m. We incorporated two sources of error (spatial estimates of
693 soil input properties and PTFs' coefficients) in the uncertainty analysis carried out with first order Taylor
694 series analysis. The continuous computing and statistical developments will allow improving the
695 quantification of AWC uncertainty with a feasible computing time in future studies, for example with
696 stochastic simulations or Bayesian simulations (Poggio et al., 2016; Beguin et al., 2017; Huang et al.,
697 2017), which would ultimately allow the characterization of the probability distribution of AWC
698 estimates on a pixel base. Overall, this study provides the first estimate of AWC uncertainty, by soil
699 layers or at the whole soil profile level, for mainland France that can be incorporated into ecological and
700 agricultural modelling. The end-users of the AWC maps will be essential for evaluating the usefulness
701 of the maps for assessing the provision of ecosystem services and modelling ecological processes, and
702 to indicate limitations in their exploitation due to the AWC prediction uncertainty.

703 The reproducible modelling framework allows replacing each component of the AWC calculation
704 (PTFs, soil input properties) when more accurate maps are developed thanks to the selection of
705 covariates that characterize better the processes driving the spatial distribution of soil input properties,
706 the implementation of new regression algorithms, and the acquisition of new soil profile data. Indeed,
707 key aspects for improving AWC estimates are expanding the calibration data on coarse elements
708 (especially for deeper layers), incorporating the nature and soil hydraulic properties of coarse elements
709 into the calculation, improving the estimates of clay and sand and their mineralogy, and improving the
710 prediction of soil depth. According to our initial objectives, a major output of this study is the method
711 we developed to estimate the uncertainty of AWC predictions by taking into account both uncertainties

712 linked to the soil input variables and to the PTFs' coefficients. The second major output is the prediction
713 of AWC and its uncertainty for mainland France according to international specifications, which
714 provides this country a nearly complete set of the mandatory attributes to be predicted according to the
715 GlobalSoilMap initiative.

716

717 **Acknowledgements**

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

726

727 **References**

- 728 Achache, J., Debeglia, N., Grandjean, G., Guillen, A., Le Bel, L., Ledru, P., et al., 1997.
729 GEOFRANCE 3D: l'imagerie géologique et géophysique 3D du sous-sol de la France. *Mém.*
730 *Société Géologique Fr.* 53–71.
- 731 Aitchison, J., 1982. The statistical analysis of compositional data. *J. R. Stat. Soc. Ser. B Methodol.* 44
732 (2), 139–177.
- 733 Akpa, S.I.C., Odeh, I.O.A., Bishop, T.F.A., Hartemink, A.E., 2014. Digital mapping of soil particle-
734 size fractions for Nigeria. *Soil Sci. Soc. Am. J.* 78 (6), 1953–1966.
735 doi:10.2136/sssaj2014.05.0202.
- 736 Al Majou, H., Bruand, A., Duval, O. 2008a. Use of in situ volumetric water content at field capacity to
737 improve prediction of soil water retention properties. *Can. J. Soil Sci.*, 88(4), 533–541.
- 738 Al Majou, H., Bruand, A., Duval, O., Le Bas, C., Vautier, A., 2008b. Prediction of soil water retention
739 properties after stratification by combining texture, bulk density and the type of horizon. *Soil*
740 *Use Manage.* 24(4), 383–391. doi:10.1111/j.1475-2743.2008.00180.x.
- 741 Angelini, M.E., Heuvelink, G.B., 2018. Including spatial correlation in structural equation modelling of
742 soil properties. *Spat. Stat.*, 25, 35–51. doi.org/10.1016/j.spasta.2018.04.003.

- 743 Angelini, M.E., Heuvelink, G.B., Kempen, B., 2017. Multivariate mapping of soil with structural
744 equation modelling. *Eur. J. Soil Sci.*, 68, 575–591. doi:10.1111/ejss.12446.
- 745 Arnold, J. G., and J. R. Williams. 1987. Validation of SWRRB: Simulator for water resources in rural
746 basins. *J. Water Resour. Plan. Manage. ASCE*113(2): 243-256.
- 747 Arnold, J. G., and N. Fohrer. 2005. SWAT2000: Current capabilities and research opportunities in
748 applied watershed modeling. *Hydrol. Process.* 19(3): 563-572.
- 749 Arrouays, D., Hardy, R., Schnebelen, N., Le Bas, C., Eimberck, M., Roque, J., et al., 2004. Le
750 programme inventaire gestion et conservation des sols en France. *Etude Gest. Sols*, 11 (3),
751 187–197. <https://prodinra.inra.fr/record/75841>.
- 752 Arrouays, D., Grundy, M.G., Hartemink, A.E., Hempel, J.W., Heuvelink, G.B.M., Hong, S.Y., et al.,
753 2014. GlobalSoilMap: toward a fine-resolution global grid of soil properties. *Adv. Agron.* 125
754 (125), 93–134. doi.org/10.1016/B978-0-12-800137-0.00003-0.
- 755 Arrouays, D., Saby, N.P.A., Thioulouse, J., Jolivet, C., Boulonne, L., Ratié, C., 2011. Large trends in
756 French topsoil characteristics are revealed by spatially constrained multivariate analysis.
757 *Geoderma* 161, 107–114. doi.org/10.1016/j.geoderma.2010.12.002.
- 758 Ballabio, C., Panagos, P., Monatanarella, L., 2016. Mapping topsoil physical properties at European
759 scale using the LUCAS database. *Geoderma* 261, 110–123.
760 doi.org/10.1016/j.geoderma.2015.07.006.
- 761 Beguin, J., Fuglstad, G.A., Mansuy, N., Paré, D., 2017. Predicting soil properties in the Canadian
762 boreal forest with limited data: Comparison of spatial and non-spatial statistical approaches.
763 *Geoderma*, 306, 195–205. doi.org/10.1016/j.geoderma.2017.06.016.
- 764 Bertran, P., Liard, M., Sitzia, L., Tissoux, H., 2016. A map of Pleistocene aeolian deposits in Western
765 Europe, with special emphasis on France. *J. Quaternary Sci.*, 31: 844-856 e2909.
766 doi:10.1002/jqs.2909.
- 767 Bishop, T.F.A., McBratney, A.B., Laslett, G.M., 1999. Modelling soil attribute depth functions with
768 equal-area quadratic smoothing splines. *Geoderma*, 91 (1–2), 27–45. doi.org/10.1016/S0016-
769 7061(99)00003-8.
- 770 Bouma, J., 1989. Using soil survey data for quantitative land evaluation. *Adv. Soil Sci.* 9, 177–213.
771 doi.org/10.1007/978-1-4612-3532-3_4.
- 772 Breiman, L., 2001. Random forests. *Machine learning*, 45(1), 5–32.
773 doi.org/10.1023/A:1010933404324.
- 774 Brisson, N., Mary, B., Ripoche, D., Jeuffroy, M. H., Ruget, F., Nicoullaud, B., et al., 1998. STICS: a
775 generic model for the simulation of crops and their water and nitrogen balances. I. Theory and
776 parameterization applied to wheat and corn. *Agron.*, 18(5–6), 311–346.
- 777 Bruand, A., Pérez Fernández, P., Duval, O., 2003. Use of class pedotransfer functions based on texture
778 and bulk density of clods to generate water retention curves. *Soil Use Manage.* 19, 232–242.
- 779 Bruand, A., Duval, O., Cousin, I., 2004. Estimation des propriétés de rétention en eau des sols à partir
780 de la base de données SOLHYDRO: Une première proposition combinant le type d'horizon,
781 sa texture et sa densité apparente. *Étude et Gestion des Sols*, 11 (3), 323 – 332. Brus, D.J.,
782 Kempen, B., Heuvelink, G.B.M., 2011. Sampling for validation of digital soil maps. *Eur. J.*
783 *Soil Sci.* 62 (3), 394–407. doi:10.1111/j.1365-2389.2011.01364.x.
- 784 Buchanan, S., Triantafilis, J., Odeh, I.O.A., Subansinghe, R., 2012. Digital soil mapping of
785 compositional particle-size fractions using proximal and remotely sensed ancillary data.
786 *Geophysics* 77:WB201–WB211. doi:10.1190/geo2012-0053.1.

- 787 Carré, F., McBratney, A.B., Mayr, T., Montanarella, L., 2007. Digital soil assessments: Beyond DSM.
788 *Geoderma*, 142 (1–2), 69–79. doi.org/10.1016/j.geoderma.2007.08.015.
- 789 Cerdan, O., Govers, G., Le Bissonnais, Y., Van Oost, K., Poesen, J., Saby, N.P.A., et al., 2010. Rates
790 and spatial variations of soil erosion in Europe: A study based on erosion plot data.
791 *Geomorphol.*, 122 (1–2), 167–177. doi.org/10.1016/j.geomorph.2010.06.011.
- 792 Ciampalini, R., Martin, M.P., Saby, N.P.A., Richer-de-Forges, A.C., Arrouays, D., Nehlig, P., Martelet,
793 G., 2014. Soil texture GlobalSoilMap products for the French region “Centre”, in: Arrouays,
794 D., McKenzie, N., Hempel, J., Richer de Forges, A.C., McBratney, A. (Eds.), *Globalsoilmap:*
795 *Basis of the Global Spatial Soil Information System*. CRC Press, pp. 121–126.
- 796 Coleman, K., Jenkinson, D., Crocker, G.J., Grace, P.R., Klir, J., Körschens, M., et al., 1997.
797 *Simulating trends in soil organic carbon in long-term experiments using RothC-26.3.*
798 *Geoderma*, 81(1–2), 29–44. doi.org/10.1016/S0016-7061(97)00079-7.
- 799 Cousin, I., Labidi, A., Le Bris, X., Champolivier, L., Bernicot, M.H., Bessard Duparc, P., Bouthier, A.,
800 2016. Evaluating the Available Water Content of soils at large scale to improve the estimation
801 of soil services. 2nd Global Soil Security Conference, 05–06/12/2016.
- 802 Chrétien, J., 1971. Essai de caractérisation des sables en tant que squelette minéral du sol. *Ann.*
803 *Agron.*, 22 (6), 615–654.
- 804 Didan, K., 2015. MOD13A1 MODIS/Terra Vegetation Indices 16-Day L3 Global 500m SIN Grid
805 V006. NASA EOSDIS LP DAAC. doi: 10.5067/MODIS/MOD13A1.006.
- 806 Dietze, M.C., 2017. *Ecological forecasting*. Princeton University Press. Princeton, New Jersey, USA.
- 807 Dominati, E., Patterson, M., Mackay, A., 2010. A framework for classifying and quantifying the
808 natural capital and ecosystem services of soils. *Ecol. Econ.*, 69(9), 1858–1868.
809 doi.org/10.1016/j.ecolecon.2010.05.002.
- 810 Durand, Y., Brun, E., Mérindol, L., Guyomarc’h, G., Lesaffre, B., Martin, E., 1993. A meteorological
811 estimation of relevant parameters for snow models. *Ann. Glaciol.*, 18, 65–71.
812 <https://doi.org/10.3189/S0260305500011277>.
- 813 Durand, Y., Laternser, M., Giraud, G., Etchevers, P., Lesaffre, B., Mérindol, L., 2009: Reanalysis of 44
814 Yr of Climate in the French Alps (1958–2002): Methodology, Model Validation, Climatology,
815 and Trends for Air Temperature and Precipitation. *J. Appl. Meteor. Climatol.*, 48, 429–449,
816 <https://doi.org/10.1175/2008JAMC1808.1>.
- 817 Efron, B., Tibshirani, R.J., 1993. *An introduction to the bootstrap*. Monogr. Stat. Appl. Probab. 57.
818 CRC Press, Boca Raton, FL.
- 819 European Environmental Agency, 2007. CLC2006 technical guidelines. Technical Report No.
820 17/2007, European Environmental Agency, Copenhagen, Denmark.
- 821 Faroux, S., Kaptué Tchuenté, A.T., Roujean, J.L., Masson, V., Martin, E., Moigne, P.L., 2013.
822 ECOCLIMAP-II/Europe: A twofold database of ecosystems and surface parameters at 1 km
823 resolution based on satellite information for use in land surface, meteorological and climate
824 models. *Geosci. Model Dev.*, 6(2), 563–582. doi.org/10.5194/gmd-6-563-2013.
- 825 Folberth, C., Skalský, R., Moltchanova, E., Balkovič, J., Azevedo, L. B., Obersteiner, M., Van Der
826 Velde, M., 2016. Uncertainty in soil data can outweigh climate impact signals in global crop
827 yield simulations. *Nat. Commun.*, 7, 11872. doi.org/10.1038/ncomms11872.
- 828 Garrett, R.G., 2015. rgr: applied geochemistry EDA. R Package Version 1.1.10 ([http:// CRAN.R-](http://CRAN.R-project.org/package=rgr)
829 [project.org/package=rgr](http://CRAN.R-project.org/package=rgr)).

