Oblique geographic coordinates as covariates for digital soil mapping

Anders Bjørn Møller[1], Amélie Marie Beucher[1], Nastaran Pouladi[1], Mogens Humlekrog Greve[1]

[1]Department of Agroecology, Aarhus University, Tjele, 8830, Denmark

5 *Correspondence to*: Anders Bjørn Møller (anbm@agro.au.dk)

Abstract. Decision tree algorithms such as Random Forest have become a widely adapted method for mapping soil properties in geographic space. However, implementing explicit spatial~~geographic~~ trends~~relationships~~ into these algorithms~~methods~~ has proven problematic. Using x- and y-coordinates as covariates gives orthogonal artefacts in the maps, and alternative methods using distances as covariates can be inflexible and difficult to interpret. We propose instead the use

10 of coordinates along several axes tilted at oblique angles to provide an easily interpretable method for obtaining a realistic prediction surface. We test the method on four spatial datasets and compare it to similar methods ~~for mapping topsoil organic matter contents in an agricultural field in Denmark~~. The results show that the method provides accuracies better than or on par with the most reliable alternative methods, namely kriging and distance-based covariates ~~the use of buffer distances to the training points~~. Furthermore, the proposed method is highly flexible, scalable and easily interpretable. This makes it a

15 promising tool for mapping soil properties with complex spatial variation. ~~We believe that the method will be highly useful for mapping soil properties in larger areas, and testing it for this purpose is a logical next step.~~

1 Introduction

Machine learning has become a frequently applied means for mapping soil properties in geographic space. The most common approach is to train models from soil observations and covariates in the form of geographic data layers. The models

20 can often provide reliable predictions of soil properties. Many researchers have used decision tree algorithms, as they are computationally efficient, do not rely on assumptions about the distributions of the input variables, and can use both numeric and categorical data ~~and are immune to correlated and redundant covariates~~ (Quinlan, 1996, Mitchell, 1997, Rokach and Maimon, 2005, Tan et al., 2014). Additionally, they effectively handle nonlinear relationships and complex interactions (Strobl et al., 2009).

25 However, a disadvantage of decision tree models is that they do not explicitly take into account spatial trends in the data~~relationships~~. Unlike geostatistical methods, such as kriging, the predictions can therefore contain spatial bias.~~es in the form of spatially autocorrelated residuals.~~

A number of studies have applied regression-kriging (RK) as a solution (Knotters et al., 1995, Odeh et al., 1995, Hengl et al., 2004). By kriging the residuals of the predictive model and adding the kriged residuals to the prediction surface, this

1

30   approach can account for spatial trends and achieve higher accuracies.~~By kriging the residuals of the predictive model and adding the kriged residuals to the prediction surface, soil mappers have been able to reduce or remove spatial biases.~~ A disadvantage of RK~~this approach~~ is that the combination of two models hinders the combination of spatial trends~~relationships~~ with the other covariates. Spatial trends~~relationships~~ therefore remain disconnected from other statistical relationships in the analysis, leading to difficulties in interpreting the model and its associated uncertainties.

35   An obvious solution to this problem would be to use the x- and y-coordinates of the soil observations as covariates. However, results have shown that this approach can lead to unrealistic orthogonal artefacts in the output maps when used in conjunction with decision tree algorithms (Behrens et al., 2018, Hengl et al., 2018, Nussbaum et al., 2018). The cause of this problem lies in the splitting procedure of decision tree algorithms, as they use only one covariate for each split. Therefore, a dataset containing only the x- and y-coordinates will force the algorithm to make orthogonal splits in geographic space.

40   Several researchers have proposed solutions to this problem. Behrens et al. (2018) proposed the use of Euclidean distance fields (EDF) in the form of distances to the corners and middle of the study area as well as the x- and y-coordinates ~~of the soil observations~~. Their results showed that this approach efficiently integrated spatial trends ~~relationships~~ and that accuracies were better than or on par with other methods for integrating spatial context.

On the other hand, Hengl et al. (2018) suggested an approach referred to as spatial Random Forest (RFsp). This method

45   consists of calculating data layers with buffer distances to each of the soil observations in the training dataset. It then trains a Random Forest model, using the buffer distances as covariates, combined with auxiliary data or on their own. One of the main advantages of this approach is that it incorporates distances between observations in a similar manner to geostatistical models. The authors assessed the use of ~~demonstrated the method~~RFsp on a large number of spatial prediction problems and showed that it effectively eliminated spatial ~~autocorrelation~~trends in ~~from~~ the residuals.

50   Although these two methods are able to integrate spatial trends~~relationships~~ in machine learning models, they can be difficult to interpret ~~are not without shortcomings. Firstly, t~~. The distances used in EDF ~~both methods~~ depend ~~either~~ on the geometry of the study area, ~~in the case of EDF, or~~and for RFsp, they depend on the locations of the soil samples~~, in the case of RFsp~~. The meaning and interpretation of the distances therefore varies depending on the study area and the soil observations. ~~Another shortcoming relating to~~ EDF and RFsp also have limited flexibility, as ~~is that~~ both methods specify the number of

55   geographic data layers a priori. For EDF, the number of distance fields is seven, and for RFsp, the number of buffer distances is equal to the number of soil observations. This means that there is no straightforward way to increase the number of spatially explicit covariates, if the number is insufficient to account for spatial trends~~relationships~~. ~~In addition, v~~Vice versa, there is no way to decrease the number of spatially explicit covariates, even if a smaller number would suffice. The latter is especially relevant for RFsp, as the method is computationally unfeasible for datasets with a large number of

60   observations (Hengl et al., 2018).

In this study, we propose an alternative method for including spatially explicit covariates for mapping soil properties. With the method, we aim to address directly the cause of the orthogonal artefacts produced with~~arising from the use of~~ x- and y-

2

coordinates as covariates in decision tree models. Furthermore, we aim to improve upon the shortcomings of previous methods by developing a method that is both flexible and easily interpretable.

65 We refer to the method as Oblique Geographic Coordinates (OGC). In short, it works by calculating coordinates for the observations along a series of axes, tilted at ~~various~~several oblique angles relative to the x-axis. By including oblique coordinates as covariates, we enable the decision tree algorithm to make oblique splits in geographic space. As this is not possible with only x- and y-coordinates as covariates, this addition should allow the model to produce a more realistic prediction surface. Furthermore, the number of oblique angles is adjustable, and soil mappers can therefore choose a number

70 that suits their purpose. Some mapping tasks may require a higher number of oblique angles than others, and soil mappers can therefore increase the number as necessary. Alternatively, if a small number of oblique angles suffices, soil mappers can reduce their number and thereby shorten computation times.~~Furthermore, as the number of oblique angles is adjustable, it should be possible to optimize it, both in terms of accuracy and computational efficiency.~~

We test the method on four spatial datasets. Firstly, we test it for predicting soil organic matter contents in a densely sampled

75 agricultural field in Denmark, located in northern Europe. Secondly, we test it on three publicly available spatial datasets (*meuse, eberg* and Swiss rainfall). We hypothesise that OGC can provide accuracies on par with previous methods for including explicitly spatial covariates. We also hypothesize that it is possible to adjust the number of oblique angles in order to optimize accuracy, and that the results allow meaningful interpretations.

2 Materials and methods

80 ### 2.1 Study areas

We test OGC and compare it to other methods based on four spatial datasets. Firstly, we test it for a predicting soil organic matter (SOM) for an agricultural field in Denmark (Vindum). Secondly, we test it on three publicly available datasets. For Vindum, we will present methods and results in detail. For the other three datasets, we will present methods and results in brief, while Appendix A contains a detailed presentation of the methods and results for these datasets.

85 #### 2.1.1 Vindum

This study area is a 12-ha agricultural field located in Denmark in northern Europe (9.568°E; 56.375°N, ETRS 1989) (Figure 1). It lies in a kettled moraine landscape 55 – 66 m above sea level. The parent materials in the field include clay till, glaciofluvial sand and peat. The climate is temperate coastal, with mean monthly temperatures ranging from 1°C in January to 17°C in July and a mean annual precipitation of 850 mm (Wang, 2013). The field contains 285 measurements of soil

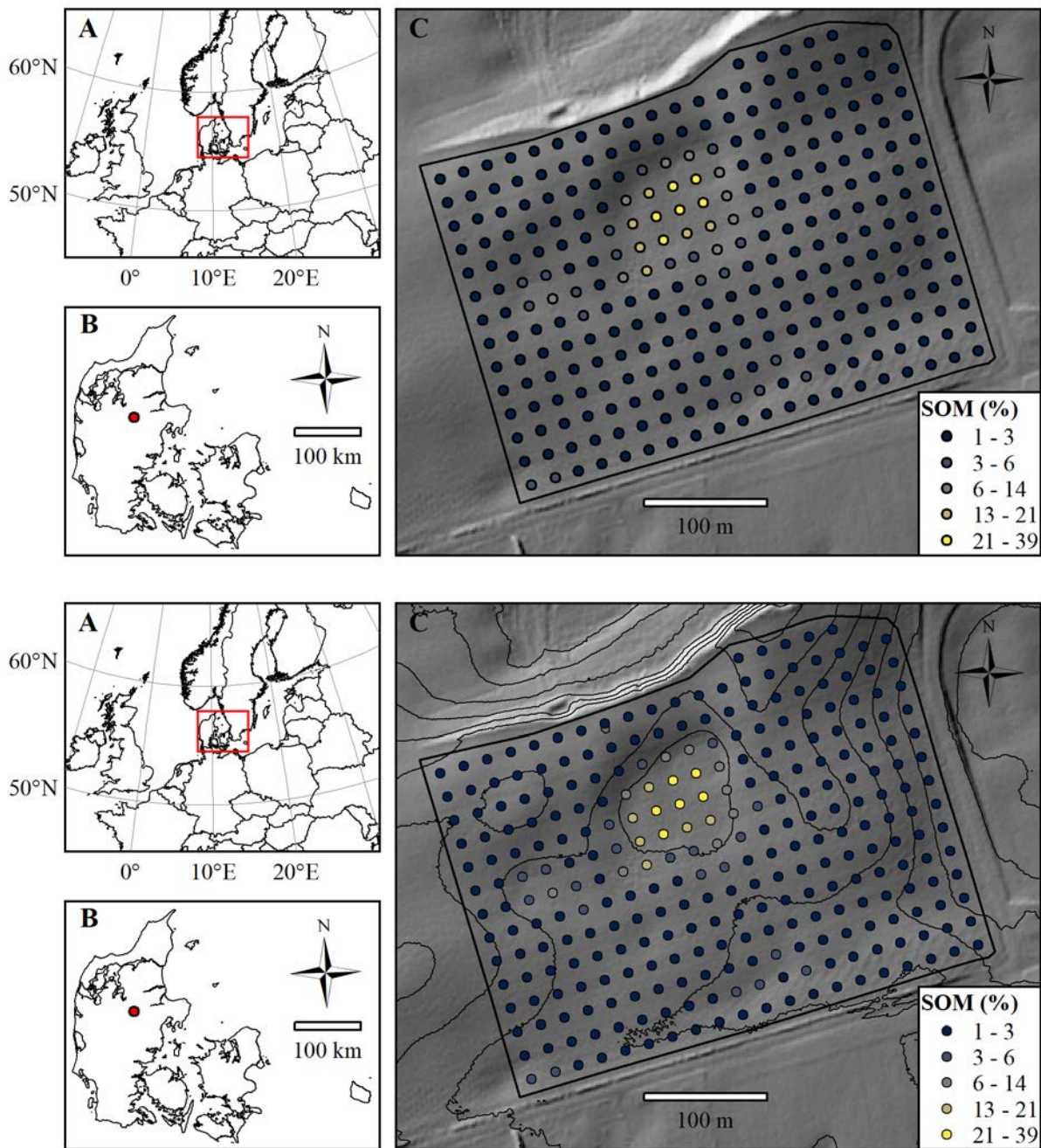90 organic matter (SOM) from the depth interval 0 – 25 cm, located in a 20 m grid.

Figure 1: A: Location of Denmark in northern Europe. B: Location of the Vindum field study area within Denmark. C: Map of the Vindum fieldstudy area, including locations of the samples extracted for soil organic matter (SOM) measurements. The thin black lines are 2 m contour lines. The background shows hill shade (northwest, 45° altitude) based a digital elevation model (DEM) in 1.6x1.6 m resolution (National Survey and Cadastre, 2012).

95

The SOM contents of the topsoil in the field range from 1.3% to 38.8% with a mean value of 3.5% and a median of 2.2%. The values have a strong positive skew of 4.7 and are leptokurtic with a kurtosis of 26.9. Logarithmic transformation reduces skewness (2.9) and kurtosis (11.1). Pouladi et al. (2019) described that spatial structure of the data with a stable variogram with 139 m range, nugget of 0 and sill of 23.8.

2.1.2 Additional datasets

For additional analyses, we included the *meuse* dataset, the *eberg* dataset and the Swiss rainfall dataset. The *meuse* dataset, available through the R package *sp* (Pebesma et al., 2020), contains 155 measurements of soil heavy metal concentrations from a 5-km^2 flood plain of the Meuse river near the village of Stein in the Netherlands. For this dataset, we mapped zinc concentrations. The *eberg* dataset, available through the R package *plotKML* (Hengl et al., 2020) contains 3,670 soil observations from a 100-km^2 area in Ebergötzen near the city Göttingen in Germany. For this dataset, we mapped soil types. Lastly, the *Swiss rainfall* dataset contains 476 rainfall measurements from May 8, 1986 in Switzerland (Dubois et al., 2003). Although this is not a soil dataset, we included it because of the high anisotropy of the data, which makes it useful for comparing methods on their ability to account for anisotropic spatial problems. We describe these three datasets in more detail in Appendix A.

2.2 Oblique geographic coordinates

The method that we propose consists of calculating coordinates along a number of axes titled at various oblique angles, relative to the x-axis. In the following, we show that it is possible to calculate the coordinate of a point *(b₁, a₁)* along an axis tilted at an angle θ relative to the x-axis, based on θ and the x- and y-coordinates of *(b₁, a₁)*. ~~We also show that it is possible to derive the calculation using basic trigonometry.~~ Equation~~s~~ (1) ~~, (2), (3) and (4)~~ show~~s~~ the ~~derivation of~~ the calculation of the oblique geographic coordinate, using Figure 2 for illustration.

$$\cancel{b_2 = c * \cos A_2} \tag{1}$$

$$\cancel{c = \sqrt{a_1{}^2 + b_1{}^2}} \tag{2}$$

$$\cancel{A_2 = \theta - A_1}$$

$$\cancel{A_1 = \tan^{-1} \frac{a_1}{b_1}} \tag{3}$$

$$\cancel{b_2 = \sqrt{a_1{}^2 + b_1{}^2} * \cos\left(\theta - \tan^{-1} \frac{a_1}{b_1}\right)} \tag{4}$$

$$\underline{OGC = b_2 = \sqrt{a_1{}^2 + b_1{}^2} * \cos\left(\theta - \tan^{-1} \frac{a_1}{b_1}\right)} \tag{1}$$

, where θ is the angle of the titled axis relative to the x-axis; ~~A_1 is the angle between the x-axis and the line c between the origin of the coordinate system and the point (b_1, a_1); A_2 is the difference between θ and A_1;~~ a_1 is the y-coordinate of *(b₁, a₁)*; b_1 is the x-coordinate of *(b₁, a₁)*; ~~b_2 is a line with the angle θ between the origin of the coordinate system its intersection with~~

~~a_2; a_2 is a line perpendicular to b_2 going from (b_1, a_1) to its intersection with b_2. The length of~~ b_2 (or "*OGC*") is ~~equal to~~ the coordinate of (b_1, a_1) along an axis tilted with the angle θ relative to the x-axis.
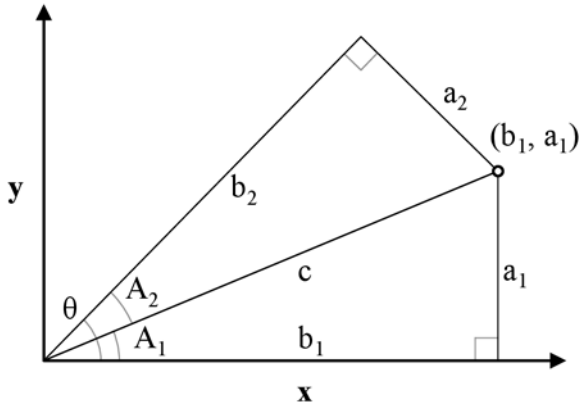
Figure 2: Illustration for the derivation of the oblique geographic coordinate for the point (b_1, a_1) along an axis tilted with the angle θ from the x-axis. The coordinate is equal to the length of b_2. Triangles a_1b_1c and a_2b_2c are right triangles with the same hypotenuse c. The sides a_1 and b_1 are the x- and y-coordinates of the point (b_1, a_1), respectively. A_1 is the angle between the x-axis and the line c between the origin of the coordinate system and the point (b_1, a_1); A_2 is the difference between θ and A_1.

As the x- and y-coordinates of soil observations are known, and θ is given, it is possible to calculate coordinates at oblique angles for all soil observations in a dataset. Likewise, as the x- and y-coordinates of the cells in a geographic raster layer are known, it is possible to calculate oblique coordinates for the cells. Our approach relies on calculating coordinates along n axes tilted at angles ranging from 0 to $\pi((n-1)/n)$ with increments of π/n between the angles. θ should not be π or greater, as coordinates along axes tilted at these angles will correlate with coordinates along axes tilted at angles of 0 to $\pi((n-1)/n)$. For example, coordinates along an axis with $\theta = 0.25\pi$ (northeast) perfectly correlate with coordinates along an axis with $\theta = 1.25\pi$ (southwest). Figure 3 shows coordinates along axes tilted at six different angles relative to the x-axis for the Vindum study area. The coordinate rasters A and D are equivalent to the x- and y-coordinates, respectively, while the coordinate rasters B, C, E and F show coordinates at oblique angles.
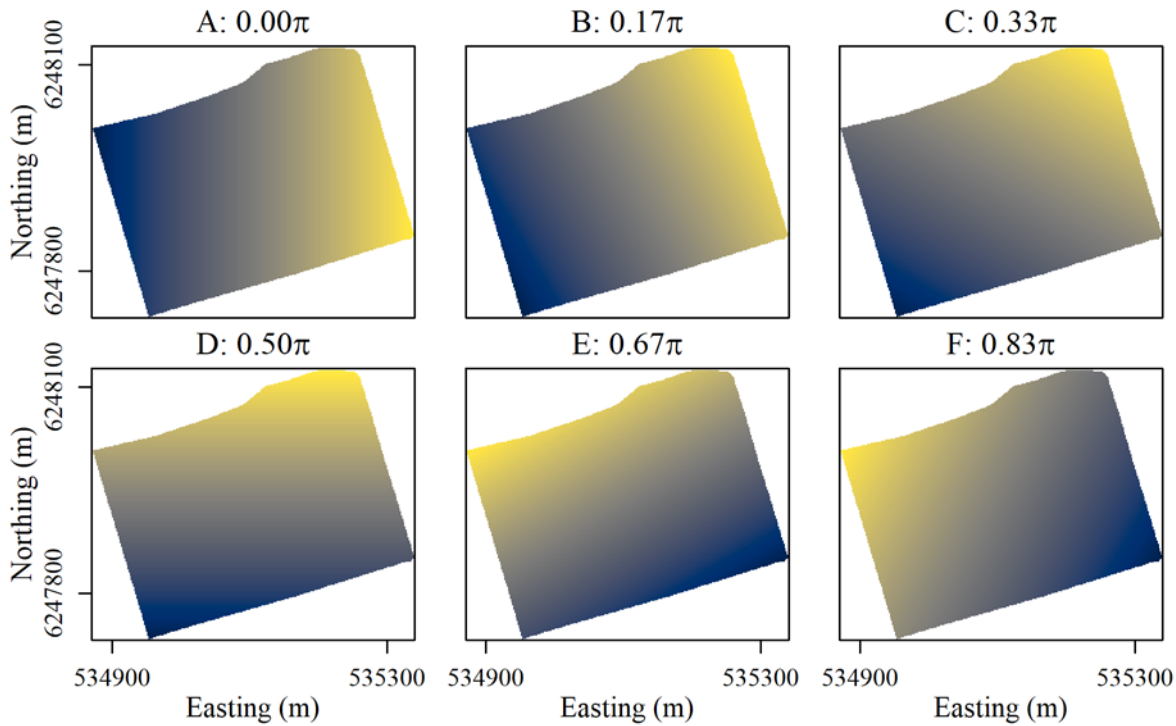
Figure 3: Examples of rasters with coordinates tilted at six different angles for the Vindum study area. Easting and northing for UTM Zone 32N, ETRS 1989.

