



HAL
open science

Evaluating the impact of using digital soil mapping products as input for spatializing a crop model: The case of drainage and maize yield simulated by STICS in the Berambadi catchment (India)

Philippe Lagacherie, Samuel Buis, Julie Constantin, Subramanian Dharumarajan, Laurent Ruiz, M. Sekhar

► To cite this version:

Philippe Lagacherie, Samuel Buis, Julie Constantin, Subramanian Dharumarajan, Laurent Ruiz, et al.. Evaluating the impact of using digital soil mapping products as input for spatializing a crop model: The case of drainage and maize yield simulated by STICS in the Berambadi catchment (India). *Geoderma*, 2022, 406, 10.1016/j.geoderma.2021.115503 . hal-03401773

HAL Id: hal-03401773

<https://hal.inrae.fr/hal-03401773>

Submitted on 16 Oct 2023

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 Attribution - NonCommercial 4.0 International License

1 **Evaluating the impact of using digital soil mapping products as input for**
2 **spatializing a crop model: The case of drainage and maize yield simulated**
3 **by STICS in the Berambadi Catchment (India)**
4

5 Lagacherie¹, P.; Buis², S.; Constantin³, J. ; Dharumarajan⁴, S. ; Ruiz^{5,6,8}, L., Sekhar⁷, M.

7 ¹ LISAH, University Montpellier, IRD, INRAE, Montpellier SupAgro, 34060 Montpellier,
8 France

9 ² EMMAH, INRAE Avignon, France

10 ³ AGIR, INRAE, Toulouse, France

11 ⁴ ICAR-National Bureau of Soil Survey and Land Use Planning, Hebbal, Bangalore, 560024,
12 India

13 ⁵ Géosciences Environnement Toulouse, Université Paul-Sabatier, IRD, CNRS, 31400
14 Toulouse, France; laurent.ruiz@inra.fr

15 ⁶ Indo-French Cell for Water Sciences, IRD, Indian Institute of Science, Bangalore 560012,
16 India

17 ⁷ Indian Institute of Science, Civil Engineering Department, Bangalore 560012, India

18 ⁸ SAS, INRAE, Agrocampus Ouest, Rennes, France

19
20
21 Corresponding Author: P. Lagacherie : philippe.lagacherie@inrae.fr
22

23 **Abstract**
24

25 Digital Soil Mapping (DSM) can be an alternative data source for spatializing crop models
26 over large areas. The objective of the paper was to evaluate the impact of DSM products and
27 their uncertainties on a crop model's outputs in an 80 km² catchment in south India. We used
28 a crop model called STICS and evaluated two essential soil functions: the biomass production
29 (through simulated yield) and water regulation (via calculated drainage). The simulation was
30 conducted at 217 sites using soil parameters obtained from a DSM approach using either
31 Random Forest or Random Forest Kriging. We first analysed the individual STICS
32 simulations, i.e., at two cropping seasons for 14 individual years, and then pooled the
33 simulations across years, per site and crop season. The results show that i) DSM products
34 outperformed a classical soil map in providing spatial estimates of STICS soil parameters, ii)
35 although each soil parameters were estimated separately, the correlations between soil
36 parameters were globally preserved, ii) Errors on STICS' yearly outputs induced by DSM
37 estimations of soil parameters were globally low but were important for the few years with
38 high impacts of soil variations, iii) The statistics of the STICS simulations across years were
39 also affected by DSM errors with the same order of magnitude as the errors on soil inputs and
40 iv) The impact of DSM errors was variable across the studied soil parameters. These results
41 demonstrated that coupling DSM with a crop model could be a better alternative to the

42 classical Digital Soil Assessment techniques. As such, it will deserve more work in the
43 future.

44

45 Keywords: Soil mapping, soil functions, digital soil assessment, crop model, machine
46 learning, uncertainty analysis

47

48

49 **1. Introduction**

50

51 Spatializing a crop model consists of applying the model over an area much larger
52 than those over which it was developed (Faivre et al, 2004). Various reviews (Hartkamp et al,
53 1999, Faivre et al, 2004, Ginaldi et al, 2019) provide examples of crop model spatialisation
54 for a large range of purposes and at a large range of scales (from field to continental scales).
55 Crop model spatialisation has been widely applied for assessing agricultural production and
56 potentialities (e.g. Lal et al,1993), testing scenario about the impact of agricultural practices
57 on water quality (e.g. Beaujouan et al, 2001), evaluating irrigation requirements (Sousa &
58 Santos Pereira, 1999) or the impact of climate change on crop production (Wassenaar et al.,
59 2000; Deryng et al., 2016). Several crop models were used in these spatialisations, the most
60 common being DSSAT (Jones et al. 2003), APSIM (Keating et al. 2003), CropSyst
61 (Stöcke et al. 2003), EPIC (Williams et al. 1989) and STICS (Brisson et al, 2003).

62 Among the different sources of uncertainty that may affect the spatialization of crop models,
63 the estimation of the soil inputs is one of the most critical. Currently, the most common
64 source of soil data used as input of crop models has been those provided by traditional
65 choropleth soil maps (see e.g. table 1, Faivre et al., 2004). It is assumed that the best
66 estimation of any soil property, and thus of soil inputs of the crop models, at an unvisited site
67 covered by a soil map, is given by the soil mapping unit mean at all sites, often estimated by
68 one so-called ‘representative’ soil profile. However, this assumption may generate a
69 substantial uncertainty on soil inputs (Leenhardt et al., 1994a) that further propagates to the
70 outputs of crop models (Leenhardt et al., 1994b). Alternative sources of soil data such as
71 kriging of soil properties have been proposed (Faivre et al., 2004). However, to our
72 knowledge, there has been no study in the literature using kriging of soil properties for
73 spatializing a crop model beyond the field scale.

74 Digital Soil Mapping (DSM) can be an alternative to choropleth soil maps for providing the
75 crop model soil inputs. DSM was developed as an alternative to conventional soil survey for

76 mapping soil properties at limited costs (McBratney et al., 2003, Lagacherie et al., 2007).
77 McBratney et al. (2003) proposed the equation $S = f(s,c,o,r,p,a,n)$ for summarizing the
78 general principle of DSM. According to this equation, a soil property (S) can be predicted by
79 a spatial inference function (f) using, as input, the existing soil information (s), the spatial
80 covariates that map the different factors of soil formation early defined by Jenny (1941)
81 (c,o,r,p,a, standing for climate, organisms, relief, parent material and age respectively) and
82 the geographical location (n) that can highlight any spatial trends missed by the other
83 covariates. Although DSM is as much limited as conventional soil survey by the availability
84 of the existing soil information, it has several important advantages over conventional soil
85 survey: i) it exploits a large range of spatial data on landscape provided by the spatial data
86 infrastructures, ii) it provides a local estimation of the uncertainty of predictions which allows
87 making a realistic use of the outputs and iii) its outputs can be easily updated if new data are
88 collected. DSM has moved rapidly to operationality thanks to the launch of the
89 GlobalSoilMap projects (Arrouays et al, 2014), which enhanced the release of a number of
90 GSM products providing soil property predictions at fine resolution at the national (e.g.
91 Mulder et al., 2016, Adhikari et al., 2013), continental (e.g. Ballabio et al., 2019) and global
92 (Hengl et al., 2007) scales.

93 Despite the rapid development of operational Digital Soil Mapping across the globe, it
94 has rarely been envisaged yet as a possible source of soil inputs for spatializing mechanistic
95 models (Tavares Wahren et al., 2016; van Tol et al., 2020) and, to our knowledge, has never
96 been envisaged for spatializing crop models.

97 This paper aims to take the first step in filling this gap by evaluating the uncertainty of
98 the outputs a crop model called STICS when it is simulated using soil inputs provided by a
99 DSM approach compared to observed ones. The test was performed in an 80 km² catchment,
100 located in south India. The study compared STICS outputs using observed soil parameters
101 from a set of 217 soil profiles in the catchment, or by using the parameters estimated from
102 DSM (Random Forest and Random Forest Kriging). We first analysed individual STICS
103 simulations, i.e. at two crops seasons for 14 individual years and then pooled simulations
104 across years, per site and crop season. We analysed two STICS outputs, which are essential
105 for soil functions and ecosystem Services assessments at local sites (Adhikari & Hartemink,
106 2016): the biomass production (through yield) and the water regulation (through drainage).

107

108 **2. Methods**

109

2.1. Digital Soil Mapping

A classical Digital Soil Mapping approach was applied to map the soil properties required as inputs for the STICS soil crop model (denoted further the STICS soil inputs). This approach consisted in building a prediction function produced by a Random Forest (RF) (Breiman et al, 2001) or by a Regression Kriging algorithm (RK) (Hengl et al, 2004) with RF as regression algorithm (Vaysse et al., 2017), from a limited number of sites at which both the soil properties and the spatial data envisaged for to predicting soil properties are available. In the following, we give some details on the algorithms and on the validation approach.

2.1.1. Random Forest

Recent performance testing has found that the Random Forest was among the best models for predicting soil properties (Nussbaum et al., 2018), which confirmed a test performed on a wider range of machine learning applications (Caruana et al., 2006). We summarize hereafter the main principles of Random forest, using excerpts of Meinshausen (2006). More details are given on Random Forests in Breiman et al (2001) and Meinshausen (2006).

Let Y be a real-valued response variable and X be a covariate or predictor variable that is likely high-dimensional. A standard goal of statistical analysis is to infer the relationship between Y and X . Random Forests grow a large (>500) ensemble of trees using n independent observations (Y_i, X_i) , $i = 1, \dots, n$. Each tree grows via a recursive partitioning of the source set using one predictor variable X . At each step, the source set is split into two subsets following a test on the value of X . When Y is a quantitative variable, the selected test is the one that minimizes the within-subset variance of Y (Breiman et al., 1984). The recursive partitioning is limited by a stopping rule, and the subsets are produced by the last split being the leaves of the tree. The ensemble of trees is produced by using a random sample of the training data and a random subset of the predictor variables for each tree.