- 830 GDAL/OGR contributors (2015). GDAL/OGR Geospatial Data Abstraction software Library. Open
831 Source Geospatial Foundation. URL <http://gdal.org>.
- 832 Gis Sol, 2011. L'état des sols de France. Groupement d'intérêt scientifique sur les sols (188 pp.).
- 833 Goulard, M., Voltz, M., 1992. Linear coregionalization model: Tools for estimation and choice of
834 cross-variogram matrix. *Math. Geol.* 24: 269. <https://doi.org/10.1007/BF00893750>.
- 835 GRASS Development Team, 2016. Geographic Resources Analysis Support System (GRASS)
836 Software, Version 7.0. Open Source Geospatial Foundation. Electronic document:
837 <http://grass.osgeo.org>.
- 838 Hengl, T., Heuvelink, G.B.M., Rossiter, D.G., 2007. About regression-kriging: from equations to case
839 studies. *Comput. Geosci.* 33 (10), 1301–1315. doi.org/10.1016/j.cageo.2007.05.001.
- 840 Heuvelink, G.B.M., Burrough, P.A., Stein, A., 1989. Propagation of errors in spatial modelling with
841 GIS. *Int. J. Geogr. Inf. Syst.* 3, 303–322.
- 842 Heuvelink, G.B.M., Kros, J., Reinds, G.J., De Vries, W., 2016. Geostatistical prediction and simulation
843 of European soil property maps. *Geoderma Reg.*, 7(2), 201–215.
844 doi.org/10.1016/j.geodrs.2016.04.002.
- 845 Heuvelink, G.B.M., Pebesma, W. J., 1999. Spatial aggregation and soil process modelling. *Geoderma*,
846 89: 47–65. [doi.org/10.1016/S0016-7061\(98\)00077-9](https://doi.org/10.1016/S0016-7061(98)00077-9).
- 847 Hong, S. Y., Minasny, B., Han, K. H., Kim, Y., Lee, K., 2013. Predicting and mapping soil available
848 water capacity in Korea. *PeerJ*, 1:e71 <https://doi.org/10.7717/peerj.71>.
- 849 Huang, J., Malone, B.P., Minasny, B., McBratney, A.B., Triantafilis, J., 2017. Evaluating a Bayesian
850 modelling approach (INLA-SPDE) for environmental mapping. *Sci. Total Environ.*, 609, 621–
851 632. doi.org/10.1016/j.scitotenv.2017.07.201.
- 852 Huang, J., Subasinghe, R., Triantafilis, J., 2014. Mapping particle-size fractions as a composition
853 using additive log-ratio transformation and ancillary data. *Soil Sci. Soc. Am. J.*, 78(6), 1967–
854 1976. [doi:10.2136/sssaj2014.05.0215](https://doi.org/10.2136/sssaj2014.05.0215).
- 855 Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the
856 radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens.*
857 *Environ.*, 83(1–2), 195–213. DOI: [10.1016/S0034-4257\(02\)00096-2](https://doi.org/10.1016/S0034-4257(02)00096-2).
- 858 Institut National de l'Information Géographique et Forestière, 2012 Base de Données Forêt. In : BD
859 Forêt ®.
- 860 Jolivet, C., Arrouays, D., Boulonne, L., Ratié, C., Saby, N., 2006. Le Réseau de Mesures de la Qualité
861 des Sols de France (RMQS). État d'avancement et premiers résultats. *Étude Gest. Sols* 13 (3),
862 149–164.
- 863 Jolivet C., Almeida-Falcon J.L., Berché P., Boulonne L., Fontaine M., Gouny L., et al., 2018. Manuel
864 du Réseau de mesures de la qualité des sols. RMQS2: deuxième campagne métropolitaine,
865 2016–2027, Version 3, INRA, US 1106 InfoSol, Orléans, France.
- 866 King, D., Burrill, A., Daroussin, J., Le Bas, C., Tavernier, R., Van Ranst, E., 1995. The EU soil
867 geographical database, in: King, D., Jones, R.J.A. Jones, Thomasson A.J. (Eds), *European*
868 *land information systems for agro-environmental monitoring*. Joint Research Centre, Ispra,
869 Italy, pp. 43–60.
- 870 Lacoste, M., Mulder, V.L., Richer-de-Forges, A.C., Martin, M.P., Arrouays, D., 2016. Evaluating large-
871 extent spatial modeling approaches: A case study for soil depth for France. *Geoderma Reg.*,
872 7(2), 137–152. doi.org/10.1016/j.geodrs.2016.02.006.

- 873 Lagacherie, P., McBratney, A.B., 2006. Chapter 1 Spatial Soil Information Systems and Spatial Soil
874 Inference Systems: Perspectives for Digital Soil Mapping, in: Lagacherie, P., McBratney,
875 A.B., Voltz, M. (Eds.), *Developments in Soil Science*, 31. Elsevier, Amsterdam, The
876 Netherlands, pp. 3–22. doi.org/10.1016/S0166-2481(06)31001-X.
- 877 Lark, R.M., Bishop, T.F.A., 2007. Cokriging particle size fractions of the soil. *Eur. J. Soil Sci.* 58 (3),
878 763–774. doi.org/10.1111/j.1365-2389.2006.00866.x.
- 879 Laroche, B., Richer-de-Forges, A., Leménager, S., Arrouays, D., Schnebelen, N., Eimberck, M., et al.,
880 2014. Le programme Inventaire Gestion Conservation des Sols de France : volet Référentiel
881 Régional Pédologique. *Étude Gest. Sols*, 21 (1), 25–36. https://prodinra.inra.fr/record/282113.
- 882 Leenaars, J.G.B., Claessens, L., Heuvelink, G.B.M., Hengl, T., Ruiperez González, M., van Bussel,
883 L.G.J., Guilpart, N., Yang, H., Cassman, K.G., 2018. Mapping rootable depth and root zone
884 plant-available water holding capacity of the soil of sub-Saharan Africa, *Geoderma*, 324, 18–
885 36. doi.org/10.1016/j.geoderma.2018.02.046.
- 886 Lin L.I., 1989. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 45, 255–
887 268. DOI : 10.2307/2532051.
- 888 Marchant, B., Viscarra Rossel, R., Webster, R., 2013. Fluctuations in method-of-moments variograms
889 caused by clustered sampling and their elimination by declustering and residual maximum
890 likelihood estimation. *Eur. J. Soil Sci.*, 64: 401–409. doi:10.1111/ejss.12029.
- 891 Marchetti, R., Donatelli, M. and Spallacci, P., 1997. Testing denitrification functions of dynamic crop
892 models. *J. Environ. Qual.*, 26(2), pp. 394–401.
893 doi:10.2134/jeq1997.00472425002600020009x.
- 894 Mardhel, V., Gravier, A., 2005. Carte de vulnérabilité intrinsèque simplifiée des eaux souterraines du
895 Bassin Seine-Normandie (Rapport BRGM/RP-54148-FR).
- 896 McBratney, A.B., Santos, M.L.M., Minasny, B., 2003. On digital soil mapping. *Geoderma* 117 (1–2),
897 3–52. doi.org/10.1016/S0016-7061(03)00223-4.
- 898 Meinshausen, N., 2006. Quantile regression forests. *J. Mach. Learn. Res.*, 7(Jun), 983–999.
- 899 Minasny, B., McBratney, A.B., Bristow, K.L., 1999. Comparison of different approaches to the
900 development of pedotransfer functions for water-retention curves. *Geoderma* 93(3–4), 225–
901 253. doi.org/10.1016/S0016-7061(99)00061-0.
- 902 Mulder, V.L., Lacoste, M., Richer-de-Forges, A.C., Arrouays, D., 2016. GlobalSoilMap France: High-
903 resolution spatial modelling the soils of France up to two meter depth. *Sci. Total Environ.*, 73,
904 1352–1369. doi.org/10.1016/j.scitotenv.2016.07.066.
- 905 Nemes, A., Schaap, M.G., Wösten, J.H.M., 2003. Functional evaluation of pedotransfer functions
906 derived from different scales of data collection. *Soil Sci. Soc. Am. J.* 67(4), 1093–1102.
907 doi:10.2136/sssaj2003.1093.
- 908 Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J., 1994. Spatial prediction of soil properties from
909 landform attributes derived from a digital elevation model. *Geoderma* 63(3–4), 197–214.
- 910 Odeh, I.O.A., Todd, A.J., Triantafilis, J., 2003. Spatial prediction of soil particle-size fractions as
911 compositional data. *Soil Sci.* 168 (7), 501–515. doi: 10.1097/01.ss.0000080335.10341.23.
- 912 O'Leary, G.J., Li Liu, D., Ma, Y., Li, F.Y., McCaskill, M., Conyers, M., et al., 2016. Modelling soil
913 organic carbon 1. Performance of APSIM crop and pasture modules against long-term
914 experimental data. *Geoderma*, 264, 227–237. doi.org/10.1016/j.geoderma.2015.11.004.

- 915 Padarian, J., Minasny, B., McBratney, A.B., Dalglish, N., 2014. Predicting and mapping the soil
 916 available water capacity of Australian wheatbelt. *Geoderma Reg.*, 2–3, 110–118.
 917 doi.org/10.1016/j.geodrs.2014.09.005.
- 918 Parton, W.J., Schimel, D.S., Cole, C.V., Ojima, D.S., 1987. Analysis of factors controlling soil organic
 919 matter levels in Great Plains Grasslands 1. *Soil Sci. Soc. Am. J.*, 51(5), 1173–1179.
 920 doi:10.2136/sssaj1987.03615995005100050015x.
- 921 Piedallu, C., Gegout, J.C., Bruand, A., Seynave, I., 2011. Mapping soil water holding capacity over
 922 large areas to predict potential production of forest stands. *Geoderma* 160(3–4), 355–366.
 923 doi.org/10.1016/j.geoderma.2010.10.004.
- 924 Poggio, L., Gimona, A., 2017. 3D mapping of soil texture in Scotland. *Geoderma Reg.*, 9, 5–16.
 925 doi.org/10.1016/j.geodrs.2016.11.003.
- 926 Poggio, L., Gimona, A., Brown, I., Castellazzi, M., 2010. Soil available water capacity interpolation
 927 and spatial uncertainty modelling at multiple geographical extents. *Geoderma*, 160(2), 175–
 928 188. doi.org/10.1016/j.geoderma.2010.09.015.
- 929 Poggio, L., Gimona, A., Spezia, L., Brewer, M.J., 2016. Bayesian spatial modelling of soil properties
 930 and their uncertainty: The example of soil organic matter in Scotland using R-INLA.
 931 *Geoderma*, 277, 69–82. doi.org/10.1016/j.geoderma.2016.04.026.
- 932 Quinlan, J.R., 1992. Learning with continuous classes. *Proceedings of the 5th Australian Joint*
 933 *Conference On Artificial Intelligence*, pp. 343–348.
- 934 Quintana-Seguí, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., et al., 2008.
 935 Analysis of near-surface atmospheric variables: validation of the SAFRAN analysis over
 936 France. *J. Appl. Meteorol. Climatol.* 47, 92–107. doi.org/10.1175/2007JAMC1636.1.
- 937 R Core Team (2015). R: A language and environment for statistical computing. R Foundation for
 938 Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- 939 Román Dobarco, M., Cousin, I., Le Bas, C., Martin, M.P., 2019. Pedotransfer functions for predicting
 940 available water capacity in French soils, their applicability domain and associated uncertainty.
 941 *Geoderma*, 336, 81–95. doi.org/10.1016/j.geoderma.2018.08.022.
- 942 Román Dobarco, M., Saby, N.P.A., Bourennane, H., Cousin, I., Arrouays, D., Martin, M.P., 2018.
 943 Digital soil mapping and uncertainty propagation of available water capacity for metropolitan
 944 France. *TERRAENVISION Abstracts Vol. 1, TNV2018-ECO1-396, TERRAenVISION*, 1
 945 February 2018, Barcelona, Spain.
- 946 Sanchez, P.A., Ahamed, S., Carré, F., Hartemink, A.E., Hempel, J., Huising, J., et al., 2009. Digital soil
 947 map of the world. *Science*, 325(5941), 680–681. doi: 10.1126/science.1175084.
- 948 Schwärzel, K., Feger, K.-H., Häntzschel, J., Menzer, A., Spank, U., Clausnitzer, F., Köstner, B.,
 949 Bernhofer, C., 2009. A novel approach in model-based mapping of soil water conditions at
 950 forest sites. *For. Ecol. Manage.* 258 (10), 2163–2174.
- 951 Shrestha, D.L., Solomatine, D.P., 2006. Machine learning approaches for estimation of prediction
 952 interval for the model output. *Neural Netw.* 19, 225–235.
- 953 Silva, B.M., Silva É.A., Oliveira, G.C., Ferreira, M.M., Serafim, M.E., 2014. Plant-available soil water
 954 capacity: estimation methods and implications. *Rev. Bras. Cienc. Solo* 38, 464–475.
- 955 Styc, Q., Lagacherie, P., 2018. What is the best spatial soil inference system for mapping Available
 956 water capacity? A test in Languedoc-Roussillon (France). *Geophys. Res. Abstr.*, 20,
 957 EGU2018-6602-1, EGU General Assembly 2018, 9 April 2018, Vienna, Austria.