2.3 ~~Experiments~~Method comparison

2.3.1 Vindum

We use the 285 SOM observations from the Vindum study area in order to test the accuracy of predictions made by Random Forest models using OGC as covariates. In addition to OGC, we also employed 19 data layers with auxiliary data, which Pouladi et al. (2019) derived from a 1.6 m DEM, satellite imagery and electromagnetic induction. Topographic variables included the sine and cosine of the aspect, depth of sinks, plan and profile curvature, elevation, flow accumulation, valley bottom flatness, mid-slope position, standard and modified topographic wetness index, slope gradient, slope length and valley depth. Satellite imagery included normalized difference, absolute difference, ratio and soil-adjusted vegetation indices. Lastly, we used the apparent electrical conductivity from a DUALEM 1 sensor in perpendicular mode.

Table 1: Auxiliary data variables used as covariates in the study, including name, description, the mean value and the range. Pouladi et al. (2019) describe the derivation of the variables.

Predictor variable	Description	Mean (range)
DEM		
Cos aspect	Cosine of surface aspect	-0.1 (-1.0 - 1.0)

7

Sin aspect	Sine of surface aspect	0.32 (-1.0 - 1.0)
Depth of sinks	Depth of sinks (m)	0.1 (0.0 - 1.1)
Plan curvature	Shape of the surface in the horizontal plane	0 (-34 - 15)
Profile curvature	Shape of the surface in the vertical plane	0.00 (-0.06 - 0.04)
Elevation	Elevation from DEM; m above sea level	60.8 (54.6 - 66.2)
Flow accumulation	Number of upslope cells	74 (3 - 8969)
MRVBF	Multiresolution index of valley bottom flatness	1.5 (0.0 - 4.9)
Mid-slope position	Covers the warmer zones of slopes	0.5 (0.0 - 1.0)
SAGA wetness index	SAGA GIS modified topographic wetness index	4.0 (2.2 - 8.6)
Slope gradient	Local slope gradient (degrees)	4.9 (0.0 - 17.5)
SL	Slope length factor	0.4 (0.0 - 2.3)
TWI	Topographic wetness index	6.6 (3.7 - 14.6)
Valley depth	Depth of valleys (m)	1.4 (0.1 - 8.1)
Sentinel 2		
DVI	Difference vegetation index	1735 (1202 - 3294)
NDVI	Normalized difference vegetation index	0.5 (0.3 - 0.7)
RVI	Ratio vegetation index	2.8 (2.0 - 6.4)
SAVI	Soil-adjusted vegetation index	0.7 (0.5 - 1.1)
DUALEM 1mPRP		
ECa	Apparent electrical conductivity	8.9 (4.9 - 16.0)

In order to optimize the number of raster layers for OGC, we generated datasets with 2 – 100 coordinate rasters. We then trained Random Forest models from each dataset, both with and without auxiliary data. In order to assess predictive accuracy, we used 100 repeated splits on the SOM observations, each using 75% of the observations for model training and a 25% holdout dataset for accuracy assessment. We trained models using the R package *ranger* (Wright and Ziegler, 2015) and parameterized the models using the R package *caret* (Kuhn, 2008). For each split, we tested five different values for *mtry*, minimum node sizes of 1, 2, 4 and 8, and two different splitting rules *variance* and *extratrees*. We mainly adjusted *mtry* and the minimum node size in order to avoid overfitting. We tested *mtry* values at even intervals between 2 and the total number of covariates, including both auxiliary data and spatially explicit covariates. The tested *mtry* values therefore varied depending on the number of covariates. The *extratrees* splitting rule generates random splits, as opposed to the *variance* splitting rule, which chooses optimal splits. Per default, *extratrees* generates one random split for each covariate and then chooses the random split that gives the largest variance reduction (Geurts et al., 2006). It therefore leads to a greater degree of randomization~~The *extratrees* splitting rule allows suboptimal splits, which can increase randomization relative to~~

160

165

8

170 ~~the default *variance* splitting rule (Geurts et al., 2006)~~. We selected the setup that provided the lowest RMSE for the out-of-bag predictions on the training data, and used this setup for predictions on the 25% holdout dataset.

We used the same 100 repeated splits for each number of coordinate rasters, with and without auxiliary data. We calculated accuracy based on Pearson's R^2, RMSE and Lin's concordance criterion (ccc), and subsequently used the number of coordinate rasters that yielded the lowest RMSE. We selected a different number of coordinate rasters with and without

175 auxiliary data.

We then compared the accuracies obtained with the optimal numbers of coordinate rasters, with and without auxiliary data, to the accuracies obtained with other methods. We tested kriging, Random Forest models trained only on the auxiliary data and Random Forest models trained using EDF and RFsp, with and without auxiliary data. We trained the Random Forest models using the same procedure outline~~ds~~ above. For kriging, we used variograms automatically fitted on logarithmic-

180 transformed SOM observations using the *autofitVariogram* function of the R package *automap* (Hiemstra, 2013). A previous study using the same dataset showed that kriging predicted SOM more accurately than regression-kriging (Pouladi et al., 2019). We therefore omitted regression-kriging from the analysis, although, without this previous finding, it would have been relevant to include it.

We used the same 100 repeated splits for assessing the accuracies of all methods. This allowed us to carry out pairwise t-

185 tests between the accuracies of the methods. We used the results of the pairwise t-tests to rank the methods according to their accuracies according to each of the metrics. If there was no statistical difference ($p > 0.05$) between the accuracies of two or more methods, these methods received the same rank. We calculated separate ranks for the methods for each accuracy metric, resulting in three different sets of ranks.

In order to illustrate the results, we produced maps of SOM with each method, using models trained from all the data.

190 ~~Furthermore, w~~We also investigated the covariate importance of models trained with OGC and tested all methods ~~the results~~ for spatially autocorrelated residuals using experimental variograms. To produce sample variograms of the residuals, we produced maps with each method using all observations. We converted both observations and predictions to natural logarithmic scale. We then subtracted the predictions from the observations and calculated variograms for these residuals. For this purpose, we used the function *variogram* from the R package *gstat* (Pebesma and Graeler, 2020) with its default

195 parameters.

2.3.1 Additional datasets

We also compared OGC to other methods based on the three additional datasets *meuse*, *eberg* and Swiss rainfall. The methods in the comparison depended on the dataset. For the *meuse* dataset, we tested all the methods tested on the Vindum dataset, with the addition of RK using Random Forest models for regression. For the *eberg* dataset, we tested Random Forest

200 models based on auxiliary data (AUX), EDF and OGC, as well as the combined methods (EDF + AUX and OGC + AUX). For the Swiss rainfall dataset, we tested only purely spatial methods, including ordinary kriging (OK), EDF, RFsp and OGC. As for the Vindum dataset, we tested each method based on 100 splits into training and test data and carried out pairwise t-

9

tests on the resulting accuracies. Appendix A gives additional details on the methods for each dataset. For the three additional datasets, we focused on the accuracies and maps produced with each method. We therefore omitted analyses of
205   the residuals and covariate importance for these datasets.

3 Results and discussion

### 3.1 Optimal number of coordinate rasters

3.1.1 Vindum

210   For the Vindum dataset, ~~Without auxiliary data,~~ accuracies of predictions obtained with OGC, without auxiliary data, increased with the number of coordinate rasters up to an optimum at seven coordinate rasters (Figure 4). However, with more than seven coordinate rasters, accuracies deteriorated slightly with the number of coordinate rasters. This pattern was the same for all three metrics. On the other hand, with OGC in combination with auxiliary data, accuracies generally increased with the number of coordinate rasters. The increase was greatest when the number of coordinate rasters was small,
215   while the effect of more coordinate rasters decreased for larger numbers of coordinate rasters. With auxiliary data, the optimal number of coordinate rasters was 94 for Pearson's R^2, 80 for RMSE and 89 for ccc. Accuracies with auxiliary data were almost invariably higher than accuracies achieved without auxiliary data.
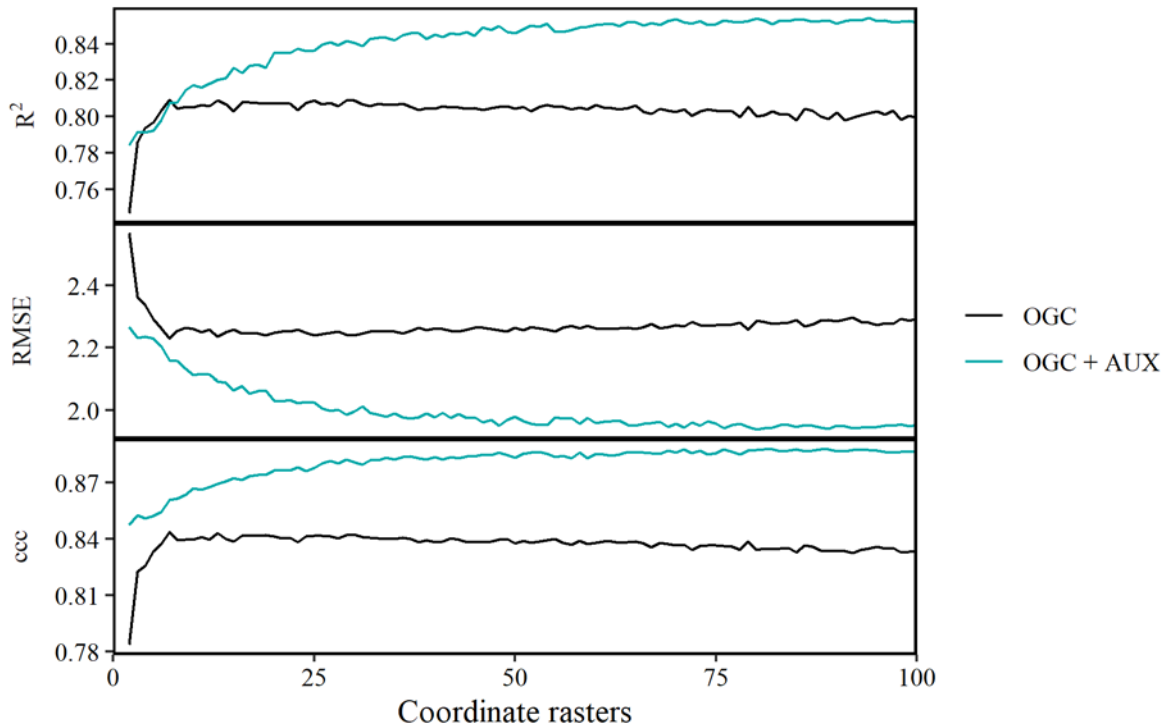
Figure 4: Effects of the number of coordinate rasters on the accuracy of SOM predictions on the Vindum dataset, calculated as Pearson's R^2, root mean square error (RMSE) and Lin's concordance criterion (ccc). We calculated effects for Random Forest models trained on only coordinate rasters (OGC) and with coordinate rasters in combination with auxiliary data (OGC + AUX). The lines represent mean values obtained from 100 repeated splits (75% training dataset, 25% test dataset) for each number of coordinate rasters.

Figure 5 shows SOM contents mapped for Vindum with increasing ~~the effect of increasing the~~ numbers of coordinate rasters, without auxiliary data. The predictions with only two coordinate rasters showed a pattern very typical of predictions with x- and y-coordinates with very visible orthogonal artefacts. As the number of coordinate rasters increased, the patterns of the artefacts changed. With coordinate rasters at three different angles, the artefacts had a hexagonal pattern, and with coordinate rasters at four different angles, the artefacts gained an octagonal pattern. Furthermore, as the number of coordinate rasters increased, the artefacts became less pronounced. Although some artefacts were visible with coordinate rasters at seven different angles, they were much less visible than the artefacts in the map produced with only two coordinate rasters.

11
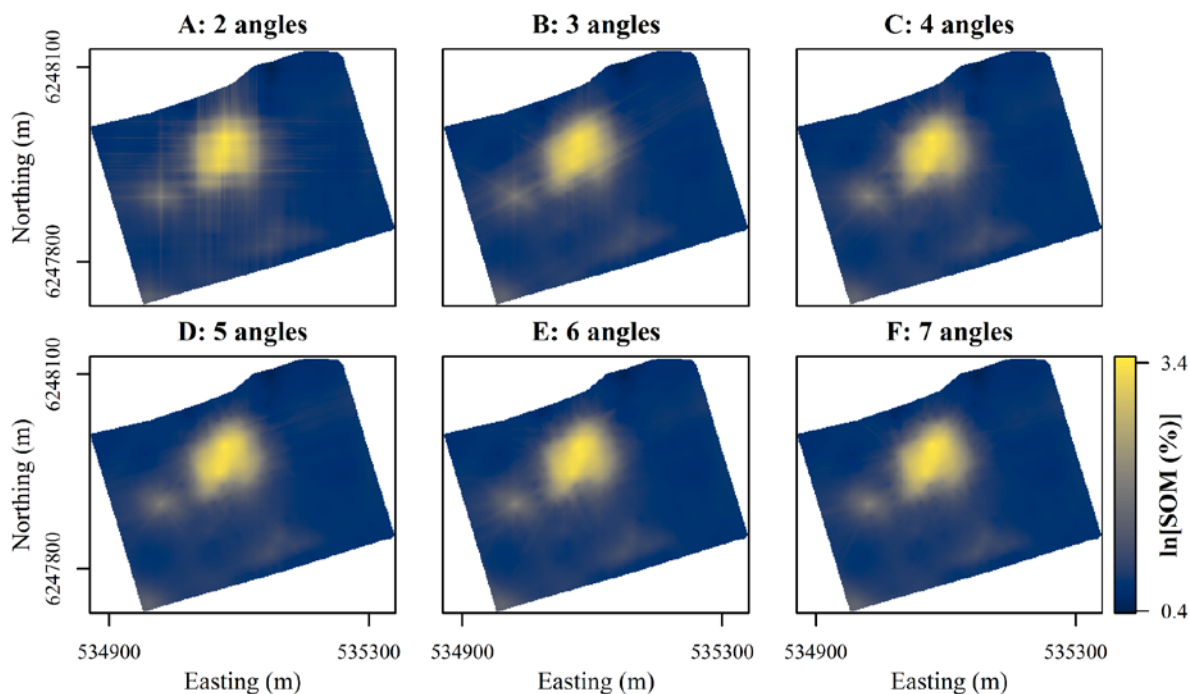
Figure 5: Maps of soil organic matter (SOM) contents in the topsoil at Vindum predicted using Random Forest models trained with coordinate rasters at two to seven different angles as covariates. Easting and northing for UTM Zone 32N, ETRS 1989.

235   With auxiliary data, the effect of increasing the number of coordinate rasters was less clearly visible for the Vindum dataset (Figure 6Figure 6). Even with only two coordinate rasters, the predictions had no orthogonal artefacts. However, they contained noisy patterns and sharp boundaries in some areas. This is most likely an artefact from the auxiliary data. For example, using a high-resolution DEM may have created noise in the predictions. However, with coordinate rasters at 80 different angles, the spatial pattern of the predicted SOM contents became substantially smoother, with a reduction both in

240   noise and in sharp boundaries. Furthermore, some areas with moderately high SOM contents became more clearly visible and coherent, for example in the area approximately one third of the way from the western to the northern corner of study area. The predicted patterns with a higher number of coordinate rasters were therefore not only more accurate, but also more realistic.
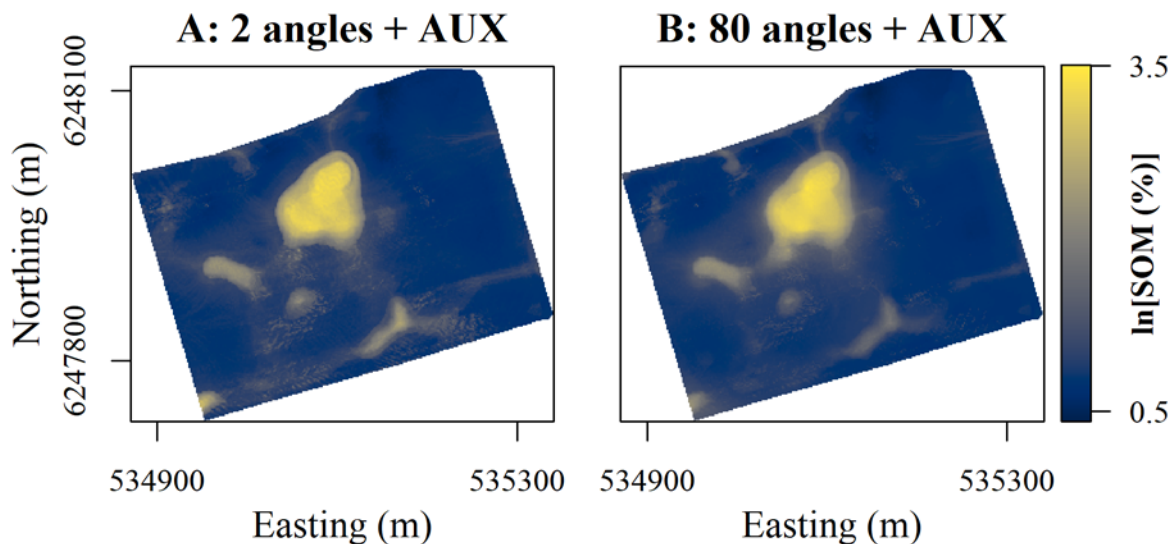
Figure 6: Maps of soil organic matter (SOM) contents in the topsoil at Vindum predicted using Random Forest models trained using auxiliary data in conjunction with coordinate rasters at (A) two and (B) 80 different angles as covariates. Easting and northing for UTM Zone 32N, ETRS 1989.

3.1.2 Additional datasets

For the three additional datasets, the effect of increasing the number of coordinate rasters without auxiliary data was generally the same as for the Vindum dataset. In all three cases, there was relatively little, if any increase in accuracy after an initially very steep increase. For the *meuse* dataset, the optimal number of coordinate rasters was six or eight, depending on the accuracy metric (Figure A1). For the *eberg* dataset, the optimal number was 91, but there was only limited improvement in accuracy with more than five coordinate rasters (Figure A3). For the Swiss rainfall dataset, the optimal number of coordinate rasters was 33 or 50, depending on the accuracy metric (Figure A5).

As for the Vindum dataset, the optimal number of coordinate rasters was generally larger in combination with auxiliary data than without auxiliary data. For the *meuse* dataset, the optimal number of coordinate rasters in combination with auxiliary data was 11 or 13, depending on the accuracy metric. For the *eberg* dataset, the optimal number of coordinate rasters in combination with auxiliary data was 22. However, unlike the results for the Vindum dataset, accuracies for these two datasets gradually decreased when the number of coordinate rasters was larger than the optimal value.

In summary, the combination of OGC with auxiliary data generally increased the optimal number of coordinate rasters. Furthermore, in some cases, accuracy deteriorates when the number of coordinate rasters surpasses an optimal value, while in other cases it reaches a plateau. The decrease in accuracy past the optimum may be due to correlation between the coordinate rasters. Coordinates x and y are perfectly uncorrelated, but the coordinate rasters become increasingly correlated as their number increases. The optimal value may therefore be a trade-off between the increased ability of the model to

account for spatial trends and the adverse effect of increasingly correlated covariates. It is therefore likely that it depends on the complexity of the spatial distribution of the target variable as well as the number of observations.

With OGC in combination with auxiliary data, the process-based covariates in the auxiliary data most likely help to reduce the effect of correlation between the coordinate rasters. Furthermore, in this case, the number of coordinate rasters also affects the relative weighting between the auxiliary data and the coordinate rasters. When *mtry* is smaller than the total number of covariates, a higher number of coordinate rasters increases the chance that a coordinate raster will be available for a split. The optimal number of coordinate rasters may therefore depend on the optimal weighting between process-based and explicitly spatial covariates. This optimal weighting may depend on the number of covariates in the auxiliary data as well as the strength of the relationship between the target variable and the auxiliary data.

At present, several factors could therefore explain the optimal number of coordinate rasters for each dataset, with and without auxiliary data. The exact interplay between these factors is unclear, and the best option may therefore be to experiment with different numbers of coordinate rasters.

~~3.2 Comparison with other methods~~

3.2 Method comparison

3.2.1 Predictive accuracy

For all four datasets, t~~T~~here were large overlaps in the accuracies of the methods, as accuracies varied across the 100 repeated splits (Figure 7, Figure A2, Figure A4, Figure A6)~~, especially for RMSE~~. However, an analysis on the Vindum dataset revealed that the accuracies generally correlated between the methods across the repeated splits. The mean correlation coefficient (Pearson's R) was 0.52 (0.19 – 0.88) for R^2, 0.71 (0.65 – 0.71) for RMSE and 0.65 (0.41 – 0.89) for ccc. This shows that some holdout datasets yielded consistently high accuracies, while others yielded consistently low accuracies. Furthermore, especially for R^2 and ccc, a few holdout datasets yielded much lower accuracies than the other holdout datasets, leading to long negative tails (Figure 7, Figure A2, Figure A6).

14
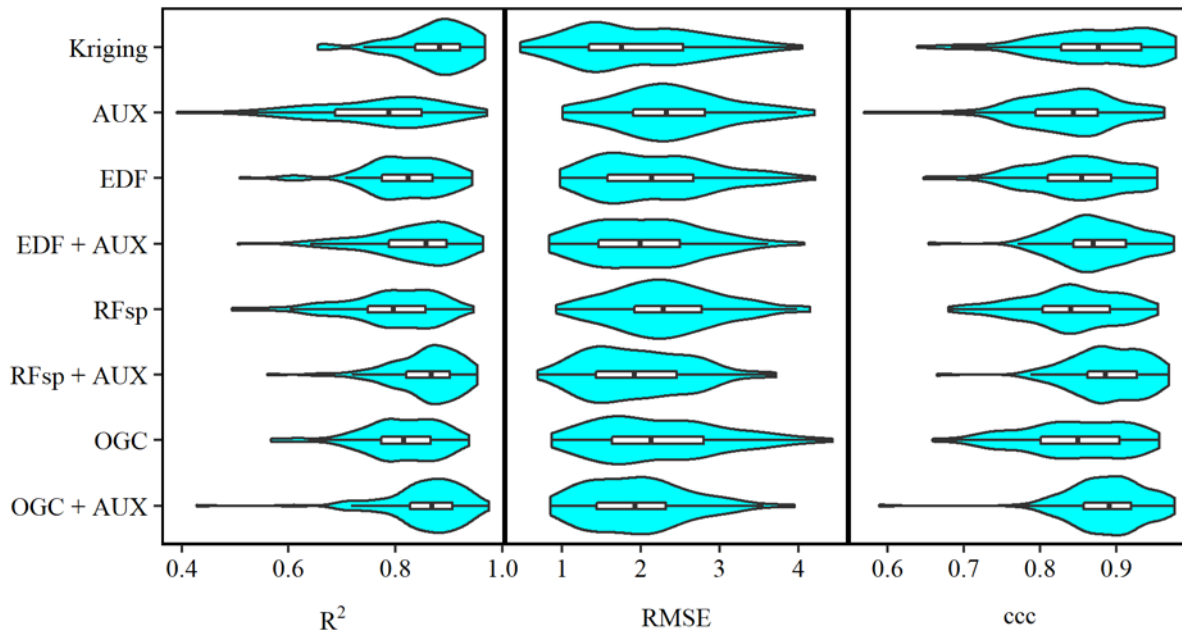
Figure 7: Violin plots showing accuracies of soil organic matter predictions on the Vindum dataset with kriging, and Random Forest models trained using either auxiliary data (AUX), Euclidean distance fields (EDF), distances to observations (RFsp), oblique geographic coordinates (OGC) or EDF, RFsp or OGC in conjunction with AUX. The figure shows Pearson's R^2, room mean square error (RMSE) and Lin's concordance obtained from 100 repeated splits (75% training dataset, 25% test dataset).

For the Vindum dataset, ~~K~~kriging achieved the highest rank for R^2 (Table 2). For RMSE, kriging shared the highest rank with EDF, RFsp and OGC in combination with auxiliary data. Lastly, OGC and RFsp in combination with auxiliary data shared the highest rank for ccc. In short, kriging, RFsp with auxiliary data and OGC with auxiliary data all had the highest rank for two accuracy metrics out of three. We therefore regard these three methods as the most accurate methods for the Vindum dataset. Furthermore, we regard these three methods as equally accurate for this dataset, as none of them was universally more accurate than the other two methods. ~~We therefore regard these three methods as most accurate.~~

Auxiliary data used on their own, as well as RFsp without auxiliary data had the lowest rank for all three accuracy metrics on the Vindum dataset. Furthermore, OGC without auxiliary data had the same rank as EDF without auxiliary data for all three accuracy metrics.

15

Table 2: Ranks for the accuracies of the methods <u>on the Vindum dataset</u>, calculated as <u>Pearson's</u> R^2, RMSE and ccc, respectively. Methods for which a pairwise t-test did not give a significant difference in accuracy (p > 0.05) received equal ranks for the metric in question. Ranks for the methods therefore differ between the three metrics. AUX: Auxiliary data. EDF: Euclidean distance fields. OGC: Oblique geographic coordinates. RFsp: <u>Distances between observations.</u>~~Spatial Random Forest.~~

Rank	R^2 Method	Mean	RMSE Method	Mean	ccc Method	Mean
1	Kriging	0.87	EDF + AUX	2.0	OGC + AUX	0.89
			Kriging	2.0	RFsp + AUX	0.89
			OGC + AUX	1.9		
			RFsp + AUX	1.9		
2	OGC + AUX	0.85	EDF	2.2	EDF + AUX	0.87
	RFsp + AUX	0.86	OGC	2.2	Kriging	0.87
3	EDF	0.82	AUX	2.4	AUX	0.84
	EDF + AUX	0.83	RFsp	2.3	EDF	0.85
	OGC	0.81			OGC	0.84
					RFsp	0.84
4	AUX	0.77				
	RFsp	0.79				

Pouladi et al. (2019) tested several methods for predicting SOM <u>on the Vindum dataset</u>~~within the field~~, including kriging and the machine learning algorithms Cubist and Random Forest, with and without kriged residuals. The authors found that kriging provided the most accurate predictions of SOM. The results ~~in this study~~<u>for Vindum</u> affirm the high accuracy of kriging predictions, but they also show that Random Forest models combining auxiliary data with spatial <u>trends</u>~~relationships~~ can achieve similar accuracies.

<u>For the *meuse* dataset, OGC in combination with auxiliary data achieved the highest rank for R^2 and RMSE (Table 3). For ccc, OGC in combination with auxiliary data shared the highest rank with EDF in combination with auxiliary data. Without auxiliary data, OGC received third rank for RMSE and fourth rank with R^2 and ccc. OGC without auxiliary data was generally on par with models based only on auxiliary data and with EDF. It was less accurate than combined methods and OK (R^2 and ccc). RFsp without auxiliary data was the least accurate method.</u>

Table 3<u>: Ranked accuracies obtained with each method on the *meuse* dataset, calculated as Pearson's R^2, RMSE and ccc. Methods received shared ranks if a pairwise t-test revealed no statistically significant difference between their accuracies for the metric in question. Each t-test used accuracies obtained with 100 repeated splits into training and test datasets. AUX: Auxiliary data. EDF: Euclidean distance fields. RFsp: Distances to observations. OGC: Oblique geographic coordinates. OK: Ordinary kriging. RK: Regression-kriging.</u>

Rank	R^2 Method	Mean	RMSE Method	Mean	ccc Method	Mean
1	OGC + AUX	0.68	OGC + AUX	202	EDF + AUX	0.78
					OGC + AUX	0.78

Rank	Method	Accuracy	Method		Method	
2	EDF + AUX	0.67	EDF + AUX	204	RFsp + AUX	0.77
	RFsp + AUX	0.66	RFsp + AUX	206		
3	OK	0.63	AUX	224	OK	0.76
	RK	0.63	EDF	226	RK	0.76
			OGC	220		
			OK	215		
			RK	216		
4	AUX	0.61	RFsp	250	AUX	0.74
	EDF	0.59			OGC	0.74
	OGC	0.61				
5	RFsp	0.50			EDF	0.71
6					RFsp	0.63

For the *eberg* dataset, OGC in combination with auxiliary data was the most accurate method (Table 4). Without auxiliary data, OGC had the third rank. It was less accurate than EDF combined with auxiliary data, but more accurate than EDF without auxiliary data and models based only on auxiliary data. Models based only on auxiliary data yielded the lowest accuracies.

Table 4: Ranks of the accuracies (percent cases correctly predicted) obtained with each method on the *eberg* dataset. Pairwise t-tests showed that differences between the accuracies of the methods were all statistically significant ($p < 0.05$). AUX: Auxiliary data. EDF: Euclidean distance fields. OGC: Oblique geographic coordinates.

Rank	Method	Accuracy
1	OGC + AUX	0.39
2	EDF + AUX	0.38
3	OGC	0.37
4	EDF	0.37
5	AUX	0.35

For the Swiss rainfall dataset, OGC was the most accurate method for all three metrics (Table 5). RFsp was the second-most accurate method, followed by EDF. OK was the least accurate method.

Table 5: Ranked accuracies on the Swiss rainfall dataset for each method. Pairwise t-test showed statistically significant ($p < 0.05$) differences between the methods for all three metrics. The ranks are the same for all three metrics. EDF: Euclidean distance fields. RFsp: Distances to observations. OGC: Oblique geographic coordinates. OK: Ordinary kriging.

Rank	Method	R^2	RMSE	ccc
1	OGC	0.831	4.7	0.902
2	RFsp	0.822	4.8	0.893
3	EDF	0.818	4.9	0.891
4	OK	0.804	5.0	0.887

In summary, for Vindum, *meuse* and *eberg*, OGC combined with auxiliary data was either the most accurate method or one

340 of the most accurate methods. Without auxiliary data, OGC was not one of the most accurate methods for these datasets. However, for the Swiss rainfall dataset, OGC was the most accurate method, even though we used no auxiliary data. It is important to consider that in most cases all methods yielded acceptable accuracies. Although the differences between the accuracies of the methods were in many cases statistically significant, they were generally small. However, the results show that OGC compares well with other methods for integrating spatial trends in machine learning models.

345 **3.2.2 Maps**

For the Vindum dataset, Kkriging produced a smooth prediction surface, which is very common for this method (Figure 8A). The prediction surface with EDF was mostly smooth, but it also contained a distinct "rings in the water" artefact caused by the use of the raster with the distance to the middle of the study area as a covariate (Figure 8B). The prediction surface with RFsp was smoother than the prediction surface produced by kriging (Figure 8C). The predictions with only auxiliary data

350 were very similar to the predictions made with x- and y-coordinates in combination with auxiliary data (compare Figure 8C to Figure 6Figure 6A). In combination with auxiliary data, both EDF and RFsp produced smoothing effects similar to the effect seen with OGC in combination with auxiliary data (compare Figure 8E and Figure 8F to Figure 6Figure 6B). However, for EDF the smoothing was less visible than with OGC, and for RFsp it was more visible than with OGC.
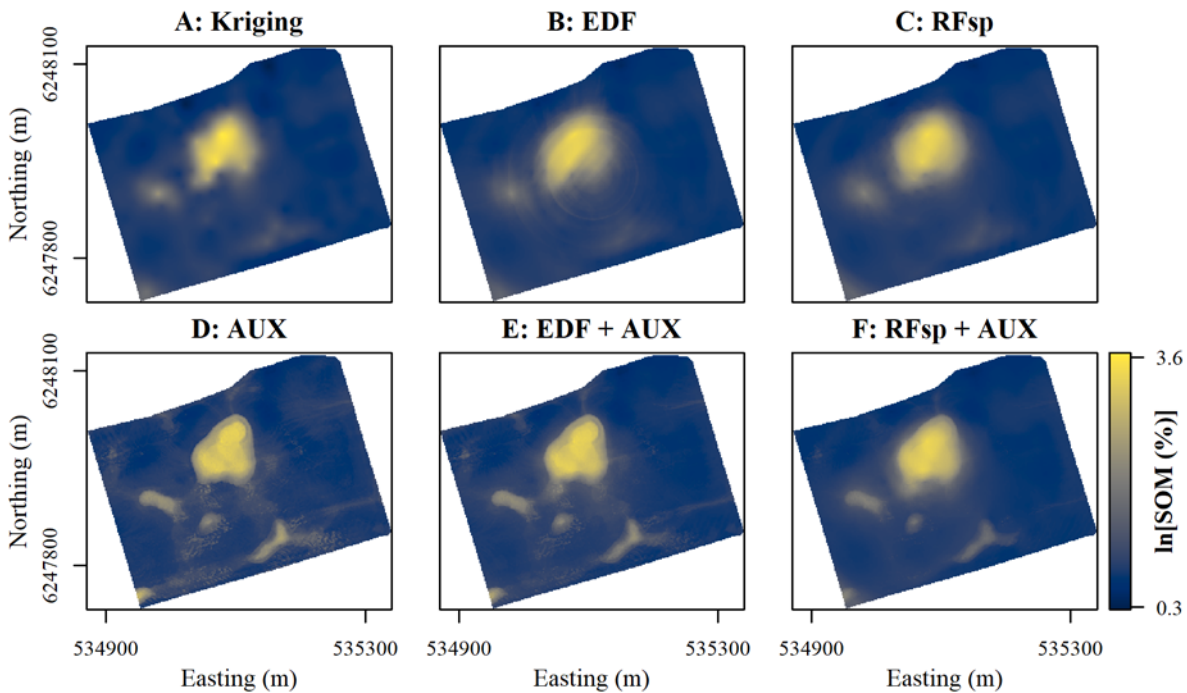

355 **Figure 8: Prediction of soil organic matter (SOM) contents for the topsoil at Vindum using A: Kriging, or Random Forest models trained with B: Euclidean distance fields (EDF), C: Distances to observations (RFsp), D: Auxiliary data (AUX), E: EDF in conjunction with AUX, or F: RFsp in conjunction with AUX. Easting and northing for UTM Zone 32N, ETRS 1989.**

18

For the *meuse* dataset, OK, EDF and RFsp produced smooth prediction surfaces (Figure 9). However, OGC without auxiliary data produced a prediction surface with several abrupt, angular artefacts. The accuracy of OGC without auxiliary data was on par with some of the other methods, but the maps revealed that the predictions were not realistic. Predictions with the combined methods (RK, EDF + AUX, RFsp + AUX and OGC + AUX) were mostly similar to predictions with only auxiliary data. However, in some places these methods smoothed out the spatial patterns produced with only auxiliary data (for example in the northern part of the study area), and in other places they made them more distinct (for example south-west of the middle of the study area). In this regard, the results are similar to the results from Vindum.
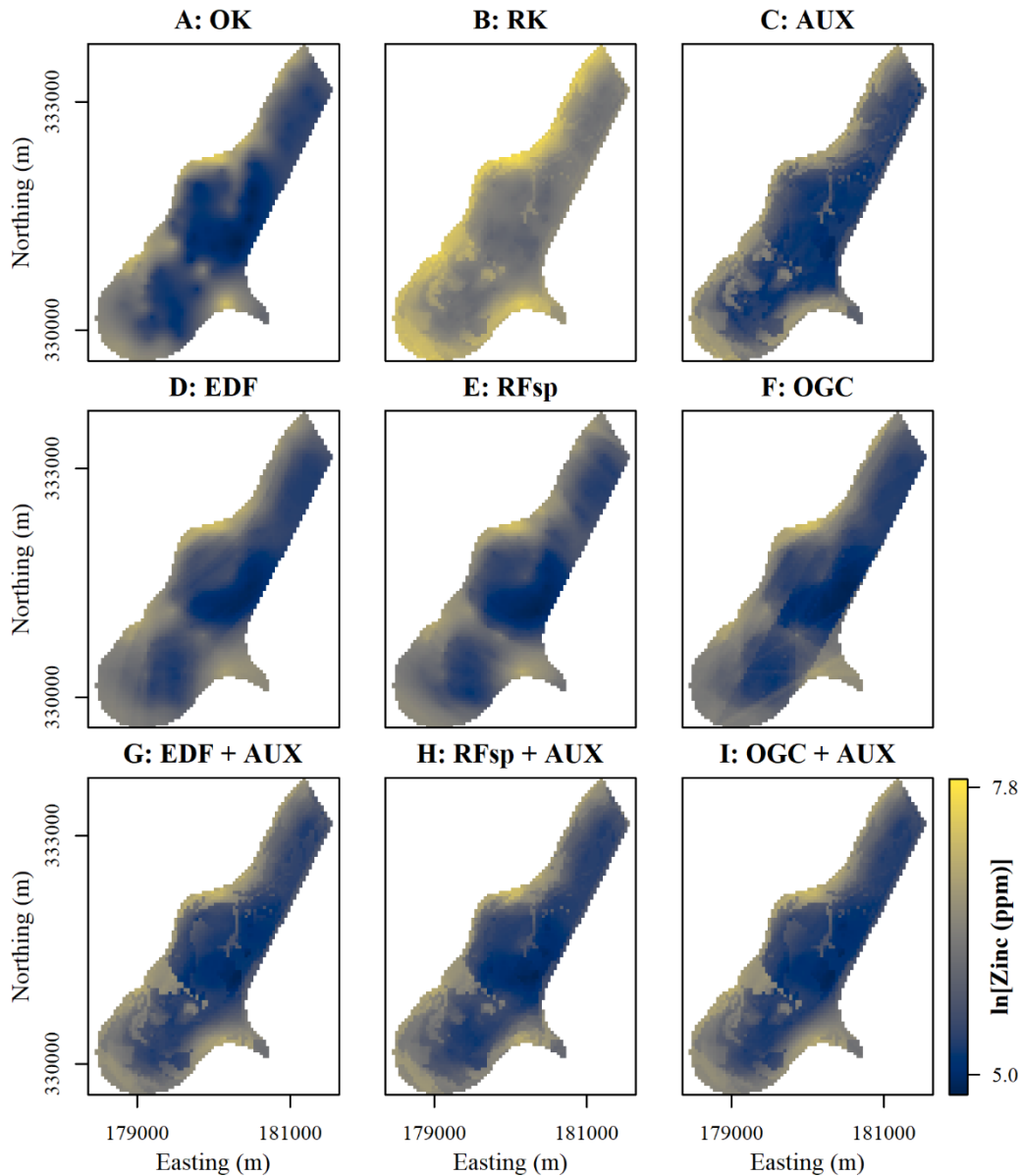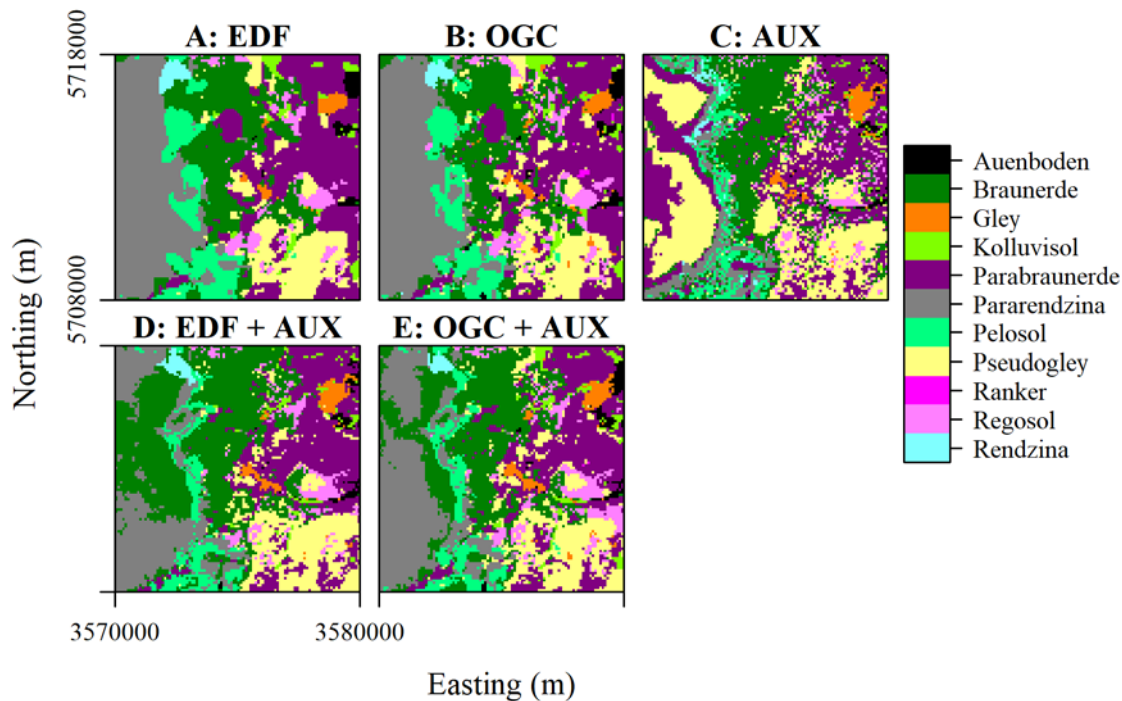
370   For the *eberg* dataset, predictions based only on auxiliary data showed a very noisy spatial pattern with many soil types occupying small incoherent areas (Figure 10C). The spatial patterns produced with OGC and especially EDF were much smoother and contained several large, rounded areas with little internal variation in soil types (Figure 10A and Figure 10B). The predictions obtained with the combined methods were similar to the spatial pattern obtained with only auxiliary data. However, they were much smoother, as the soil types occupied mostly coherent areas. The effect for predictions of soil types

375   therefore appears similar to the effect for numeric variables seen for Vindum and *meuse*.

380   For the Swis rainfall dataset, OK produced a smooth, highly anisotropic prediction surface (Figure 11A). The prediction surfaces of EDF, RFsp and OGC also showed anisotropy, but they were generally smoother and more rounded. For example, with OK, some individual observations showed an effect on the prediction surface as elongated spots in the direction of the anisotropy. With the other three methods, a few individual observations showed an effect in the prediction surface, but their effects are more rounded and less distinct. The predictions with EDF, RFsp and OGC therefore appear more general than the

385   OK predictions. Moreover, the prediction surfaces of EDF, RFsp and OGC appear very similar.
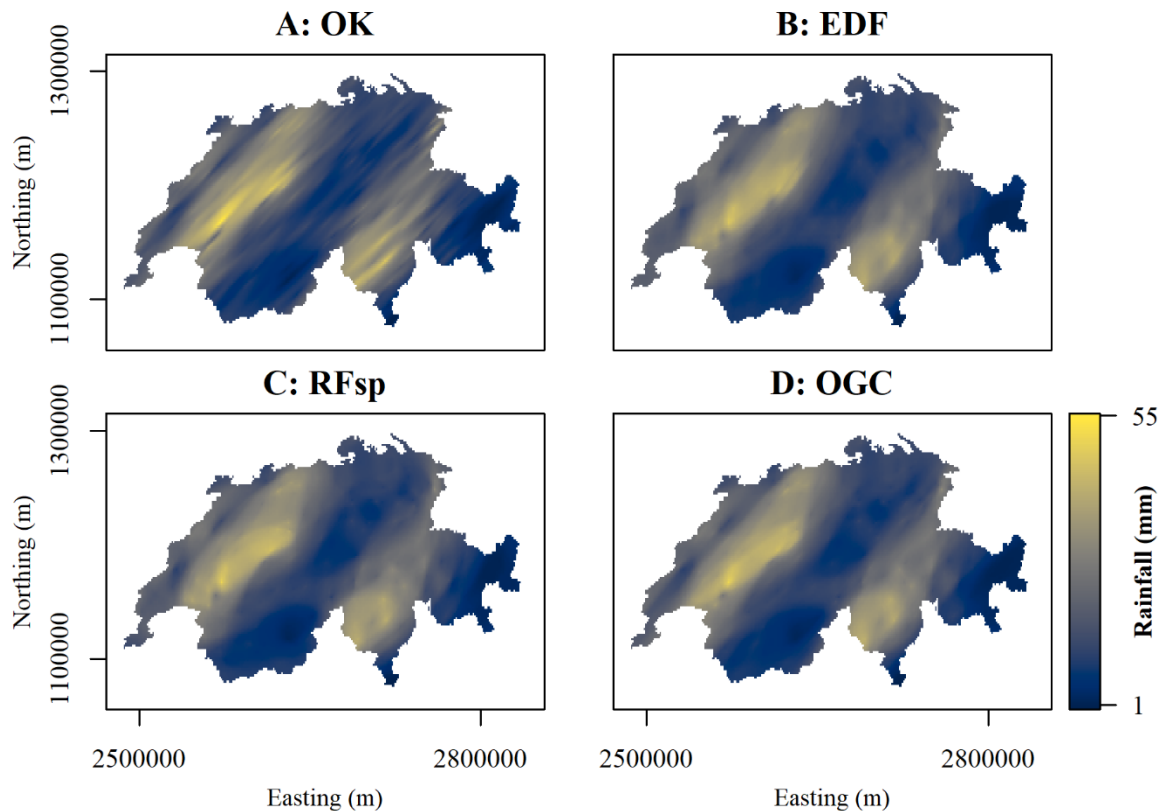
Figure 11: Maps of rainfall on 1986-05-08 in Switzerland predicted with each method. Northing and easting are coordinates according to the Swiss coordinate system LV95. A: Ordinary kriging (OK). B: Euclidean distance fields (EDF). C: Distances to observations (RFsp). D: Oblique geographic coordinates (OGC).

3.2.3 Residuals

For the Vindum dataset, the residuals of the SOM predictions~~For all methods except kriging, the residuals of the SOM predictions~~ had some degree of spatial dependence for all methods except kriging (Figure **12**~~Figure 9~~). This finding contrasts with Hengl et al. (2018) who found that there was no spatial trend in the residuals of predictions with RFsp. EDF, RFsp and OGC used without auxiliary data had the most spatially dependent residuals.~~, with nugget to sill ratios of 0.40. On the other hand, EDF and OGC in combination with auxiliary data had the least spatially dependent residuals after kriging, with nugget to sill ratios of 0.65 and 0.64, respectively.~~ However, the residuals of the combined methods (EDF + AUX, RFsp + AUX and OGC + AUX) had less spatial dependence than the residuals of models based only on auxiliary data. OGC + AUX was the machine learning method with the least spatially dependent residuals, although the residuals still had more spatial dependence than kriging residuals.
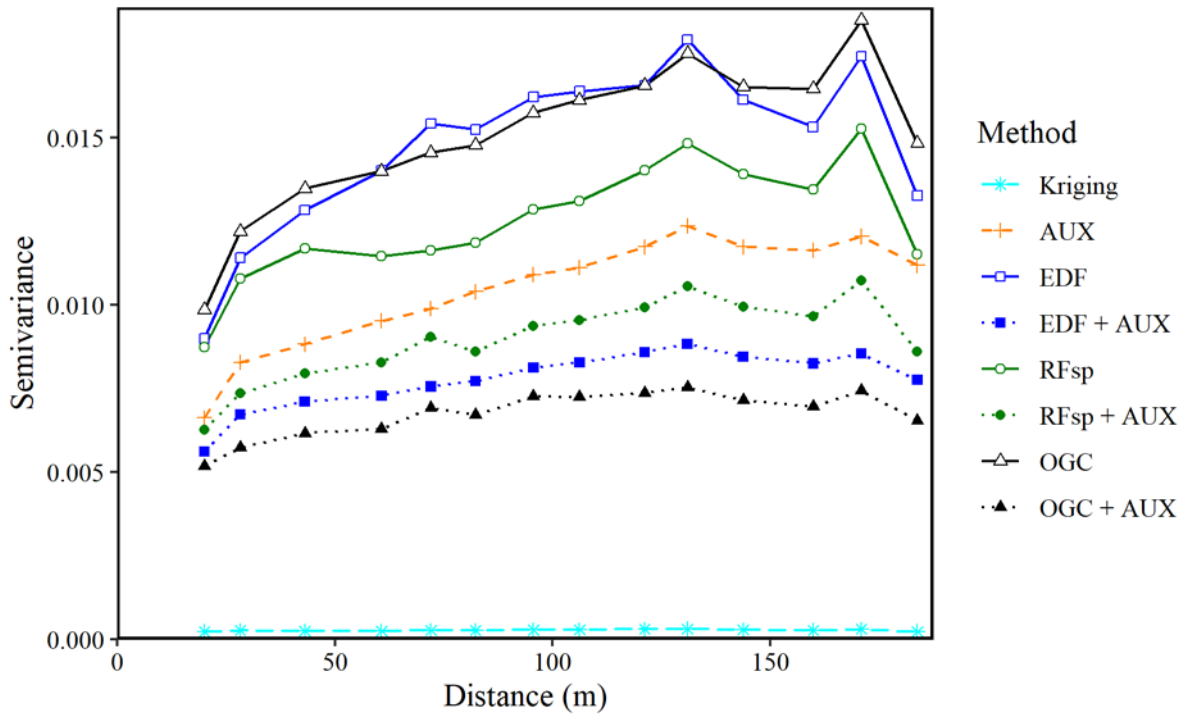
21
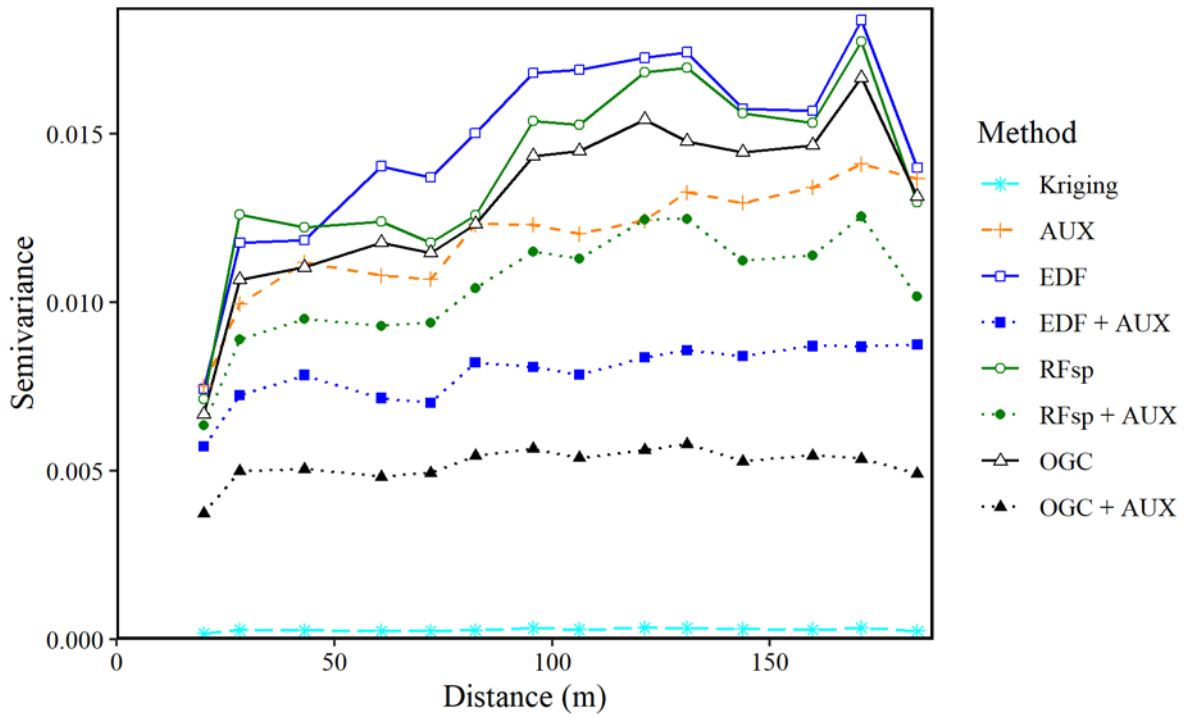
400

Figure 12**: Experimental variograms for the residuals of the SOM predictions made with each method <u>for the Vindum dataset</u>. The variograms use residuals from natural logarithmic-transformed SOM measurements and predictions. AUX: Auxiliary data. EDF: Euclidean distance fields. RFsp: Spatial Random Forest. OGC: Oblique geographic coordinates.**

405 **3.3 Covariate importance**

<u>For the Vindum dataset,</u> ~~T~~the most important covariate from the auxiliary data was the depth of sinks (<u>Table 6</u>~~Table 3~~). The most likely reason for its high importance is the presence of a large sink with very high SOM contents northwest of the middle of th<u>is</u>~~e~~ study area (Figure 1). As sinks trap surface runoff, they often have wet conditions, which give rise to peat accumulation.

410

Table 6: Covariate importance for the model using OGC in combination with auxiliary data for Vindum. The importance for OGC represents the sum of the importance of the coordinate rasters at 80 different angles.
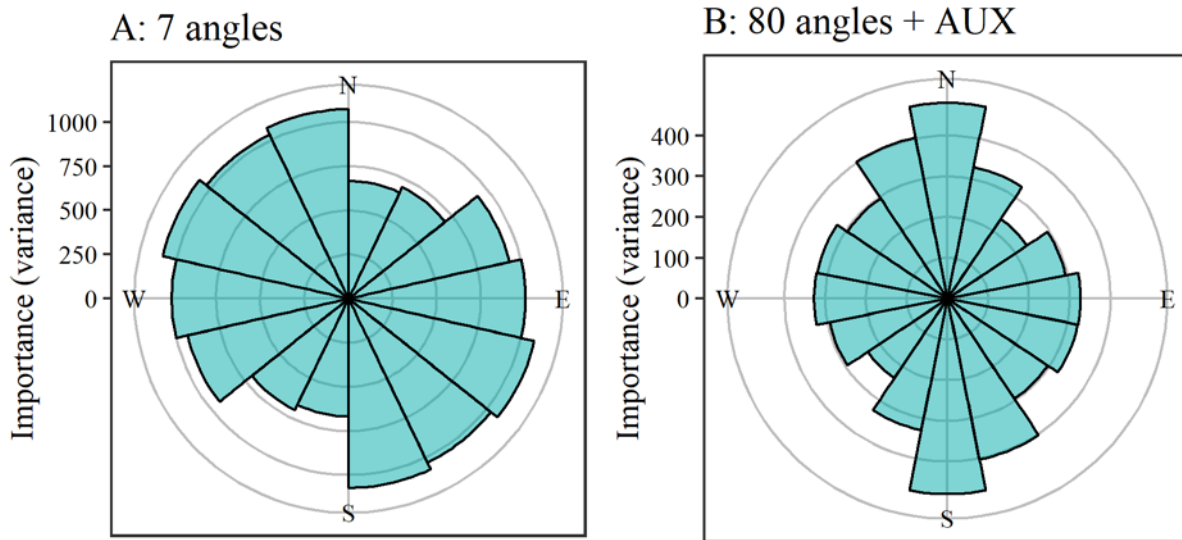
Covariate	Importance (variance)
OGC	2689
Depth of sinks	2003
MRVBF	476
SAGA wetness index	170
Elevation	166
Valley depth	157
EC$_a$	123
Slope gradient	101
Mid-slope position	84
NDVI	76
Plan curvature	74
SL	64
SAVI	58
Cos aspect	44
DVI	42
TWI	38
RVI	34
Flow accumulation	32
Sin aspect	32
Profile curvature	21

When used in combination with the auxiliary data, the importance of the individual coordinate rasters varied from 0.6% to 3.1% of the importance of the depth of sinks, with mean value of 1.7%. The most important coordinate raster had $\theta = 0.48\pi$ (close to a north-south axis) and was the 12th most important covariate. The sum of the importance of the coordinate rasters was equal to 134.3% of the importance of the depth of sinks (Table 6Table 3). Therefore, with coordinate rasters at 80 different angles, the effect of the individual rasters on the predictions was subtle, but their combined effect was strong. Figure 13Figure 10 shows the importance of the coordinate rasters relative to θ, in a way similar to a wind rose. The plots repeat the bars for $\theta \geq \pi$, as the importance for a given angle is directionless. For example, the importance of $\theta = 0$ (East) is equal to the importance of $\theta = \pi$ (West).

Without auxiliary data, the most important coordinate rasters had a general northwest to southeast angle (Figure 13Figure 10). On the other hand, the coordinate rasters with angles between a north-south and a northeast-southwest axis had low importance. The most likely reason for this pattern is the location of the sink with very high SOM contents to the northwest of the middle of thise study area. This creates a large difference in the SOM contents of the northwestern and southeastern parts of the study area, giving large importance to covariates that can explain this difference. Additionally, the northwest side

of the sink has a very steep slope, creating a steep gradient in SOM contents in this direction. A stable variogram showed anisotropy along a north-north-east to south-south-west axis ($\theta = 0.34\pi$) with a major range of 136 m and a minor range of 118 m. The direction of the anisotropy therefore coincided with the direction of the least important coordinate rasters.

430

Figure 13: Covariate importance of the coordinate rasters at various angles for Vindum. A: Importance of coordinate rasters at seven different angles. B: Importance of coordinate rasters at 80 different angles used with auxiliary data (importance of auxiliary data not shown). The sizes of the bars show the importance of the coordinate rasters at a given angle. Bars in B show the sum of the importance for coordinate rasters aggregated into 0.125π intervals.

435 On the other hand, with OGC in combination with auxiliary data, the most important coordinate rasters had tilt angles close to a north-south axis ($\theta = 0.5\pi$). At the same time, the least important coordinate rasters had tilt angles close to a northeast-southwest axis ($\theta = 0.25\pi$). The residuals from the predictions with only auxiliary data only also displayed a degree of anisotropy. A stable variogram showed anisotropy along a northeast to southwest axis ($\theta = 0.21\pi$), with a major range of 52 m and a minor range of 38 m. Again, the angle of the anisotropy coincided with the angle of the least important coordinate

440 rasters. The spatial pattern of the residuals therefore differed from the spatial pattern of the SOM contents in the Vindum study area. Apparently, there are unaccounted-for processes decreasing the spatial variation along a northeast-southwest axis relative to other angles.

A possible cause of the anisotropy in the residuals may be the ploughing direction. The main ploughing direction in the Vindum study area is along an east-north-east to west-south-west axis ($\theta = 0.18\pi$). This angle is nearly parallel to the angle

445 of the least important coordinate rasters (Figure 14Figure 11). The ploughing direction, combined with the topography, has a large impact on soil movement, as ploughing displaces soil both along and across its direction (Lindstrom et al., 1990, De Alba, 2003, Heckrath et al., 2006). Most of the study area has the same ploughing direction, irrespective of the topography, resulting in up-, down- and cross-slope ploughing in various parts of the field. This creates in a complex pattern of soil

redistribution, which likely affects the SOM contents of the topsoil. As downslope soil movement is strongest in the

450 ploughing direction, variation in soil properties parallel to this direction is likely to be smaller than the variation

perpendicular to the ploughing direction. This corresponds to the low importance of coordinate rasters with angles close to

the ploughing direction. However, none of the auxiliary data accounted for the ploughing direction. This indicates that OGC

can add information on the most likely processes affecting soil properties in an area.
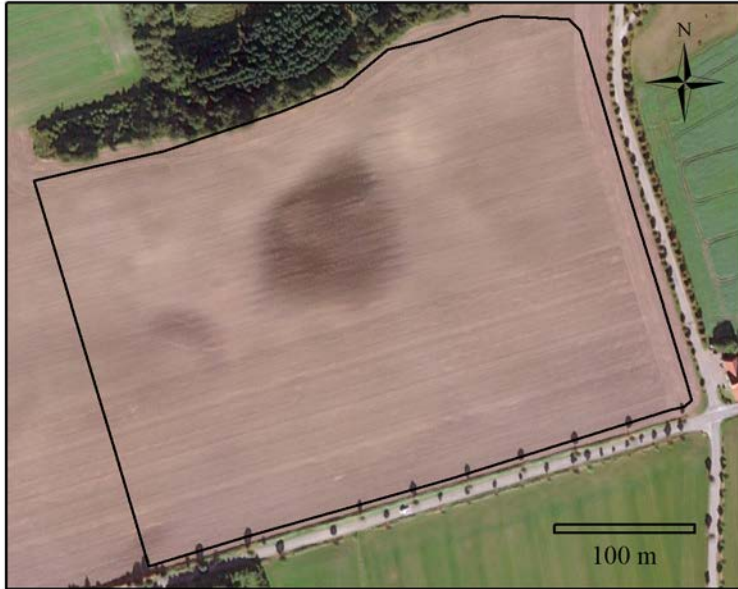
455 **Figure 14: Orthophoto of the study area from September 27, 2016 (Esri, 2019). Sources: Esri, DigitalGlobe, Earthstar Geographics, CNES/Airbus DS, GeoEye, USDA FSA, USGS, Aerogrid, IGN, IGP, and the GIS User Community.**

3.4 Choice of method

At Vindum, Tthe three most accurate methods were kriging, RFsp with auxiliary data and OGC with auxiliary data. For *meuse*, OGC and EDF combined with auxiliary data were most accurate, and for *eberg*, OGC combined with auxiliary data

460 was most accurate. For the Swiss rainfall dataset, OGC was the most accurate method.
Although kriging was in most cases less accurate than other methods, Manysome soil mappers would probably still choose krigingit for mapping SOM in this fieldsoil properties, givendue to its computational efficiency and conceptual simplicity. However, aside from accuracy, anThe advantage of the methods based on machine learning instead lies in the fact that they provide larger amounts of information than geostatistical modelstheir interpretability. Kriging in itself does not provide

465 information on the processes that control spatial variation in soil properties, but researchers can interpret machine learning models can include covariates related to soil processes, providing information on the processes that are most likely to affect in order to discover the most likely processes affecting the spatial distribution of a soil property.
With spatial approaches such as EDF, RFsp and OGC, researchers can incorporate feature space and geographic space in a machine learning model. Of the previously used approaches, OGC is most similar to EDF, as it used the x- and y-

26

470 coordinates, and the distances to the corners of the study area resemble coordinates. On the other hand, RFsp is more similar to geostatistical models, as it relies on distances between observations. However, this similarity comes at the cost of calculating a large number of distance rasters.

~~The benefit~~ One advantage of using spatially explicit covariates (EDF, RFsp or OGC) is that researchers can interpret local and spatial effects at once. In this regard, OGC has an advantage over EDF and RFsp, as it is clear what the coordinate

475 rasters represent. ~~On the other hand, i~~It is less clear how researchers should interpret distances to the corners of the study area or the distance to a specific observation. We have also shown that it is straightforward to illustrate covariate importance for OGC.

~~While kriging, RFsp with auxiliary data and OGC with auxiliary data yielded equally accurate predictions, it is likely that it is due to the high sampling density of the study area. For larger, less densely sampled areas, OGC and RFsp with auxiliary~~

480 ~~data are likely to provide higher relative accuracies.~~

Furthermore, an advantage of OGC relative to RFsp is that OGC required fewer covariates to achieve the same accuracy. In fact, without auxiliary data, OGC achieved a higher accuracy with a smaller number of covariates for the data sets Vindum, *meuse* and Swiss rainfall. This demonstrates a clear advantage of OGC, as it is possible to adjust the number of coordinate rasters. EDF and RFsp do not presently have similar options.

485 We will stress that, as a rule, soil mappers should not use machine learning models relying only on spatial trends~~relationships~~, as EDF, RFsp and OGC all yielded low~~er~~ accuracies without auxiliary data for the soil datasets (Vindum, *meuse* and *eberg*). Moreover, surprisingly, these methods had the most spatially autocorrelated residuals for the Vindum dataset, although they relied exclusively on spatial trends~~relationships~~. The results therefore suggest that soil mappers should primarily use these methods in combination with auxiliary data, and not on their own~~are best suited for~~

490 ~~integrating spatial relationships with auxiliary data~~. If no auxiliary data are available, kriging is likely to be a better option.~~If the purpose requires a purely spatial method, kriging is a better option.~~ However, results from the Swiss rainfall dataset show that, for other spatial problems, auxiliary data may be unnecessary.

4 Conclusions

We have shown in this study that the use of oblique geographic coordinates (OGC) is a reliable method for integrating

495 auxiliary data with spatial trends~~relationships~~ for modelling and mapping soil properties. In most cases, the method eliminated the orthogonal artefacts that arise from the use of x- and y-coordinates and achieved higher accuracies than maps created with only two coordinate rasters. However, for *meuse*, without auxiliary data, OGC still produced abrupt angular artefacts in the final map. ~~It~~Soil mappers should therefore combine OGC with auxiliary data, as this gives higher accuracies and spatial patterns with a higher degree of realism.

500 OGC is more interpretable than previous similar approaches, and more flexible, as it is possible to adjust the number of coordinate rasters. This should allow soil mappers to find a good compromise between accuracy and computational

27

efficiency for mapping soil properties, as the optimal number of coordinate rasters may vary depending on the study area and the soil property in question.

At this point, we have only tested the method for ~~one soil property in one area~~three soil datasets and one meteorological dataset. ~~, and i~~It will therefore be highly relevant to test the method for other soil properties and areas. It will especially be relevant to test the method in larger, less densely sampled areas. Previous studies have shown that machine learning is likely to provide higher accuracies in such areas (Zhang et al., 2008, Greve et al., 2010, Keskin et al., 2019), and it will be relevant to test if this is also the case for oblique geographic coordinates~~as it may have its greatest relative advantages in these cases~~. ~~Furthermore, the~~ ~~r~~Results from the Vindum and the Swiss rainfall datasets also suggest that the method can be useful for ~~predicting~~ mapping ~~properties~~ variables with anisotropic spatial distributions, and it will therefore be relevant to test it on datasets with a high degree of anisotropy. Lastly, one should note that we carried out this study for relatively small areas using "flat" coordinate systems. Using OGC for larger areas and other coordinate systems may require alterations to the method.

We call upon researchers within digital soil mapping to aid us in testing oblique geographic coordinates as covariates for additional datasets~~this endeavour~~, and we have therefore made the function for generating oblique geographic coordinates available as an R package. Moreover, to allow other researchers to test methods on the ~~dataset that we used~~Vindum dataset, we have made it available as well as part of the same package.

5 Code and data availability

The function for generating oblique geographic coordinates is available as an R package at https://bitbucket.org/abmoeller/ogc/src/master/rPackage/OGC/. The package also contains the SOM observations and auxiliary data ~~used in this study.~~from the Vindum dataset.

Furthermore, we have made the R code used in this study available in a public repository at ~~http://dx.doi.org/10.5281/zenodo.3496935~~http://dx.doi.org/10.5281/zenodo.3820068.

6 Appendix A: Methods and results for additional datasets

6.1 Methods

6.1.1 meuse

We mapped zinc contents for the *meuse* dataset (155 points). The *meuse* dataset contains covariates including the flooding frequency and the distance to the river. We added two covariates in the form of a digital elevation model (DEM, https://www.ahn.nl/) and surface water occurrence (Pekel et al., 2016). We converted the categorical raster of flooding frequency to indicator variables and transformed all the covariates to principal components. This resulted in six principal components.

28

We tested all the methods applied to the Vindum dataset, with the addition of regression-kriging (RK). We used Random Forest models trained on the auxiliary data for regression and then kriged the residuals using the function *krige.conv* from the R package *geoR* (Ribeiro Jr et al., 2020). As for the Vindum dataset, we tested each method with 100 repeated splits into training (75%) and test (25%) data. For each split, we calculated Pearson's R^2, RMSE and ccc. We carried out pairwise t-test on the accuracies obtained with each method in order to asses if the differences between their accuracies were statistically significant. We also produced maps with each of the nine methods in order to compare results.

6.1.2 eberg

We mapped soil types for the *eberg* dataset. The *eberg* dataset contains 3,670 soil observations. We removed points outside the coverage of the covariates and points without a soil type classification. Furthermore, we removed the soil types "Moor" and "HMoor", as only one observation was available for each soil type. This reduced the dataset to 2,552 observations. The *eberg* dataset contains covariates including the parent material, a DEM, the SAGA GIS topographic wetness index and the Thermal Infrared reflectance from satellite imagery. We converted the parent material classes to indicators and converted all covariates to principal components. This resulted in 11 principal components.

The dataset is highly clustered, which is likely to affect accuracy assessments, as some areas have much higher point densities than others. To counter this effect, we organized the data in 100 groups using k-means clustering on their coordinates. We then produced 100 splits into training and test data based on these groups. In each split, the training data contained observations from 75 groups, and the test data contained observations from the remaining 25 groups. As we aimed to predict a categorical variable, we did not use kriging. Furthermore, due to the large size of the dataset, we did not use RFsp, as this would require us to produce more than 2,000 raster layers with buffer distances. Hengl et al. (2018) avoided this by calculating only buffer distances to each soil type. However, we did not choose this solution, as it would create problems for accuracy assessment. If a raster layer contains distances to test observations as well as training observations, the result would be circular logic, invalidating the accuracy assessment. Buffer distances based only on the training data would be less problematic. However, as we used 100 repeated splits, this was not an option.

We therefore tested only five methods for the *eberg* dataset: Models based on auxiliary data (AUX), Euclidean distance fields (EDF), OGC, as well as EDF and OGC combined with auxiliary data. Due to the large size of the dataset, model training was slower than for the other datasets. We therefore tuned a Random Forest model only once for each method and used the resulting parameterization for all 100 data splits. For each split, we calculated accuracy on the test data as the fraction of observations correctly predicted. We carried out pairwise t-tests on the accuracies obtained with each method in order to assess if the differences between their accuracies were statistically significant.

We produced maps of soil types with each of the five methods in order to compare results.

6.1.3 Swiss rainfall

The Swiss rainfall dataset contains 467 rainfall observations from Switzerland from May 8, 1986. We did not use any covariates for this dataset, and we therefore tested only purely spatial methods. We tested ordinary kriging with correction for anisotropy, EDF, RFsp and OGC. As for the Vindum dataset, we tested each method with 100 repeated splits into training data (75%) and test data (25%). For each split, we calculated Pearson's R^2, RMSE and ccc. We carried out pairwise t-tests on the accuracies obtained with each method in order to assess if the differences between their accuracies were statistically significant. Lastly, we produced maps of rainfall with each of the four methods in order to compare results.

6.2 Results

6.2.1 meuse

For the *meuse* dataset, the accuracy of OGC combined with auxiliary data was consistently higher than the accuracy of OGC without auxiliary data, irrespective of the accuracy metric and the number of coordinate rasters (Figure A1). The accuracy of OGC initially increased quickly with the number of coordinate rasters up to an optimum, after which there was no further improvement. For OGC + AUX, the increase in accuracy was more gradual, up to an optimum, after which accuracy deteriorated slightly. The optimal number of coordinate rasters without auxiliary data was six for RMSE and ccc and eight for R^2. With auxiliary data, the optimal number of coordinate rasters was 11 for RMSE and 13 for R^2 and ccc. In the subsequent analysis, we used six coordinate rasters for OGC without auxiliary data and 11 coordinate rasters for OGC with auxiliary data.
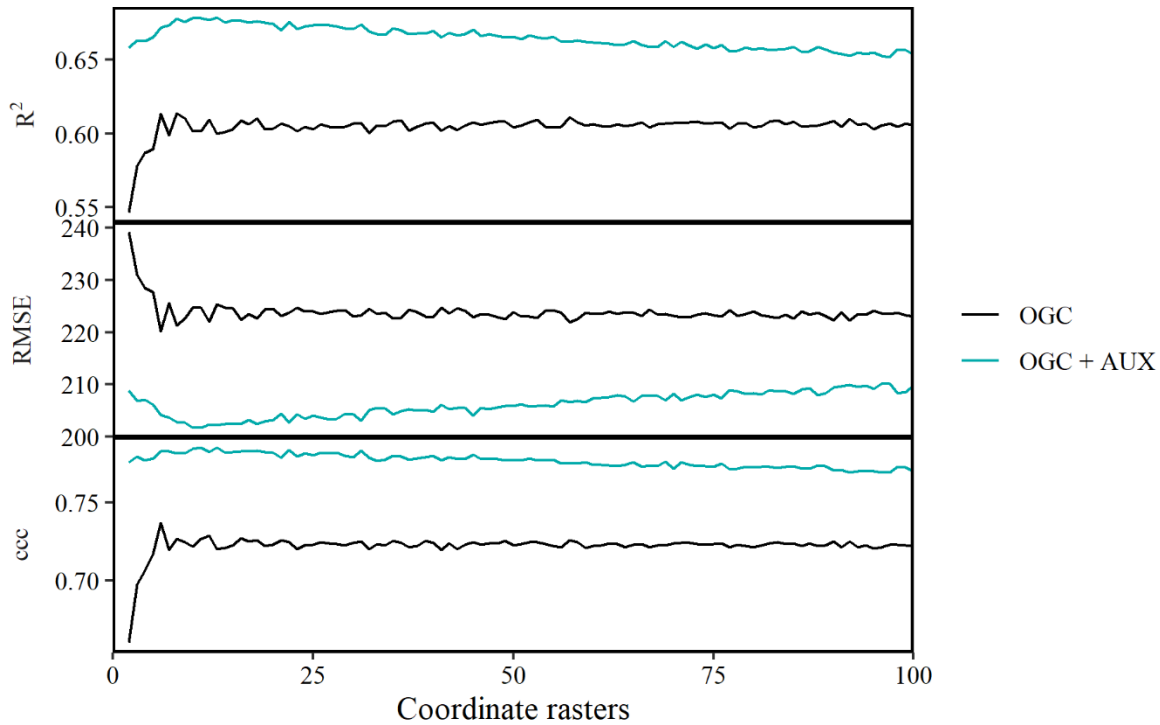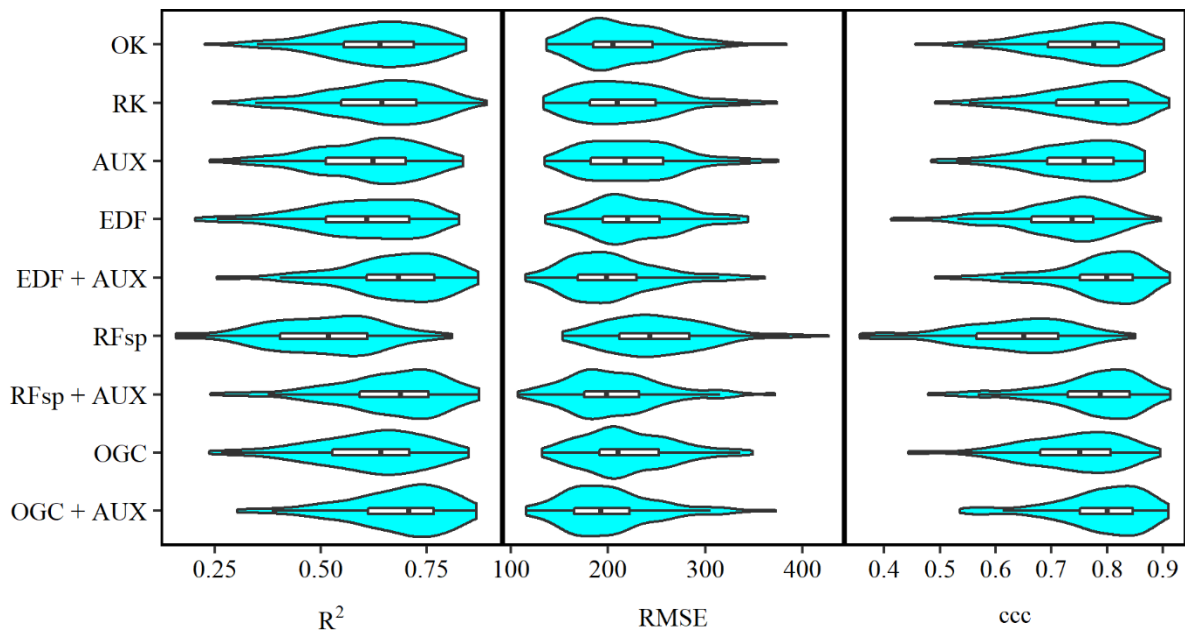
Figure A1: Accuracy of predictions on the *meuse* dataset (zinc contents) versus the number of coordinate rasters with oblique geographic coordinates, with and without auxiliary data. The values are averages obtained with 100 splits into training and test data.

For the *meuse* dataset, as for the Vindum dataset, the differences between the accuracies of the methods was relatively small relative to the variation in accuracy between the test splits (Figure A2). Furthermore, most methods had long tails with lower accuracies.
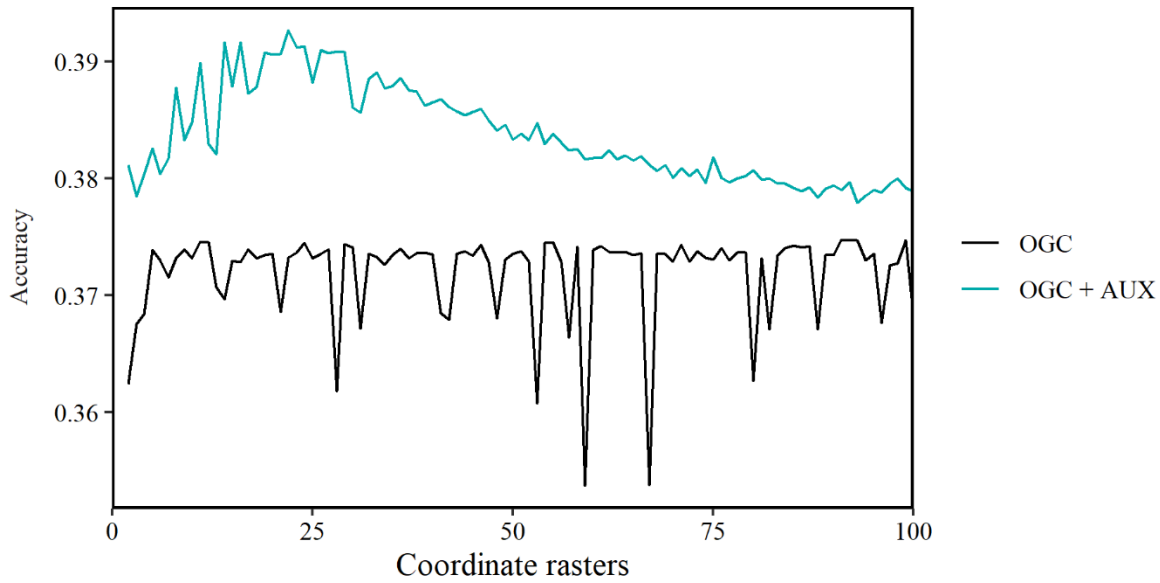
6.2.2 eberg

For the *eberg* dataset, accuracy for OGC without auxiliary data first increased sharply up to five coordinate rasters. Past this point, there was little improvement in accuracy, and some numbers of coordinate rasters produced sharp, irregular drops in accuracy (Figure A3). Combined with auxiliary data, the accuracy of OGC increased up to 22 coordinate rasters, after which it gradually declined. Without auxiliary data, the optimal number of coordinate rasters was 91. However, the highly irregular pattern of the accuracies did not justify any number past the initial increase, and we therefore used only five coordinate rasters. For OGC combined with auxiliary data, we used 22 coordinate rasters.

For the *eberg* dataset, as for the Vindum dataset, variation in accuracy between the splits into training and test data was in most cases greater than variation between the methods (Figure A4). However, unlike the other datasets, the distributions of the accuracies were mostly symmetric.
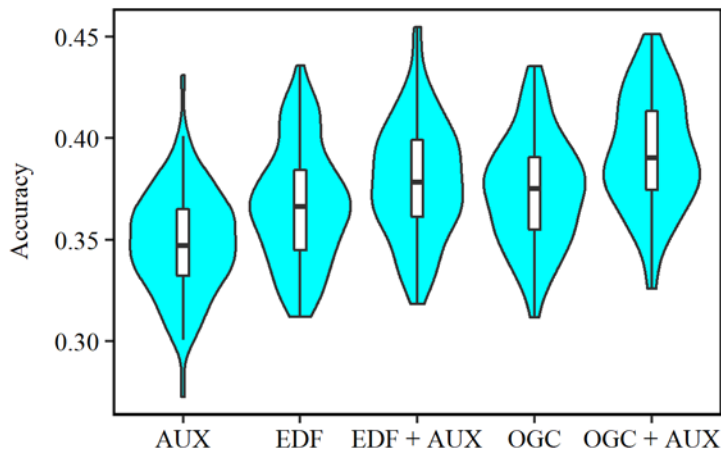
6.2.3 Swiss rainfall

For the Swiss rainfall dataset, the accuracy of OGC generally increased with the number of coordinate rasters (Figure A5). The increase in accuracy was steep at first, then gradual. For Pearson's R^2, the optimal number of coordinate rasters was 33, and for RMSE and ccc, it was 50. There was little change in accuracy past the optimal number of coordinate rasters.
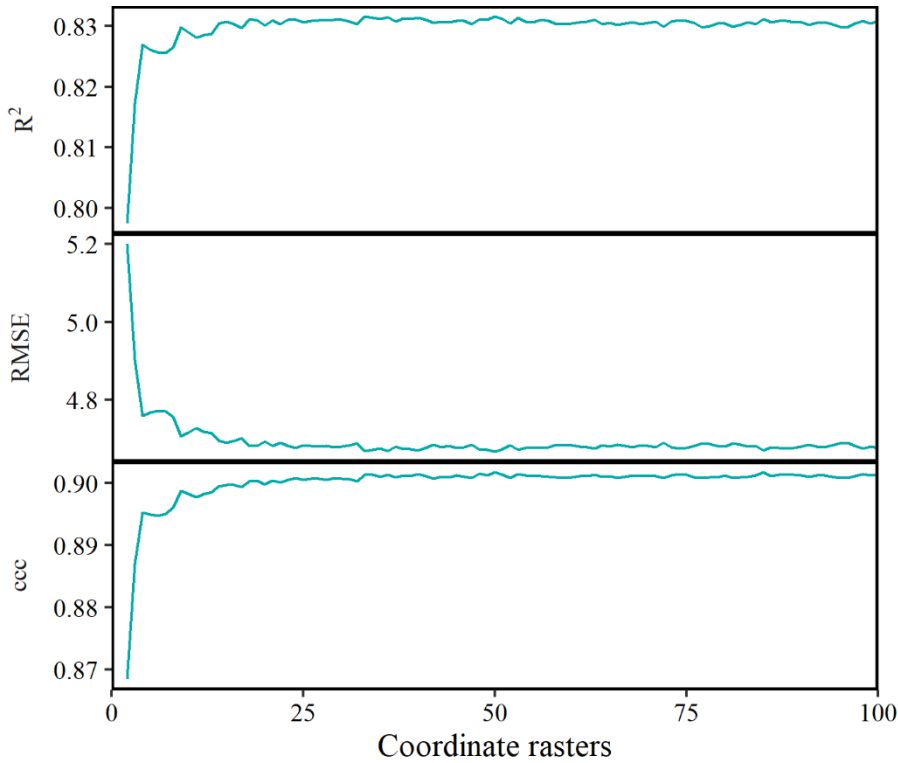
Figure A5: Accuracy of predictions on the Swiss rainfall dataset versus the number of rasters with oblique geographic coordinates. The values are averages obtained with 100 splits into training and test data.

As for the other datasets, variation in accuracies on the Swiss rainfall dataset was greater between the splits into training and test data than between the methods (Figure A6). The distributions of RMSE were mostly symmetric, but the distributions of R^2 and ccc had long negative tails, as some splits yielded much lower accuracies than other splits.
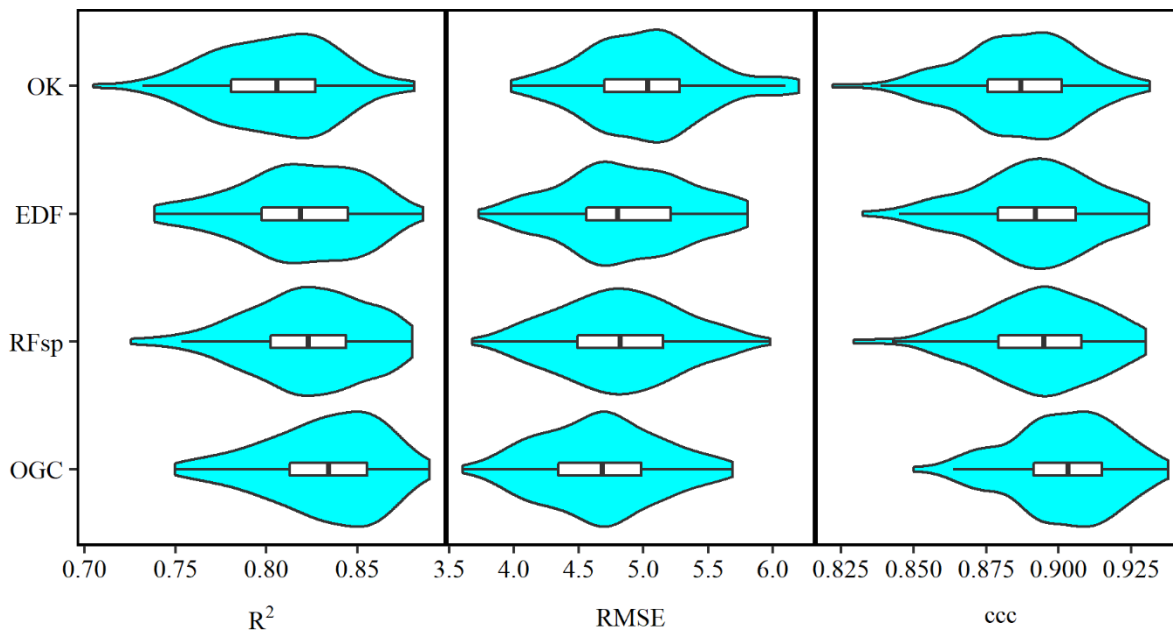
7 Author contribution

Anders Bjørn Møller and Nastaran Pouladi prepared the data. Anders Bjørn Møller carried out the analyses and prepared the manuscript with inputs from all co-authors.

8 Competing interests

The authors declare that they have no conflict of interest

9 Acknowledgements

References

Behrens, T., Schmidt, K., Viscarra Rossel, R., Gries, P., Scholten, T. and MacMillan, R. Spatial modelling with Euclidean distance fields and machine learning. Eur. J. Soil Sci. 69(5), 757-770. http://dx.doi.org/10.1111/ejss.12687, 2018.

De Alba, S. Simulating long-term soil redistribution generated by different patterns of mouldboard ploughing in landscapes of complex topography. Soil Tillage Res. 71(1), 71-86. http://dx.doi.org/10.1016/s0167-1987(03)00042-4, 2003.

Dubois, G., Malczewski, J. and De Cort, M. Mapping radioactivity in the environment: Spatial interpolation comparison 97. Office for Official Publications of the European Communities, 2003.

635   Esri. World Imagery. Scale not given. September 27, 2016. https://www.arcgis.com/home/item.html?id=10df2279f9684e4a9f6a7f08febac2a9 (accessed 19-06-2019), 2019.

Geurts, P., Ernst, D. and Wehenkel, L. Extremely randomized trees. Mach. Learn. 63(1), 3-42. http://dx.doi.org/10.1007/s10994-006-6226-1, 2006.

Greve, M.H., Greve, M.B., Kheir, R.B., Bøcher, P.K., Larsen, R. and McCloy, K. Comparing decision tree modeling and

640   indicator kriging for mapping the extent of organic soils in Denmark, in: Boettinger, J.L., Howell, D.W., Moore, A.C., Hartemink, A.E. and Kienast-Brown, S. (Eds.), Digital soil mapping: Bridging research, environmental application, and operation. Springer Netherlands, Dordrecht, 267-280. http://dx.doi.org/10.1007/978-90-481-8863-5_22, 2010.

Heckrath, G., Halekoh, U., Djurhuus, J. and Govers, G. The effect of tillage direction on soil redistribution by mouldboard ploughing on complex slopes. Soil Tillage Res. 88(1-2), 225-241. http://dx.doi.org/10.1016/j.still.2005.06.001, 2006.

645   Hengl, T., Heuvelink, G.B.M. and Stein, A. A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma 120(1-2), 75-93. http://dx.doi.org/10.1016/j.geoderma.2003.08.018, 2004.

Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B. and Gräler, B. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 6, e5518. http://dx.doi.org/10.7717/peerj.5518, 2018.

Hengl, T., Roudier, P., Beaudette, D., Pebesma, E. and Blaschek, M. Package 'plotKML': Visualization of spatial and

650   spatio-temporal objects in Google Earth. https://cran.r-project.org/web/packages/plotKML/plotKML.pdf (accessed 21-04-2020), 2020.

Hiemstra, P. Package 'automap': Automatic interpolation package. R package version 1.0-14. https://cran.r-project.org/web/packages/automap/index.html (accessed 15-08-19), 2013.

Keskin, H., Grunwald, S. and Harris, W.G. Digital mapping of soil carbon fractions with machine learning. Geoderma 339,

655   40-58. http://dx.doi.org/10.1016/j.geoderma.2018.12.037, 2019.

Knotters, M., Brus, D.J. and Oude Voshaar, J.H. A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. Geoderma 67(3-4), 227-246. http://dx.doi.org/10.1016/0016-7061(95)00011-c, 1995.

Kuhn, M. Building predictive models in R using the caret package. J. Stat. Softw. 28(5), 1-26.

660   http://dx.doi.org/10.18637/jss.v028.i05, 2008.

Lindstrom, M.J., Nelson, W.W., Schumacher, T.E. and Lemme, G.D. Soil movement by tillage as affected by slope. Soil Tillage Res. 17(3-4), 255-264. http://dx.doi.org/10.1016/0167-1987(90)90040-k, 1990.

Mitchell, T. Decision tree learning, in: Machine Learning. McGraw Hill, New York, 52-80. 1997.

National Survey and Cadastre. Danmarks Højdemodel 2007, DHM-2007/Terræn. National Survey and Cadastre, 2012.

665   Nussbaum, M., Spiess, K., Baltensweiler, A., Grob, U., Keller, A., Greiner, L., Schaepman, M.E. and Papritz, A. Evaluation of digital soil mapping approaches with large sets of environmental covariates. SOIL 4(1), 1-22. http://dx.doi.org/10.5194/soil-4-1-2018, 2018.

Odeh, I.O.A., McBratney, A.B. and Chittleborough, D.J. Further results on prediction of soil properties from terrain attributes: Heterotopic cokriging and regression-kriging. Geoderma 67(3-4), 215-226. http://dx.doi.org/10.1016/0016-

670   7061(95)00007-b, 1995.

Pebesma, E., Bivand, R., Rowlingson, B., Gomez-Rubio, V., Hijmans, R.J., Sumner, M., MacQueen, D., Lemon, J., O'Brien, J. and O'Rourke, J. Package 'sp': Classes and methods for spatial data. https://cran.r-project.org/web/packages/sp/sp.pdf (accessed 21-04-2020), 2020.

Pebesma, E. and Graeler, B. Package 'gstat': Spatial and spatio-temporal geostatistical modelling, prediction and simulation.

675   R package version 2.0-5. https://cran.r-project.org/web/packages/gstat/gstat.pdf (accessed 16-04-2020), 2020.

Pekel, J.F., Cottam, A., Gorelick, N. and Belward, A.S. High-resolution mapping of global surface water and its long-term changes. Nature 540(7633), 418-422. http://dx.doi.org/10.1038/nature20584, 2016.

Pouladi, N., Møller, A.B., Tabatabai, S. and Greve, M.H. Mapping soil organic matter contents at field level with Cubist, Random Forest and kriging. Geoderma 342, 85-92. http://dx.doi.org/10.1016/j.geoderma.2019.02.019, 2019.

680   Quinlan, J.R. Learning decision tree classifiers. ACM Comput. Surv. 28(1), 71-72. http://dx.doi.org/10.1145/234313.234346, 1996.

Ribeiro Jr, P.J., Diggle, P.J., Christensen, O., Schlather, M., Bivand, R. and Ripley, B. Package 'geoR': Analysis of geostatistical data. https://cran.r-project.org/web/packages/geoR/geoR.pdf (accessed 21-04-2020), 2020.

Rokach, L. and Maimon, O. Decision trees, in: Data Mining and Knowledge Discovery Handbook. Springer, 165-192. 2005.

685   Strobl, C., Malley, J. and Tutz, G. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychol. Method 14(4), 323-348. http://dx.doi.org/10.1037/a0016973, 2009.

Tan, P.-N., Steinbach, M. and Kumar, V. Classification: Basic concepts, decision trees, and model evaluation, in: Introduction to Data Mining. Pearson Education, Limited, 2014, 145-205. 2014.

690   Wang, P.R. Referenceværdier: Døgn-, måneds- og årsværdier for regioner og hele landet 2001 - 2010, Danmark for temperatur, relativ luftfugtighed, vindhastighed, globalstråling og nedbør. Teknisk Rapport 12-24. Danish Meteorological Institute, 2013.

Wright, M.N. and Ziegler, A. ranger: a fast implementation of random forests for high dimensional data in C++ and R. J. Stat. Softw. 77(1). http://dx.doi.org/10.18637/jss.v077.i01, 2015.

695   Zhang, X., Lin, F., Jiang, Y., Wang, K. and Wong, M.T. Assessing soil Cu content and anthropogenic influences using decision tree analysis. Environ. Pollut. 156(3), 1260-1267. http://dx.doi.org/10.1016/j.envpol.2008.03.009, 2008.

Replies to comments on our manuscript "Oblique geographic coordinates as covariates for digital soil mapping"

We here present our replies to the comments on our manuscript from the Editor, from the two anonymous referees and from Dr. Alaxandre Wadoux. In addition to our replies, we also outline the changes in the manuscript occasioned by each comment. We have previously presented most of these replies in the open discussion. However, we have changed some of the replies due to the revision requested by the editor.

1 Editor's comments

COMMENT 1

We have now studied the reports of 3 reviewers. All of them are very positive and agree that your approach is a novel and interesting addition to DSM. There are many smaller issues raised that will require your attention while revising your ms. I would like to highlight that the first 2 reviewers strongly suggest to demonstrate that your approach has 'global' merit and we would like to see this reflected in the revised ms. Thanks for considering SOIL and we are looking forward receiving your updated study!

K Van Oost

REPLY

We thank the editor for taking the time to consider our manuscript. We acknowledge the consensus that our manuscript should include additional datasets, and we consider this in the revised manuscript. Furthermore, we will address the smaller issues as described in our replies to the referees and to Dr. Alexandre Wadoux.

CHANGES

In addition to the Vindum dataset, the revised manuscript will include three other, publicly available datasets:

1. *meuse*: For this dataset, we map soil zinc contents using all methods applied to Vindum, with the addition of regression kriging.
2. *eberg*: For this dataset, we map soil types. As this is a classification problem, we omit kriging. Furthermore, due to the large number of observations (>2000), we omit RFsp.
3. *Swiss rainfall*: For this dataset, we map rainfall on 1986-05-08 in Switzerland. This is not a soil dataset, but we include it due to its high anisotropy. We use no auxiliary data. We therefore compare only purely spatial approaches (Kriging, EDF, RFsp and OGC).

In the revised manuscript, we mainly focus on the Vindum dataset in order to provide an in-depth demonstration of OGC. We present a detailed account of the methods and results from

the additional datasets in Appendix A. However, we include a short overview of the results from the additional datasets in the main results section in order to compare methods.

2 Comments from Referee #1

We thank Referee #1 for the well-thought and qualified comments on our manuscript. In the following, we will address the referee's comments and describe the changes that they have occasioned in the manuscript.

COMMENT 1

The present paper is well written and structured. Moreover, the study aims to make a contribution to the field of DSM by providing a novel methodological framework based on the usage of coordinates. This is something that can be considered as rather 'out of the box thinking', because most attention in the international literature goes either to the use of advanced geostatistical methods (e.g. capturing the spatial autocorrelation through kriging) or external drift fitting based on '(environmental) co-variates' or a combination of both. Hence, the work certainly merit respect for its originality and the methodological framework seems to provide useful thoughts to be considered in future DSM-studies. However, I also see some shortcomings which should be addressed/considered in order to maximize its potential to be applied widely, and as such, I am looking forward receiving the authors replies on the associated comments and suggestions presented below.

REPLY

We thank the referee for seeing the value in our research. We will address the shortcomings listed by the referee in the following.

COMMENT 2

Major Comments: I believe that the main issue with this research is that it considers only one rather small field characterized by 1 remarkable / specific spatial structure as regards the variation of SOM (i.e. one spot/area with clearly higher values) in order to test the validity of the present new methodology, whereas the authors claim that the method will be highly useful for mapping soil properties in larger areas. Hence, I believe that the present methodology requires further testing by considering larger areas (e.g. catchment-regional scale) with more complex spatial patterns in SOM in order to prove the validity of the statements that have been made in this respect. Moreover, it would also be interesting to consider other key soil variables (besides SOM) to check whether the usages of oblique geographic coordinates as covariates could be seen as a universal DSM approach. In this context, I believe that using a national soil inventory database could be a good way forward. I may understand that this might not be possible in this study, but I still believe that this should be mentioned clearly (as a critical note) in the discussions (and maybe be picked up by the authors in future research).

REPLY

We agree with the referee that it is a shortcoming that we only tested the method on one dataset.

CHANGES

As stated in the reply to the editor, we have included three additional datasets in the revised manuscript.

COMMENT 3

When I have a look at the performance of the different mapping methods (as presented in the Violin plots in figure 7), it seems to me that your new OGC (+AUX) method only results in (very) small improvements as compared to some other (more commonly used) methods such as Kriging. Hence, I was wondering whether this improvement is statistically significant? And if this might still be the case when either (i) another field (characterized by a different spatial pattern), (ii) another soil variable or (iii) larger geographical extent are considered?

REPLY

Some of the differences in accuracy are statistically significant and some are not. We used the same 100 repeated training/test splits for all methods, and this allowed us to carry out pairwise t-tests between the accuracies of the methods. We then ranked the methods using the results of these t-tests. Methods that did not have statistically significant differences for a given metric received the same rank for that metric, but methods with statistically significant accuracies received different ranks. For example, for the Vindum dataset, OGC + AUX and RFsp + AUX always received the same rank, as there were no statistically significant differences in the accuracies. Meanwhile, kriging always received a higher rank than AUX, because the differences in their accuracies were statistically significant.

For Vindum, kriging, RFsp + AUX and OGC + AUX all received the highest rank for two out of three accuracy metrics. We therefore regard these three methods as equally accurate. We already state in the manuscript that we regard these three methods as most accurate. However, we see that we have not explicitly stated that we regard them as equally accurate.

In the revised manuscript, we have carried out pairwise t-tests for the three additional datasets. For meuse, the tests show that EDF + AUX and OGC + AUX share the highest rank. For eberg, OGC + AUX has the highest rank, as its accuracy is statistically significantly higher than the accuracies of the other methods. Lastly, for the Swiss rainfall dataset, OGC has the highest rank, with statistically significant differences to the other methods.

CHANGES

We have rephrased L210 to state explicitly that we regard the three methods as equally accurate:

"We therefore regard these three methods as the most accurate methods for the Vindum dataset. Furthermore, we regard these three methods as equally accurate for this dataset, as none of them was universally more accurate than the other two methods."

We will also include the results of the t-tests for the additional datasets in the revised manuscript.

COMMENT 4

Minor Comments: I'm not too sure if it is entirely appropriate to use R2 as a measure to compare the different methods, because (i) a very high R2 value may also mean an 'overfit' and (ii) each method has it own degree of (model) complexity. Hence, I guess that it could be a good idea to take (also) another statistical measure into consideration that specifically aims to evaluate the methods' performance taking into account its complexity (in order to avoid overfitting)?

REPLY

We understand the referee's concern, as a very high R^2 on a training dataset can indicate overfitting of a model. However, we report R^2 for 25% holdout datasets not used in the models. Our R^2 values therefore indicate the predictive capabilities of the models rather than their fit on the training data. Furthermore, we are not aware of any measures of accuracy that account for complexity in Random Forest models. We are even less aware of any accuracy measures capable of comparing complexities of conceptually very different models, such as Random Forest and kriging. We think most readers will be aware that kriging is much simpler method than Random Forest. In fact, we explicitly stated this in the first version of the manuscript (L295 – L296).

COMMENT 5

Figure 1 - Subpanel C: Showing hill shade is not enough to give the reader an insight into the topographical configuration of the field. Hence, I suggest adding contour lines.

REPLY

We thank the referee for this helpful comment. We agree that adding contour lines improves the visualization of the topography of the study area.

CHANGES

Due to the referee's comment, we have prepared a new version of Figure 1, where we have added 2 m contour lines. We will include this updated figure in the final version of the manuscript:
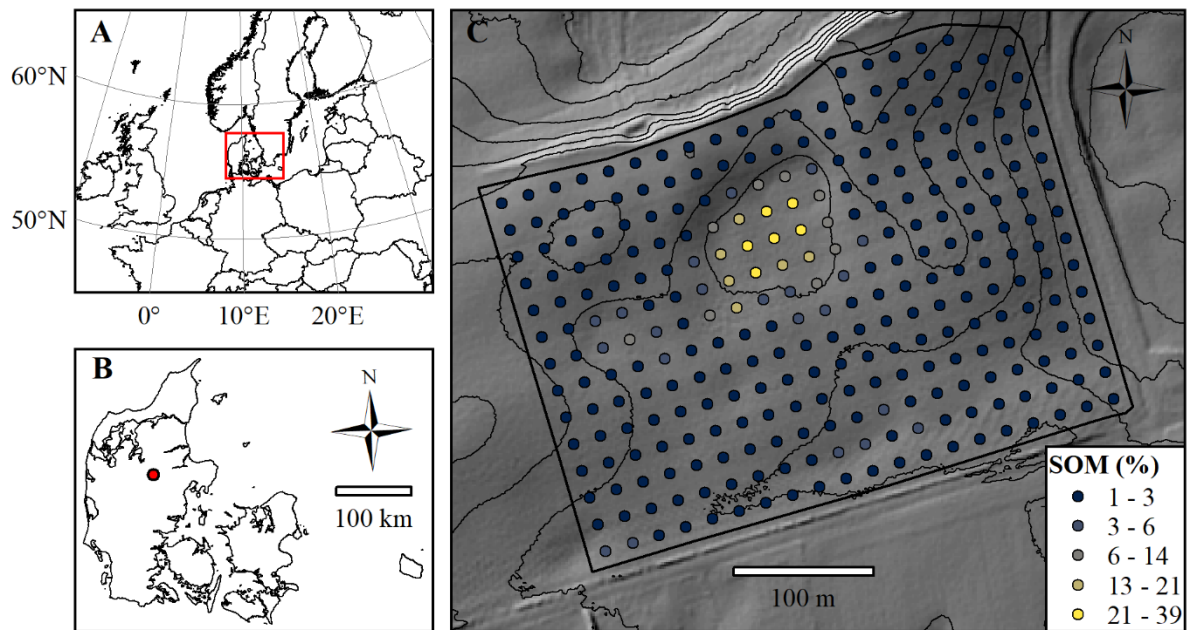
3 Comments from Referee #2

We thank the referee for the qualified and insightful comments on our manuscript. In the following, we will address the referee's comments and describe the changes that we have made to the manuscript because of the comments.

COMMENT 1

The manuscript "Oblique geographic coordinates as covariates for digital soil mapping" from Møller et al. presents a valuable contribution to integrate predictor information on spatial position into machine learning approaches for digital soil mapping. It, thereby, seeks to overcome the known problem of orthogonal artefacts sometimes introduced by the usage of xy-coordinates as covariates in recursive partitioning algorithms. While commonly applied covariates usually relate to site characteristics that approximate the soil forming factors, the inclusion of coordinates provides a chance to reflect further spatial patterns we are not necessarily aware of. The authors show that the usage of a multitude of oblique spatial coordinates reflects spatial anisotropy. Major spatial axes identified through predictor importance measures may then give a hint on the geographic direction of the underlying processes as the authors demonstrate. The article compares the new approach (OGC) to existing approaches such as Euclidean distance fields (EDF) and spatial Random Forest (RFsp). The article is written using adequate language and it follows a clear structure. The figures are well prepared. Furthermore, it is a rare, but highly welcome choice of the authors to provide the R code of their approach. While the authors very clearly demonstrate the power of their approach particularly due to the clear figures and the comparison to similar approaches, certain aspects would require reconsideration:

REPLY

We thank the referee for the support for our manuscript. Furthermore, we are happy that the referee appreciates our choice to share the code for our study. We will consider the issues that the referee raises in the following.

COMMENT 2

- I do not understand why the OGC+AUX approach is not directly compared to regression kriging, but to ordinary kriging. Ordinary kriging would require a stationary mean which is not given in this particular research setting. Accordingly, a regression model would first have to be fitted to model the trend from covariate data, while then spatial autocorrelation in the residuals will be accounted for by ordinary kriging of the residuals. While the regression model is fitted by random forest, this would also allow for direct comparability. The authors provide rather vague arguments against regression kriging (lines 27-32).

REPLY

Our study focuses on one-step methods, as one of the goals in developing OGC is to create a feasible one-step method. Two-step approaches such as regression-kriging require researchers to interpret two models at once, which can confound analyses of uncertainty and the processes that govern the spatial distribution of soil properties. We believe that this is a relevant consideration, but it is not our main reason for omitting regression-kriging. Our first reason for this choice is that a previous study carried out in the same area showed that kriging predicted SOM more accurately than regression-kriging using both Cubist and Random Forest models (Pouladi et al., 2019). When a relatively simple method outperforms complex approaches, we believe that it is right to consider the complex approaches as redundant. Without this previous finding, we believe that it would have been relevant to include regression-kriging in the comparison.

CHANGES

We see that the manuscript does not clearly state our reasons for omitting regression-kriging. We will therefore add the following paragraph to section 2.3:

"A previous study using the same dataset showed that kriging predicted SOM more accurately than regression-kriging (Pouladi et al., 2019). We therefore omitted regression-kriging from the analysis, although, without this previous finding, it would have been relevant to include it."

COMMENT 3

The data in this study display spatial autocorrelation. Specifically, a range of 139 m is mentioned. This is not surprising due to the high spatial data density. Furthermore, the authors mention a couple of processes that may have caused this spatial dependency. However, this aspect is not accounted for in the evaluation approach. 100 random splits 75/25 (training/ test set) make it very likely that spatially autocorrelated sampling points will end up in the test and training set for the majority of the 100 splits. As a consequence, the test sets are not independent of the training sets and will lead to overly optimistic error values. This aspect at least needs to be mentioned. Particularly, in the context of Figure 7. The argumentation line of the introduction requires some improvement.

REPLY

Our main priority in the study is to compare the accuracies of several methods, not to assess their accuracies in absolute terms. Furthermore, we do not consider the issue of spatial

autocorrelation to be as grievous as to warrant attention in the manuscript. Firstly, geostatistical approaches such as kriging would be useless if there was no spatial autocorrelation in the data. Secondly, the sample distribution in the field is very even, and as a result, only very few areas in the field are more than 20 meters from the nearest sample, and all areas are within the range of spatial autocorrelation. Therefore, having training and test samples within the range of spatial autocorrelation actually represents the general conditions in the field quite well. We therefore do not believe that our accuracy metrics are very much overly optimistic. If we were to extrapolate our results to a larger area, spatial autocorrelation would be an issue to consider, but this is not the goal of our study.

COMMENT 4
Certain aspects need to be better clarified:

- The main advantage of OGC+AUX over using only XY+AUX is the high number of coordinates, as the usage of only two oblique coordinates would lead to similar artefacts as demonstrated in the results. - The usage of coordinates as predictors in a regression model differs from fitting a geostatistical model to the residuals of a regression model. The approach closest to fitting a semivariogram is RFsp, as it accounts for the distance between points. However, it comes at the cost of introducing a high number of covariates as the authors state, correctly. It is important that the authors also compare their approach to RFsp, but the difference in calculating a different set of coordinates and taking the distance between points into account should be explicitly mentioned. In contrast, OGC+AUX and EDF+AUX really follow a similar approach in calculating a set of different coordinates. OGC+AUX is demonstrated to be superior to EDF+AUX. - Overall, whether it is worse to make the effort of calculating a high number of oblique coordinates could only be decided while being compared to regression kriging.

REPLY
We agree with the referee, and we see the need for further clarification. We will add several statements to the final version of the manuscript for this purpose.

CHANGES
We will add the following statements:

L44: "One of the main advantages of this approach [RFsp] is that it incorporates distances between observations in a similar manner to geostatistical models".

L301: "Of the previous approaches, OGC is most similar to EDF, as it used the x- and y-coordinates, and the distances to the corners of the study area resemble coordinates. On the other hand, RFsp is more similar to geostatistical models, as it relies on distances between observations. However, this similarity comes at the cost of calculating a large number of distance rasters."

L319: "The method eliminated the orthogonal artefacts that arise from use of x- and y-coordinates and also achieved higher accuracies than maps created with only two coordinate

rasters. However, for meuse, without auxiliary data, OGC still produced angular artefacts in the final map."

COMMENT 5
There are a couple of statements that are problematic. Please consider rephrasing:

- lines 19-21 "…decision tree algorithms…are immune to correlated and redundant covariates". There are a couple of publications that show the contrary.

REPLY
Our experience has shown that decision trees are less vulnerable to correlated and redundant covariates than other model types, such as artificial neural networks. However, we admit that this does not constitute a full immunity.

CHANGES
We see that our statement is not correct, and we will therefore remove it from the final version of the manuscript.

COMMENT 6
- line 29 "By kriging the residuals…soil mappers have been able to reduce or remove spatial bias". We usually fit a geostatistical model to explain spatial autocorrelation not to remove spatial bias. Please also correct throughout the manuscript, e.g. lines 45/46.

REPLY
We agree with the referee that our phrasing is incorrect, and we will therefore change it (see below). However, the phrasing in lines 45 – 46 is in line with the study to which we refer. We quote the authors: "Further analysis shows that in both cases there is no remaining spatial autocorrelation in the residuals […]. Hence, both methods have fully accounted for the spatial structure in the data" (Hengl et al., 2018). The authors of this study refer to a figure, which shows a pure nugget variogram for the residuals of their model.

CHANGES
We will rephrase the sentence in question:

"By kriging the residuals of the predictive model and adding the kriged residuals to the prediction surface, soil mappers have been able to explain spatial autocorrelation and achieve higher accuracies."

COMMENT 7
- line 47 "...methods are able to integrate spatial relationships…" I am not convinced that by the mere consideration of coordinates we account for spatial relationships, leave alone spatial autocorrelation. Please explain or rephrase.

REPLY
We agree that our use of the term "spatial relationships" is inaccurate.

CHANGES

We will replace the term "spatial relationships" with the term "spatial trends" throughout the manuscript.

COMMENT 8

- lines 51-56 "Another shortcoming relating to EDF and RFsp is that…" As EDF and RFsp did not intend to keep the number of coordinate covariates variable I would suggest "reduced flexibility" instead of "shortcoming".

REPLY

We agree with the referee, and we will rephrase as requested.

CHANGES

We will rephrase L51:

"EDF and RFsp also have limited flexibility as both methods specify the number of geographic data layers a priori."

COMMENT 9

- line 65-66. "…it should be possible to optimise it" Please be specific: is it possible or not? Does it make sense to optimise it? Why did the authors then merely test all numbers of coordinate covariates?

REPLY

We see that the sentence is not very clear. We will therefore rephrase it.

CHANGES

We will rephrase lines 65 – 66:

"Furthermore, the number of oblique angles is adjustable, and soil mappers can therefore choose a number that suits their purpose. Some mapping tasks may require a higher number of oblique angles than others, and soil mappers can therefore increase the number as necessary. Alternatively, if a small number of oblique angles suffices, soil mappers can reduce their number and thereby shorten computation times."

COMMENT 10

Further comments:

- Please delete equations (1) – (3). This is simple trigonometry.

REPLY

We agree with the referee.

CHANGES

We will delete equations 1 – 3.

Please also consider adapting the symbology: b2 is the knew oblique coordinate that replaces b1 (=x) and a1 (=y) not only b1 as somehow suggested by naming it b2.

REPLY
Our reason for naming b_2 is that it forms one of the sides of the right triangle a_2b_2c. We will therefore not rename it, as it would obscure interpretation of Figure 2. However, we see that the equations and the figure do not sufficiently stress the fact that the length of b_2 is equal to the new oblique coordinate.

CHANGES
We will add "OGC" to equation 4, to stress that OGC is equal to the length of b_2:

$$OGC = b_2 = \sqrt{{a_1}^2 + {b_1}^2} * \cos\left(\theta - \tan^{-1}\frac{a_1}{b_1}\right)$$

COMMENT 12
- lines 135-136. Please add the tested mtry values

REPLY
In each model, we tested five *mtry* values at even intervals between 2 and *NC*, where *NC* is the total number of covariates (counting both auxiliary data and spatially explicit covariates). The tested *mtry* values therefore depended on the method, and the number of covariates differed between methods.

CHANGES
We will add this explanation to the paragraph, starting at line 137:

"We tested *mtry* values at even intervals between 2 and the total number of covariates, including auxiliary data and spatially explicit covariates. The tested *mtry* values therefore varied depending on the number of covariates."

COMMENT 13
- line 136. Please explain how extratrees allows for suboptimal splits

REPLY
We will rephrase the sentence to better clarify how *extratrees* works.

CHANGES
We will rephrase the sentence as follows:

"The *extratrees* splitting rule generates random splits, as opposed to the *variance* splitting rule, which chooses optimal splits. Per default, *extratrees* generates one random split for each covariate and then chooses the random split that gives the largest variance reduction (Geurts et al., 2006). It therefore leads to a greater degree of randomization."

- Does the approach work on any type of coordinate system? I suppose coordinates have to be projected?

REPLY
This is a very interesting question, which we have given some though, although we have not included these thoughts in the first version of the manuscript. In the study, we use UTM coordinates, which have the advantage that the x- and y-coordinates have the same unit. Furthermore, it is reasonable to treat relatively small study areas as two-dimensional planes. In practical terms, OGC may also work reasonably well for larger areas with other coordinate systems, such as latitude/longitude systems. However, interpretation would not be as straightforward as in this study.

Using OGC at a global extent would probably require changes to the method. Because longitude is circular, points located on different sides of 180° L would have drastically different coordinates, even if the actual distances between them were short. One solution to this problem could be to replace the present version of OGC with latitudes rotated at various angles around a pair of equatorial axes. However, the implementation and testing of such an approach is far outside the scope of this study.

Due to the interest of this question, we will shortly address it in the conclusions section of the revised manuscript.

CHANGES
We will add the following statement to the conclusions section:

"One should note that we carried out this study for relatively small areas using "flat" coordinate systems. Using OGC for larger areas and other coordinate systems may require alterations to the method."

4 Comments from Dr. Alexandre Wadoux

General reply

We thank Dr. Alexandre Wadoux for the insightful comments on our manuscript "Oblique geographic coordinates as covariates for digital soil mapping" (Møller et al., 2019, Wadoux, 2019). We have found the comments very helpful in improving the manuscript, and we would like to give our replies to the comments.

We will start with a general reply to the commenter's use of "pseudocovariates" as a label for oblique geographic coordinates. We see this label as misplaced. We believe the term "pseudocovariates" is only appropriate for covariates, which are clearly unsuited for the purpose, and this is not the case for oblique geographic coordinates.

Notable examples of pseudocovariates in the statistical literature have included randomly generated covariates for testing variable selection (Wu et al., 2007, Sandri and Zuccolotto, 2008, Sandri and Zuccolotto, 2009, Ghosal et al., 2019). In the mapping literature, recent studies have used pictures projected in geographic space as cautionary tales (Fourcade et al.,

2018, Wadoux et al., 2019). The commenter correctly asserts that pseudocovariates with a spatial pattern can predict properties in geographic space with moderate success. However, we do not believe this to mean that researchers should disregard covariates that explicitly account for spatial position.

In fact, the digital soil mapping literature has a rich number of studies, which have included spatial position as a covariate. The *scorpan* approach to digital soil mapping presented by McBratney et al. (2003) explicitly includes spatial position as a component. Although most studies in the review include spatial position through kriging or regression-kriging, the authors are open to the use of covariates to account for spatial position. We quote:

"As was discussed in Section 2, soil can be predicted from spatial coordinates alone. […] This may indeed reflect some other environmental variable such as climate, and because of this it can be argued that n is not really a factor, but simply putting the coordinates is a simple way to ensure that spatial trends not included in the other environmental variables are not missed. Therefore, n could also be described by some linear or nonlinear (nonaffine) transformation of the original spatial coordinates," (McBratney et al., 2003).

Oblique geographic coordinates represent such a transformation of the spatial coordinates. As one may expect from the previous reference, several studies have included x- and y-coordinates as covariates (Poggio and Gimona, 2014, Nussbaum et al., 2018, Koch et al., 2019, Lagacherie et al., 2019). Other studies have included spatial position in the form of distance-based covariates, for example using distances to the coastline (Holmes et al., 2015) or rivers (Rudiyanto et al., 2018).

Recently, studies have included additional distance-based covariates, including distances to the corners and middle of the study area (Behrens et al., 2018), and distances to observations (Hengl et al., 2018). We hope therefore to have demonstrated that the use of covariates to account for spatial position is a theoretically sound, well-established practice, which does not warrant the label "pseudocovariates". Using covariates to include spatial position in machine learning models is in itself not new. Oblique geographic coordinates are simply a new method for doing this, with some advantages over previous methods.

In addition to this general reply, we would like to address the specific comments in the following.

Specific replies
We structure our replies by first showing the comment in question, then our reply to the comment.

COMMENT 1
This study tries to account for residuals spatial autocorrelation of a machine learning model by adding a set of pseudo-covariates. I have a few comments on the paper. I hope the authors find them useful and that it helps them to improve their manuscript. Overall, the study would benefit from a test of the method on several case studies, using different scales, different

calibration sampling designs. A single case study at local scale and predicting a single soil property is in my opinion not enough to draw general conclusions.

REPLY
We agree that it would be beneficial to test the method on additional datasets.

CHANGES
In the revised manuscript, we include three additional datasets. For details, see our reply to the editor's comment and the revised manuscript.

COMMENT 2
About the methodology:1) Any set of covariates with spatial pattern added to the original set of covariates may result in higher accuracy with a ML algorithm. This is because ML algorithms can find relevant patterns even when the covariates are meaningless and not related to any soil forming process. The increase of accuracy that the authors obtain with the RF OGC + AUX model may well be obtained by adding any set of covariates with a spatial structure (see Fourcade et al., 2018).

REPLY
We concur that it would probably be possible to obtain the accuracies obtained with OGC + AUX with other (but not just any) sets of covariates. For example, RFsp + AUX achieves similar accuracies, although with a larger number of covariates. However, we will also remind the commenter that OGC do not simply have spatial structure – they have only spatial structure and nothing more. As we have already stated in our general reply, using covariates to account for spatial position is a well-established practice. OGC account for spatial position in a clear and systematic way, which is useful for decision tree algorithms and easily yields to interpretation.

COMMENT 3
2) Spatial autocorrelation in the raw data is not a problem per se and one should rather focus on remaining spatial autocorrelation on the residuals. I am strongly in favor of using only pedologically relevant covariates in a RF model. If the residuals of a model built using pedologically relevant covariates present autocorrelation, then one should consider making a map of the residuals because he may see a clear pattern of why this happens. The authors might then see that they are missing an important spatial process not included in the analysis. In this case one can add additional pedologically relevant covariates that could explain this pattern, and refit the model.

REPLY
We agree that it is important to use pedologically relevant covariates in machine learning models when mapping soil properties. We do not intend OGC to be used on their own, but in combination with auxiliary data of this form. As we hope to have demonstrated in our general reply, several studies have used spatially explicit covariates in combination with the other six components of the scorpan concept for digital soil mapping. Other studies have accounted for spatial autocorrelation in the residuals by means of regression-kriging, another well-established practice.

The commenter's dedication to purely pedologically relevant covariates has merit. However, due to the complexity of soil-forming processes, the hunt for a set of covariates that perfectly explain spatial variation in soil properties, is in many cases likely to be fruitless.

COMMENT 4
3) In case one made the previous step and admits that there is unexplained residual variation, one could consider using additional pseudo-covariates because there is no better proxy to explain the soil spatial variation. I stress here that these pseudocovariates should not correlate with the pedological covariates because there would be redundancy (see next comment). In this case the pseudo-covariates should be covariates computed based on the remaining residuals. This would effectively tackle the problem of the residual autocorrelation and the authors would ensure that the pseudocovariates do not interfere with the pedologically relevant covariates.

REPLY
Redundancy is generally not a risk for decision tree models, as they simply choose the optimal covariate in each split (Breiman, 2001). See also our reply to the next comment.

Furthermore, we doubt if the approach, which the commenter suggests, would be useful. We are not sure how the commenter would create a covariate based on the residuals. However, the attempt would create a serious risk of circular logic, which could invalidate model fitting and the assessment of model accuracy. Models should be based on covariates, not vice versa.

COMMENT 5
4) In this study, the authors include the set of pseudo-covariates with the set of pedologically relevant covariates. This is in my opinion very harmful because they can have pseudo-covariates which integrate over several of the pedologically relevant covariates, making them in some cases even better predictors. This is unrealistic and undesirable. This also makes the model less interpretable in terms of variable importance.

REPLY
Firstly, we refer to our general reply. Secondly, we will state that we see the commenter's allegation of "harmfulness" as a misunderstanding. We see the integration of spatial and environmental covariates as one of the strengths of using oblique geographic coordinates. Firstly, it allows the machine learnings model to map complex processes characterized by spatial dependence as well as environmental effects (Behrens et al., 2018). This has an advantage over regression-kriging, where separate, mostly incomparable models treat environmental and spatial effects.

The commenter fears a scenario where a coordinate raster gains a higher importance than environmental covariates in a model. If this were the case, it would indeed be a cause of worry, but not for the reasons stated by the commenter. If a coordinate raster gains a higher importance than an environmental covariate, it suggests that the pedological process represented by the environmental covariate is probably not highly relevant for the soil property in this specific area. Therefore, if all environmental covariates turn out to be less

important than coordinate rasters, it would show that the environmental covariates did not adequately account for spatial variation in the soil property.

In our case, for the Vindum dataset, the most important coordinate raster was the 12th most important covariate. OGC only became the second most important covariate, when we summed their importance. This shows that spatial effects have a large influence on SOM in the study area. However, it also shows that the model did not discard environmental covariates when we included OGC. Instead, it successfully integrated the two sets of covariates and their combined effects.

COMMENT 6
5) It is concluded that adding a set of pseudo-covariates effectively accounts for spatial autocorrelation in the data. This is clearly not the case as shown in Fig. 9 and admitted by the authors at line 315 'the models built exclusively on spatial relationships had the most autocorrelated residuals.' The reason for this is that the covariates have a spatial pattern but are not related to the raw data and either to the residuals of the prediction made by a RF model. When the authors compared the sample variograms of kriging and RF residuals, it is visible that kriging do much better. The method would work if the sample variogram of RF OGC would be close to that of kriging. We can also see in Fig. 9 that the model with OGC covariates only have strong residual autocorrelation. The reduction in terms of residual autocorrelation in the OGC + AUX model is obtained by adding the pedologically relevant covariates. This is also a contradiction with the conclusion that OGC covariates account for the spatial autocorrelation.

REPLY
We never claim in the manuscript that oblique geographic coordinates fully account for spatial autocorrelation in the data. This comment would be more helpful if the commenter provided the lines where we allegedly state this.

We once refer to Hengl et al. (2018), who found that RFsp fully accounted for spatial autocorrelation in the data, but it is quite clear from the sentence that we refer to results in another study, not our own results. Our own results contrast with this earlier finding, and we will include a comment on this in the final paper.

Furthermore, the commenter appears to reverse the interpretation of Figure 9. We intend soil mappers to use OGC as an addition to environmental covariates, not on their own. The figure shows that the addition of OGC greatly reduces spatial autocorrelation in the residuals relative to the model relying only on environmental covariates. We mainly include OGC, EDF and RFsp on their own to demonstrate more clearly the effects of these sets of covariates. We do not recommend that researchers use them on their own.

CHANGES
In the revised manuscript, we clearly state that our results contrast with earlier findings:

"This finding contrasts with Hengl et al. (2018) who found that there was no spatial trend in the residuals of predictions with RFsp."

COMMENT 7

6) Fig. 9 shows that there is still autocorrelation in the residuals of the RF model. This violates the assumption made in RF modelling, i.e. independence between the data points. Since this assumption is not satisfied, the calibrated RF model is potentially flawed. The authors have potentially missed important soil processes which could be added to the model as covariates. I would be interested to see a measure of the bias in the prediction.

REPLY

We believe that it is quite an overstatement to say that any Random Forest model with spatially autocorrelated residuals is potentially "flawed". Such a conclusion would most likely invalidate a very large portion of Random Forest models used in digital soil mapping. However, we agree that it is not an optimal situation, and that it might be useful to add more environmental covariates.

As per the commenter's request, we have calculated bias as mean error (ME) for each method. We have based this calculation on residuals from models using all observations:

Method	ME
Kriging	-0.011
AUX	0.040
EDF	0.041
EDF + AUX	0.042
RFsp	0.011
RFsp + AUX	0.028
OGC	0.029
OGC + AUX	0.036

The values show that kriging has lower bias than the other methods except RFsp, but all methods have low bias.

COMMENT 8

Other considerations: Nugget to sill ratio should not be used to compare sample variograms, see Section 3.3. in https://doi.org/10.1016/j.catena.2013.09.006

REPLY

We see the error. In the final paper, we will remove mentions of the nugget-to-sill ratio when comparing the variograms.

CHANGES

We have removed mention of the nugget-to-sill ratio from the final version of the manuscript.

COMMENT 9

Very surprised to read at Line 297 that the advantage of ML algorithms is their interpretability.

I think the authors refer to the variable importance of the RF algorithm for the interpretability of the ML models. There is in my opinion a misunderstanding of the difference between ML and geo-statistical methods such as kriging. In ML you do not do inference and so you should not directly interpret the fitted model, or at least with caution. In geostatistics you can interpret because you make inference on the process that generated the data.

ML are also mostly black boxes. For example, is it impossible to interpret all the trees in a RF model, or all the neurons in a neural network model. This is in consequence not justified to claim that ML algorithms have the advantage to be interpretable.

REPLY
This comment is confusing. The commenter appears to assert that (1) machine learning models are not interpretable, but that, on the other hand, (2) geostatistical models are interpretable. The commenter seems to conflate interpretation and inference, but we believe that one should understand these two as separate terms.

Furthermore, it gives the impression of a contradiction when the commenter states that machine learning models should not include spatial relationships, but also states that geostatistical models are interpretable. Likewise, the statement that machine learning models are not interpretable contrasts with the commenter's insistence that they should only contain covariates that represent pedological processes. If spatial position matters, even to the point where a geostatistical model is exclusively interpretable, why should we not use it in a model? Moreover, if we cannot interpret a machine learning model, then why does it matter what sort of covariates we use?

In themselves, geostatistical models only inform us on the spatial structure of the data. We agree that this can be useful, but any sort of interpretation would rely almost exclusively on the user's knowledge of the target variable and the processes that affect it. On the other hand, machine learning models are potentially far more informative.

Researchers should exert caution when interpreting any form of statistical model, but we agree with the commenter that it is especially relevant for machine learning models. Machine learning models are more complex than geostatistical models, and their interpretation is therefore also more complex and requires a higher level of abstraction. Tools to interpret machine learning models include covariate importance, which we use, but other tools exist, for example partial dependency plots (Friedman, 2001). Irrespective of the tools that researchers use, it is important that they critically use their knowledge of soils and the study area as well as the machine learning algorithm.

We can see that our statement that geostatistical models and machine learning models differ in interpretability is misleading. In the final paper, we will change the phrasing to state that the difference lies in the information content provided by the models.

CHANGES
We have rephrased part of the discussion to reflect the fact that the difference between geostatistical models and machine learning lies in the information that they provide:

"The advantage of the methods based on machine learning instead lies in the fact that they provide larger amounts of information than geostatistical models. Kriging in itself does not provide information on the processes that control spatial variation in soil properties, but machine learning models can include covariates related to soil processes, providing information on the processes that are most likely to affect the spatial distribution of a soil property."

COMMENT 10
L 305: I would disagree with this conclusion; this would need to be justified by the literature or comparison between different case studies.

REPLY
We cannot see why the commenter would outright disagree with this conclusion, as the commenter also states that spatial coverage sampling favors kriging. However, we do see the need for justification from the literature. Several studies have shown that machine learning models using environmental covariates are more accurate than geostatistical models for large, less densely sampled areas, including Zhang et al. (2008), Greve et al. (2010) and Keskin et al. (2019). We will include these references in the final paper.

CHANGES
We have rephrased part of the conclusion:

"Previous studies have shown that machine learning is likely to provide higher accuracies in large, sparsely sampled areas (Zhang et al., 2008, Greve et al., 2010, Keskin et al., 2019), and it will be relevant to test if this is also the case for oblique geographic coordinates."

COMMENT 11
L 313: It is quite high accuracy a minimum CCC = 0.83.

REPLY
We agree. In the revised manuscript we will rephrase this sentence.

CHANGES
In the revised manuscript, we have rephrased this sentence: "as EDF, RFsp and OGC all yielded lower accuracies without auxiliary data".

COMMENT 12
L. 315. The authors have contradictory statements in the last paragraph of the Discussion.

REPLY
We do not see the contradiction, but we agree that the sentences are not quite clear enough.

CHANGES
In the revised manuscript, we have rephrased the last two sentences of the discussion to improve clarity:

"The results therefore suggest that soil mappers should primarily use these methods in combination with auxiliary data, and not on their own. If no auxiliary data are available, kriging is likely to be a better option."

COMMENT 13
The last sentence is not very clear. Dealing with spatial data, which are auto correlated, a spatial methods is always needed otherwise you miss an important process and the fitted model is probably flawed because of the i.i.d assumption of the errors.

REPLY
We agree on the lack of clarity. Please see our reply to the previous comment.

COMMENT 14
How did the authors compute the R2? A R2 can either indicate the closeness of the predicted values to the fitted regression line or the proportion of variance explained by the predictors. Authors should check that the R-square was computed against the 1:1 line and not against the fitted linear regression between observed and predicted, see https://doi.org/10.5194/soil-4-1-2018, Section 3.8 where the authors called it a skill score.

REPLY
We used Pearson's R^2, this is, closeness to a fitted regression line. We see that we did not include this information in the manuscript, and we will make sure to include it in the final paper.

We will not change the way we calculate R^2, as Pearson's R^2 indicates if the predictions have the same trend as the observations, which we believe is relevant in itself. We rely on several accuracy metrics, including also RMSE and CCC. CCC gives information on closeness to a 1:1 line, which the commenter requests. Furthermore, the skill score, to which the commenter refers, uses on the mean square error (MSE) of the predictions, and the variance in the dataset. It is very useful for comparing accuracies across different regression problems. However, for any single regression problem, as in our study, the variance in the dataset will be constant, and variation in the skill score will depend only on variation in MSE. As we already provide RMSE, this information would be redundant.

CHANGES
In the revised manuscript, we clearly state that we use Pearson's R^2.

COMMENT 15
Impact of the sampling design is not considered. A spatial coverage design is very poor for random forest, while it is very efficient for kriging (assuming the variogram parameters are known). You should also consider that the sampling designs affect greatly the way the sample variograms are computed.

REPLY
We agree that the sampling design favors kriging. In fact, we already state in the manuscript that an earlier study in the same area (Pouladi et al., 2019) found that kriging yielded higher

accuracies than machine learning models. It is therefore quite remarkable that OGC + AUX and RFsp + AUX allow Random Forest models to achieve accuracies on par with kriging.

COMMENT 16

How did the authors compute the sample variograms? The authors gave no information about it.

REPLY

Firstly, we produced maps with each method using all observations. Secondly, we converted both observations and predictions to natural logarithmic scale. We then subtracted the predictions from the observations and calculated variograms for these residuals. For this purpose, we used the function 'variogram' from the R package 'gstat' with its default parameters.

CHANGES

In the revised manuscript we include information on the method that we used for the sample variograms.

Furthermore, we have discovered an error in our code, which caused us to use only 75% of the observations when calculating the variograms. We have therefore recalculated the variograms using all observations and produced a new version of Figure 9. We will include this updated figure in the revised manuscript:
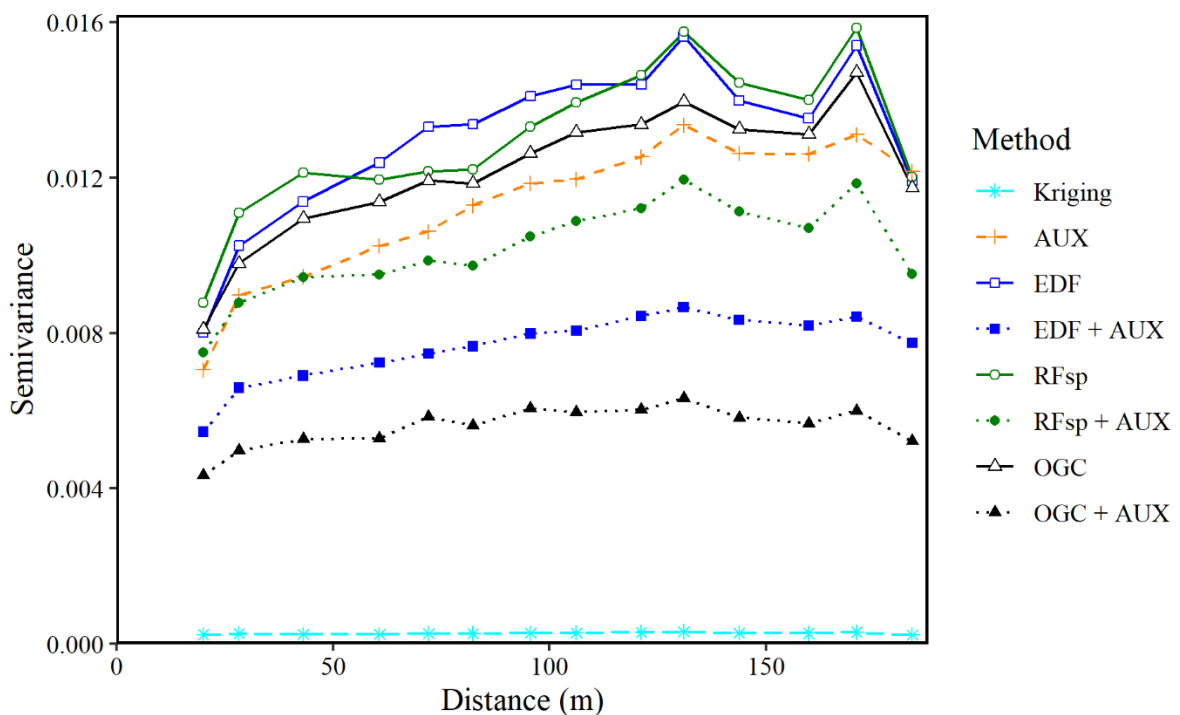
COMMENT 17

It seems that the sample variogram for ordinary kriging is not at the same scale. It is either a much better model or the authors did not back-transformed the log-transformed observations.

The authors mentioned that they log-transformed the observations prior to variogram fitting, it is not clear whether they also did it for the RF model.

REPLY
The variograms are all on the same scale. Kriging has smaller residuals than the other methods, as the variogram had a very small nugget, but we do not believe that this shows it to be a "better" model. For example, with inverse distance weighting interpolation, the residuals would be zero, but it would not necessarily by a very good model. We will also point out to the commenter that the residuals for OGC + AUX show nearly no trend. So the residuals are larger, but they have very little spatial dependence.

References

Behrens, T., Schmidt, K., Viscarra Rossel, R., Gries, P., Scholten, T. and MacMillan, R. Spatial modelling with Euclidean distance fields and machine learning. Eur. J. Soil Sci. 69(5), 757-770. http://dx.doi.org/10.1111/ejss.12687, 2018.

Breiman, L. Statistical modeling: The two cultures. