

# Spatial downscaling of soil prediction models based on weighted generalized additive models in smallholder farm settings

Yiming Xu · Scot E. Smith · Sabine Grunwald ·  
Amr Abd-Elrahman · Suhas P. Wani · Vimala D. Nair

Received: 17 February 2017 / Accepted: 28 August 2017 / Published online: 11 September 2017  
© Springer International Publishing AG 2017

**Abstract** Digital soil mapping (DSM) is gaining momentum as a technique to help smallholder farmers secure soil security and food security in developing regions. However, communications of the digital soil mapping information between diverse audiences become problematic due to the inconsistent scale of DSM information. Spatial downscaling can make use of accessible soil information at relatively coarse spatial resolution to provide valuable soil information at

relatively fine spatial resolution. The objective of this research was to disaggregate the coarse spatial resolution soil exchangeable potassium ( $K_{ex}$ ) and soil total nitrogen (TN) base map into fine spatial resolution soil downscaled map using weighted generalized additive models (GAMs) in two smallholder villages in South India. By incorporating fine spatial resolution spectral indices in the downscaling process, the soil downscaled maps not only conserve the spatial information of coarse

---

Y. Xu (✉)  
Department of Environmental Science and Engineering, Beijing  
Technology and Business University, Beijing 100048, China  
e-mail: xuyimi@ufl.edu

Y. Xu · S. E. Smith · S. Grunwald · V. D. Nair  
School of Natural Resource and Environment, University of  
Florida, 103 Black Hall, PO Box 116455, Gainesville, FL 32611,  
USA

S. E. Smith  
e-mail: sesmith@ufl.edu

S. Grunwald  
e-mail: sabgru@ufl.edu

V. D. Nair  
e-mail: vdn@ufl.edu

Y. Xu · S. E. Smith · A. Abd-Elrahman  
School of Forest Resources and Conservation – Geomatics  
Program, University of Florida, 301 Reed Lab, PO Box 110565,  
Gainesville, FL 32611-0565, USA

A. Abd-Elrahman  
e-mail: aamr@ufl.edu

S. Grunwald  
Pedometrics, Landscape Analysis and GIS Laboratory, Soil and  
Water Sciences Department, University of Florida, 2181 McCarty  
Hall, PO Box 110290, Gainesville, FL 32611, USA

A. Abd-Elrahman  
Gulf Coast REC/School of Forest Resources and Conservation –  
Geomatics Program, University of Florida, 1200 N. Park Road,  
Plant City, FL 33563, USA

S. P. Wani  
International Crops Research Institute for the Semi-Arid Tropics  
(ICRISAT), Patancheru, Hyderabad 502324, India  
e-mail: s.wani@cgiar.org

V. D. Nair  
Soil and Water Sciences Department, University of Florida, 2181  
McCarty Hall, PO Box 110290, Gainesville, FL 32611, USA

spatial resolution soil maps but also depict the spatial details of soil properties at fine spatial resolution. The results of this study demonstrated difference between the fine spatial resolution downscaled maps and fine spatial resolution base maps is smaller than the difference between coarse spatial resolution base maps and fine spatial resolution base maps. The appropriate and economical strategy to promote the DSM technique in smallholder farms is to develop the relatively coarse spatial resolution soil prediction maps or utilize available coarse spatial resolution soil maps at the regional scale and to disaggregate these maps to the fine spatial resolution downscaled soil maps at farm scale.

**Keywords** Spatial downscaling · Soil nutrients · Digital soil mapping · Generalized additive models · Remote sensing · Geographic information system · Smallholder farms

## Introduction

Providing smallholder farmers with services, inputs, and marketing arrangements can help reduce poverty and allow them to compete with larger farmers in increasingly demanding markets (Wiggins et al. 2010). As a cutting-edge discipline applied widely around the world, digital soil mapping (DSM) helps a large population of farmers secure soil and food security in smallholder farms in developing regions such as India and Africa. However, communications of the digital soil mapping information between diverse audiences become problematic due to the inconsistent scale of DSM information (Chakrabarti et al. 2015; Malone et al. 2013).

Several large-scale soil data sets, such as Harmonized World Soil Database (HWSD) (Nachtergaele et al. 2008) (spatial resolution: 1 km), soil organic carbon map of the USA (Odgers et al. 2012) (spatial resolution: 100 m), and Australian Soil and Landscape Grid (<http://www.clw.csiro.au/aclep/soilandlandscapegrid/>), were available around the world. However, those coarse spatial resolution soil maps have limited ability to guide the soil nutrient management in smallholder farms in developing countries such as India. The smallholder farmers in food-insecure regions are more vulnerable to climate change, water shortages, and market volatility, as they lack enough technical and financial support to increase their grain products and alleviate their vulnerability (Lobell et al. 2008). Those fine spatial

resolution maps are more appropriate for the soil management in small farmland compared with coarse spatial resolution maps. As spatial downscaling techniques can utilize available soil maps and disaggregate them to fine spatial resolution maps, downscaling methods have high potential to help farmers especially smallholder farmers manage their soil nutrients. For example, some free and available soil maps with 100 m spatial resolution have little significance to help manage soil nutrients in small farmland. In this case, agricultural scientists or extension workers can utilize those coarse spatial resolution maps as the base maps and apply the downscaling method to disaggregate those base maps to fine spatial resolution downscaled maps.

Area to point kriging (Kyriakidis 2004), area and point regression kriging (Goovaerts 2010), and downscaling cokriging (Pardo-Igúzquiza et al. 2006) have been popular downscaling geostatistical methods in soil science. Besides those geostatistical methods, various data mining-based downscaling methods have also been applied in DSM and remote sensing domains. Liu and Pu (2008) set up a linear downscaling method to disaggregate the simulated 990-m thermal infrared radiance data to a 90-m resolution. Taylor et al. (2013) utilized this method to derive evapotranspiration data required for digital soil mapping from 90 to 15 m pixel. Malone et al. (2012) presented a new downscaling algorithm using weighted generalized additive models (GAMs) based on Liu and Pu (2008). It combines nonparametric smoothing or fitting functions to model nonlinear relationships between soil organic carbon (SOC) and the available covariates. This method used the gridded covariates data at finer scale (90 m) to drive the downscaling process. The resampled classification trees method developed by Odgers et al. (2014) can disaggregate the soil class map into a number of realizations of the potential soil class distribution and quantify their respective degrees of confusion. Subburayalu et al. (2014) used the possibilistic decision trees method to disaggregate the Soil Survey Geographic Database (SSURGO) soil series data.

For the extension of digital soil mapping techniques in fine scale areas such as smallholder farm settings, it is important to test the feasibility of transferring the soil information at coarse spatial resolution to fine spatial resolution. Most research have already performed the disaggregation of soil organic carbon maps before (Brus et al. 2014; Kerry et al. 2012; Lobell et al. 2008; Malone et al. 2012, 2017). On contrast, soil nutrients such as soil

total nitrogen (TN) and soil exchangeable potassium ( $K_{ex}$ ) are also important indicators of soil fertility (Castrignanò et al. 2012; Kerry et al. 2012; Ouyang et al. 2013), and few research have analyzed the down-scaling behavior of those soil nutrients before. This research disaggregated base soil TN and  $K_{ex}$  maps of two smallholder farms, namely Kothapally and Masuti in South India, and evaluated the concordance between the coarse and fine spatial resolution base soil maps and the fine spatial resolution downscaled soil maps.

## Materials and methods

### Description of the study areas

Kothapally (latitude 17° 20' to 17° 24' N and longitude 78° 5' to 78° 8' E) is a smallholder village located in Ranga Reddy District, Telangana, India. It is characterized by an undulating topography with an average slope of 2.5%. The Vertisols and associated soils make up 90% of the area. According to Sreedevi et al. (2004), the soil depth ranges from 30 to 120 cm, the annual rainfall is 802 mm, and the annual mean temperature is 26.6 °C. The main plantation system during the rainy season are cotton (*Gossypium hirsutum*) and rice (*Oryza sativa*). Sorghum (*Sorghum bicolor*) is the predominant crop type during the dry season. There are approximately 1500 people in the village, and the average landholding per household is 1.4 ha (Sreedevi et al. 2004).

Masuti (latitude 16° 28' to 16° 33' N and longitude 75° 45' to 75° 50' E) is a smallholder village located in Bijapur District, Karnataka, India. The major soil type in the northern part of the village is Entisols, and the major soil type in the southern part of the village is Vertisols. Two dams were located in the north central of the village. An irrigation canal was built in the southwestern area of the village. In Masuti, 80% of the farmlands were irrigated using wells. Sorghum (*S. bicolor*), tomato (*Lycopersicon esculentum* var. *esculentum*), and onion (*Allium cepa*) are three major crops in the dry season. Cotton (*G. hirsutum*), rice (*O. sativa*), and maize (*Zea mays*) are major crops in the rainy season.

### Soil sampling

A total of 255 soil samples at 0–15 cm were collected in Kothapally in May 2012 (Xu et al. 2017a), and a total of 259 soil samples at 0–15 cm were collected in Masuti in

February to March 2013 by International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) and University of Florida (Fig. 1). All the soil samples were measured by Trimble GEOXT2005 (Trimble Navigation Ltd., Sunnyvale, California, USA). Global Positioning System (GPS) post-correction was performed by Aimil Ltd. ([www.aimil.com](http://www.aimil.com)) in Hyderabad, India. Each soil sample in two study areas was analyzed for  $K_{ex}$  (Thomas 1982) and TN (Krom 1980) in the soil laboratory at ICRISAT. Those soil samples were utilized to build soil base maps in Kothapally and Masuti.

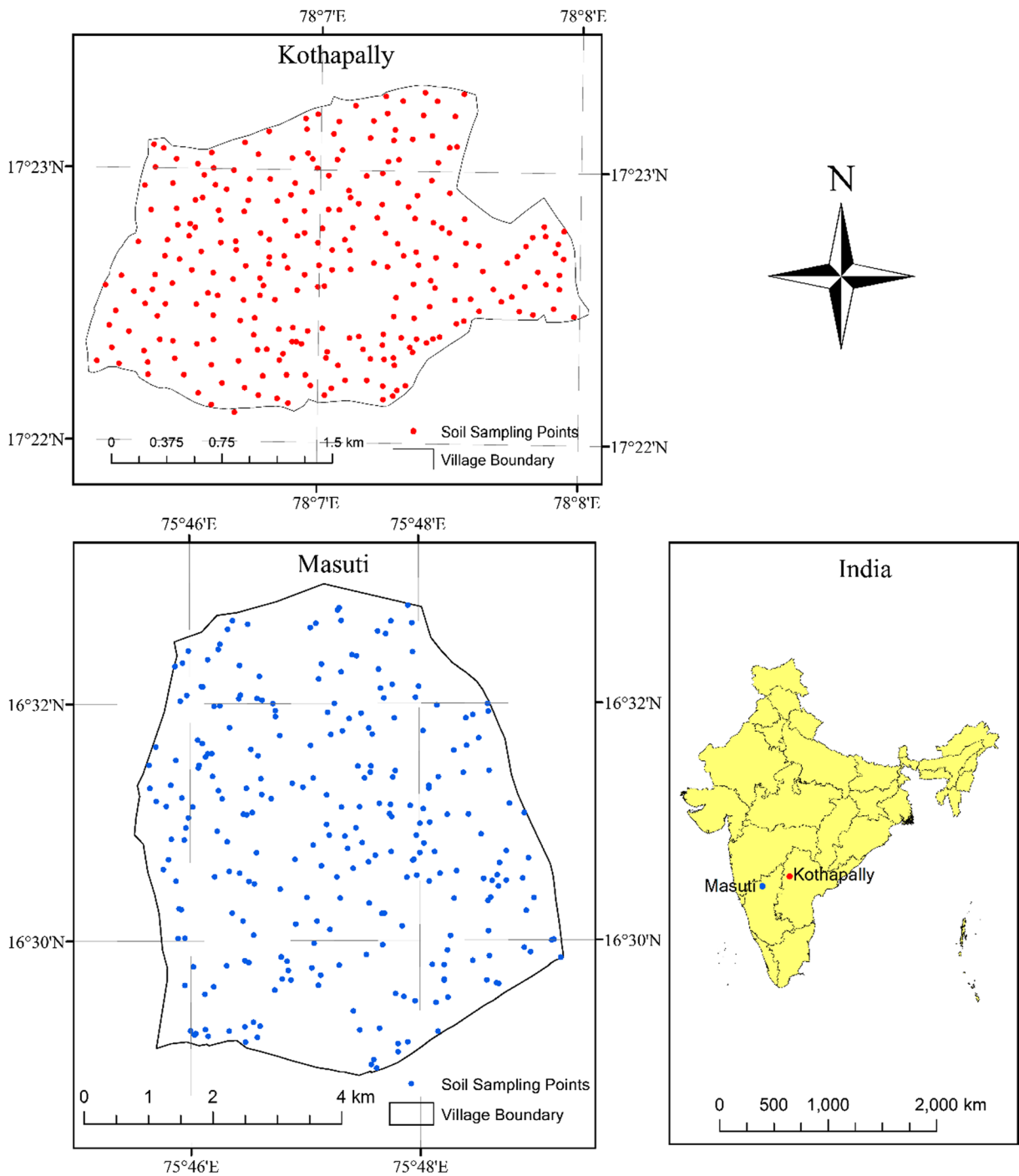
### Soil base maps

Three soil  $K_{ex}$  base maps based on three soil prediction models in Kothapally, namely model KLT (spatial resolution: 30 m) built by Landsat 8 spectral indices (Xu et al. 2017a), model KRE (5 m) built by RapidEye spectral indices, and model KWG (2 m) built by WorldView-2 and GeoEye-1 spectral indices (Xu et al. 2017a), were utilized as the base maps in this research. Three soil TN base maps based on three soil prediction models in Masuti, namely model NLT (spatial resolution: 30 m) built by Landsat 8 spectral indices, model NRE (5 m) built by RapidEye spectral indices, and model NWG (2 m) built by WorldView-2 and Pleiades-1A spectral indices, were also utilized as the base maps in this research. Random Forest was utilized to develop soil TN and soil  $K_{ex}$  prediction models by incorporating multiple spectral indices from the remote sensing images (Xu et al. 2017a).

### Weighted generalized additive models

The target soil property value (e.g., total nitrogen) at each coarse resolution pixel (e.g., Landsat 8 pixel) is defined as  $\hat{T}_k, k = 1, \dots, B$ ;  $B$  is the total number of coarse pixels across the extent of a study area, and  $\hat{t}_m, m = 1, \dots, D$ , denotes the estimate of the target soil property value (e.g., total nitrogen) at each fine resolution pixel (e.g., WorldView-2 pixel). This algorithm is a two-stage algorithm including initialization and iteration (Malone et al. 2012).

Initialization: a weighted generalized additive model (Hastie and Tibshirani 1990) was used:



**Fig. 1** The soil sampling sites in Kothapally and Masuti Village, India

$$\hat{t}_m = \alpha + f_1(x_1) + f_2(x_2) + \dots + f_p(x_p) \quad (1)$$

where  $\alpha$  is a constant,  $x_1, x_2, \dots, x_p$  are each of the covariate data sources, and  $f_j$  are nonparametric

smoothing splines that relate  $\hat{t}_m$  to the covariates. All  $f_j$  are computed through an iterative backfitting algorithm, which are obtained by means of a smoothing of the dependent variable  $\hat{t}_m$  against the covariates  $x_j$ .

Iteration: at the  $l$ -th iteration,  $\hat{t}_m^{l-1}$  are updated to  $\hat{t}_m^l$  using Eq. (2) in order to make the average of  $\hat{t}_m^l$  estimates of finer resolution grid cells equal to the value of their encapsulating coarse resolution grid cell (i.e., to equal  $\hat{T}_k$ ):

$$\hat{t}_m^l = \hat{t}_m^{l-1} \times \frac{\hat{T}_k}{\left(\frac{1}{m}\right) \sum \hat{t}_m^{l-1}} \quad (2)$$

The average of  $\hat{t}_m^l$  estimates  $\left(\left(\frac{1}{m}\right) \sum \hat{t}_m^l\right)$  is denoted as  $\hat{t}_k^l$ . With the newly adjusted value, a new weighted nonlinear regression model (GAM) between the  $\hat{t}_m^l$  and the suite of available covariates is fitted to all the grid cells. When  $\frac{1}{D} \sum |\hat{t}_m^l - \hat{t}_m^{l-1}|$  becomes equal to or less than a given stopping criterion value called SCV (the weights remain constant), the iterations stop.

The “dissever” algorithm was scripted in the R programming environment and included in a package called “ithir”.

### Description of downscaling process

Table 1 described the environmental variables from different remote sensing images utilized in the downscaling processes. The environmental variables used in the downscaling process are fine spatial resolution remote sensing spectral indices that had strong relevance with the target soil property identified by the Boruta algorithm (Kursa and Rudnicki 2010). In Kothapally, the “Dissever” algorithm utilizing the weighted GAMs (Malone et al. 2012) was utilized to disaggregate the model KLT (30 m) to downscaled model KLTRE (5 m) by incorporating spectral indices from RapidEye images, the model KLT (30 m) to downscaled model KLTWG (2 m) by incorporating spectral indices from the WorldView-2/GeoEye-1 images, and the model KRE (5 m) to downscaled model KREWG (2 m) by incorporating spectral indices from WorldView-2/GeoEye-1 images.

In Masuti, the “Dissever” algorithm was also utilized to disaggregate the model NLT (30 m) to downscaled model NLTRE (5 m) by incorporating spectral indices from RapidEye images, the model NLT (30 m) to downscaled model NLTWP (2 m) by incorporating spectral indices from WorldView-2/Pleiades-1A images, and the model NRE (5 m) to downscaled model NREWP (2 m) by incorporating spectral indices from WorldView-2/Pleiades-1A images.

**Table 1** Environmental variables in the downscaling process

Downscaling model	Environmental variables	Spatial resolution transformation (m)
<b>K<sub>ex</sub></b>		
Model KLTRE	RE <sub>b</sub> Green, RE <sub>b</sub> PCA1, RE <sub>b</sub> Rededge, RE <sub>b</sub> ARVI, RE <sub>b</sub> Rededge, RE <sub>a</sub> REB, RE <sub>b</sub> PCA5, RE <sub>a</sub> NIR	30 to 5
Model KLTWG	GEARVI, GEGreen, WV <sub>a</sub> ARVI, WV <sub>a</sub> N2RE, WV <sub>a</sub> Red, WV <sub>a</sub> N2R	30 to 2
Model KREWG	GEARVI, GEGreen, WV <sub>a</sub> ARVI, WV <sub>a</sub> N2RE, WV <sub>a</sub> Red, WV <sub>a</sub> N2R	5 to 2
<b>TN</b>		
Model NLTRE	RE <sub>d</sub> REB, RE <sub>c</sub> PCA3, RE <sub>d</sub> RB, RE <sub>c</sub> RB, RE <sub>d</sub> CI, RE <sub>c</sub> Rededge, RE <sub>a</sub> PCA1	30 to 5
Model NLTWP	PLPCA3, WV <sub>b</sub> RB, WV <sub>b</sub> CI, PLRB, PLCI, PLRG, WV <sub>b</sub> PCA3	30 to 2
Model NREWP	PLPCA3, WV <sub>b</sub> RB, WV <sub>b</sub> CI, PLRB, PLCI, PLRG, WV <sub>b</sub> PCA3	5 to 2

K<sub>ex</sub> model in Kothapally: RE<sub>b</sub>Green the green reflectance from RapidEye (2013-2-24), RE<sub>b</sub>PCA1 principal component 1 of bands in RapidEye (2013-2-24), RE<sub>a</sub>Rededge red edge reflectance from RapidEye (2010-4-19), RE<sub>b</sub>ARVI atmospherically resistant vegetation index (Kaufman and Tanré 1996) from RapidEye (2013-2-24), RE<sub>b</sub>Rededge red edge reflectance from RapidEye (2013-2-24), RE<sub>a</sub>REB band ratio of red edge to blue from RapidEye (2010-4-19), RE<sub>b</sub>PCA5 principal component 5 of bands in RapidEye (2013-2-24), RE<sub>a</sub>NIR near-infrared reflectance from RapidEye (2010-4-19), GEARVI atmospherically resistant vegetation index from GeoEye-1 (2012-1-21), GEGreen green reflectances from GeoEye-1, WV<sub>a</sub>ARVI atmospherically resistant vegetation index from WorldView-2 (2011-12-14), WV<sub>a</sub>N2RE the band ratio of near-infrared band 2 to red edge from WorldView-2, WV<sub>a</sub>Red red reflectances from WorldView-2, WV<sub>a</sub>N2R the band ratio of near infrared band 2 to red band from WorldView-2

TN model in Masuti: RE<sub>d</sub>REB the band ratio of red edge to blue from RapidEye (2013-1-5), RE<sub>c</sub>PCA3 principal component 3 of bands in RapidEye image (2012-12-11), RE<sub>d</sub>RB the band ratio of red to blue from RapidEye (2013-1-5), RE<sub>c</sub>RB the band ratio of red to blue from RapidEye (2012-12-11), RE<sub>d</sub>CI crust index (Karnieli 1997) from RapidEye (2013-1-5), RE<sub>c</sub>Rededge red edge reflectance from RapidEye (2012-12-11), RE<sub>a</sub>PCA1 principal component 1 of bands in RapidEye image (2013-1-5), PLPCA3 principal component of Pleiades-1A (2013-3-3), WV<sub>b</sub>RB the band ratio of red to blue from WorldView-2 image (2011-2-28), WV<sub>b</sub>CI crust index from WorldView-2 image (2011-2-28), PLRB the band ratio of red to blue from Pleiades-1A, PLCI crust index from Pleiades-1A, PLRG the band ratio of red to green from Pleiades-1A, WV<sub>b</sub>PCA3 principal component 3 of bands in WorldView-2



## Validation metrics

The coefficient of determination ( $R^2$ ) and mean absolute errors (MAEs) were evaluated between the base maps and downscaled maps produced by the GAM method:

$$MAEs = \frac{1}{n} \sum |\hat{y}_i - y_i| \quad (3)$$

where  $\hat{y}_i$  is the prediction value of downscaled maps, and  $y_i$  is the value of base maps.

## Results

### Downscaling of soil $K_{ex}$ model by GAMs in Kothapally

Figure 2 shows the three  $K_{ex}$  base maps and downscaled maps in Kothapally. Overall, the six maps in Fig. 2 did not demonstrate significantly different spatial patterns of soil  $K_{ex}$ . After incorporating the remote sensing spectral indices with fine spatial resolution in the downscaling process, the soil  $K_{ex}$  downscaled maps generally enhanced their capabilities to characterize soil  $K_{ex}$  distribution in the fine scale farmland compared with the coarse spatial resolution soil  $K_{ex}$  base maps. For example, the 30-m  $K_{ex}$  base map by model KLT (Fig. 2a) showed a more homogenous spatial pattern of  $K_{ex}$ . In contrast, the 2-m  $K_{ex}$  downscaled map from 30 m by model KLTWV (Fig. 2f) showed a relatively low linear  $K_{ex}$  pattern in the road and a more fragmented  $K_{ex}$  spatial pattern in the farmland.

Comparing Fig. 3a, b, Fig. 3c, d, and Fig. 3e, f, the differences between the fine spatial resolution  $K_{ex}$  downscaled maps and the fine spatial resolution  $K_{ex}$  base maps are smaller than those between the coarse spatial resolution soil  $K_{ex}$  base maps and fine spatial resolution soil  $K_{ex}$  base maps. In general, 67.72% of the differences between the 5-m  $K_{ex}$  base map and 30-m  $K_{ex}$  base map (Fig. 3a), 59.53% of the differences between the 2-m  $K_{ex}$  base map and 5-m  $K_{ex}$  base map (Fig. 3c), and 46.77% of the differences between the 2-m  $K_{ex}$  base map and 30-m  $K_{ex}$  base map (Fig. 3e) have an absolute difference value smaller than  $50 \text{ mg kg}^{-1}$  (approximately 10% of maximum  $K_{ex}$  concentration). In contrast, 79.94% of the differences between the 5-m  $K_{ex}$  base map and 5-m  $K_{ex}$  downscaled map from 30 m (Fig. 3b), 79.62% of the differences between the 2-m  $K_{ex}$  base map and 2-m  $K_{ex}$  downscaled map from 5 m (Fig. 3d), and 79.05% of the differences between the

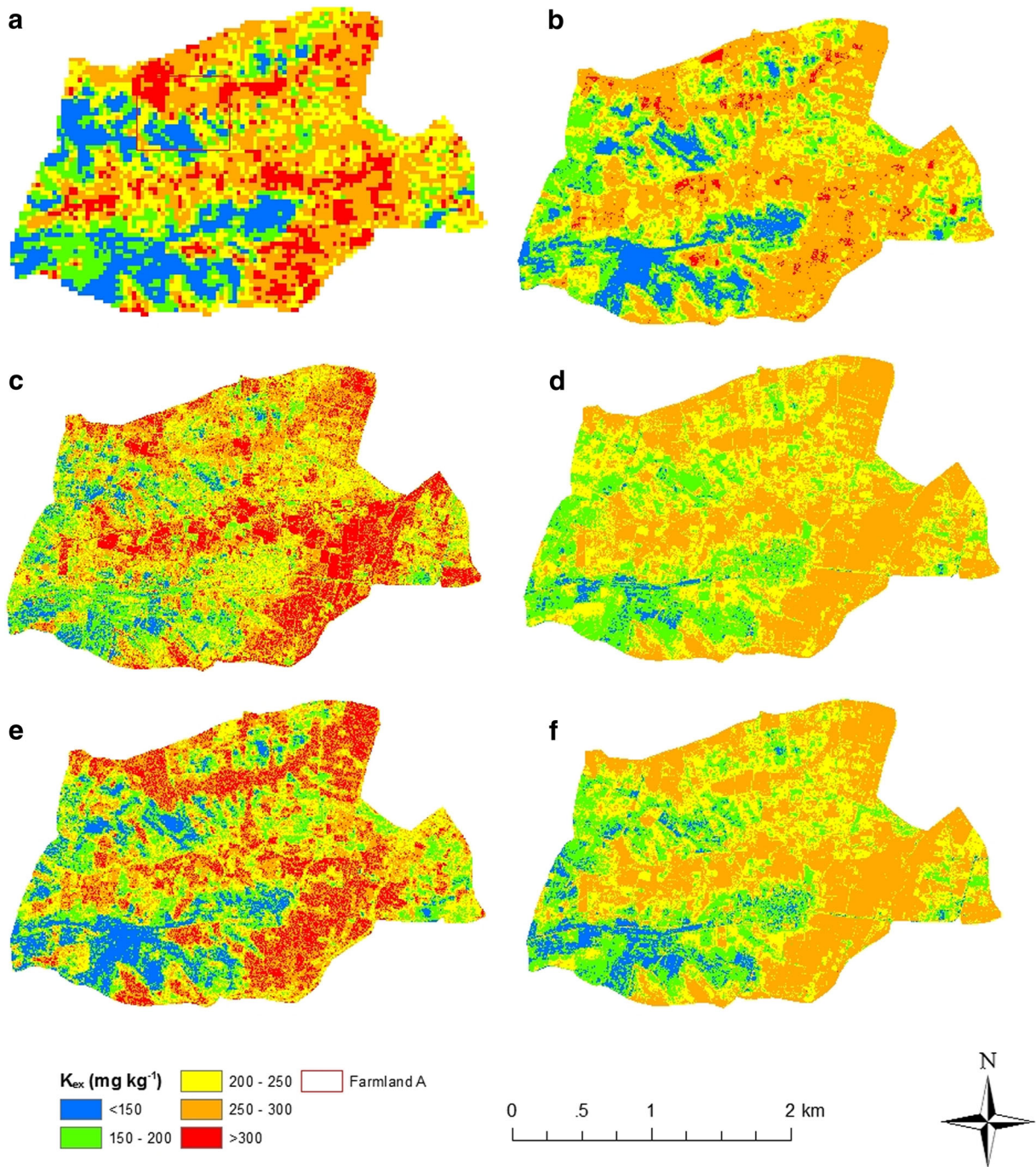
2-m  $K_{ex}$  base map and 2-m  $K_{ex}$  downscaled map from 30 m have an absolute value smaller than  $50 \text{ mg kg}^{-1}$  (Fig. 3f).