For the regression, the prediction $\hat{Y}_\theta(x)$ of a single tree θ of a Random Forest for a new data point x can be represented as the weighted average of the original observations Y_i , $i = 1, \dots, n$:

$$\hat{Y}_\theta(x) = \sum_{i=1}^n w_{\theta i}(x, \theta) Y_i \quad [4]$$

where $w_{\theta i}(x, \theta)$ is the weight vector given by a positive constant that is 1 if the observation

143 Y_i is part of the same leaf and 0 otherwise.

144 By using Random Forests, the prediction is the average prediction of k single trees that were
145 constructed as described above.

$$147 \quad \widehat{Y}_T(x) = \sum_{i=1}^n w_{Ti}(x) Y_i \quad [5]$$

146 [5]

148 with $w_{Ti}(x) = k^{-1} \sum_{t=1}^k w_{\theta_i}(x, \theta)$

149

150 The Random Forest Algorithm has three hyperparameters that must be fitted to obtain the
151 best possible result, i) the number of observations drawn randomly for each tree, ii) the
152 number of variables drawn randomly for each split and iii) the minimum number of samples
153 that a node must contain. For the present study, the ranger package (Wright and Ziegler,
154 2017) was used for running the QRF algorithm in R environment.

155

156 **2.1.2. Regression kriging with Random Forest**

157 In Regression Kriging (RK), the estimated value at an unvisited site $\hat{Z}(s_0)$ is given by
158 summing the predicted value from regression $\hat{m}(s_0)$ and the residuals (i.e. the regression
159 errors) interpolated by ordinary kriging $\hat{e}(s_0)$.

160

$$161 \quad \hat{Z}(s_0) = \hat{m}(s_0) + \hat{e}(s_0) \quad [6]$$

162

163 In this application of RK, the trend function was obtained by calibrating a Random
164 Forest (see above). The residuals were calculated at each observation point. An
165 experimental variogram was then built from the residuals. If the visual inspection of the
166 experimental variogram gave evidence of spatial structure, RK was applied for mapping the
167 STICS soil input and the residuals were further interpolated onto the grid covering the study
168 area by ordinary kriging. Otherwise, the prediction of RF was considered.

169

170 **2.2. Crop modelling with STICS**

171 STICS is a soil-plant simulation model, developed at INRA (France) in 1996 (Brisson et al,
172 2003). STICS can simulate the water, C and N balances of various types of crops, both
173 annual and perennial, herbaceous and woody. STICS represents the plant-soil system
174 dynamics depending on soil characteristics, climate and crop management. It simulates water,
175 nitrogen and carbon balances at a daily time step within the soil-plant system. The soil is
176 modelled using a tipping-bucket approach. The crop development is mainly driven by
177 temperature, while crop biomass accumulation depends strongly on radiation interception by
178 leaves. For water balance, the approach uses potential evapotranspiration (PET) and crop
179 Leaf Area Index (LAI) to determine maximum evaporation and transpiration; the actual ones
180 depending also on the soil water available for root uptake. When requirements are not met,
181 water stress affects crop growth and yield depending on the severity and timing of the
182 stresses. In this study, the impact of N stress on crop was not considered, therefore soil
183 parameters that impact the model output are the depth of the soil layers, and for each layer the
184 bulk density, water content at field capacity and at wilting point and rock fragment
185 proportion. The 0-30cm surface layer is also characterized by its clay content that influences
186 actual evaporation.

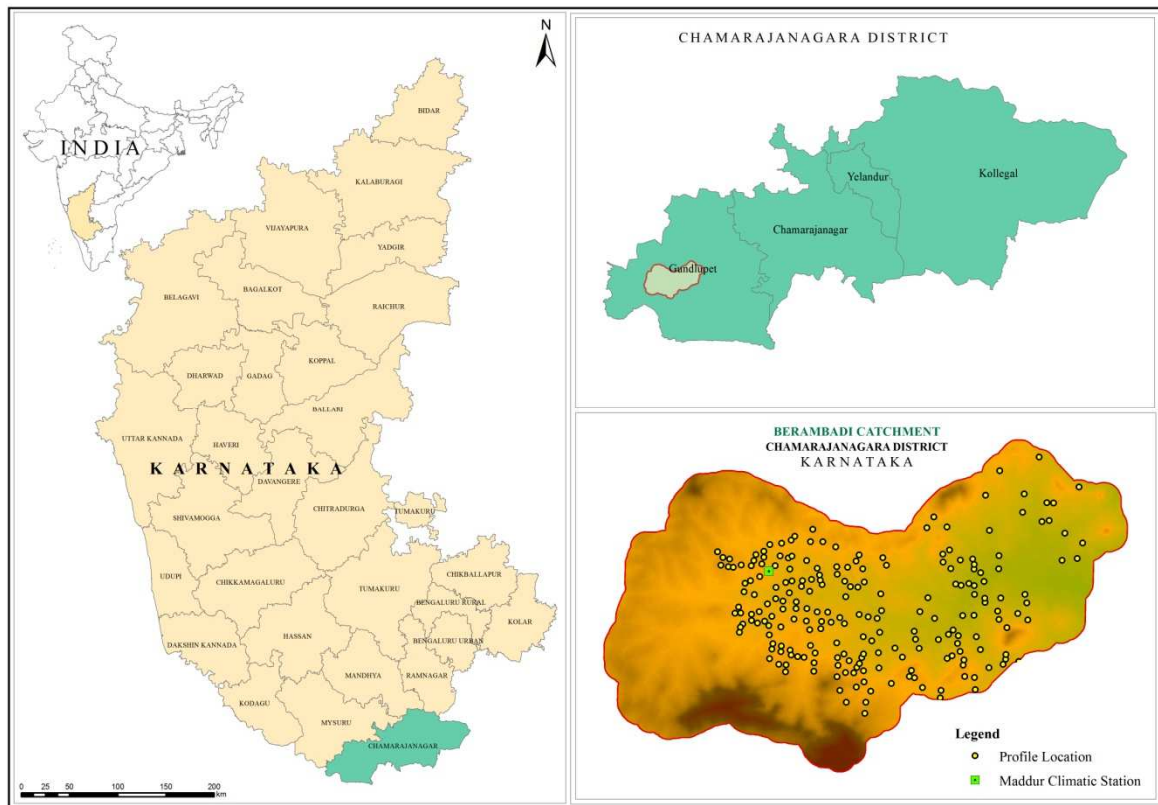
187

188 **2.3. The case study**

189 **2.3.1. Berambadi catchment**

190 The study was carried out in the Berambadi catchment located in the Deccan Plateau
191 of South India (Figure 1), and spread over 84 km² (11°43'49" to 11°48'11" N Latitude and
192 76°32'31" to 76°36'14"). It belongs to the Kabini Critical Zone Observatory (CZO) (Sekhar et
193 al., 2016), the Indian site of the SNO M-TROPICS (<https://mtropics-fr.obs-mip.fr/>), which is
194 part of the OZCAR Research Infrastructure (Gaillardet et al. 2018; <https://www.ozcar-ri.org>).
195 The elevation of the catchment ranges from 720 to 1362 m above mean sea level. The annual
196 precipitation (2005-2018) is 993 mm, in which around 55% rainfall is received during the
197 southwest monsoon (June to September) and 25% during the north-east monsoon (October to
198 December). According to this rainfall distribution, agricultural practices are organized around
199 two main cropping seasons: Kharif (May-August) and Rabi (September-December) (Sharma
200 e al., 2018). Indeed, in the Berambadi catchment, the early start of the Kharif cropping season
201 (before the onset of the SW monsoon) is possible because frequent convective storms occur
202 in April and May. The average annual temperature is 22.6 °C. The soil moisture regime is
203 ustic and soil temperature regime is isohyperthermic. The bedrock is granitic gneiss. Red

204 soils (Alfisols) developed from Granitic parent material are found in uplands and hillslopes
205 while Black soils (vertisols and its intergrades) are found in the valleys (Shivaprasad, 1998).
206



207
208 *Figure 1. Study area, sampling locations (yellow dots) and climatic station (green dot)*

2.3.2. Datasets

2.3.2.1. Soil data

212 We used soil observations from 217 soil profiles in the Berambadi catchment in sites selected
213 for representing at best the soil variations (Figure 1, bottom right). A part of the sites was
214 provided by the Sujala III project (http://watershed.kar.nic.in/SujalaIII_Doc.htm). The spatial
215 sampling was completed by a soil survey carried out during April-May 2017 to fill the gaps.
216 All the soil observations were performed following the protocol that is currently applied in
217 India (NBSS&LUP staff, 2016). Each site was documented with the required soil properties for
218 calculating the considered STICS soil inputs (see details below).

2.3.2.2. Soil Covariates

221 In Digital Soil Mapping, soil properties are predicted from a set of spatial data that are
222 available over the whole study area, namely the soil covariates. A set of candidate soil

223 covariates was provided as input of the DSM regression algorithm (here RF) (Table.1). A
 224 Digital elevation model (DEM) with 10 m resolution was obtained from Cartosat-1 and
 225 processed using ArcGIS10 data management toolbox. The first and secondary derivates of
 226 DEM like elevation, slope, aspect, curvatures (plan and profile) and topographic position
 227 index (TPI), Multi-resolution Index of Valley Bottom Flatness (MrVBF) and Multi-resolution
 228 Ridge Top Flatness (MrRTF) were derived by using Saga-GIS 2.3.1 version. Landform and
 229 soil order are derived from Legacy soil datasets of Chamrajnagar district (1: 50,000 scale).
 230 Normalized Difference Vegetation Index (NDVI) was extracted from satellite data for
 231 modelling. In addition, 10 bands of sentinel-2 data (Date of Acquisition : 24th April 2017)
 232 were used as covariates.
 233

Source	name	Description
Cartosat DEM (10m)	elevation	Elevation
	Slope	Local Slope gradient
	Aspect	Local Slope Aspect
	Northeastness	Sin (Aspect + 45°)
	Northwestness	Cos(Aspect + 45°)
	Profile Curv	Curvature parallel to slope direction
	Plan Curvature	Curvature perpendicular to slope direction
	MRVBF	Multi Resolution Valley Bottom Flatness Index (Gallant & Dowling, 2003),
	MRRTF	Multi Resolution ridge Top Flatness Index (Gallant & Dowling, 2003),
	TPI	Topographical Position Index (mean difference of elevation with neighbouring pixels)
TWI	Topographical Wetness Index (Sorensen et al, 2006) : $\ln a / \tan b$, with a the upstream area and b the slope gradient	
Resourcesat-2 data (5.8m)	NDVI	Normalised Difference Vegetation index
Sentinel 2	Bands 2 to 8A and 11, 12	Different bands from Visible to NIR (10-20 m resolution)
1:50 000 Soil Map (KSRSAC, 2016)	landform	Landform
	Soil Order	USDA soil order

234 *Table 1. Covariates used for DSM Model*

235

236

237 **2.3.2.3. Climate data**

238 Daily values of rainfall, air humidity, wind velocity, maximum, minimum and mean air
239 temperatures, precipitation and global radiation, that are required by STICS, were obtained
240 from an automatic weather station (CIMEL, type ENERCO 407 AVKP from 2005 to 2015,
241 and OTT type WS501 from 2016 to 2018) located in the study area (Figure 1).