- 958 Tetegan, M., de Forges, A.C.R., Verbeque, B., Nicoullaud, B., Desbourdes, C., Bouthier, A., Arrouays,
959 D., Cousin, I., 2015. The effect of soil stoniness on the estimation of water retention properties
960 of soils: A case study from central France. *Catena* 129, 95–102.
- 961 Tetegan, M., Nicoullaud, B., Baize, D., Bouthier, A., Cousin, I., 2011. The contribution of rock
962 fragments to the available water content of stony soils: Proposition of new pedotransfer
963 functions. *Geoderma* 165(1), 40–49.
- 964 Tóth, B., Weynants, M., Nemes, A., Mako, A., Bilas, G., Tóth, G., 2015. New generation of hydraulic
965 pedotransfer functions for Europe. *Eur. J. Soil Sci.* 66(1), 226–238. doi:10.1111/ejss.12192.
- 966 Tóth, B., Weynants, M., Pásztor, L., Hengl, T. 2017. 3D Soil Hydraulic Database of Europe at 250 m
967 resolution. *Hydrological Processes*, 31:2662–2666. doi.org/10.1002/hyp.11203.
- 968 Ugbaje, S.U., Reuter, H.I., 2013. Functional digital soil mapping for the prediction of available water
969 capacity in Nigeria using legacy data. *Vadose Zone J.*, 12(4). doi:10.2136/vzj2013.07.0140.
- 970 USGS, 2004. Shuttle Radar Topography Mission, 1 Arc Second Scene SRTM_u03_n008e004,
971 Unfilled Unfinished 2.0, Global Land Cover Facility. University of Maryland, College Park,
972 Maryland (February 2000).
- 973 Vanderlinden, K., Giráldez, J. V., Van Meirvenne, M., 2005. Soil water-holding capacity assessment in
974 terms of the average annual water balance in southern Spain. *Vadose Zone J.*, 4, 317–328.
- 975 Van Ranst E., Vanmechelen L., Thomasson A.J., Daroussin J., Hollis J.M., Jones R.J.A., Jamagne M.,
976 King D., 1995. Elaboration of an extended knowledge database to interpret the 1:1,000,000
977 EU soil map for environmental purposes, in: King D., Jones R.J.A., Thomasson A.J. (Eds.).
978 European Land Information Systems for Agro-environmental monitoring, CEC, Luxembourg.
979 286 pages.
- 980 Vaysse, K., Lagacherie, P., 2017. Using quantile regression forest to estimate uncertainty of digital soil
981 mapping products. *Geoderma*, 291, 55–64. doi.org/10.1016/j.geoderma.2016.12.017.
- 982 Veihmeyer, F.J., Hendrickson, A.H. 1927. The relation of soil moisture to cultivation and plant growth.
983 *Soil Sci.* 3, 498–513.
- 984 Viscarra Rossel R.A., Chen C., Grundy M.J., Searle R., Clifford D., Campbell P.H., 2015. The
985 Australian three-dimensional soil grid: Australia's contribution to the GlobalSoilMap project.
986 *Soil Res.* 53, 845–864. doi.org/10.1071/SR14366.
- 987 Wösten, J.H.M., Lilly, A., Nemes, A., Le Bas, C., 1999. Development and use of a database of
988 hydraulic properties of European soils. *Geoderma* 90(3–4), 169–185. doi.org/10.1016/S0016-
989 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_p : 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.

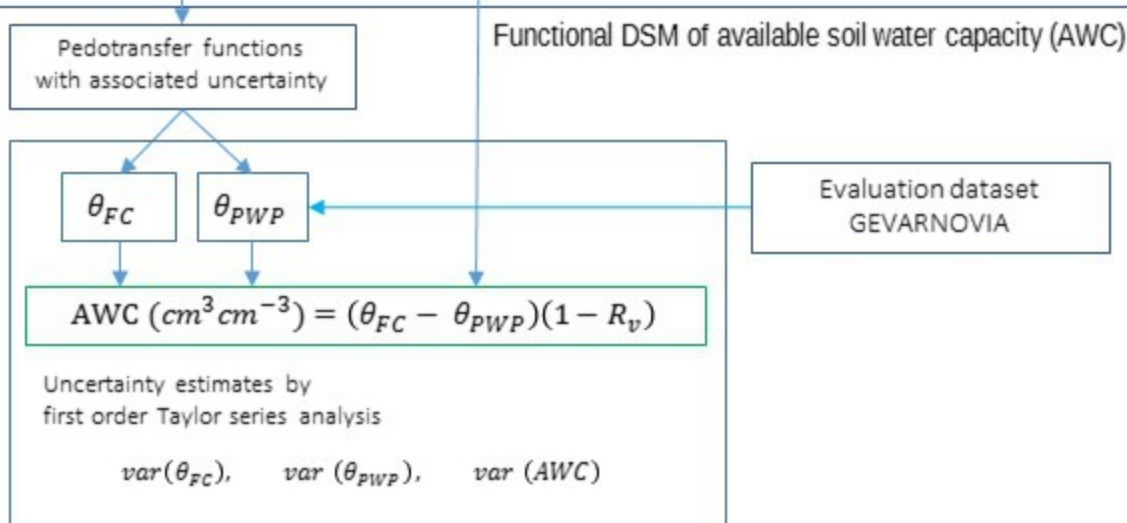
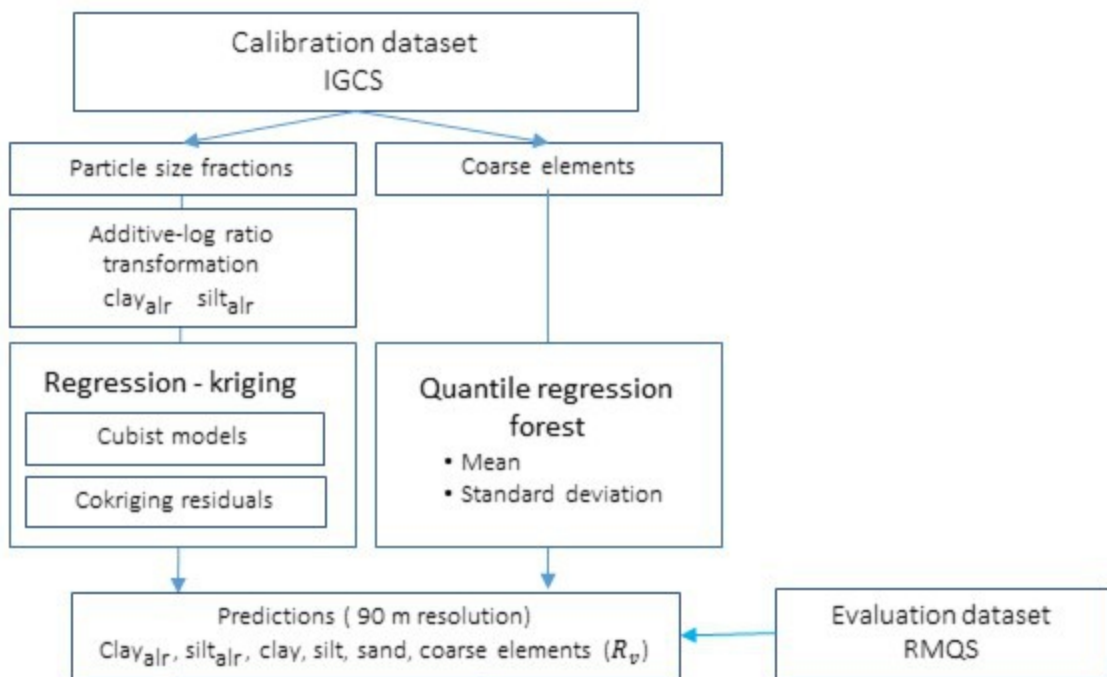
Figure 7: a) Soil thickness (cm), b) total available water capacity (AWC) (mm) to a maximum depth of 2 m, c) standard deviation (SD) of total AWC (mm), and d) relative error of AWC (coefficient of variation, CV).

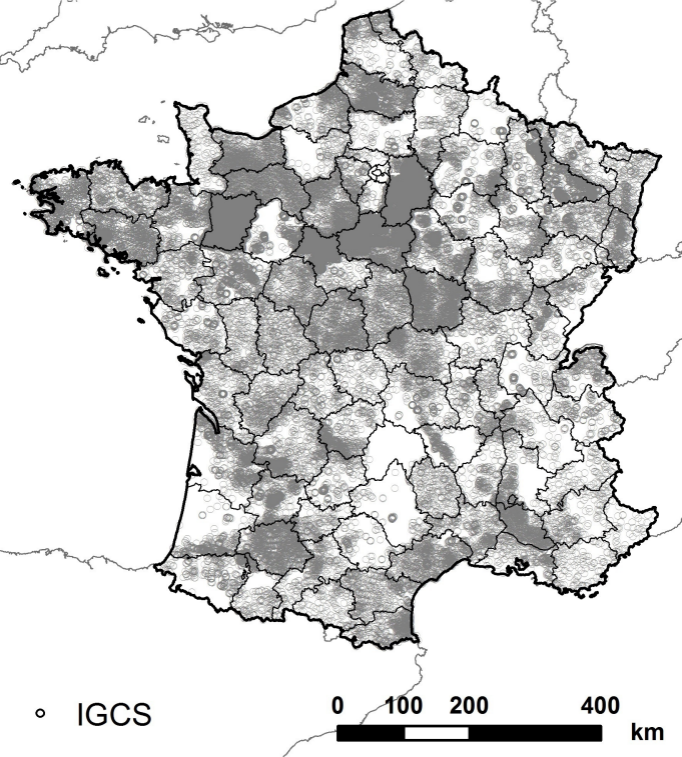
Figure 8: Soil properties used as input for the calculation of AWC and their standard deviation (SD) (15 - 30 cm). a) Clay_{alr}, b) Silt_{alr}, c) Coarse elements (%), d) Clay_{alr} SD, e) Silt_{alr} SD, and f) Coarse elements SD (%).

Figure 9: Components of the variance of elementary soil moisture at field capacity ($\text{cm}^3 \text{cm}^{-3}$) 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 ($\text{cm}^3 \text{cm}^{-3}$) 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 ($3.80 \cdot 10^{-5} \text{cm}^6 \text{cm}^{-6}$) (not included in the figure).

Digital Soil Mapping (DSM) of soil input properties

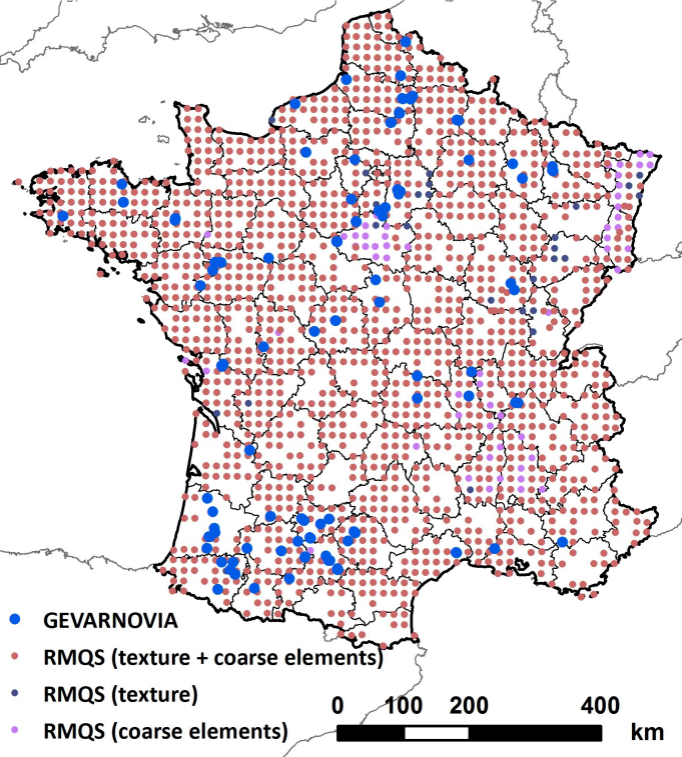


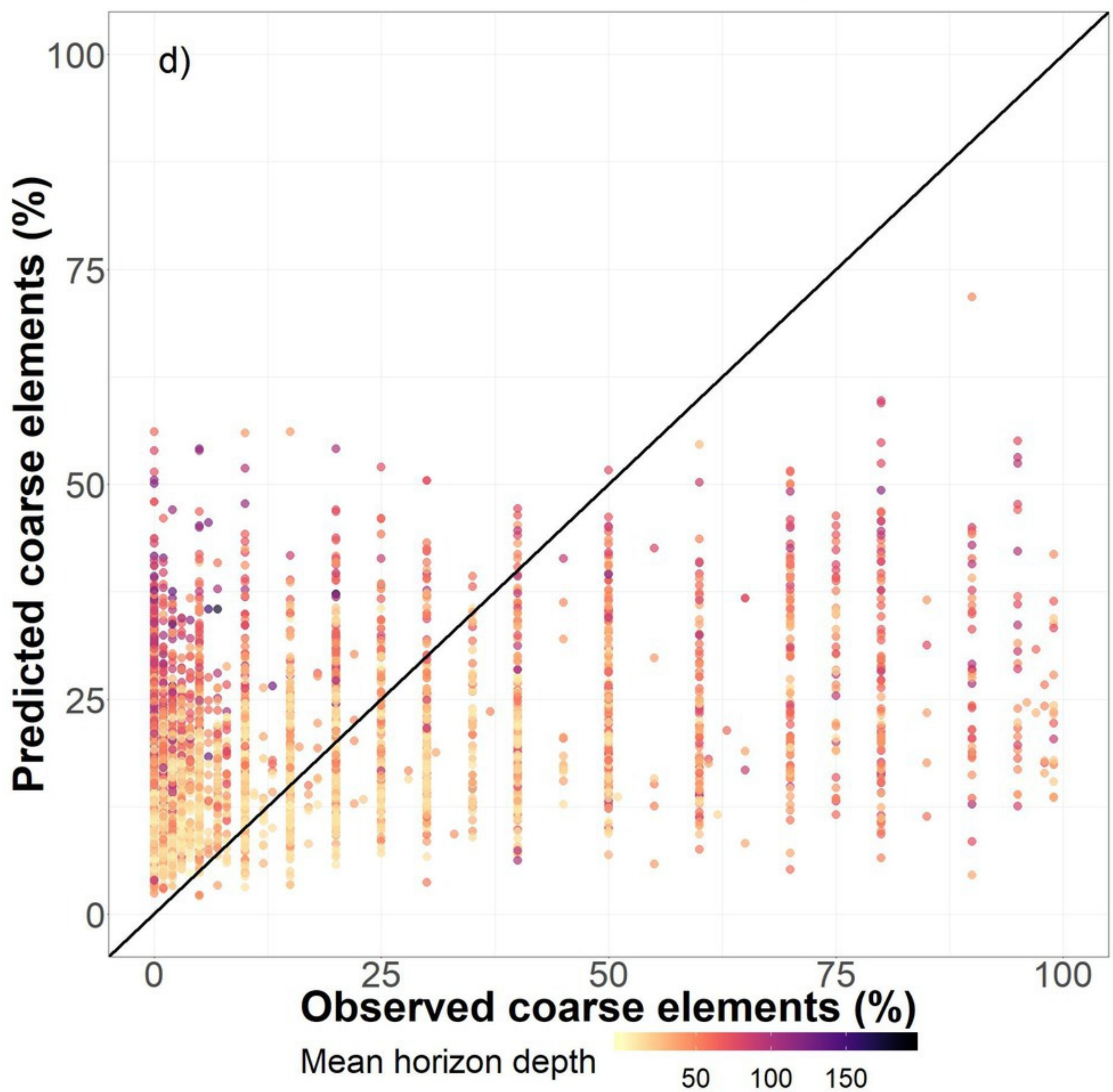
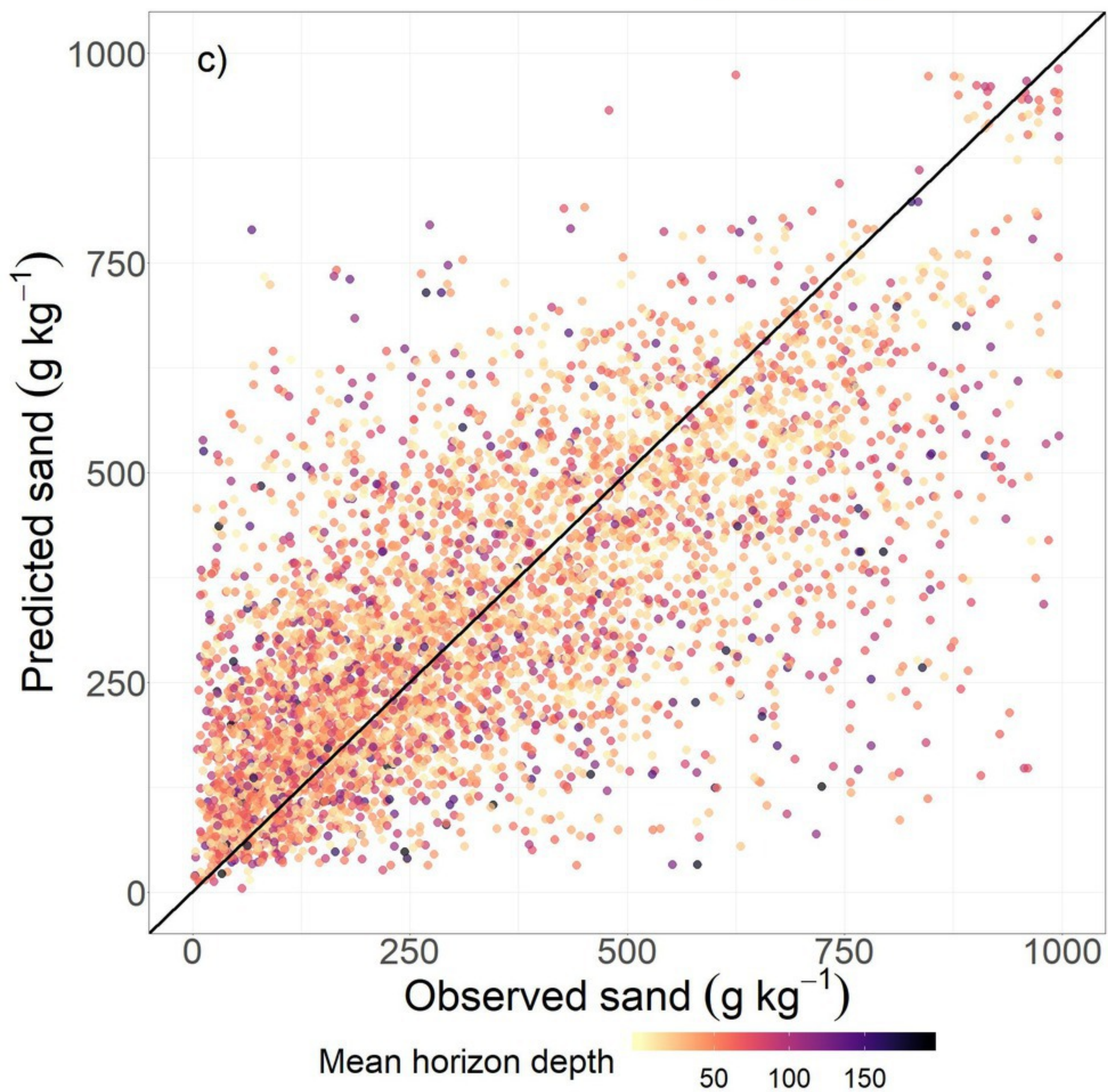
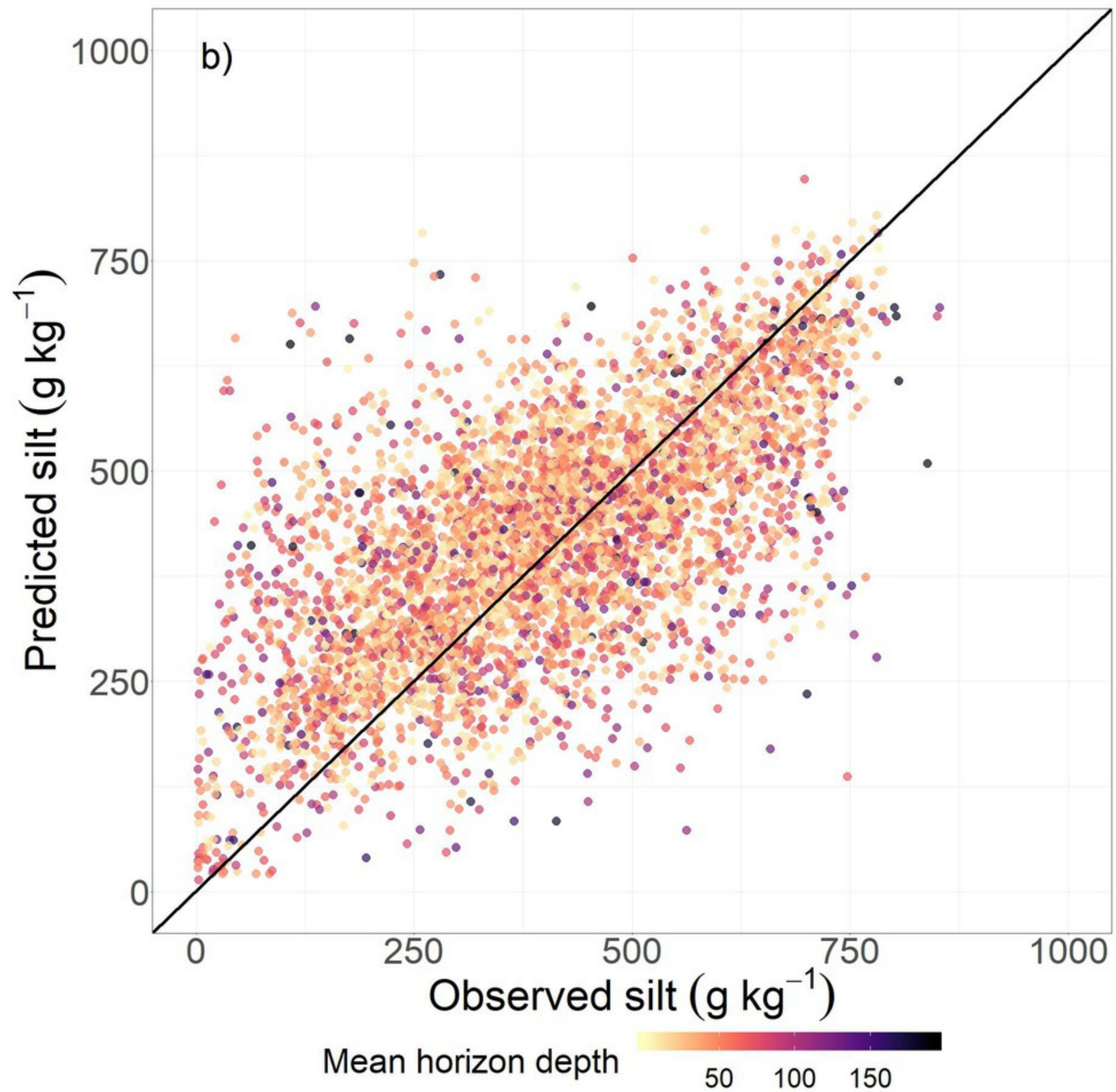
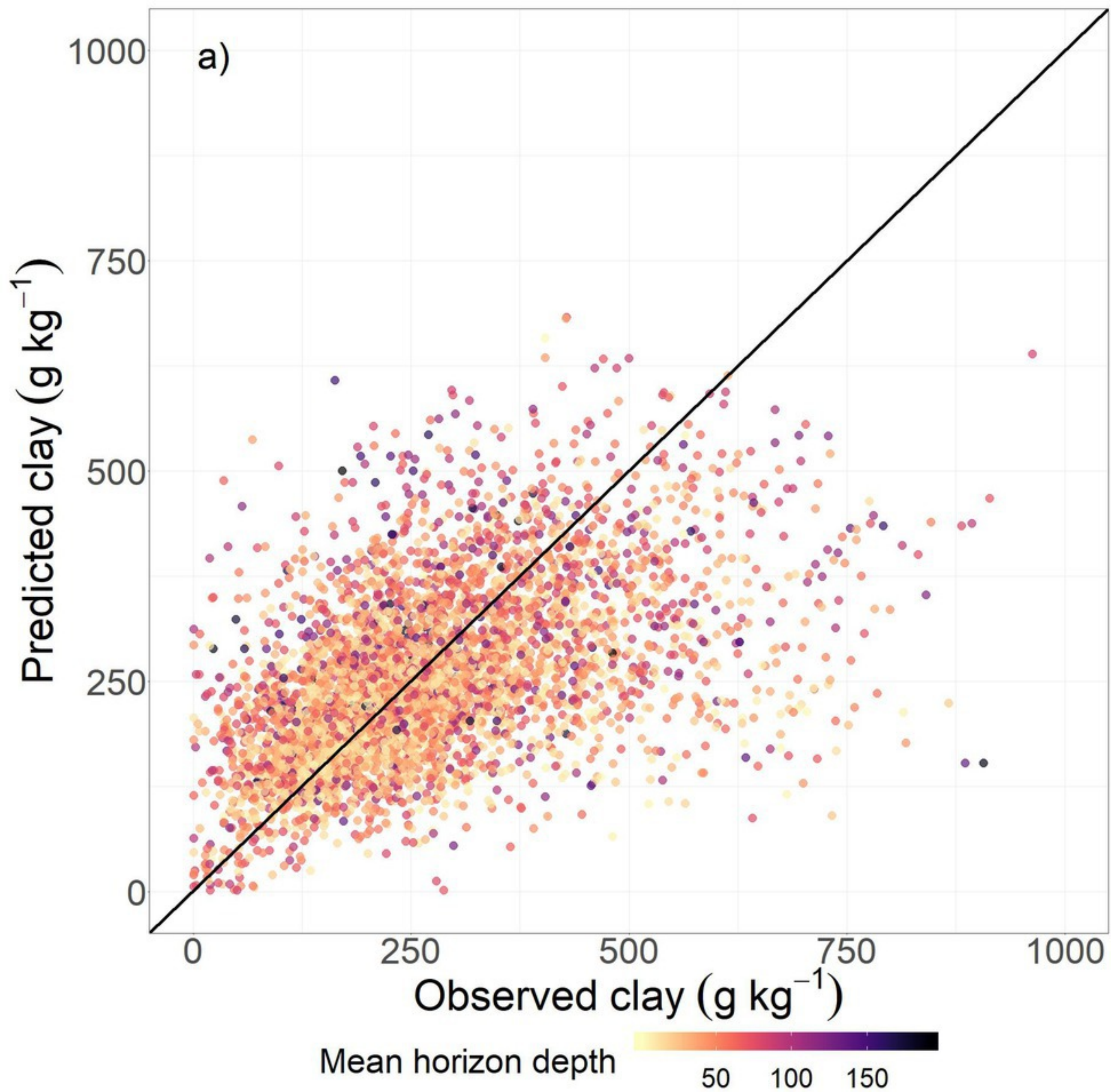


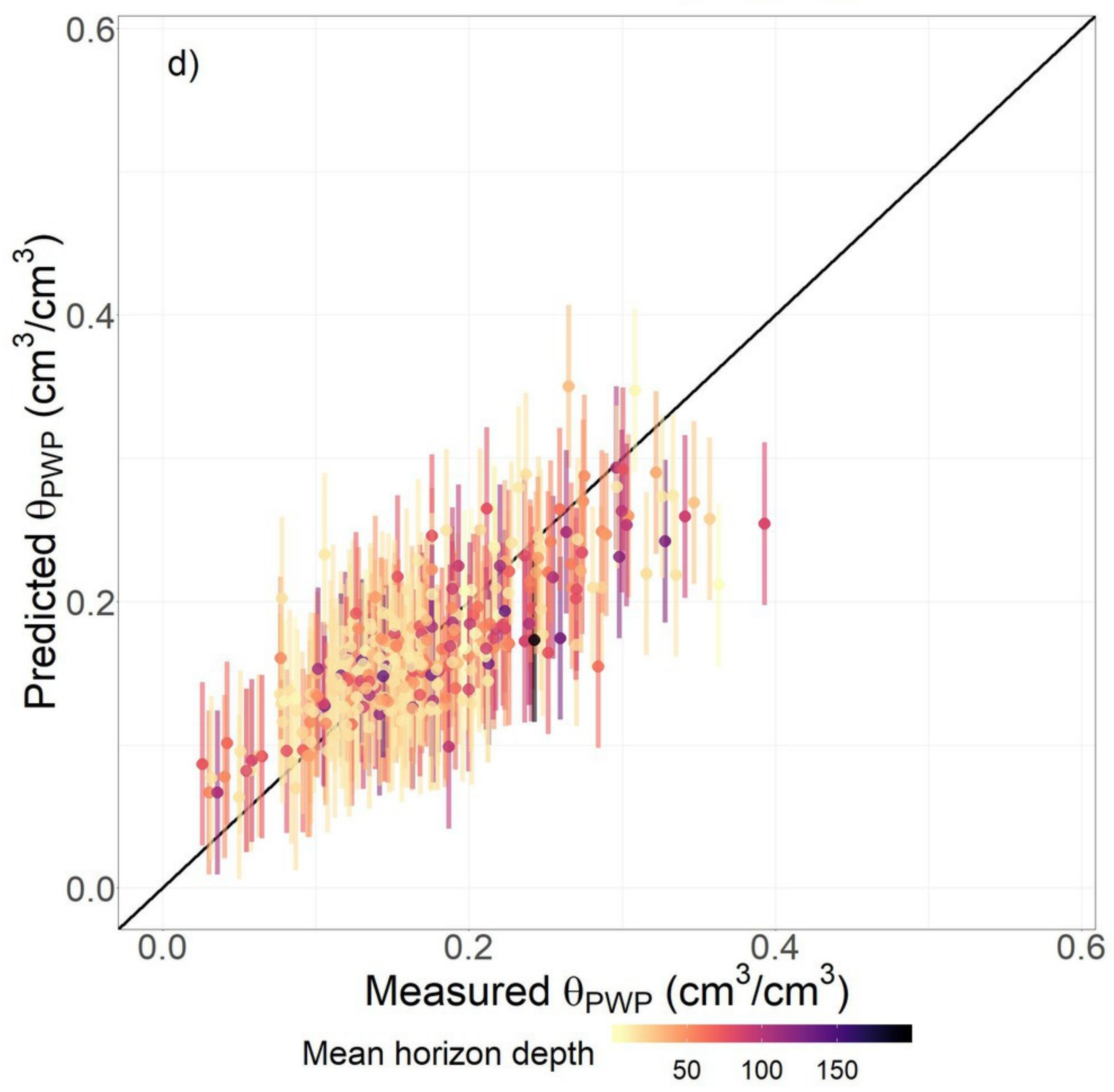
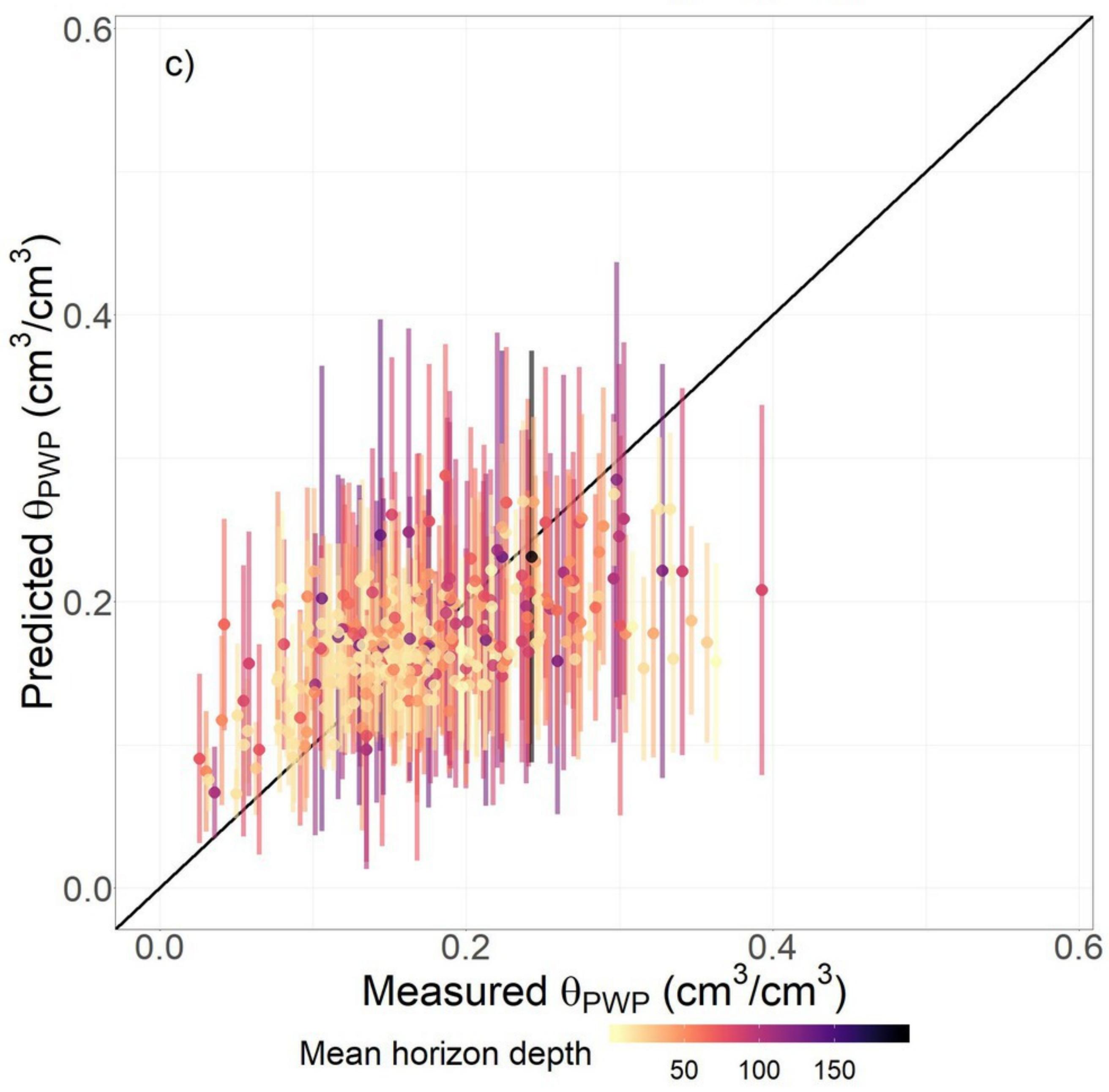
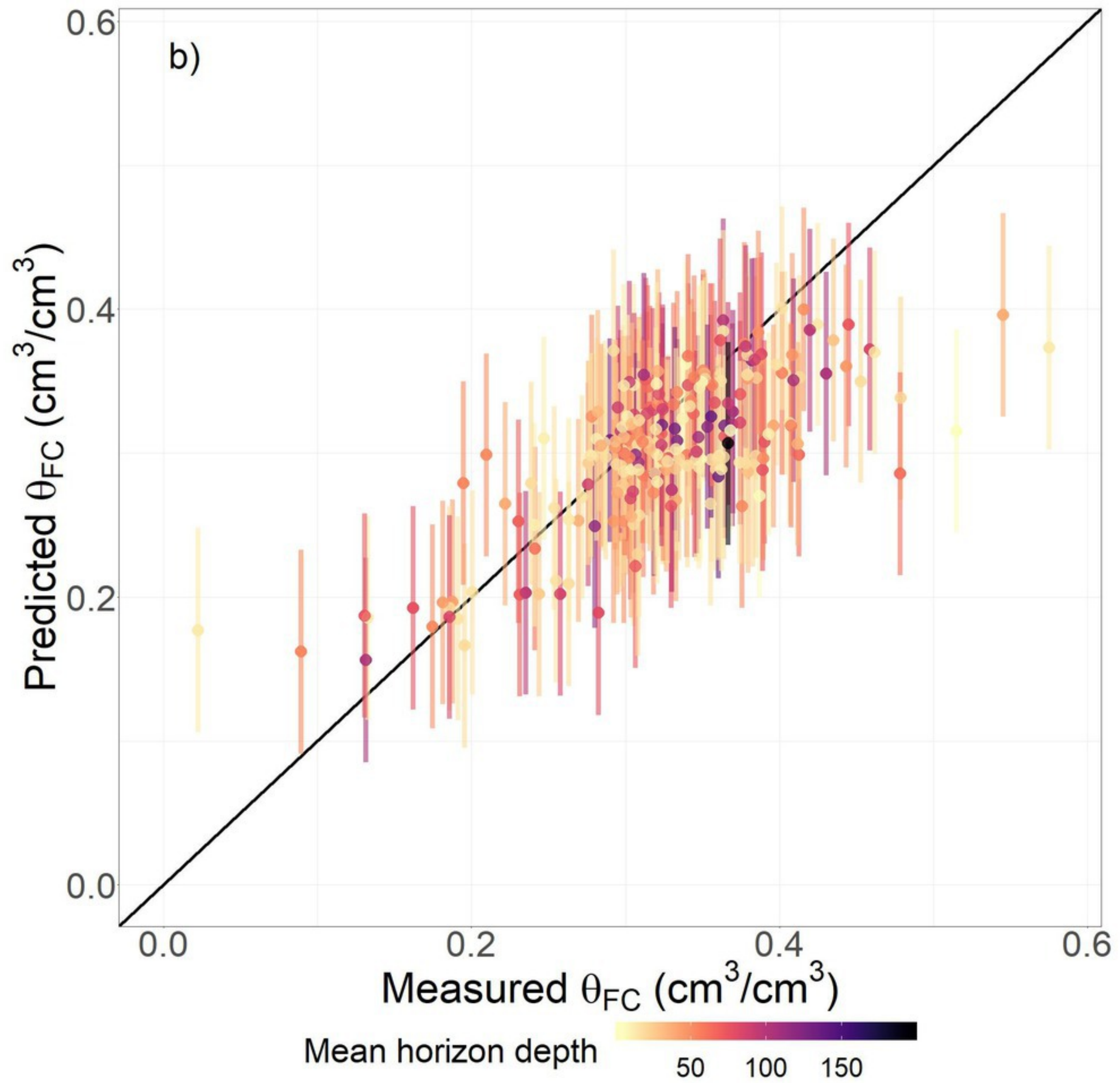
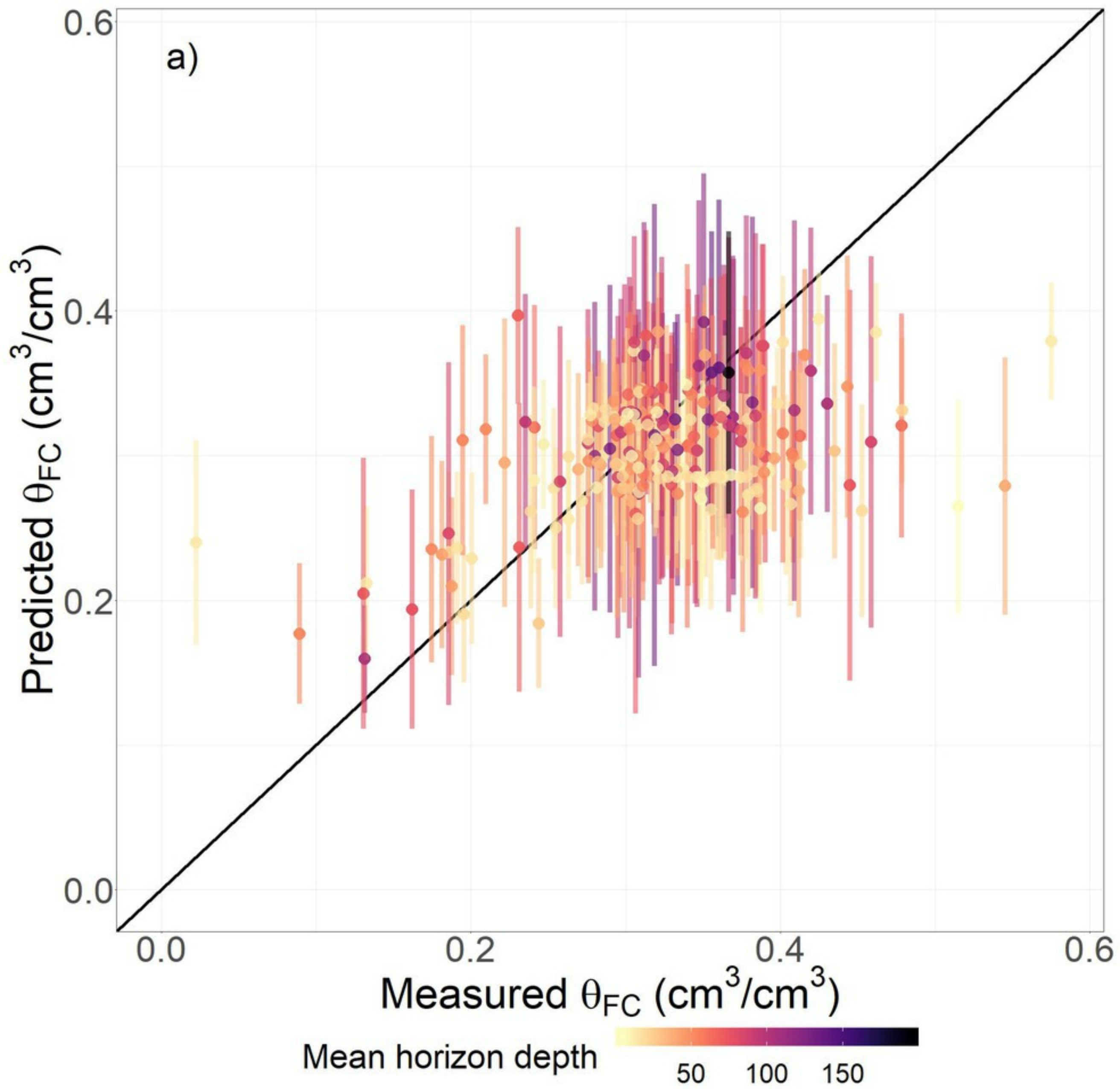
○ IGCS

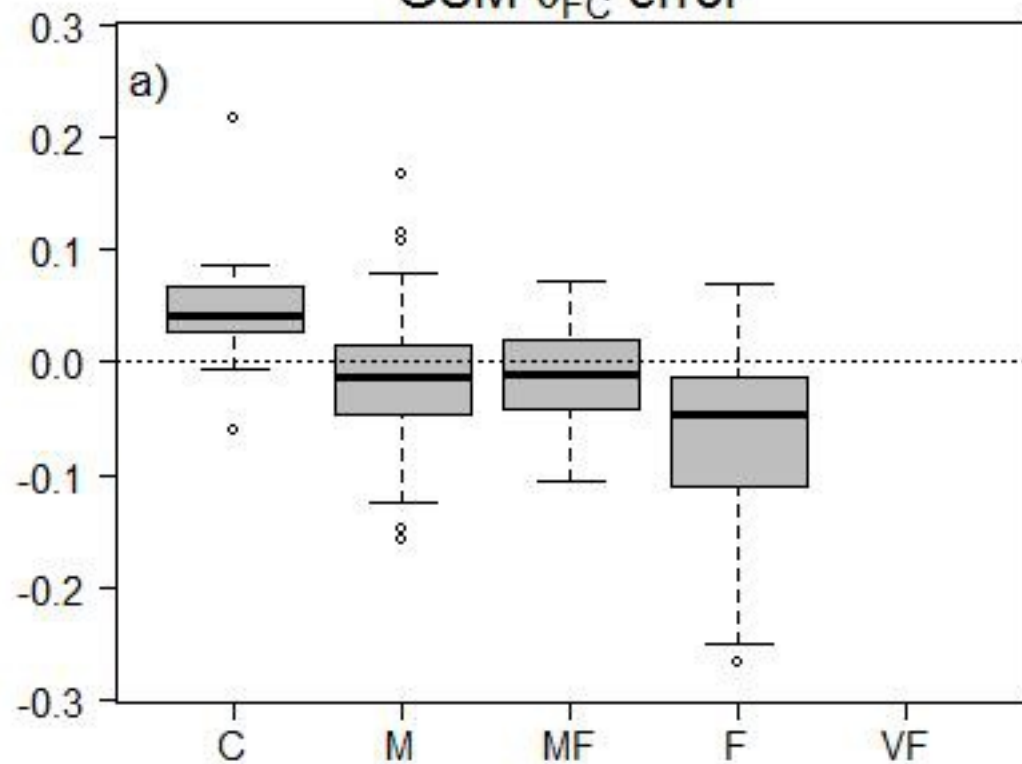
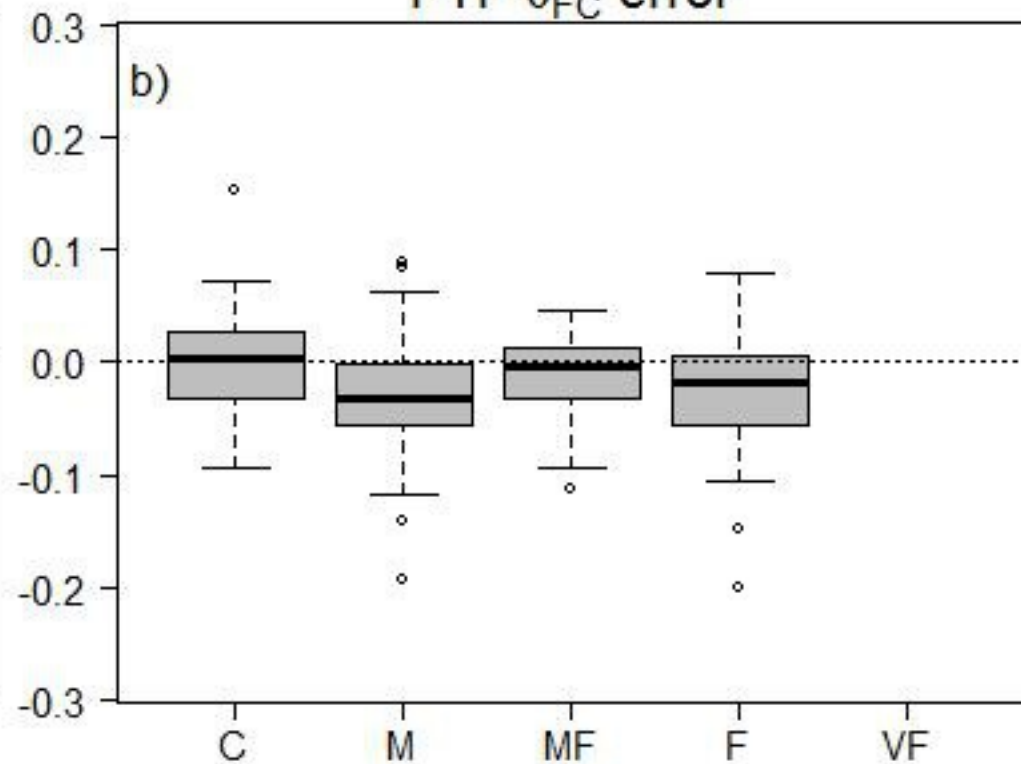
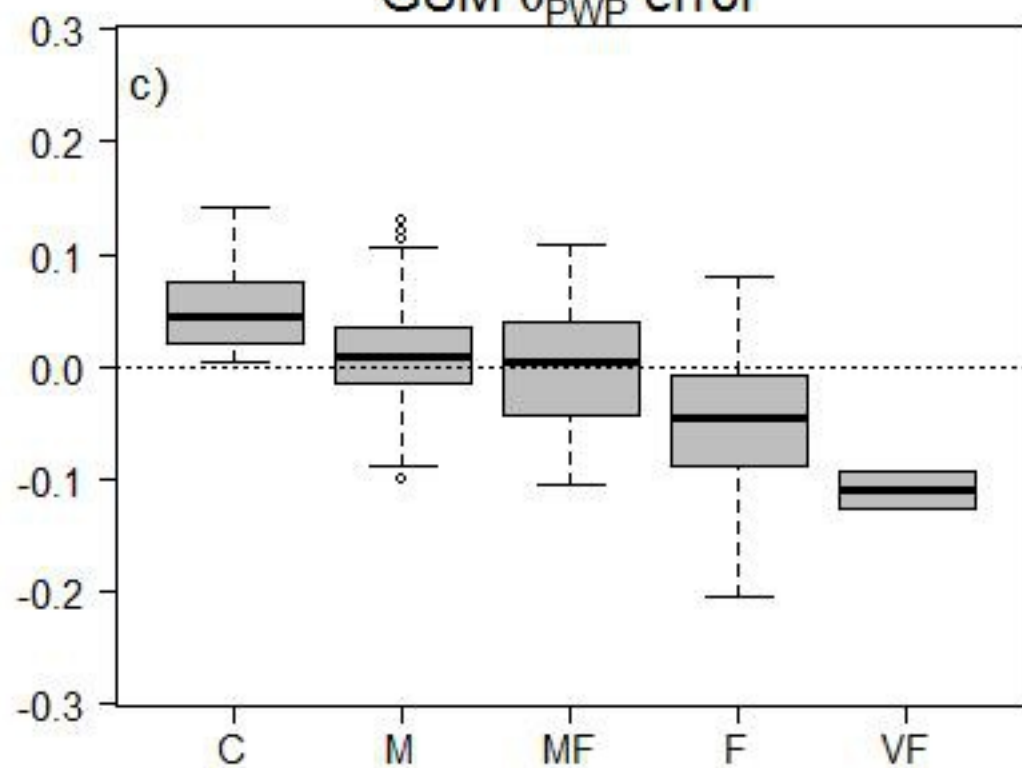
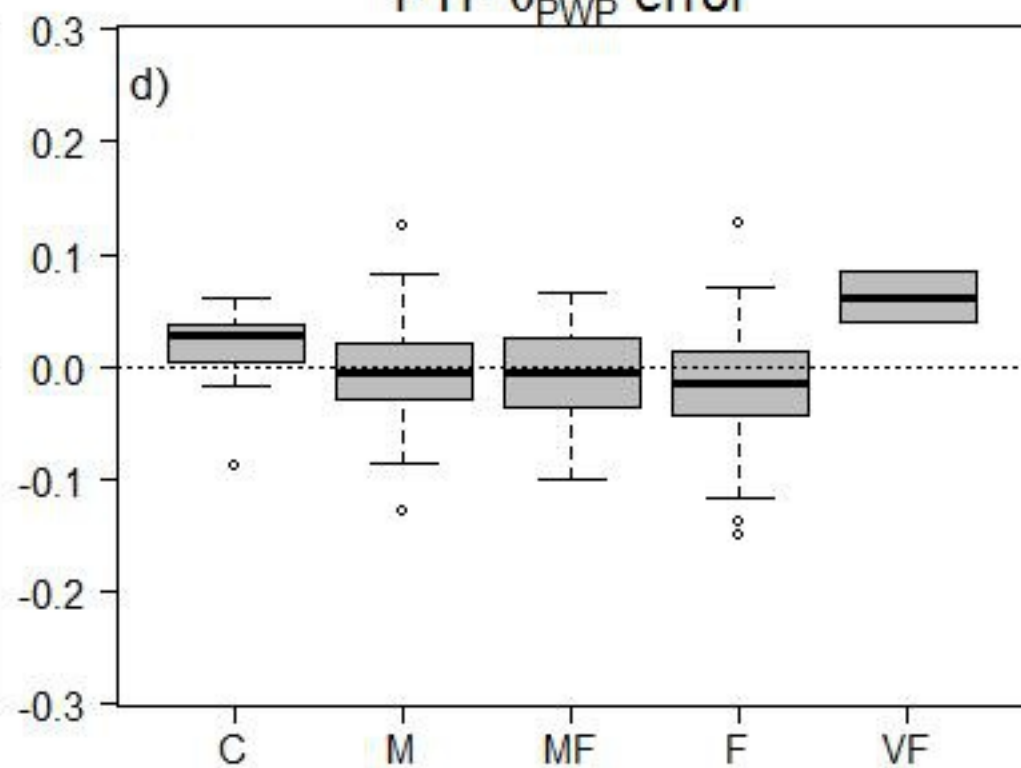
0 100 200 400

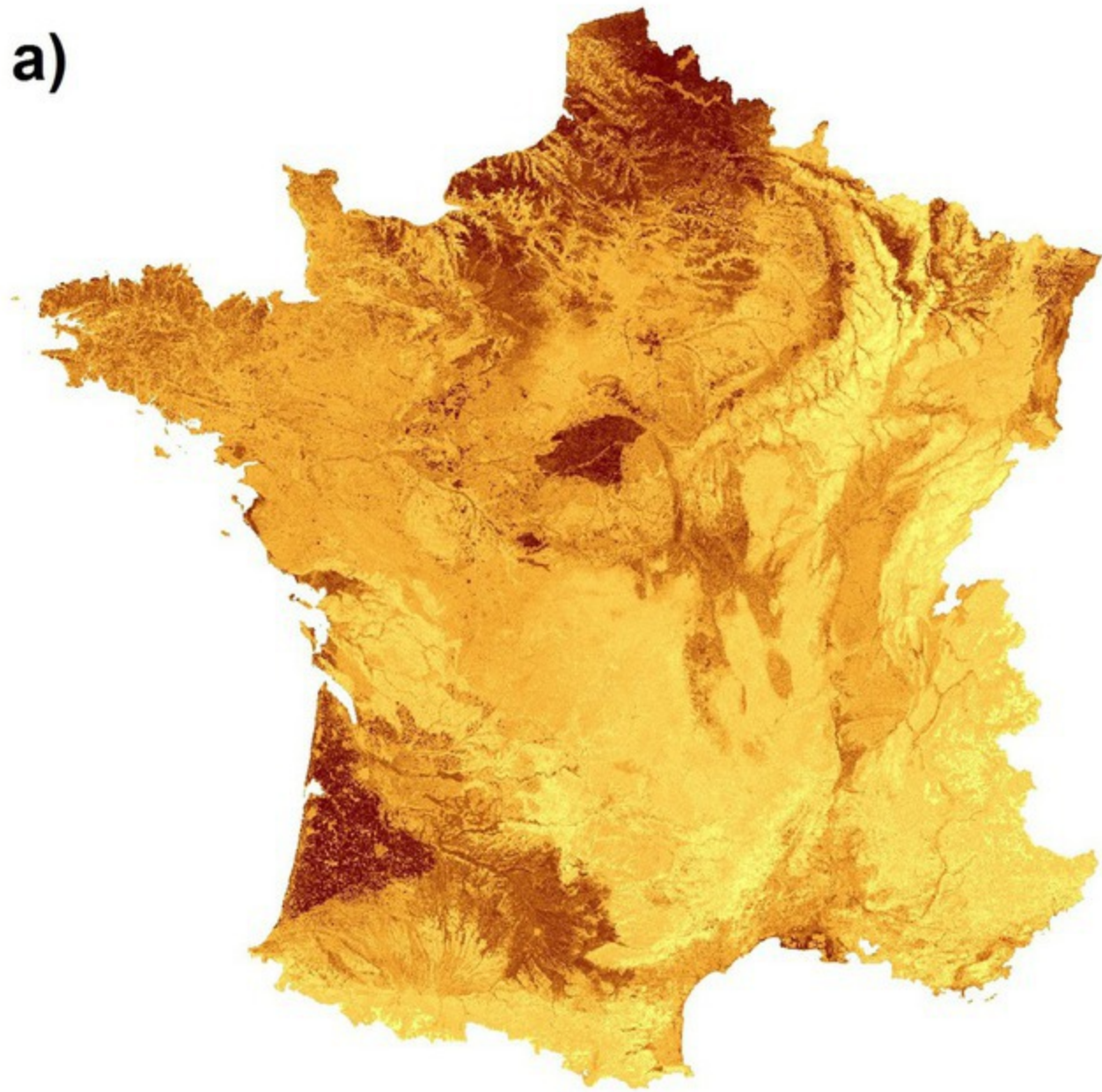
km



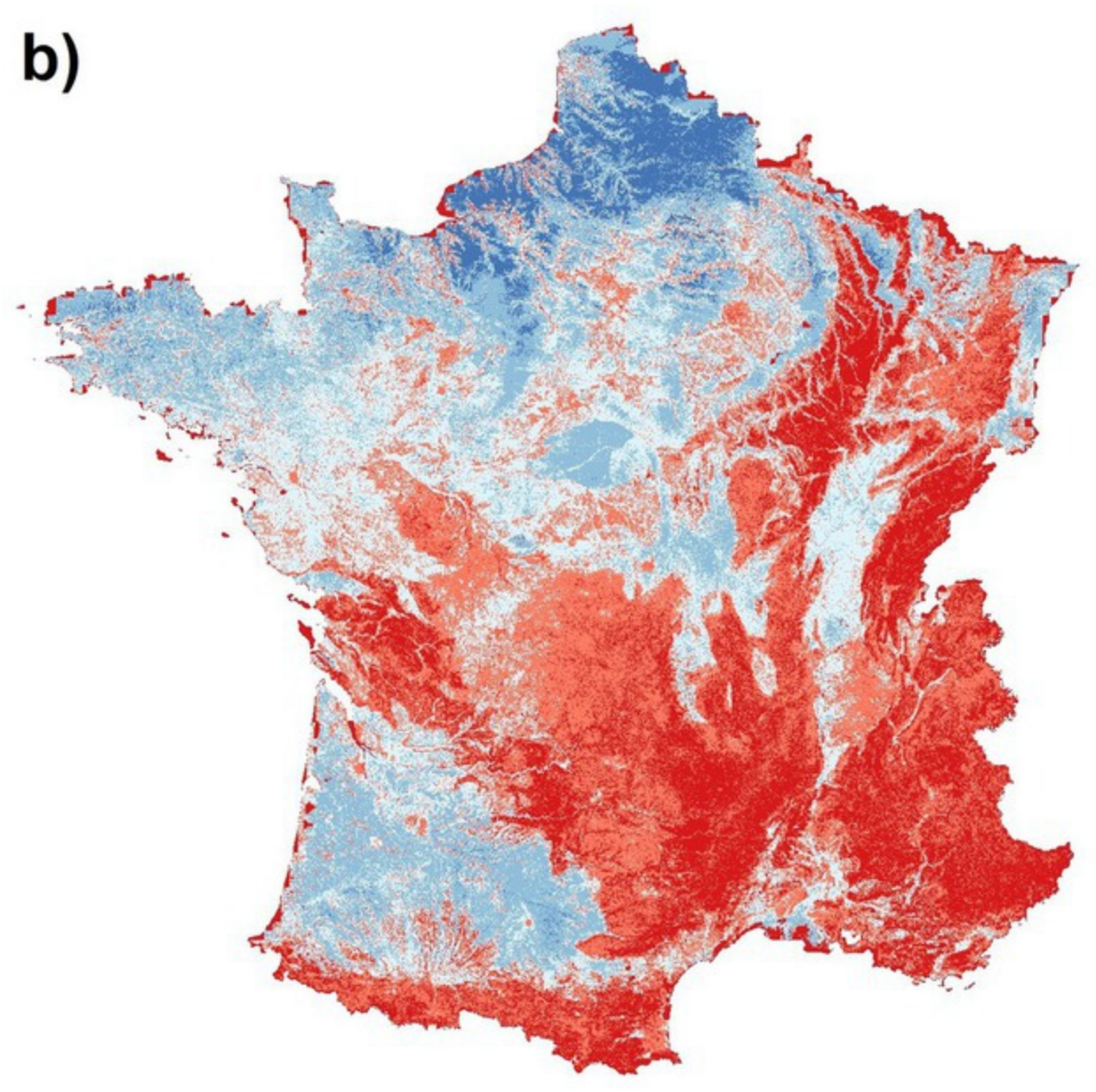
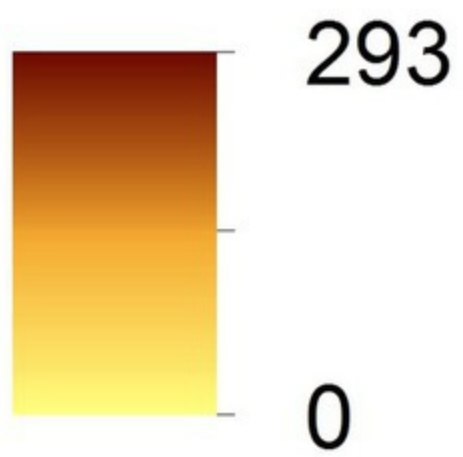




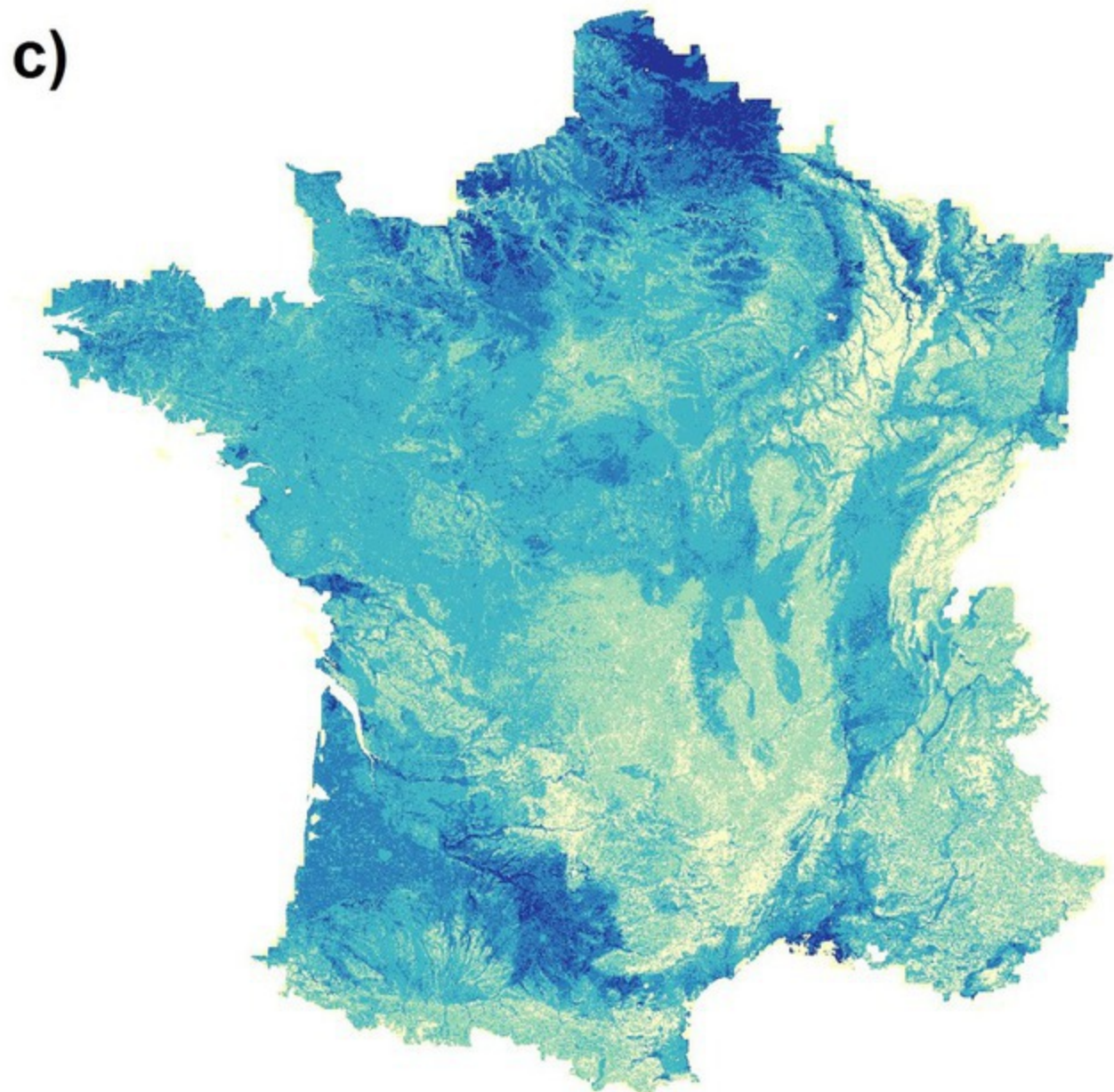
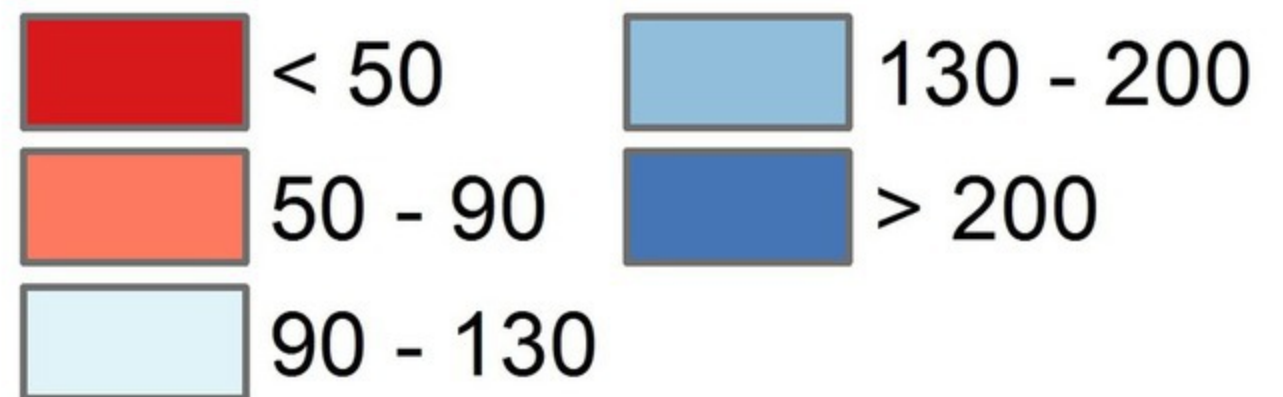
GSM θ_{FC} errorPTF θ_{FC} errorGSM θ_{PWP} errorPTF θ_{PWP} error