Figure 4 demonstrated the soil  $K_{ex}$  maps in a smaller site in the Kothapally village (Farmland A). First, fine spatial resolution soil  $K_{ex}$  base maps showed more heterogeneous spatial patterns of  $K_{ex}$  compared with soil  $K_{ex}$  downscaled maps. For example, the 5-m  $K_{ex}$  base map (Fig. 4b) showed a more heterogeneous spatial pattern of  $K_{ex}$  compared with the 5-m  $K_{ex}$  downscaled map from 30 m (Fig. 4d). Second,  $K_{ex}$  downscaled maps showed more heterogeneous spatial patterns of  $K_{ex}$  compared with coarse spatial resolution  $K_{ex}$  base maps. For example, the 2-m  $K_{ex}$  downscaled map from 30 m (Fig. 4f) displayed more subtle and detailed soil  $K_{ex}$  characterizations compared with the 30-m  $K_{ex}$  base map (Fig. 4a). The 2-m downscaled  $K_{ex}$  map from 30 m (Fig. 4f) showed evident linear and point  $K_{ex}$  patterns in the southwestern area of Farmland A, while the 30-m  $K_{ex}$  base map (Fig. 4a) generalized the soil  $K_{ex}$  distribution. Third, soil  $K_{ex}$  downscaled maps had similar spatial patterns of  $K_{ex}$  with coarse spatial resolution  $K_{ex}$  base maps. For example, both the 2-m  $K_{ex}$  downscaled map from 5 m (Fig. 4e) and 5-m  $K_{ex}$  base maps (Fig. 4b) showed comparably high  $K_{ex}$  in the northern areas of Farmland A. Fourth, soil  $K_{ex}$  downscaled maps had similar spatial patterns of  $K_{ex}$  with fine spatial resolution  $K_{ex}$  base maps. For example, the 2-m  $K_{ex}$  downscaled map from 5 m (Fig. 4e) shared a similar  $K_{ex}$  pattern with 2-m  $K_{ex}$  base maps (Fig. 4c) in terms of fragmented and heterogeneous spatial distribution of  $K_{ex}$ .

From Fig. 5, the pixel values of the  $K_{ex}$  downscaled maps and the fine spatial resolution  $K_{ex}$  base maps showed relatively high coefficients of determination ( $R^2 > 0.7$ ) and MAE smaller than  $35 \text{ mg kg}^{-1}$  in Kothapally. The 2-m  $K_{ex}$  base map (Fig. 2e) and 2-m  $K_{ex}$  downscaled map from 30 m (Fig. 2f) showed the highest coefficient of determination ( $R^2 = 0.78$ ). The results suggested that  $K_{ex}$  downscaled maps disaggregated by the GAMs method have relatively high concordance with fine spatial resolution  $K_{ex}$  base maps in Kothapally.

### Downscaling of soil TN model by GAMs in Masuti

Figure 6 showed three soil TN base maps and three soil TN downscaled maps in Masuti. Fine spatial resolution



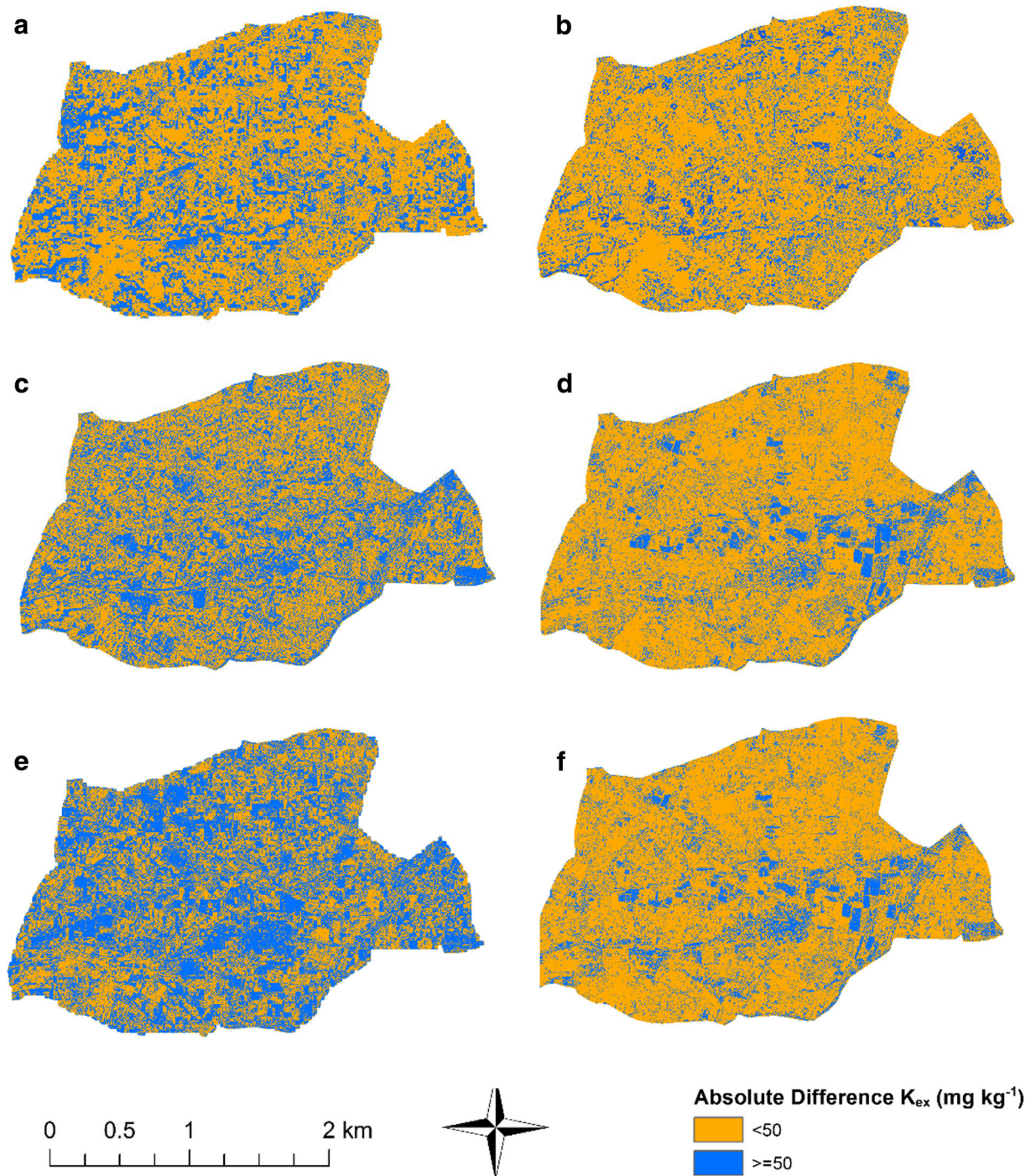
**Fig. 2** **a** 30-m K<sub>ex</sub> base map (Model KLT). **b** 5-m K<sub>ex</sub> downscaled map from 30 m (Model KLTRE). **c** 5-m K<sub>ex</sub> base map (Model KRE). **d** 2-m K<sub>ex</sub> downscaled map from 5 m (Model KREWG). **e**

2-m K<sub>ex</sub> base map (Model KWT). **f** 2-m K<sub>ex</sub> downscaled map from 30 m (Model KLTWG) in Kothapally

soil TN downscaled maps showed stronger capabilities to distinguish the subtle variations of TN. For example, the 2-m TN downscaled map by model NLTWP (Fig. 6f) can demonstrate more heterogeneous, patchy, and

fragmented TN spatial patterns compared with the 30-m TN base map by model NLT (Fig. 6a). In addition, the prediction ranges of TN base maps were also generally larger than those of TN downscaled maps.



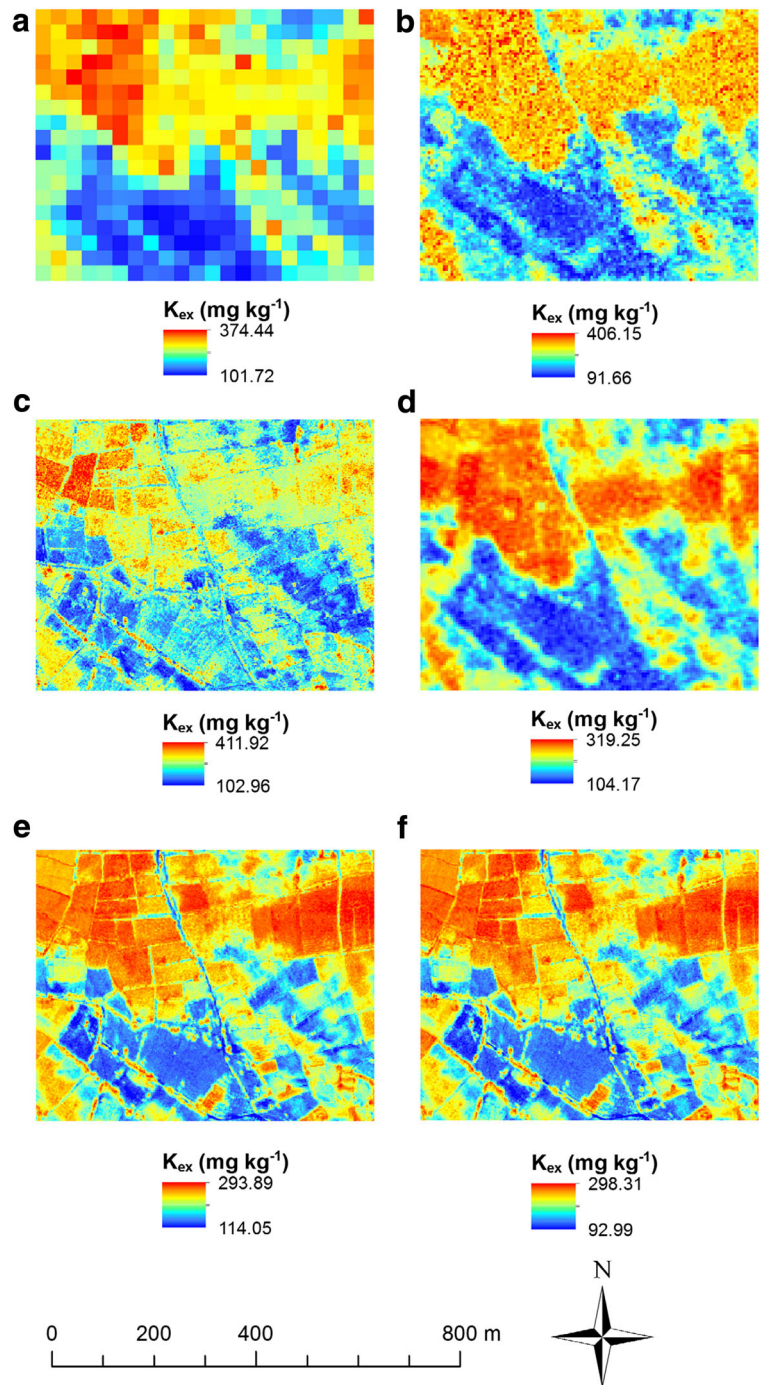


**Fig. 3** **a** Map of differences between the 30-m  $K_{ex}$  base map and 5-m  $K_{ex}$  base map. **b** Map of differences between the 5-m  $K_{ex}$  base map and 5-m  $K_{ex}$  downscaled map from 30 m. **c** Map of differences between the 2-m  $K_{ex}$  base map and 5-m  $K_{ex}$  base map. **d** Map of differences between the 2-m  $K_{ex}$  base map and 2-m  $K_{ex}$

downscaled map from 5 m. **e** Map of differences between the 2-m  $K_{ex}$  base map and 30-m  $K_{ex}$  base map. **f** Map of differences between the 2-m  $K_{ex}$  base map and 2-m  $K_{ex}$  downscaled map from 30 m in Kothapally

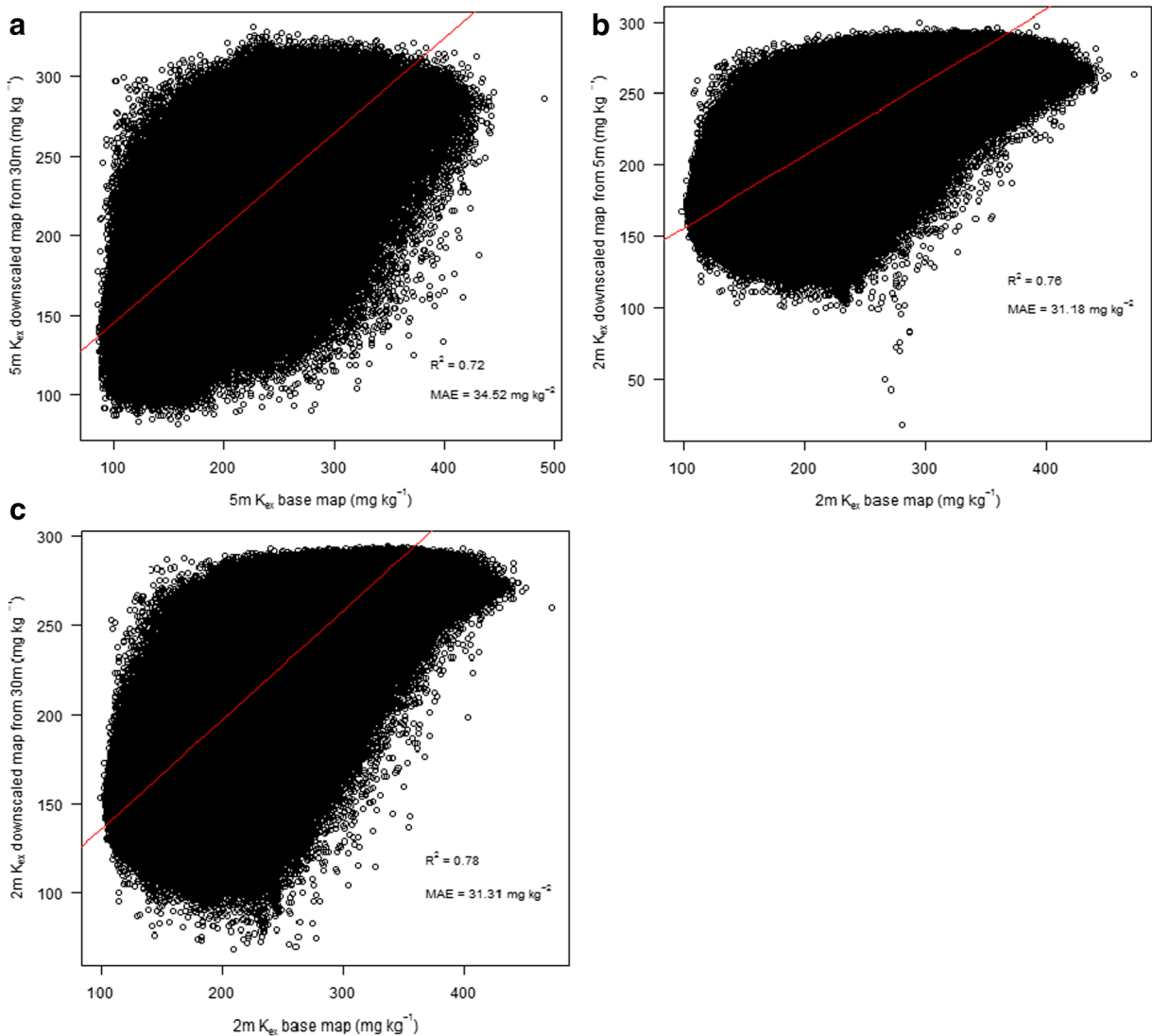


**Fig. 4** **a** 30-m  $K_{ex}$  base map (Model KLT). **b** 5-m  $K_{ex}$  base map (Model KRE). **c** 2-m  $K_{ex}$  base map (Model KWG). **d** 5-m  $K_{ex}$  downscaled map from 30 m (Model KLTRE). **e** 2-m  $K_{ex}$  downscaled map from 5 m (Model KREWG). **f** 2-m  $K_{ex}$  downscaled map from 30 m (Model KLTWG) in Farmland A



In Fig. 7, the differences between the fine spatial resolution TN downscaled maps and the fine spatial resolution TN base maps are smaller than those between the coarse spatial resolution TN base maps and fine spatial resolution TN base maps. In general, 65.24% of the differences between the 30-m TN base map and 5-m

TN base map (Fig. 7a), 68.73% of the differences between the 5-m TN base map and 2-m TN base map (Fig. 7c), and 66.07% of the differences between the 30-m TN base map and 2-m TN base map (Fig. 7e) have an absolute difference value smaller than 80 mg kg<sup>-1</sup> (approximately 10% of maximum TN concentration). In



**Fig. 5** **a** Concordance plot between the 5-m  $K_{ex}$  base map and 5-m  $K_{ex}$  downsampled map from 30 m. **b** Concordance plot between the 2-m  $K_{ex}$  base map and 2-m  $K_{ex}$  downsampled map from 5 m. **c**

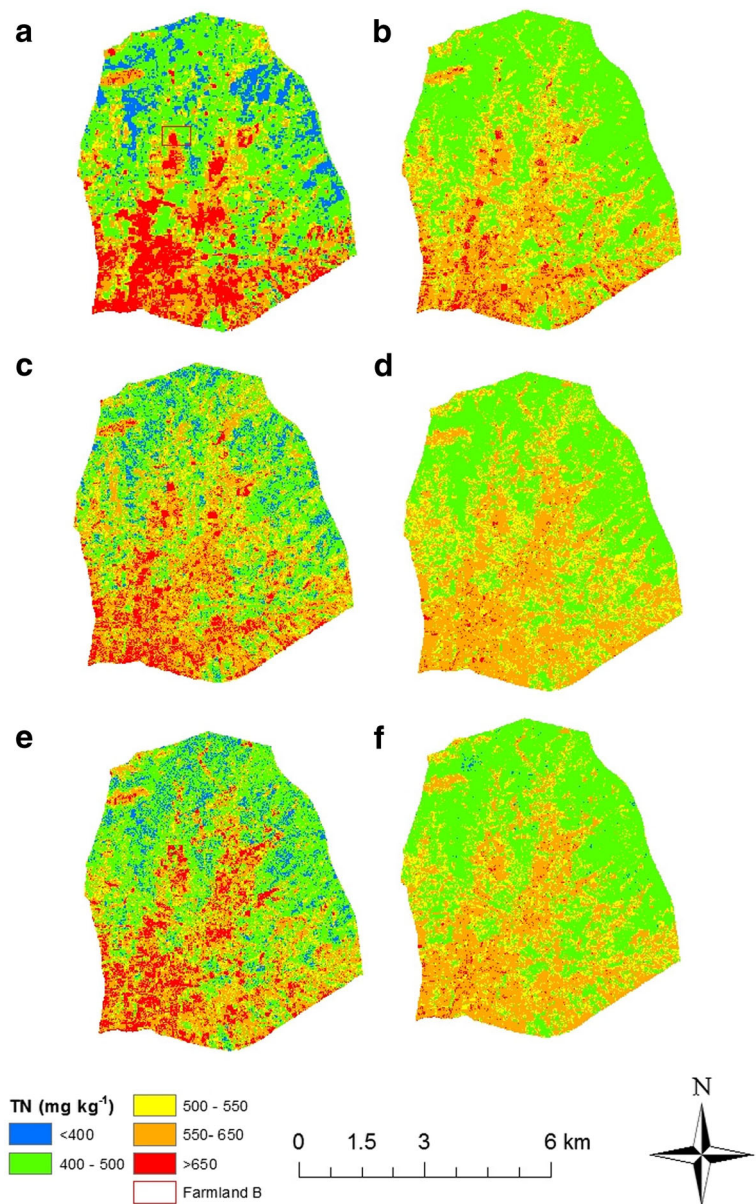
Concordance plot between the 2-m  $K_{ex}$  base map and 2-m  $K_{ex}$  downsampled map from 30 m in Kothapally

contrast, 94.69% of the differences between the 5-m TN base map and 5-m TN downsampled map from 30 m (Fig. 7b), 86.72% of the differences between the 2-m TN base map and 2-m TN downsampled map from 5 m (Fig. 7d), and 90.80% of the differences between the 2-m TN base map and 2-m TN downsampled map from 30 m have an absolute value smaller than  $80 \text{ mg kg}^{-1}$  (Fig. 7f).

Total nitrogen maps of Farmland B, a small farm site in Masuti (Fig. 8), can demonstrate the subtle differences of the base TN maps and downsampled TN maps. The downscaling behavior of the soil TN model in Masuti is similar to that of the soil  $K_{ex}$  model in

Kothapally. First, fine spatial resolution TN base maps showed more fragmented spatial patterns of TN compared with TN downsampled maps. For example, the 5-m TN base map (Fig. 8b) showed a more heterogeneous spatial pattern of TN compared with the 5-m TN downsampled map from 30 m (Fig. 8d). Second, TN downsampled maps showed more fragmented spatial patterns of TN compared with coarse spatial resolution TN base maps. In addition, TN downsampled maps had similar spatial patterns of TN with both coarse spatial resolution TN base maps and fine spatial resolution TN base maps. Interestingly, the 2-m TN downsampled map from 5 m

**Fig. 6** **a** 30-m TN base map (Model NLT). **b** 5-m downscaled TN map from 30 m (Model NLTRE). **c** 5-m TN base map (Model NRE). **d** 2-m downscaled TN map from 5 m (Model NREWP). **e** 2-m TN base map (Model NWP). **f** 2-m downscaled TN map from 30 m (Model NLTWP) in Masuti



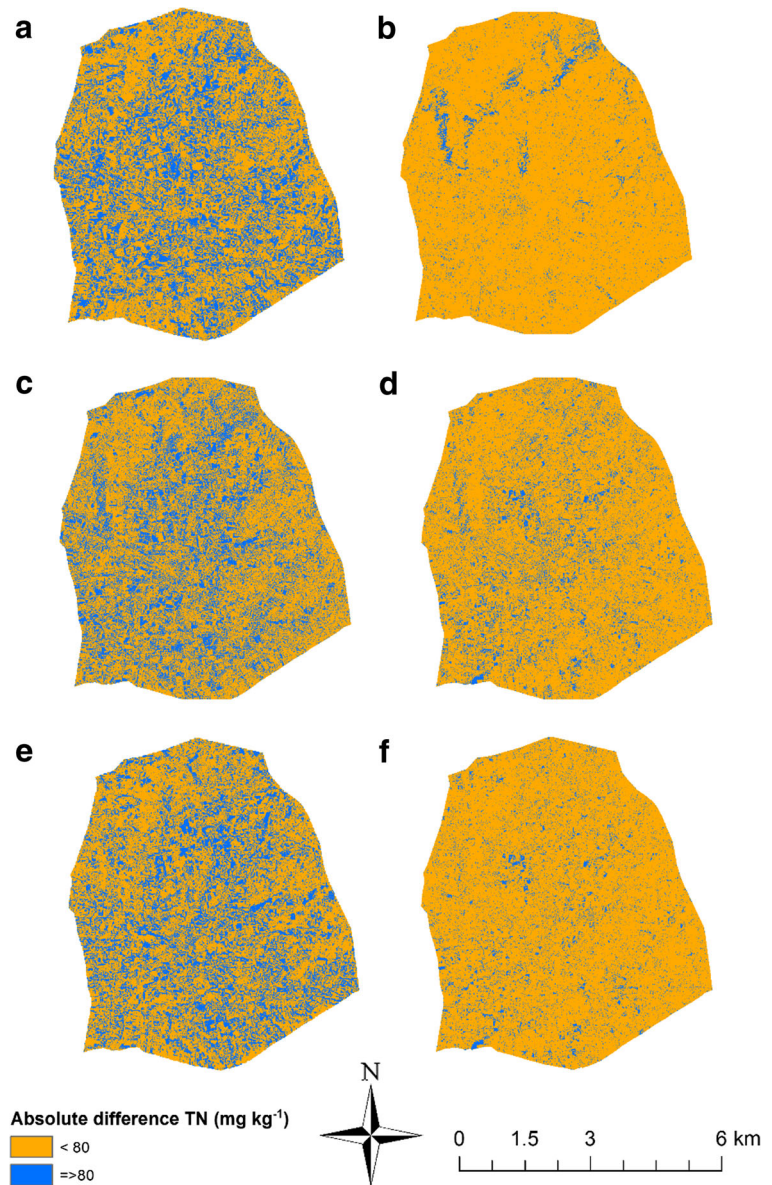
based on model NREWP (Fig. 8e) has higher variations and more heterogeneous TN distributions than the 2-m TN downscaled map from 30 m based on the model NLTWP (Fig. 8f). The downscaling process from model NLT to model NLTWP and from model NRE to model NREWP utilized the same fine spatial resolution spectral indices (Table 1). However, there are more heterogeneous spatial patterns of TN in Fig. 8e by model NREWP compared with that in Fig. 8f by model NLTWP. This may be explained by the fact that the

spatial resolution of 5-m base map by model NRE is finer than 30-m base map by model NLT.

From Fig. 9, the pixel values of the downscaled TN maps and the fine spatial resolution TN base maps showed very high correlations ( $R^2 > 0.85$ ) and MAE smaller than  $42 \text{ mg kg}^{-1}$  in Masuti. The 5-m downscaled TN map from 30 m (Fig. 6b) and 5-m base TN map (Fig. 6c) showed the highest correlation ( $R^2 = 0.90$ ) and MAE ( $33.08 \text{ mg kg}^{-1}$ ). Similar to the results from soil  $K_{\text{ex}}$  downscaling model in Kothapally, the results suggested



**Fig. 7** **a** Map of differences between the 5-m TN base map and 30-m TN base map. **b** Map of differences between the 5-m TN base map and 5-m TN downscaled map from 30 m. **c** Map of differences between the 2-m TN base map and 5-m TN base map. **d** Map of differences between the 2-m TN base map and 2-m TN downscaled map from 5 m. **e** Map of differences between the 2-m TN base map and 30-m TN base map. **f** Map of differences between the 2-m TN base map and 2-m TN downscaled map from 30 m in Masuti



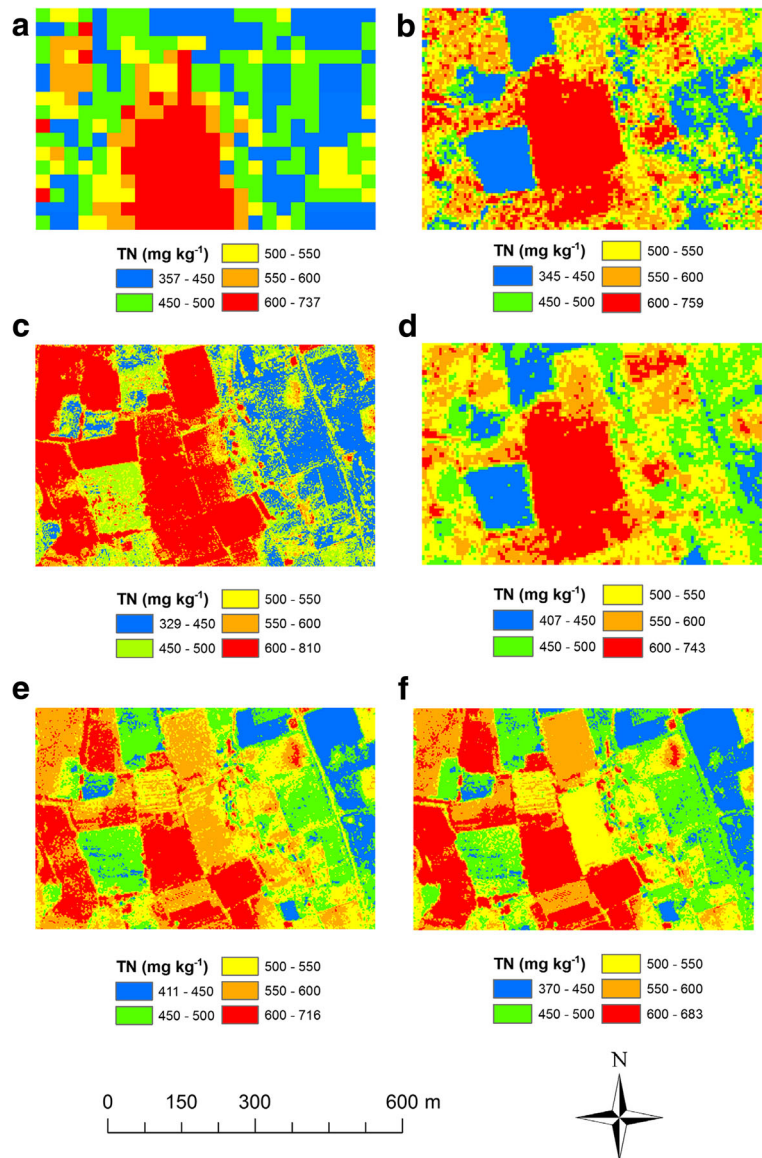
that TN downscaled maps using the GAMs method have relatively high concordance with fine spatial resolution TN base maps in Masuti.

## Discussion

By comparing different soil maps in both Kothapally and Masuti, the similarity between the downscaled maps and fine spatial resolution base maps is higher than the similarity between coarse spatial resolution base maps and fine spatial resolution base maps. Generally, the

priori information of the downscaling method is extracted from expert knowledge and data mining analysis. In this research, random forest-based algorithm “*Boruta*” (Kursa and Rudnicki 2010) helps identify the most relevant fine resolution spectral indices with the target soil properties. The “*Dissever*” algorithm predicts the target variable (soil property) based on an additive combination of nonlinear functions of the environmental variables (Malone et al. 2012). The results from Xu et al. (2017b) showed that coarse spatial resolution soil prediction models based on Landsat 8 images (Model KLT) can attain comparable model performance as fine

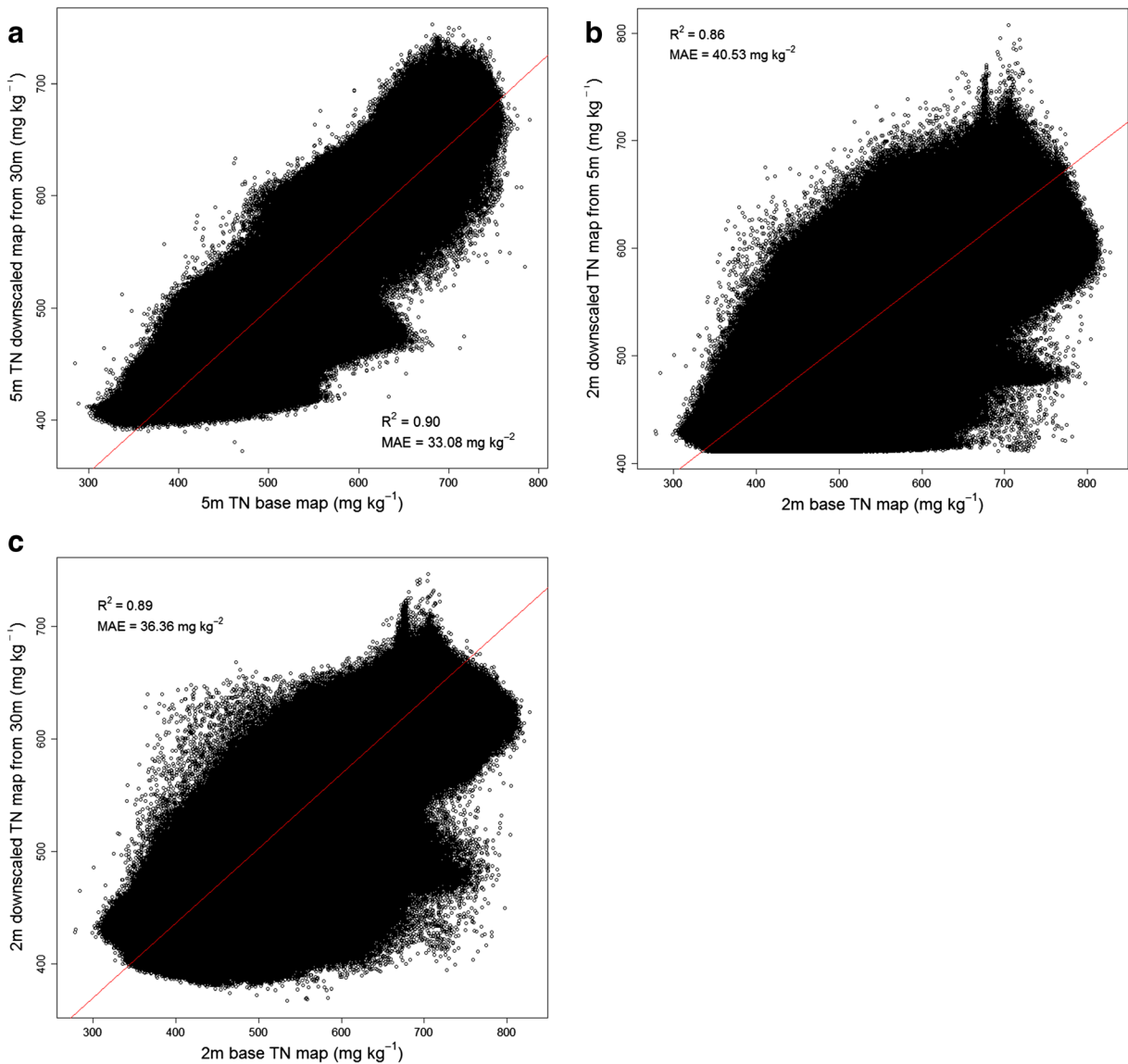
**Fig. 8** **a** 30-m TN base map (Model NLT). **b** 5-m TN base map (Model NRE). **c** 2-m TN base map (Model NWP). **d** 5-m TN downscaled map from 30 m (Model NLTRE). **e** 2-m TN downscaled map from 5 m (Model NREWP). **f** 2-m TN downscaled map from 30 m (Model NLTWP) in Farmland B



spatial resolution soil prediction models based on WorldView-2 and GeoEye-1 images (Model KWG). On the one hand, Landsat 8-based coarse spatial resolution maps such as Figs. 2a and 6a have limited capability to identify the spatial variation of soil properties in fine scale farmland. On the other hand, RapidEye-based or WorldView-2-based fine spatial resolution soil prediction maps such as Figs. 2c and 6e require the purchase of remote sensing images and more processing time, although those models can provide more site-specific soil management recommendations to smallholder farmers.

The downscaling methods build a bridge between soil prediction models with different spatial resolutions. According to Xu et al. (2017a), the incorporation of the image pansharpened spectral indices not only increased the spatial resolution of the soil prediction maps but also enhanced the prediction accuracy of soil prediction models. To some extent, the downscaling application in this research is a form of spectral fusion or image fusion between the coarse spatial resolution soil prediction model and fine spatial resolution spectral indices.

The appropriate and economical strategy to extend the Digital Soil Mapping in the smallholder farms in



**Fig. 9** **a** Concordance plot between the 5-m TN base map and 5-m TN downscaled map from 30 m. **b** Concordance plot between the 2-m TN base map and 2-m TN downscaled map from 5 m. **c**

Concordance plot between the 2-m TN base map and 2-m TN downscaled map from 30 m in Masuti

developing countries is to develop the relatively coarse spatial resolution (e.g., 30 m) soil prediction maps or utilize available coarse spatial resolution soil maps at the regional scale, and to downscale the relatively coarse spatial resolution soil prediction maps to the fine spatial resolution soil prediction maps (e.g., 5 and 2 m) at a specific smallholder farm scale. Another strategy to downscale the soil model is to disaggregate the inputs of soil models such as coarse spatial resolution spectral indices to the fine spatial resolution spectral indices before the

model establishment (McBratney 1998; Taylor et al. 2013). Most research utilizes different algorithms to downscale the remote sensing products to the target support (Kim and Barros 2002; Liu and Journel 2009). However, the problem of this strategy lies in the following: (1) it does not take advantage of the available soil prediction models at coarse spatial resolution and (2) the downscaling of multiple model inputs such as spectral indices is relatively time-consuming in that it is not efficient and economical for large-scale implementation promotion.



Some research have already downscaled the soil maps from 1000 to 90 m (Malone et al. 2012), 100 to 10 m (Malone et al. 2017), and 90 to 15 m pixels (Taylor et al. 2013). However, few researchers have disaggregated the soil map from medium spatial resolution (30 m) to very fine spatial resolution (e.g., 5 or 2 m) before. This research incorporated the up-to-date very fine resolution (VFR) image spectral indices in the downscaling process, and those downscaled soil maps (5 and 2 m) can provide more field-specific nutrient information in the small farmland compared with other research. Comparing with the geostatistical-based downscaling methods such as area-to-point kriging (Brus et al. 2014) and Bayesian area-to-point kriging (Truong et al. 2014), the weighted generalized additive method in this research can utilize the fine spatial resolution spectral indices and have advantages in downscaling grid soil maps.

The significance of this research is that it analyzed the characteristics of downscaled soil models from medium spatial resolution (30 m) to fine spatial resolution (5 m) and super fine spatial resolution (2 m), and from fine spatial resolution (5 m) to super fine spatial resolution (2 m). This research also makes the batch production of fine spatial resolution soil maps at smallholder farm settings possible. The soil downscaled maps not only depict the spatial details of soil properties in fine spatial resolution but also conserve spatial information from the coarse spatial resolution soil base map. Endmember users such as smallholder farmers or agricultural experts can download the free soil maps from the internet such as FAO–UNESCO soil maps (Hartemink et al. 2013) and the GlobalSoilMap project (Arrouays et al. 2014), and utilize updated downscaling techniques such as GAMs to disaggregate these available free soil maps into fine spatial resolution soil maps. This can help smallholder farmers formulate site-specific and sustainable soil management scheme in smallholder farm settings.

## Conclusions

The results suggested that the differences between the downscaled soil maps by GAMs method and fine spatial resolution base soil maps were lower than those between the coarse spatial resolution base soil maps and fine spatial resolution base soil maps. To a certain extent, the downscaling process can be considered a data fusion

method between the coarse spatial resolution soil map and fine spatial resolution spectral indices. The downscaled soil map not only captures the soil property variation in fine spatial resolution but also preserves the information of soil property distribution in the coarse resolution map.

**Acknowledgements** This work is supported by the grant award no. 1201943 “Development of a Geospatial Soil-Crop Inference Engine for Smallholder Farmers” EAGER National Science Foundation and Research Foundation for Youth Scholars of Beijing Technology and Business University. The soil analysis was performed in the soil laboratory at the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) in Patancheru/Hyderabad, India. We thank Christopher M. Clingensmith at the University of Florida, and ICRISAT staff members and villagers of Kothapally and Masuti for the support with field sampling. We also thank Yiming Xu’s PhD committee member Dr. Thomas K. Frazer for his commitment and guidance. A matching assistantship for Yiming Xu was provided by the School of Natural Resources and Environment, University of Florida, and China Scholarship Council.

## References

- Arrouays, D., Grundy, M. G., Hartemink, A. E., Hempel, J. W., Heuvelink, G. B. M., Hong, S. Y., et al. (2014). Chapter Three—GlobalSoilMap: Toward a fine-resolution global grid of soil properties. In Donald L. Sparks (Ed.), *Advances in Agronomy* (Vol. 125, pp. 93–134). Academic Press. <http://www.sciencedirect.com/science/article/pii/B9780128001370000030>. Accessed 16 August 2014.
- Brus, D. J., Orton, T. G., Walvoort, D. J. J., Reijneveld, J. A., & Oenema, O. (2014). Disaggregation of soil testing data on organic matter by the summary statistics approach to area-to-point kriging. *Geoderma*, 226–227, 151–159. <https://doi.org/10.1016/j.geoderma.2014.02.011>.
- Castrignanò, A., Wong, M. T. F., Stelluti, M., De Benedetto, D., & Sollitto, D. (2012). Use of EMI, gamma-ray emission and GPS height as multi-sensor data for soil characterisation. *Geoderma*, 175–176, 78–89. <https://doi.org/10.1016/j.geoderma.2012.01.013>.
- Chakrabarti, S., Bongiovanni, T., Judge, J., Nagarajan, K., & Principe, J. C. (2015). Downscaling satellite-based soil moisture in heterogeneous regions using high-resolution remote sensing products and information theory: a synthetic study. *IEEE Transactions on Geoscience and Remote Sensing*, 53(1), 85–101. <https://doi.org/10.1109/TGRS.2014.2318699>.
- Goovaerts, P. (2010). Combining areal and point data in geostatistical interpolation: applications to soil science and medical geography. *Mathematical Geosciences*, 42(5), 535–554. <https://doi.org/10.1007/s11004-010-9286-5>.
- Hartemink, A. E., Krasilnikov, P., & Bockheim, J. G. (2013). Soil maps of the world. *Geoderma*, 207–208, 256–267. <https://doi.org/10.1016/j.geoderma.2013.05.003>.

- Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized Additive Models*. CRC Press.
- Karnieli, A. (1997). Development and implementation of spectral crust index over dune sands. *International Journal of Remote Sensing*, 18(6), 1207–1220.
- Kaufman, Y. J., & Tanré, D. (1996). Strategy for direct and indirect methods for correcting the aerosol effect on remote sensing: from AVHRR to EOS-MODIS. *Remote Sensing of Environment*, 55(1), 65–79.
- Kerry, R., Goovaerts, P., Rawlins, B. G., & Marchant, B. P. (2012). Disaggregation of legacy soil data using area to point kriging for mapping soil organic carbon at the regional scale. *Geoderma*, 170, 347–358. <https://doi.org/10.1016/j.geoderma.2011.10.007>.
- Kim, G., & Barros, A. P. (2002). Downscaling of remotely sensed soil moisture with a modified fractal interpolation method using contraction mapping and ancillary data. *Remote Sensing of Environment*, 83(3), 400–413. [https://doi.org/10.1016/S0034-4257\(02\)00044-5](https://doi.org/10.1016/S0034-4257(02)00044-5).
- Krom, M. D. (1980). Spectrophotometric determination of ammonia: a study of a modified Berthelot reaction using salicylate and dichloroisocyanurate. *Analyst*, 105(1249), 305–316. <https://doi.org/10.1039/AN9800500305>.
- Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the Boruta package.
- Kyriakidis, P. C. (2004). A geostatistical framework for area-to-point spatial interpolation. *Geographical Analysis*, 36(3), 259–289. <https://doi.org/10.1111/j.1538-4632.2004.tb01135.x>.
- Liu, Y., & Journel, A. G. (2009). A package for geostatistical integration of coarse and fine scale data. *Computers & Geosciences*, 35(3), 527–547. <https://doi.org/10.1016/j.cageo.2007.12.015>.
- Liu, D., & Pu, R. (2008). Downscaling thermal infrared radiance for subpixel land surface temperature retrieval. *Sensors*, 8(4), 2695–2706. <https://doi.org/10.3390/s8042695>.
- Lobell, D. B., Burke, M. B., Tebaldi, C., Mastrandrea, M. D., Falcon, W. P., & Naylor, R. L. (2008). Prioritizing climate change adaptation needs for food security in 2030. *Science*, 319(5863), 607–610. <https://doi.org/10.1126/science.1152339>.
- Malone, B. P., McBratney, A. B., & Minasny, B. (2013). Spatial scaling for digital soil mapping. *Soil Science Society of America Journal*, 77(3), 890. <https://doi.org/10.2136/sssaj2012.0419>.
- Malone, B. P., McBratney, A. B., Minasny, B., & Wheeler, I. (2012). A general method for downscaling earth resource information. *Computers & Geosciences*, 41, 119–125. <https://doi.org/10.1016/j.cageo.2011.08.021>.
- Malone, B. P., Styc, Q., Minasny, B., & McBratney, A. B. (2017). Digital soil mapping of soil carbon at the farm scale: a spatial downscaling approach in consideration of measured and uncertain data. *Geoderma*, 290, 91–99. <https://doi.org/10.1016/j.geoderma.2016.12.008>.
- McBratney, A. B. (1998). Some considerations on methods for spatially aggregating and disaggregating soil information. *Nutrient Cycling in Agroecosystems*, 50(1–3), 51–62. <https://doi.org/10.1023/A:1009778500412>.
- Nachtergaele, F., Van Velthuizen, H., Verelst, L., Batjes, N., Dijkshoorn, K., Van Engelen, V., et al. (2008). Harmonized world soil database. *Food and Agriculture Organization of the United Nations*.
- Odgers, N. P., Libohova, Z., & Thompson, J. A. (2012). Equal-area spline functions applied to a legacy soil database to create weighted-means maps of soil organic carbon at a continental scale. *Geoderma*, 189–190, 153–163. <https://doi.org/10.1016/j.geoderma.2012.05.026>.
- Odgers, N. P., Sun, W., McBratney, A. B., Minasny, B., & Clifford, D. (2014). Disaggregating and harmonising soil map units through resampled classification trees. *Geoderma*, 214–215, 91–100. <https://doi.org/10.1016/j.geoderma.2013.09.024>.
- Ouyang, W., Xu, Y., Hao, F., Wang, X., Siyang, C., & Lin, C. (2013). Effect of long-term agricultural cultivation and land use conversion on soil nutrient contents in the Sanjiang Plain. *Catena*, 104, 243–250. <https://doi.org/10.1016/j.catena.2012.12.002>.
- Pardo-Igúzquiza, E., Chica-Olmo, M., & Atkinson, P. M. (2006). Downscaling cokriging for image sharpening. *Remote Sensing of Environment*, 102(1–2), 86–98. <https://doi.org/10.1016/j.rse.2006.02.014>.
- Sreedevi, T., Shiferaw, B., & Wani, S. (2004). *Adarsha Watershed in Kothapally Understanding the Drivers of Higher Impact: Global Theme on Agroecosystems Report no. 10*.
- Subburayalu, S. K., Jenhani, I., & Slater, B. K. (2014). Disaggregation of component soil series on an Ohio County soil survey map using possibilistic decision trees. *Geoderma*, 213, 334–345. <https://doi.org/10.1016/j.geoderma.2013.08.018>.
- Taylor, J. A., Jacob, F., Galleguillos, M., Prévot, L., Guix, N., & Lagacherie, P. (2013). The utility of remotely-sensed vegetative and terrain covariates at different spatial resolutions in modelling soil and watertable depth (for digital soil mapping). *Geoderma*, 193–194, 83–93. <https://doi.org/10.1016/j.geoderma.2012.09.009>.
- Thomas, G. W. (1982). Exchangeable cations. *Methods of soil analysis. Part 2. Chemical and microbiological properties*, (methodsofsoilan2), 159–165.
- Truong, P. N., Heuvelink, G. B. M., & Pebesma, E. (2014). Bayesian area-to-point kriging using expert knowledge as informative priors. *International Journal of Applied Earth Observation and Geoinformation*, 30, 128–138. <https://doi.org/10.1016/j.jag.2014.01.019>.
- Wiggins, S., Kirsten, J., & Llambi, L. (2010). The future of small farms. *World Development*, 38(10), 1341–1348. <https://doi.org/10.1016/j.worlddev.2009.06.013>.
- Xu, Y., Smith, S. E., Grunwald, S., Abd-Elrahman, A., & Wani, S. P. (2017a). Incorporation of satellite remote sensing pan-sharpened imagery into digital soil prediction and mapping models to characterize soil property variability in small agricultural fields. *ISPRS Journal of Photogrammetry and Remote Sensing*, 123, 1–19. <https://doi.org/10.1016/j.isprsjprs.2016.11.001>.
- Xu, Y., Smith, S. E., Grunwald, S., Abd-Elrahman, A., & Wani, S. P. (2017b). Evaluating the effect of remote sensing image spatial resolution on soil exchangeable potassium prediction models in smallholder farm settings. *Journal of Environmental Management*, 200, 423–433. <https://doi.org/10.1016/j.jenvman.2017.06.017>.