242

243 **2.4. The experiment**

244 The two STICS outputs considered to assess the uncertainty when using the soil inputs
245 provided by a DSM approach, crop yield and drainage, were selected for their importance in
246 representing important soil functions (agricultural production and water regulation) and for
247 their expected sensitivity to the soil inputs that could be mapped in the Berambadi catchment.
248 Any other sources of uncertainty of STICS outputs, i.e. related with climate, agricultural
249 practices and modelling errors, were removed from the experiment. In the following, we
250 provide some details about the experiment designed accordingly.

251

252 **2.4.1. The considered STICS soil parameters**

253 The STICS soil layers were selected as those effectively observed by the soil surveyors in the
254 field, namely, from top to bottom, the cultivated layer (A horizon), the horizon with
255 pedogenic processes (B horizon) and the regolith till the occurrence of rock or paralithic
256 contact (C horizon). For each layer, the following STICS soil inputs were considered: horizon
257 thickness, clay content and albedo at the soil surface, bulk density, water content at Field
258 Capacity (FC) and at Permanent wilting point (PWP).

259 Horizon thicknesses were directly derived from the soil profile observations. In the absence
260 of laboratory measurements, clay contents were determined from the observed soil classes by
261 taking the within-soil-class mean that were calculated from 20 soil profiles of the Berambadi
262 catchment having particle size analysis. A similar approach was applied for estimating bulk
263 density from existing datasets in the southern Deccan Plateau. Water contents at FC and PWP
264 were determined from the observed textural classes using the following local pedotransfer
265 function (PTF)s, established by using a dataset of 70 samples from the Berambadi catchment:

$$266 \quad FC_{\text{fine earth}} = -4.7744 + 0.5403 * \text{clay} \quad [7]$$

$$267 \quad PWP_{\text{fine earth}} = -0.9054 + 0.6265 * \text{clay} \quad [8]$$

268 These pedotransfer functions provided FC and PWP as percentages of the fine earth, which
269 may differ from the percentage of the whole soil layer for layers with non-null and variable
270 rock fragment contents. In the STICS model, FC and PWP expressed in percentages of fine
271 earth were converted into FC and PWP expressed in percentage of the whole soil layers,
272 involving rock fragment contents (RFC) as follows:

$$273 \quad FC_{\text{soil layer}} = FC_{\text{fine earth}} \times (100 - \text{RFC})/100 \quad [9]$$

$$274 \quad PWP_{\text{soil layer}} = PWP_{\text{fine earth}} \times (100 - \text{RFC})/100 \quad [10]$$

275

276 The soil albedo was calculated from the Munsell soil color value recorded for the first
277 described soil horizon, using the following PTF, developed by Post et al. (2000):

$$278 \quad \text{Soil albedo (0.3-2.8 } \mu\text{m)} = 0.069(\text{colour value}) - 0.114 \quad (R^2=0.93) \quad [11]$$

279

280 For all the remaining STICS parameters for which no data were available, we used the same
281 values for all the 217 sites. STICS simulations were further designed accordingly to avoid
282 any artefact related with this choice (see below). Finally, the 217 sets of STICS soil
283 parameters derived from the observed soil properties were used to run STICS simulations,
284 considered here as “ground truth”.

285

286 **2.5. Spatial estimations of STICS soil parameters**

287 The STICS soil parameters presented before were also estimated at the 217 sites with the
288 DSM approach, from the set of soil covariates (section 3.2.2.) and using Random Forests
289 (RF) or regression kriging (RK) (section 2.2.). For the STICS soil parameters derived from
290 pedotransfer functions, the soil properties used as inputs of these functions were first
291 estimated and then converted into STICS parameters by applying the corresponding PTFs.
292 Consequently, the sets of estimated STICS parameters at the 217 studied sites were obtained
293 from spatial estimations of the following properties of the three soil layers A, B and C: Soil
294 layer thickness, Clay content, Rock Fragment, Field Capacity and Permanent Wilting point.

295 In order to obtain realistic results, the estimated STICS soil parameters at a given site had to
296 be predicted from a DSM model (RF or RK) that was not calibrated from a dataset that
297 included this site. To fulfil this requirement, a 10-fold cross-validation approach was repeated
298 20 times. At each of the 217 sites, the estimated values of the soil parameters were calculated
299 as the mean of the 20 predicted values obtained when the site was not in the set of sites used
300 for calibrating the DSM model.

301

302 **2.6. STICS scenario simulations**

303 The selected cropping system scenario was maize monoculture over the two main cropping
304 seasons observed in the catchment, namely Kharif (from May to August, comprising early
305 rains from convective storms followed by the South-West monsoon) and Rabi (from
306 September to December, with rains from the North-East monsoon). We considered 14
307 climatic years available from the weather station, from 2005 to 2018. We assumed that
308 considering the large inputs of N fertilizers in the catchment (Buvaneshwari et al., 2017),
309 crops were sufficiently provided in nitrogen and we did not activate the module simulating N
310 stress in the model. We selected Maize monoculture for all the simulations for three main
311 reasons. First, it is very common in the Berambadi catchment and can be cultivated on all the
312 local soil types. Second, as a rainfed crop, maize is expected to be very sensitive to water
313 conditions that are in turn driven by the soil properties – and because of its deep rooting
314 system, it is more sensitive to soil properties at depth than shallow-rooted crops. Last but not
315 least, maize was among the few crops already calibrated in the STICS model for the same
316 area (Sreelash et al., 2017).

317 According to the selected scenario, maize was assumed to be cultivated as follows: in Kharif,
318 maize was sown on the 1st of May, and harvested on the 23rd of August. In Rabi, maize was
319 sown on the 1st of September and harvested on the 19th of December. The simulations were
320 initialized with soils at field capacity on the 1st of January of the first year, then were run
321 continuously over the 14 years to better consider the effect of inter-annual variability of
322 climate on soil water status and its impacts on crop production.

323 We considered 3038 (i.e. 14 years x 217 sites) couples of STICS outputs consisting of yield
324 (t/ha) and drainage (mm). Furthermore, STICS outputs across years were summarized at each
325 of the 217 sites by computing mean values, standard deviations, 25%, and 75% quartiles.

326 All these STICS outputs were computed twice, i.e. from the observed STICS soil parameters
327 (the ground truth) and from the estimated STICS soil parameters (see section 2.5.)

328

329 **2.7. Evaluation protocol**

330 The spatial estimations of the soil properties used to derive STICS soil parameter estimations
331 (see section 2.5) and the STICS outputs (see section 2.6.) were compared to the same values
332 obtained from observed soil properties using classical statistical indicators: Mean error (ME)
333 or bias, Root Mean Square Error (RMSE) and the Mean Square Error Skill Score (SS_{MSE} ,

334 Nussbaum et al., 2018). When positive, SS_{MSE} expresses the percentage of explained variance
 335 by the predictive model. The definitions of these statistical indicators are provided below:

$$336 \quad ME = \frac{\sum_i^n (P_i - O_i)}{n} \quad (12)$$

$$337 \quad RMSE = \sqrt{\frac{\sum_i^n (P_i - O_i)^2}{n}} \quad (13)$$

$$338 \quad SS_{mse} = \frac{\sum_i^n (P_i - O_i)^2}{\sum_1^n (O_i - \bar{O})^2} = \frac{RMSE^2}{Variance(O)} \quad (14)$$

339

340 With O the observed value and P the predicted value

341 Besides this classical evaluation method, we also evaluated the DSM models for their ability
 342 to conserve the relationships between the parameters that were used together in the STICS
 343 simulations. Indeed, predicting separately each STICS soil parameters conveyed the risk of
 344 producing “pedo-chimeras” (e.g. a horizon with PWP greater than FC) that could have
 345 hampered the simulation results. The predicted correlations between STICS soil inputs were
 346 then evaluated by comparing them with the observed ones for each couple of parameters.

347

348 **3. Results**

349

350 **3.1. STICS soil parameters**

351 **3.1.1. Observed soil data**

352 Basic statistics of the STICS soil parameters as estimated from observations or from their
 353 pedotransfer function inputs are presented in Table. 2. The most variable soil properties
 354 across the Berambadi catchment ($CV > 90\%$) were the rock fragment contents of B and C
 355 horizons, the thickness of the C horizons and the soil albedo. The least variable soil
 356 properties ($CV < 30\%$) were the thickness of A horizons and the Field capacity of B and C
 357 horizons. The variations of the STICS soil parameters that were directly observed on the soil
 358 profiles (horizon thicknesses, rock fragment contents) were globally higher than those of the
 359 STICS soil parameters derived from the pedotransfer functions (Clay, FC, PWP).

360

361 *Table 2. Summary statistics of STICS soil parameters*

Soil properties	Layer					CV (%)
	Min	Max	Mean	Stdev		
Clay content (%)	A	13.4	49.4	27.3	13.3	48.7

Soil albedo	A	0.04	0.30	0.14	0.05	100.0
Thickness (cm)	A	6.0	43.0	17.8	5.0	28.1
	B	8.0	240.0	87.0	59.0	67.8
	C	0	140.0	28.0	27.0	96.4
Water at field capacity (%vol)	A	11.4	32.4	19.6	7.7	39.3
	B	11.4	32.4	21.9	5.8	26.5
	C	11.4	32.4	19.7	4.7	23.9
Water at wilting point (%vol)	A	3.8	19.3	9.8	5.7	58.2
	B	3.8	19.3	11.6	4.8	41.4
	C	3.8	19.3	9.9	3.5	35.4
Rock fragment (%)	A	0.0	50.0	18.4	10.9	59.2
	B	0.0	80.0	19.6	20.6	105.1
	C	0.0	80.0	21.1	22.8	108.1

362

363

364

365

366

367 *Table 3. Performance of digital soil mapping approach and conventional soil mapping approach*368 *(between parenthesis) in prediction of different soil inputs. RK and QRF mean regression Kriging and*369 *Random Forest Respectively.*

Soil properties	Layer	Model	SS_{MSE}	RMSE	Bias
Clay content (%)	A	RK	0.36 (0.18)	10 (12)	-0.03
Soil albedo	A	RK	0.26 (0.1)	0.04 (0.05)	0.0002
Thickness (cm)	A	RK	0.06 (0.1)	4 (4.65)	-0.01
	B	RK	0.40 (0.08)	45 (57)	-0.03
	C	RK	0 (0.04)	28.5 (26)	-0.07
Water at field capacity (%vol)	A	QRF	0.37 (0.16)	6.1 (7.3)	-0.01
	B	QRF	0.33 (0.06)	4.8 (6.1)	-0.12
	C	QRF	0.0 (0.08)	4.8 (11.7)	-0.79
Water at wilting point (%vol)	A	QRF	0.36 (0.17)	4.5 (5.6)	0.03
	B	QRF	0.36 (0.09)	3.8 (4.7)	0.03
	C	QRF	0 (0.07)	3.6 (6.3)	-0.55
Rock fragment (%)	A	RK	0.37 (0.13)	9 (10)	-0.004
	B	RK	0.35 (0.19)	16 (19)	-0.01
	C	RK	0.41(0.25)	17 (20)	-0.002

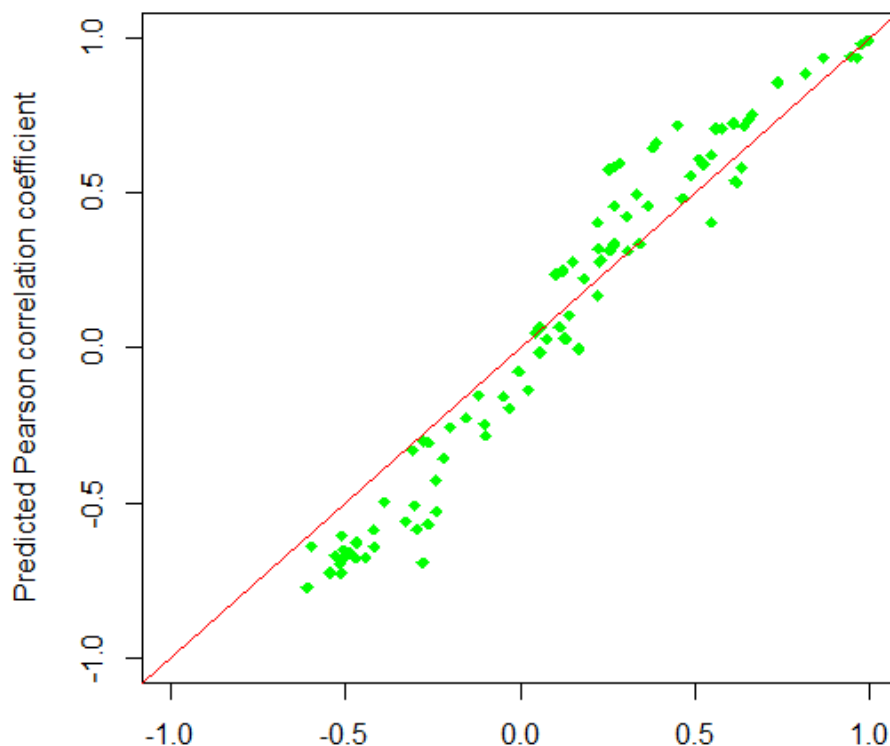
370

371

372

3.1.2. Estimated soil data by DSM

373 Random forest and Regression Kriging were applied to build prediction functions for
374 all STICS soil parameters. Their evaluations over the 217 sites with soil observation using a
375 ten-fold cross validation repeated 20 times (section 2.5.) allowed to select, for each STICS
376 soil parameters, the best function. Regression Kriging model was selected for 8 out of the 14
377 mapped STICS soil parameters. For such parameters, the spatial interpolation of spatially
378 structured residuals after RF predictions improved the results.
379 Most of the STICS soil parameters were predicted with negligible biases and moderate
380 accuracy (between 33% and 41% of explained variance for 69 % of STICS outputs). The
381 worst predictions (percentage of explained variance < 2%) were obtained for hydric
382 properties of layer C and for the thickness of layer A. It is worth noting that these parameters
383 also had the lowest standard deviations. The predictions of soil albedo had intermediate
384 performances (26% of explained variance). Nevertheless, for all soil parameters, the
385 comparisons with the R^2 and RMSE values obtained when using the mean values per soil
386 mapping unit (digits between parenthesis in table 3) showed clearly that the DSM predictions
387 clearly outperformed the soil map ones.



388
389
390
391 *Figure 2. Observed correlation coefficient vs predicted correlation coefficient for each couple of*
392 *predicted parameters*

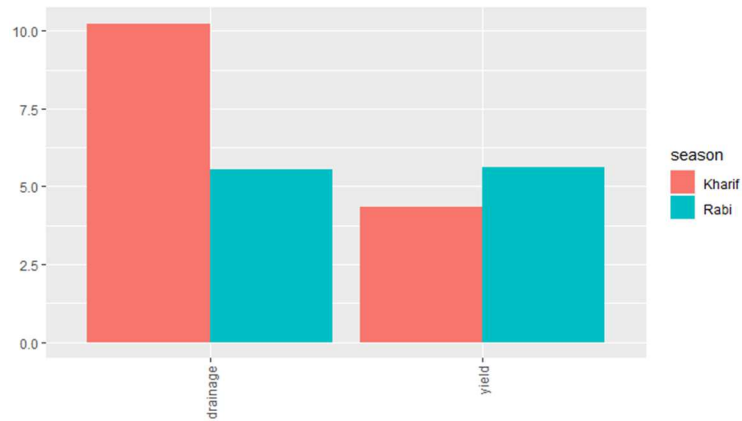
393
394 The comparisons between observed and predicted correlations between STICS soil inputs
395 revealed an overall agreement (Figure 2). This means that, although the soil properties were
396 predicted separately by different DSM models, the holistic vision of soils at each site was
397 preserved without creating “pedo-chimeras” having unrealistic associations of soil property
398 values.

399

400 **3.2. STICS output statistics**

401 We first analysed the STICS outputs obtained from observed soil inputs. Overall, the global
402 impact of soil variations on the simulated results was weak to moderate depending on the
403 output variables and seasons (figure 3). As expected, the impact of soil variability on yield
404 was strongly correlated with the impact of soil variation on water stress (figures S1 and S2,
405 supplementary materials). The mean over years of the standard deviations due to soil
406 variations varied from about 2.5% to 10% of the global mean. However, results plotted per
407 year (figure 4) show that this impact was extremely variable depending on the year. It was
408 very low for yield in Kharif 2018 (standard deviation of 0.004 t ha^{-1}) whereas it was very
409 high in Rabi 2017 (standard deviation of more than 0.8 t ha^{-1}). This impact was very low for
410 drainage in Kharif 2011 (standard deviation of about 1 mm), whereas it was very high in
411 Kharif 2007 (standard deviation of more than 25 mm in Kharif 2007). This means that, for
412 some years, the levels of yield and drainage largely depended on the soil characteristics.
413 Characterizing the years for which soil properties had a larger impact was not
414 straightforward: while there was a clear relationship between the water balance and the mean
415 and median values of yield and drainage (figure 4), it was not the case with their standard
416 deviation (reflecting the impact of variations in soil properties). The effect of soil properties
417 depended on complex intra-seasonal dynamics simulated by the model, as illustrated for the
418 years 2005 and 2011 - (Figure S3, supplementary materials), which, despite similar seasonal
419 water budgets, displayed contrasted impacts of soil properties on crop yield and drainage.

420



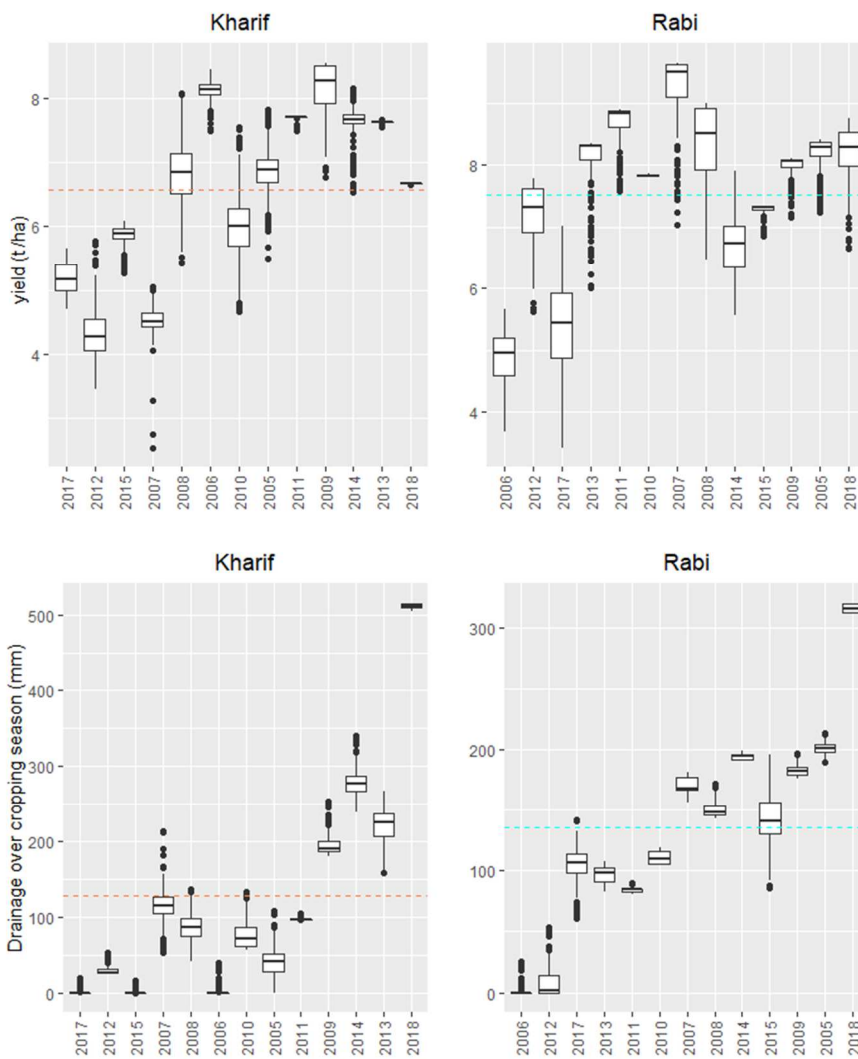
421

422 *Figure 3. Mean of standard deviations over all years of simulation divided by the global mean (in %)*
 423 *for the different simulated variables.*

424

425

426

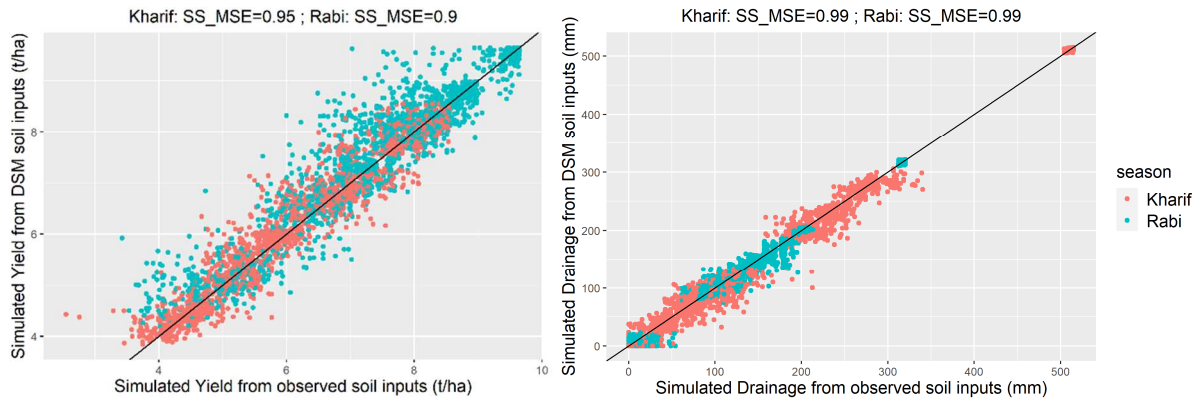


427

428

429 *Figure 4. Variations with respect to soil inputs of simulated yield (above) and drainage over the*
430 *growing season (below), per years, ordered from lowest to highest water availability, calculated as:*
431 *Soil Water Content at sowing + (Rainfall – PET) over the growing season.*

432
433



434 *Figure 5. Simulated variables obtained from DSM soil inputs versus simulated variables obtained*
435 *from observed soils. SS_MSE stands for Mean Square Error Skill Score (eq. 15).*

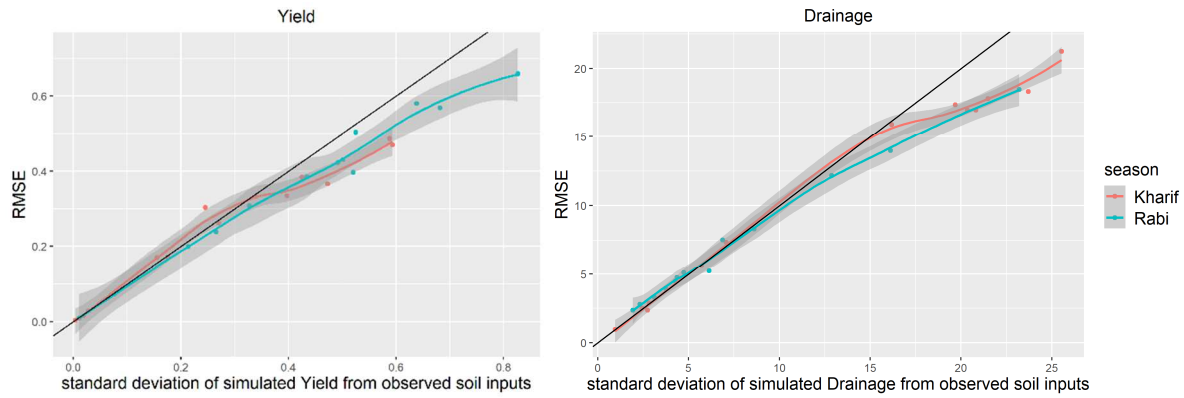
436
437
438

3.3. DSM error impacts on STICS outputs

439
440 Errors induced by DSM estimations of STICS soil inputs were globally low with respect to
441 the variations of the simulated variables (see R^2 and relative RMSE and biases given in % in
442 figure 5). However, some of the yields simulated by STICS were far from the ones obtained
443 with observed soil parameters.

444
445 To analyse this in more details, Figure 6 shows the impact of soil parameters variations (i.e.
446 standard deviations per year of simulated results obtained on observed soils) on the
447 simulation errors (RMSE per years). The RMSE was directly linked with the standard
448 deviations of the simulated reference: the higher the impact of the soil, the higher the impact
449 of soil parameters errors. For the years with the lowest soil impacts (left parts of the graph),
450 the DSM errors on STICS outputs were close to their standard deviations, which means that
451 DSM did not provide any insight on their spatial variation. However, beyond a given
452 threshold of variation, the DSM errors were lower than standard deviations which means that
453 the variations of STICS outputs related with the soil parameters were partially captured by
454 DSM estimations as soon as they became important.

455



456

457

458 *Figure 6. RMSE per year VS standard deviation of results obtained on observed soil per year. Left :*

459 *yield (T/ha), right: drainage (mm)*

460

461 3.4. Input-output errors correlations

462 We performed correlation analysis between residuals of STICS soil parameters and
 463 STICS outputs, to identify the soil parameters for which errors have the larger impact on the
 464 error of STICS outputs. Clay content on layers B and C and rock fragments were added to the
 465 analysis, although they were not direct STICS soil parameters, because they were involved in
 466 the PTF used for calculating FC and PWP, and thus were expected to strongly impact the
 467 STICS outputs values (see 4.1.).

468 The correlations between the residuals of STICS parameters and the STICS outputs
 469 were highly variable (Table 4). The water retention properties errors (water content at field
 470 capacity and wilting point) exhibited greater correlations than rock fragment errors and
 471 horizon depths errors, for all STICS outputs and the horizons. The errors on the B horizons
 472 parameters were the most correlated with STICS outputs, whatever the seasons and the
 473 outputs, with the noticeable exception of the rock fragment that exhibited a better correlation
 474 in the A horizon than in the B one. It is also apparent that the errors on drainage during the
 475 Rabi season were less correlated with STICS soil inputs than the other STICS outputs.

476 *Table 4. Correlations between absolute errors on STICS outputs for different seasons and absolute*
 477 *errors on soil STICS inputs*

Soil properties	Layer	Kharif		Rabi	
		Yield	Drainage	Yield	Drainage
Clay content (%)	A	0.38**	-0.53**	0.53**	0.06
Thickness (cm)	A	0.12	-0.11	0.09	-0.09
	B	0.37**	-0.40**	0.36**	-0.20**
	C	0.06	-0.03	-0.01	-0.07

Water at field capacity (%vol)	A	0.39**	-0.53**	0.54**	-0.04
	B	0.68**	-0.72**	0.73**	-0.44**
	C	0.40**	-0.39**	0.40**	-0.28**
Water at wilting point (%vol)	A	0.38**	-0.53**	0.53**	0.06
	B	0.58**	-0.63**	0.64**	-0.35**
	C	0.41**	-0.40**	0.41**	-0.29**
Rock fragment (%)	A	-0.21**	0.25**	-0.23**	0.00
	B	-0.11	0.10	-0.08	0.06
	C	-0.04	0.05	-0.07	0.01
Soil albedo	A	-0.12	0.20**	-0.22**	0.03

478

479

3.5. DSM error impacts on pooled STICS outputs per sites

480

481 The two STICS outputs were aggregated to provide their observed vs simulated distributions
 482 across the simulation period for each of the 217 sites. Different statistical indicators were
 483 used to describe these distributions outputs: mean, variance, 25% quartile and 75% quartile.
 484 Because of their contrasted climatic conditions, the results obtained for the two cropping
 485 seasons – Kharif and Rabi – were considered separately.

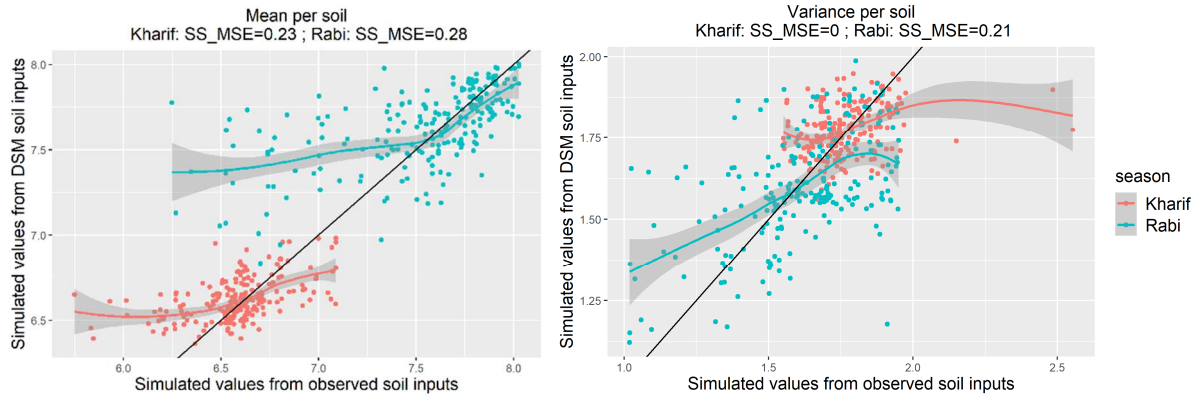
486 Figures 8 and 9 show contrasted results across seasons, STICS outputs, and statistical
 487 indicators with a limited range of R^2 (from 0.05 to 0.40) similar to those obtained for STICS
 488 soil parameters (Table 3). The scatterplots clearly show that the overall variability across sites
 489 was underestimated when the DSM estimated soil parameters were used instead of the
 490 observed ones. The lower values of all statistical parameters for yield were particularly
 491 underestimated showing that DSM did not simulate soils as limiting as to the observed ones
 492 (Figure 7). Conversely, the statistical parameters of drainage estimations were generally
 493 overestimated (Figure 8)

494

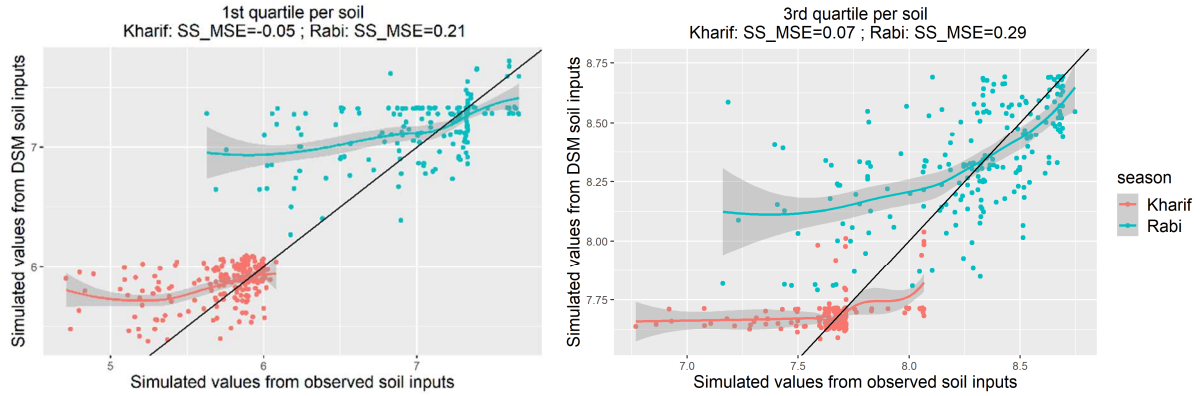
495

496

497



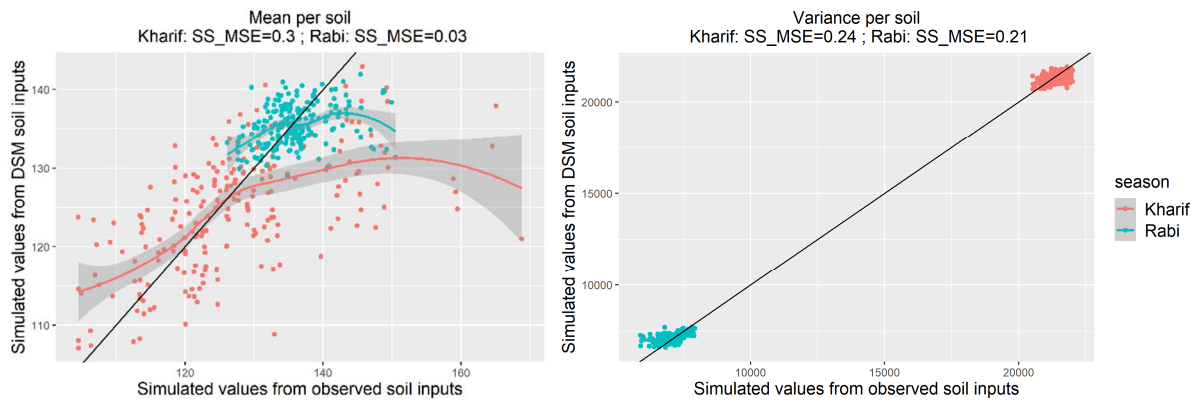
498

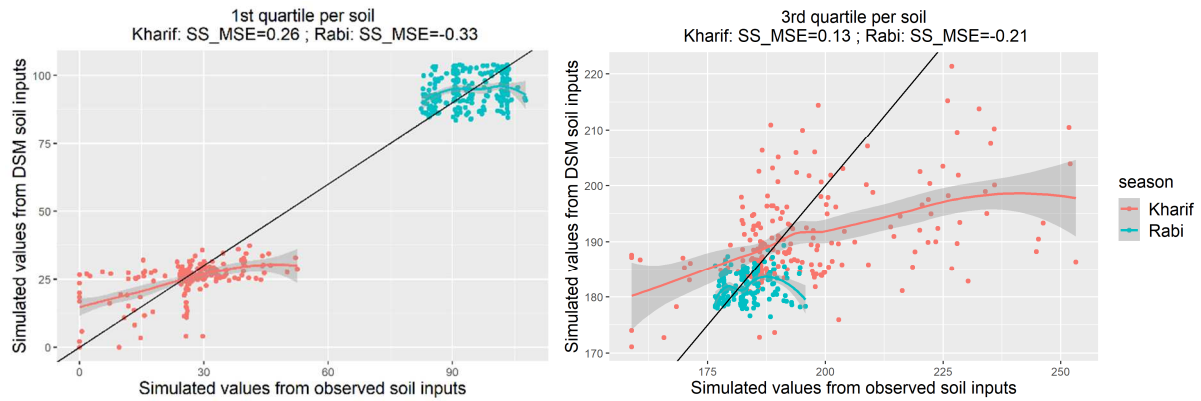


499 *Figure 7. Mean, variance, 1st and 3rd quartile computed per soil, i.e. over years, of YIELD values*
 500 *simulated from DSM soil inputs VS equivalent but computed on values simulated from observed soils.*
 501 *SS_MSE stands for Mean Square Error Skill Score (eq. 15).*

502
 503
 504
 505
 506
 507

508





509
 510 *Figure 8. Mean, variance, 1st and 3rd quartile computed per soil, i.e. over years, of drainage values*
 511 *simulated from DSM soil inputs VS equivalent but computed on values simulated from observed soils.*
 512 *SS_MSE stands for Mean Square Error Skill Score (eq. 15).*
 513

514 All the statistical indicators of yield were better predicted in Rabi seasons than in Kharif
 515 season. Drainage showed opposite results with better performances for predicting statistical
 516 indicators in Kharif season than in Rabi seasons, with, here again, the noticeable exception of
 517 variance that did not exhibit significant differences between seasons. Mean and variance were
 518 generally better predicted than quartiles, which confirmed that DSM could not generate
 519 STICS parameters corresponding to extreme soil properties.

520
 521

522 4. Discussion

523

524 4.1. Limiting factors of DSM performances

525 The present model captured only a part of the spatial variability for most of the soil
 526 properties. These results were not unexpected when looking at the DSM literature. For
 527 example, De Carvalho Jr et al. (2014) and Nussbaum et al. (2018) obtained SS_{MSE} values
 528 between 0 and 0.37 and between 0.08 and 0.28 respectively for particle size and rock
 529 fragment content predictions over Brazilian and Swiss areas of comparable extents and spatial
 530 densities of soil observations.

531 For all soil properties, the mapping errors were first due to the important sources of
 532 uncertainties that could affect the soil observations.

- 533
- 534 • Most of the observations were deduced from field determinations, without
 quantitative soil analysis, which induced uncertainty.
 - 535 • The observations were affected by the variations of soil properties that can be
 536 observed at the profile scale, and that could be imperfectly described by a single
 537 value for a given soil property.

- 538
- the sampling of soil observations was too sparse to capture the part of the soil variations that occur at short distances. Indeed, as shown by several authors (Somarathna et al., 2017, Wadoux et al., 2018, Lagacherie et al., 2020), the spatial density of soil observations strongly impacts the performances of the DSM models.
- 541
- the presence of residuals after random forest analysis that justified the use of RK showed that the covariables used for explaining the variability of the soil input data were also not sufficient. New soil covariates should be added to capture the variability of soil properties better.
- 542
- 543
- 544
- 545

546 Finally, it should be noticed that these performances were measured for a prediction of soil properties at a point level, which is rarely required by end-users. Spatial aggregation of such predictions at larger spatial supports that make sense for user's decision, (e.g. fields) would dramatically decrease such error (Bishop et al., 2015; Vaysse et al., 2017).

547

548

549

550 Although the DSM estimations of soil parameters had important errors, it must be noticed that all these DSM estimations largely outperformed the spatial estimations of STICS soil parameters obtained from the existing soil map of the Berambadi (table 3). Furthermore, despite performing separate estimations of individual soil parameters and soil layers, we observed that the correlations between soil properties and layers were correctly predicted (Figure 2), which prevented from predicting pedo-chimera. This leads to prefer Digital Soil Mapping products to classical soil maps as sources of soil parameters for spatialising soil models.

551

552

553

554

555

556

557

558

559 **4.2. Impacts of soil properties on STICS outputs.**

560 Our experiment showed that the overall impact on yearly STICS outputs of soil variations in the Berambadi catchment was low (Figure 5). This result can be explained by the fact that, on one hand, the soils of the Berambadi catchment were quite deep (about 130 cm when including layer C, see table 2), and, on the other hand, because maize is a relatively deep-rooted crop, both factors buffering the soil-related variability in STICS output under ordinary climate conditions observed in the study area. Using a shallow-rooted crop in our simulations would have probably increased the impact of soil properties of the shallow horizons but it would have reduced even more the impact of deeper ones.

561

562

563

564

565

566

567

568 Interestingly, for few simulation years, soil variations had a greater impact on STICS results (Figure 4). These impacts of soils on crop yields could be significant enough to dramatically impact farmers' livelihood, considering the socio-economic vulnerability of farming systems

569

570

571 in semi-arid regions of India (Singh et al., 2019). Moreover, we saw that such situations were
572 linked with the complex dynamics of rain distribution and crop needs within the season,
573 particularly the occurrence of drought periods early in the cropping season. Indeed, such
574 situations might become more frequent in the future, as current projections for the Indian
575 monsoon indicates increasing variability in monsoon rainfall and more frequent and severe
576 drought spells (e.g. Sharmila et al., 2015)

577 This suggests that, in our case study, while crop models do not necessarily require an
578 accurate representation of the distribution of soil properties for predicting the variations of
579 yield and drainage across years, such information could become crucial for assessing the
580 resilience of cropping systems in a changing climate.

581

582 **4.3. Impact of DSM produced soil parameters errors**

583 Our results showed that the spatialisation errors on the STICS soil parameters had an
584 overall low impact on STICS simulation errors (figure 5) because of the low overall impact
585 of soil variations on STICS outputs. However, a few simulations were strongly impacted by
586 soil variations (see above), which in turn induced, for these situations, stronger impacts of
587 DSM errors on the STICS outputs (figure 6). These impacts propagated to the statistical
588 indicators that describe the behaviour of the sites across years (figure 7). It is interesting to
589 note that the errors obtained on these statistical indicators were of the same order of
590 magnitude as the spatialisation errors of soil parameters by DSM (table 3), showing neither
591 smoothing nor amplifications of these spatialisation errors. Although the impact of soil
592 estimations on crop model outputs have already been measured when using classical
593 choropleth soil maps (Leenhardt et al., 1994b; Constantin et al., 2019; Hoffmann et al.,
594 2016), to our knowledge, this was not been done yet when using DSM products. The level of
595 error obtained with DSM being comparable to many DSM applications described in the
596 literature (see 5.1.), this study can be considered a first reference for the use of crop
597 modelling from DSM products. However, more studies will be necessary to get a complete
598 picture of the interest and limitation of this new source of soil data.

599 Yet, this study only addressed a part of such errors. First, only perennial morphological and
600 physical soil properties were addressed. The chemical soil properties that are also STICS soil
601 inputs (i.e. pH, organic nitrogen) need to be considered in future studies. Furthermore,
602 Permanent Wilting Point (PWP) and Field Capacity (FC) were not directly measured in the
603 field but estimated from pedotransfer functions (PTF). The errors associated with these

604 functions were not considered in this testing since the “observed” values of STICS outputs
605 were produced from inputs that used also these functions. However, it has been shown
606 (Roman-Dobarco et al., 2019) that the impact of PTF errors on the mapping of Soil Available
607 Water Capacity was much lower than the mapping errors and thus could be neglected in most
608 situations.

609

610 **4.4. Most impacting soil characteristics**

611 For the years for which soil variations matter (see the previous section), the same soil
612 parameter errors affected STICS outputs differently at different seasons (Table 4).
613 Correlation analysis also revealed that some STICS soil parameters errors affected the STICS
614 outputs more than others. In some cases, these differences can be easily explained. For
615 example, the predominant impact of the parameter related to the layer B on STICS parameter
616 errors is probably because it represents the main part of the soil (87 cm against 18 and 28 cm
617 for layer A and C, respectively), and the main source of water for root uptake across the
618 cropping season. As the STICS model simulates a progressive increase of rooting depth
619 across the vegetative phase of the crop, it was also expected that the parameters of horizon B
620 related to the water holding capacity (FC and WP) would have had more impact on STICS
621 outputs than its thickness. However, other differences are more difficult to interpret, such as
622 the absence of impact of soils parameters on drainage for the Rabi season or the significant
623 impact of rock fragments on both drainage and yield for horizon A but not for horizon B.

624 Overall, these results suggest that the impact of errors associated with simulated soil
625 characteristics on the studied outputs, as modelled by STICS, is particularly complex.
626 Sensitivity analysis methods may help in better understanding these impacts (Varella et al.,
627 2010). Estimating relevant sensitivity indices would require considering the multivariate
628 distributional properties of these errors and, in particular, the potentially complex dependency
629 structure between them. This is still a challenge in the sensitivity analysis (Razavi et al, 2021)
630 beyond the scope of this study.

631

632 **4.5. Coupling DSM and crop models**

633 Our study constitutes the first step in spatializing soil models using Digital Soil Mapping.
634 Maps of statistical parameters of STICS output could be provided to users by using as STICS
635 inputs the predicted parameters obtained by the DSM algorithms at the nodes of a regular
636 grid covering the study area. Such maps would constitute an interesting alternative to the

637 current maps of production functions of the soil that were derived from static combinations of
638 soil properties (Kidd et al. 2015, Harms et al., 2015). Indeed, using a crop model to
639 characterize such soil function would allow modulating soil assessment according to the
640 spatial variations of other factors than soil, e.g. climate and agricultural practices (Ellili
641 Bargaoui et al, 2021).

642 Apart from providing better soil parameter estimations than the ones provided by classical
643 soil maps, DSM has the advantage of producing ex-ante estimations of the uncertainty of the
644 soil property estimations (Heuvelink, 2014). Following error propagation techniques, such
645 estimations could be propagated through STICS simulations (Vaysse et al; 2017). This would
646 allow identifying the critical locations where the soil data are insufficient to estimate the
647 required STICS soil parameters.

648

649 **5. Conclusions**

650

651 A first experiment of spatializing a crop model (STICS) using Digital Soil mapping was
652 carried out in the Berambadi Catchment (India). The main lessons of this experiment are the
653 following:

- 654 • Digital Soil mapping outperformed soil maps for spatializing STICS soil parameters
655 while preserving correlation between soil parameters.
- 656 • Although the impact of DSM estimated soil parameters errors on yearly STICS
657 outputs were on average low, some particular years exhibited strong differences
658 between STICS outputs derived from observed vs DSM estimated soil parameters.
659 The statistics of STICS simulations across years at each site giving a view of the crop
660 production potential were also sensitive to DSM mapping errors
- 661 • The impact of DSM errors was variable across the involved soil parameters. Whereas
662 some differences of impact could be easily explained, others would need a more
663 thorough sensitivity analysis.
- 664 • Coupling DSM with a crop model represents an interesting alternative to the classical
665 Digital Soil Assessment techniques. As such, it will deserve more work in the future.

666

667

668

669 **6. Acknowledgments**

670

671 The study was funded by the project ATCHA ANR-16-CE03-0006, and the Environmental
672 Research Observatory M-TROPICS (<https://mtropics.obs-mip.fr/>), itself supported by the
673 University of Toulouse, IRD and CNRS-INSU. We thank the SUJALA III project (Karnataka
674 Watershed Development Department and World Bank) for their support. We also thank the
675 farmers of the Berambadi catchment, for their kindness and support.

676

677

678 **7. References**

679

680 Adhikari, K., Greve, M.B., Bocher, P.K., Malone, B., Minasny, B., McBratney, A.B., Greve, M.,
681 2013. High-Resolution 3-D Mapping of Soil Texture in Denmark. *Soil Sci. Soc. Am. J.* 860–
682 876.

683 Adhikari, K., Hartemink, A.E., 2016. Linking soils to ecosystem services - A global review.
684 *Geoderma* 262, 101–111.

685 Arrouays, D., Grundy, M.G., Hartemink, A.E., Hempel, J.W., Heuvelink, G.B.M., Hong,
686 S.Y., Lagacherie, P., Lelyk, G., McBratney, A.B., McKenzie, N.J., Mendonca-Santos,
687 M.D., Minasny, B., Montanarella, L., Odeh, I.O.A., Sanchez, P.A., Thompson, J.A.,
688 Zhang, G.-L., 2014. GlobalSoilMap. Toward a Fine-Resolution Global Grid of Soil
689 Properties, *Advances in Agronomy*.125.

690 Ballabio, C., Lugato, E., Fernández-ugalde, O., Orgiazzi, A., Jones, A., Borrelli, P.,
691 Montanarella, L., Panagos, P., 2019. Mapping LUCAS topsoil chemical properties at
692 European scale using Gaussian process regression. *Geoderma* 355, 113912.

693 Beaujouan V., Durand P., Ruiz L., Modelling the effect of the spatial distribution of
694 agricultural practices on nitrogen fluxes in rural catchments, *Ecol. Model.* 137 (2001) 93–
695 105.

696 Bishop, T.F.A., Horta, A., Karunaratne, S.B., 2015. Validation of digital soil maps at
697 different spatial supports. *Geoderma* 241–242, 238–249.

698 Breiman, L., 2001. Random Forests. *Mach. Learn.* 45, 5–32.

699 Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and regression

700 trees. CRC press.

701 Brisson,N., Gary,C.. Justes,E., Roche,R., Mary,B., Ripoche,D. Zimmer, D., Sierra,J.,
702 Bertuzzi,P., Burger, Bussiere,F., Cabidoche,Y.M., Cellier,P., Debaeke, P.,
703 Gaudillere,J.P., Henault, C., Maraux, F., Seguin, B., Sinoquet,H., 2003. An overview of
704 the crop model STICS. *European Journal of Agronomy*, 2003, 18 (3-4), 309-332.

705 Buvaneshwari S, Riotte J, Sekhar M, Mohan Kumar MS, Sharma AK, Duprey JL, Audry S,
706 Giriraj PR, Yerabham P, Moger H, Durand P, Braun JJ and Ruiz L (2017). Groundwater
707 resource vulnerability and spatial variability of nitrate contamination: insights from high
708 density tubewell monitoring in a hard rock aquifer. *Science of the Total Environment*, 579,
709 838-847. <http://dx.doi.org/10.1016/j.scitotenv.2016.11.017>

710 Caruana, R., Niculescu-Mizil, A., 2006. An empirical comparison of supervised learning
711 algorithms. *Proc. 23rd Int. Conf. Mach. Learn. - ICML '06* 161–168.

712 Carvalho-Jr (de), W., Lagacherie, P., Da-Silva-Chagas, C., Calderano-Filho, B., Bhering,
713 S.B., 2014. A regional-scale assessment of digital mapping of soil attributes in a tropical
714 hillslope environment. *Geoderma* 232–234, 479–486.

715 Constantin, J., Raynal, H., Casellas, E., Hoffmann, H., Bindi, M., Doro, L., Eckersten, H.,
716 Gaiser, T., Grosz, B., Haas, E., Kersebaum, K.-C., Klatt, S., Kuhnert, M., Lewan, E.,
717 Maharjan, G.R., Moriondo, M., Nendel, C., Roggero, P.P., Specka, X., Trombi, G., Villa,
718 A., Wang, E., Weihermüller, L., Yeluripati, J., Zhao, Z., Ewert, F., Bergez, J.-E., 2019.
719 Management and spatial resolution effects on yield and water balance at regional scale in
720 crop models. *Agric. For. Meteorol.* 275, 184–195.
721 <https://doi.org/10.1016/j.agrformet.2019.05.013>

722 Deryng D, Elliott J, Folberth C, Müller Pugh CTAM, Boote KJ, Conway D, Ruane AC,
723 Gerten ,D, Jones JW, Khabarov N, Olin S, Schaphoff S, Schmid E, Yang H, Rosenzweig
724 C (2016) Regional disparities in the beneficial effects of rising CO2 concentrations on
725 crop water pro- ductivity. *Nat Clim Chang* 6(8):786–790

726 Ellili-bargaoui, Y., Walter, C., Lemercier, B., Michot, D., 2021. Assessment of six soil
727 ecosystem services by coupling simulation modelling and field measurement of soil
728 properties. *Ecol. Indic.* 121, 107211.

- 729 Faivre, R., Leenhardt, D., Voltz, M., Benoit, M., Papy, F., Dedieu, G., Wallach, D., 2004.
730 Spatialising crop models. *agronomie* 24, 205–217.
- 731 Gaillardet, J., Braud, I., Hankard, F., Anquetin, S., Bour, O., Dorfliger, N., ... & Zitouna, R.,
732 2018. OZCAR: The French network of critical zone observatories. *Vadose Zone*
733 *Journal*, 17(1), 1-24.
- 734 Gallant, J.C., Dowling, T.I., 2003. A multiresolution index of valley bottom flatness for
735 mapping depositional areas. *Water Resour. Res.* 39.
- 736 Ginaldi, F., Bajocco, S., Bregaglio, S., Cappelli, G., 2019. Spatializing crop model for
737 sustainable agriculture, in: Farooq, M., Pisante, M. (Eds.), *Innovations in Sustainable*
738 *Agriculture*. Springer Nature Switzerland, pp. 599–620.
- 739 Harms, B., Brough, D., Philip, S., Bartley, R., Clifford, D., Thomas, M., Willis, R., Gregory,
740 L., 2015. Digital soil assessment for regional agricultural land evaluation. *Global Food*
741 *Security*, *Global Food Security* 5, 25–36.
742 <https://doi.org/https://doi.org/10.1016/j.gfs.2015.04.001>
- 743 Hartkamp A.D., White J.W., Hoogenboom G., *Interfacing Geo-graphic Information*
744 *Systems with Agronomic Modeling: AReview*, *Agron. J.* 91 (1999) 761–772.
- 745 Hengl, T., Heuvelink, G.B.M., Stein, A., 2004. A generic framework for spatial prediction of
746 soil variables based on regression-kriging. *Geoderma* 120, 75–93.
- 747 Hengl, T., De Jesus, J.M., Heuvelink, G.B.M., Gonzalez, M.R., Kilibarda, M., Blagotić, A.,
748 Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A.,
749 Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I.,
750 Mantel, S., Kempen, B., 2017. SoilGrids250m: Global gridded soil information based on
751 machine learning. *PLoS One* 12, 1–40.
- 752 Heuvelink, G.B.M., 2014. Uncertainty quantification of GlobalSoilMap products, in:
753 Arrouays, D., McKenzie, N.J., Hempel, J.W., Richer de Forges, A., McBratney, A.B.
754 (Eds.), *GlobalSoilMap: Basis of the Global Spatial Soil Information System - Proceedings*
755 *of the 1st GlobalSoilMap Conference2014*. CRC press/Balkema, London, pp. 335–340.

756 Hoffmann, H., Enders, A., Siebert, S., Gaiser, T., Ewert, F., 2016. Climate and soil input data
757 aggregation effects in crop models. Havard Database V3.
758 <https://doi.org/https://doi.org/10.7910/DVN/C0J5BB>

759 Jenny, H., 1941. Factors of Soil Formation, A System of Quantitative Pedology. McGraw-
760 Hill, New York.

761 Jones JW, Hoogenboom CH, Porter KJ, Boote WD, Batchelor LA, Hunt PWW, Singh U,
762 Gijssman AJ, Ritchie JT (2003) The DSSAT cropping system model. *Eur J Agron* 18(3–
763 4):235–265

764 Keating BA, Carberry PS, Hammer GL, Probert ME, Robertson MJ, Holzworth D, Huth NI,
765 Hargreaves JNG, Meinke H, Hochman Z, McLean G, Verburg K, Snow V, Dimes JP,
766 Silburn M, Wang E, Brown S, Bristow KL, Asseng S, Chapman S, McCown RL,
767 Freebairn DM, Smith CJ (2003) An overview of APSIM, a model designed for farming
768 systems simulation. *Eur J Agron* 18:267–288

769 Kidd, D., Webb, M., Malone, B., Minasny, B., Mcbratney, A., 2015. Geoderma Regional
770 Digital soil assessment of agricultural suitability , versatility and capital in. *GEODRS*,
771 *GEODRS* 6, 7–21. <https://doi.org/10.1016/j.geodrs.2015.08.005>

772 KSRSAC (2016), Karnataka GIS asset database, version 1, Karnataka State Remote Sensing
773 Applications Centre, Dept. of IT, BT and S & T, Govt. of Karnataka

774 Lagacherie, P., McBratney, A.B., Voltz, M., 2007. Digital Soil Mapping: An Introductory
775 perspective, ,. ed. Elsevier, Amsterdam.

776 Lagacherie, P., Arrouays, D., Bourennane, H., Gomez, C., Nkuba-kasanda, L., 2020.
777 Analysing the impact of soil spatial sampling on the performances of Digital Soil Mapping
778 models and their evaluation : A numerical experiment on Quantile Random Forest using
779 clay contents obtained from Vis-NIR-SWIR hyperspectral imagery. *Geoderma* 375.

780 Lal H., Hoogenboom G., Calixte J.P., Jones J.W., Beinroth F.H., (1993) Using crop
781 simulation models and GIS for regional productivity analysis, *Trans. ASAE* 36 (1993)
782 175–184
783

784 Leenhardt, D., Voltz, M., Bornand, M., Webster, R., Science, L. De, Viala, P., I, M.C., 1994a
785 Evaluating soil maps for prediction of soil water properties. *Eur. J. Soil Sci.* 45, 293–301.

786 Leenhardt, D., Voltz, M., Bornand, M., 1994b. Propagation of the error of spatial prediction
787 of soil properties in simulating crop evapotranspiration. *Eur. J. Soil Sci.* 45, 303–310.

788 McBratney, A.B., Mendonca-Santos, M.D., Minasny, B., 2003. On digital soil mapping.
789 *Geoderma* 117, 3–52.

790 Meinshausen, N., 2006. Quantile Regression Forests. *J. Mach. Learn. Res.* 7, 983–999.

791 Mulder, V.L., Lacoste, M., Richer-de-Forges, A.C., Arrouays, D., 2016. GlobalSoilMap
792 France: High-resolution spatial modelling the soils of France up to two meter depth. *Sci.*
793 *Total Environ.* 573, 1352–1369.

794 NBSS&LUP staff, 2016. Field Guide for Land Resources Inventory, Sujala III project
795 Karnataka, NBSS Publ. ICAR-NBSSLUP, Bangalore 154p.

796 Nussbaum, M., Spiess, K., Baltensweiler, A., Grob, U., Keller, A., Greiner, L., Schaepman,
797 M.E., Papritz, A., 2018. Evaluation of digital soil mapping approaches with large sets of
798 environmental covariates. *Soil* 4, 1–22.

799 Razavi, S., Jakeman, A., Saltelli, A., Prieur, C., Iooss, B., Borgonovo, E., Plischke, E., Lo
800 Piano, S., Iwanaga, T., Becker, W., Tarantola, S., Guillaume, J.H.A., Jakeman, J., Gupta,
801 H., Melillo, N., Rabitti, G., Chabridon, V., Duan, Q., Sun, X., Smith, S., Sheikholeslami,
802 R., Hosseini, N., Asadzadeh, M., Puy, A., Kucherenko, S., Maier, H.R., 2021. The Future
803 of Sensitivity Analysis: An essential discipline for systems modeling and policy support.
804 *Environmental Modelling & Software* 137, 104954.

805 Sekhar, M., Riotte, J., Ruiz, L., Jouquet, P., & Braun, J. J., 2016. Influences of climate and
806 agriculture on water and biogeochemical cycles: Kabini critical zone observatory. In *Proc.*
807 *Indian Natl. Sci. Acad.*, 82(3), 833-846.

808 Sharma, A., Hubert-Moy, L., Buvaneshwari, S., Sekhar, M., Ruiz, L., Bandyopadhyay, S., &
809 Corgne, S. (2018). Irrigation history estimation using multitemporal Landsat satellite
810 Images: application to an intensive groundwater irrigated agricultural watershed in India.
811 *Remote Sensing*, 10(6), 893. <http://dx.doi.org/10.3390/rs10060893>

- 812 Sharmila, S., Joseph, S., Sahai, A. K., Abhilash, S., & Chattopadhyay, R. (2015). Future
813 projection of Indian summer monsoon variability under climate change scenario: An
814 assessment from CMIP5 climate models. *Global and Planetary Change*, 124, 62-78.
- 815 Shivaprasad, C. R., Reddy, R. S., Sehgal, J., Velayutham. M., 1998. Soils of Karnataka for
816 optimizing land use. NBSS&LUP publication No.47b, Nagpur.
- 817 Singh, C., Solomon, D., Bendapudi, R., Kuchimanchi, B., Iyer, S., & Bazaz, A. (2019). What
818 shapes vulnerability and risk management in semi-arid India? Moving towards an agenda
819 of sustainable adaptation. *Environmental Development*, 30, 35-50.
- 820 Sørensen, R.; Zinko, U.; Seibert, J. (2006). "On the calculation of the topographic wetness
821 index: evaluation of different methods based on field observations". *Hydrology and Earth
822 System Sciences*. 10 (1): 101–112.
- 823 Somarathna, P.D.S.N., Minasny, B., Malone, B.P., 2017. More Data or a Better Model?
824 Figuring what Matters Most for the Spatial Prediction of Soil Carbon. *Soil Sci. Soc. Am.*
825 *J.* 81, 1413–1426.
- 826 Sousa V., Santos Pereira L., 1999. Regional analysis of irrigation water requirements using
827 kriging. Application to potato crop (*Solanumtuberosum* L.) at Tras-os-Montes, *Agric.*
828 *Water Manage.* 40 .21–233.
- 829 Sreelash, K., Buis, S., Sekhar, M., Ruiz, L., Tomer, S. K., &Guerif, M. (2017). Estimation of
830 available water capacity components of two-layered soils using crop model inversion:
831 Effect of crop type and water regime. *Journal of Hydrology*, 546, 166-178.
- 832 Stöckle CO, Donatelli M, Nelson R (2003) CropSyst, a cropping systems simulation model.
833 *Eur J Agron* 18(3):289–307
- 834 Tavares, F., Julich, S., Pedro, J., Gonzalez-pelayo, O., Hawtree, D., Feger, K., Jacob, J.,
835 2016. Combining digital soil mapping and hydrological modeling in a data scarce
836 watershed in north-central Portugal. *Geoderma* 264, 350–362.
- 837 van Tol, J. Van, Zijl, G. Van, Julich, S., 2020. Importance of Detailed Soil Information for
838 Hydrological Modelling in an urbanized environment. *hydrology* 1–15.

839 Varella, H., Guérif, M., Buis, S. (2010). Global sensitivity analysis measures the quality of
840 parameter estimation: the case of soil parameters and a crop model. *Environmental*
841 *Modelling & Software*, 25(3), 310-319.

842 Vaysse, K., Lagacherie, P., 2017. Using quantile regression forest to estimate uncertainty of
843 digital soil mapping products. *Geoderma* 291, 55–64.

844 Vaysse, K., Heuvelink, G.B.M., Lagacherie, P., 2017. Spatial aggregation of soil property
845 predictions in support of local land management. *Soil Use Manag.* 33.

846 Wadoux, A.M.J., Brus, D.J., Heuvelink, G.B.M., 2019. Sampling design optimization for soil
847 mapping with random forest. *Geoderma* 355.

848 Wassenaar, T., Lagacherie, P., Legros, J.P., Rounsevell, M.D.A., 1999. Modelling wheat yield
849 responses to soil and climate variability at the regional scale. *Clim. Res.* 11.

850 Williams JR, Jones CA, Kiniry JR, Spanel DA , 1989. The EPIC crop growth model. *Trans Am*
851 *Soc Agric Eng* 32:497–511

852 Wright, M.N., Ziegler, A., 2017. ranger : A Fast Implementation of Random Forests for High
853 Dimensional Data in C ++ and R. *J. Stat. Softw.* 77.

854

855