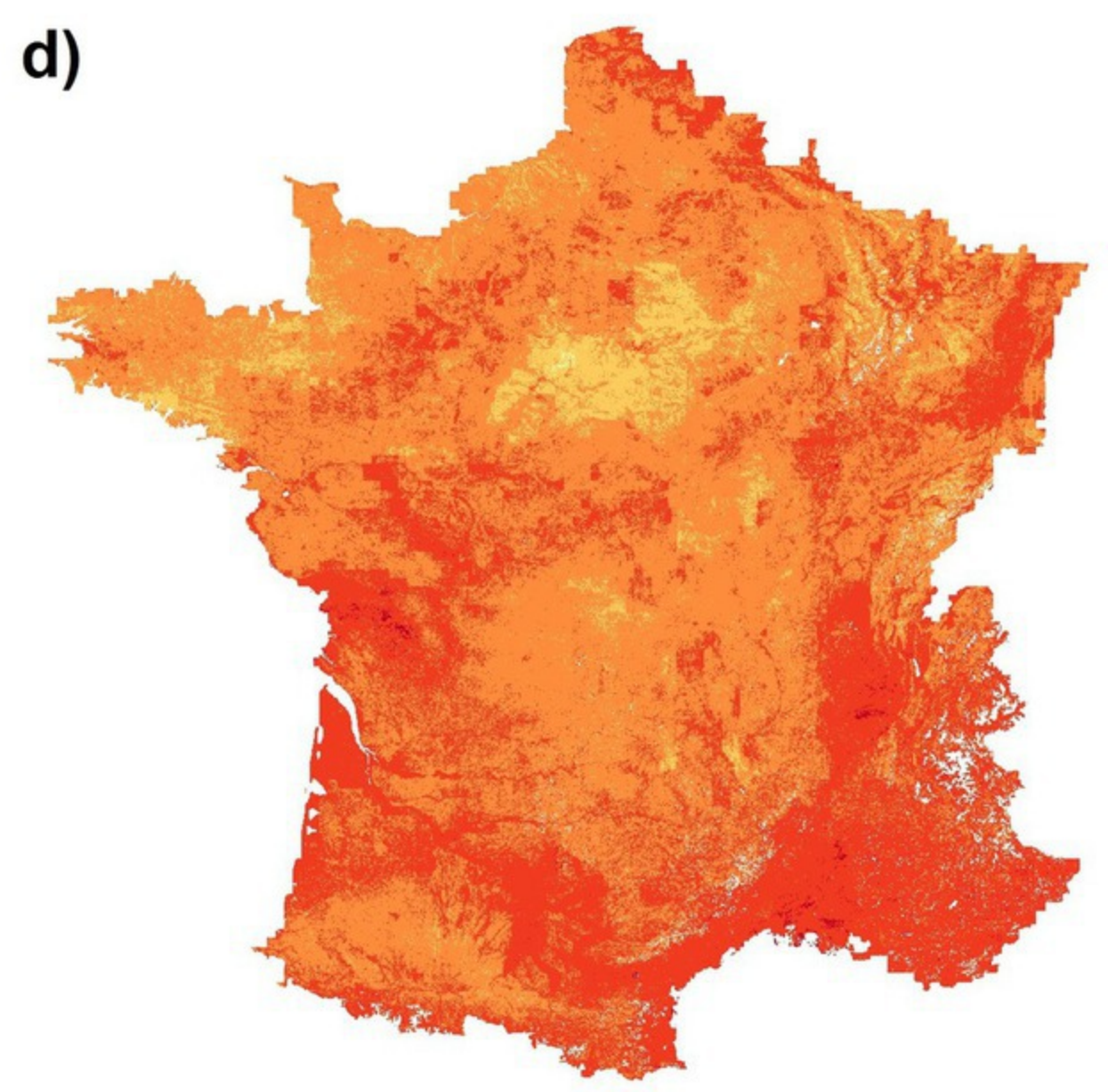
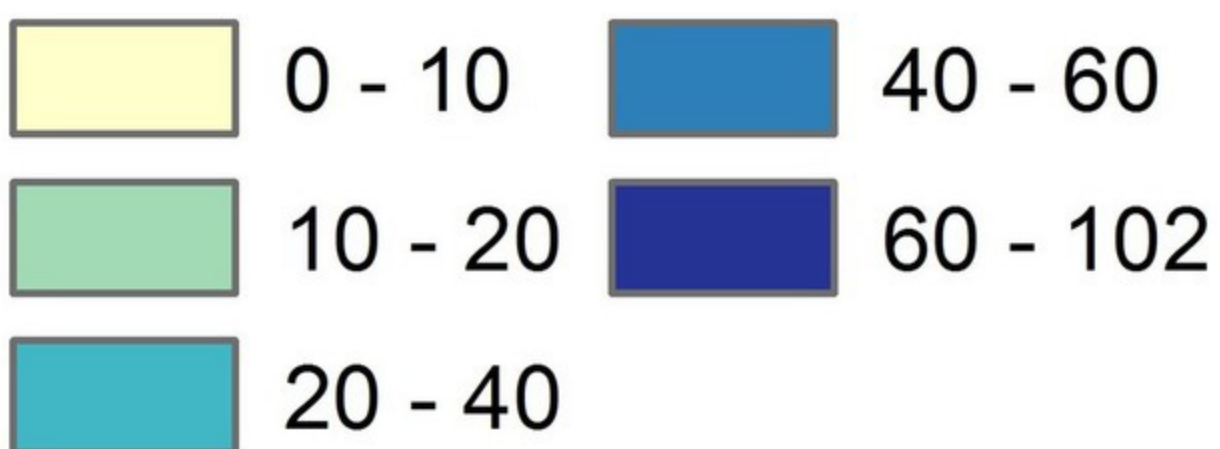
Soil thickness (cm)



AWC (mm)

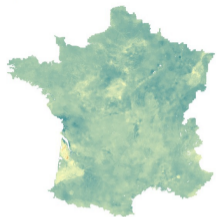
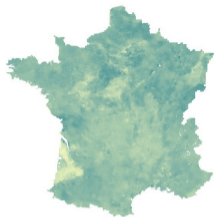
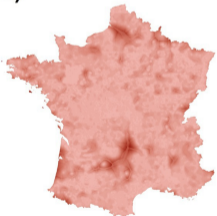
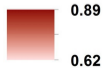
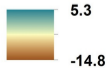


AWC SD (mm)

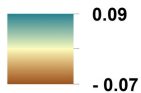


CV of AWC (%)

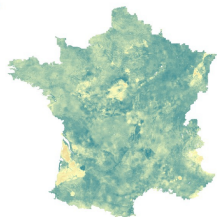


a)**b)****c)****d)****e)****f)**

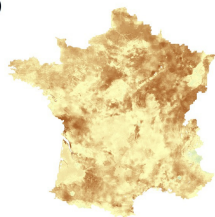
Sensitivity to soil input variables



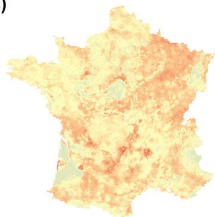
a)



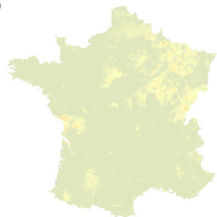
b)



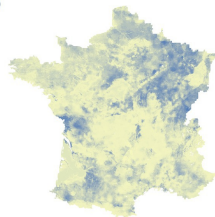
c)



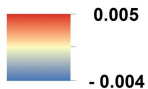
d)



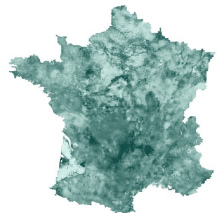
e)



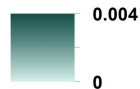
Variance terms



f)



Variance due to soil input variables



Clay
(sensitivity to PTF
clay coefficient)

a)



Sand
(sensitivity to PTF
sand coefficient)

b)



Content (%)



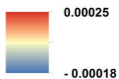
c)



d)



Variance terms



e)



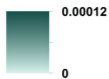
f)



g)



Variance due to
PTF coefficients



h)



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: Bouguer 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 ($\theta_{4.2}$).

$\theta_{2.0}$				$\theta_{4.2}$			
	Intercept	Clay	Sand		Intercept	Clay	Sand
Intercept	$3.80 \cdot 10^{-5}$	$-9.93 \cdot 10^{-7}$	$-3.85 \cdot 10^{-7}$	Intercept	$1.84 \cdot 10^{-5}$	$-4.07 \cdot 10^{-7}$	$-1.97 \cdot 10^{-7}$
Clay	$-9.93 \cdot 10^{-7}$	$3.17 \cdot 10^{-8}$	$7.05 \cdot 10^{-9}$	Clay	$-4.07 \cdot 10^{-7}$	$1.04 \cdot 10^{-8}$	$3.79 \cdot 10^{-9}$
Sand	$-3.85 \cdot 10^{-7}$	$7.05 \cdot 10^{-9}$	$9.09 \cdot 10^{-9}$	Sand	$-1.97 \cdot 10^{-7}$	$3.79 \cdot 10^{-9}$	$3.76 \cdot 10^{-9}$

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	
	$\text{Silt}_{\text{alr}} \times \text{Clay}_{\text{alr}}$		0.37	0.42	
5–15	Silt_{alr}	36108	0.39	0.33	178104
	Clay_{alr}		0.45	0.48	
	$\text{Silt}_{\text{alr}} \times \text{Clay}_{\text{alr}}$		0.34	0.31	
15–30	Silt_{alr}	35401	0.35	0.32	160970
	Clay_{alr}		0.40	0.50	
	$\text{Silt}_{\text{alr}} \times \text{Clay}_{\text{alr}}$		0.30	0.32	
30–60	Silt_{alr}	31494	0.58	0.48	170776
	Clay_{alr}		0.62	0.74	
	$\text{Silt}_{\text{alr}} \times \text{Clay}_{\text{alr}}$		0.48	0.46	
60–100	Silt_{alr}	24849	1.83	0.56	252306
	Clay_{alr}		1.50	0.91	
	$\text{Silt}_{\text{alr}} \times \text{Clay}_{\text{alr}}$		1.31	0.60	
100–200	Silt_{alr}	13086	3.03	1.25	167139
	Clay_{alr}		2.50	1.42	
	$\text{Silt}_{\text{alr}} \times \text{Clay}_{\text{alr}}$		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 ³ cm ⁻³)	236	0.54	0.65	0.052	-0.02	84.3
	θ_{PWP} (cm ³ cm ⁻³)	308	0.62	0.75	0.042	-0.005	85.1
GSM prediction	θ_{FC} (cm ³ cm ⁻³)	236	0.21	0.37	0.065	-0.02	71.2
	θ_{PWP} (cm ³ cm ⁻³)	308	0.29	0.47	0.057	-0.0004	76.6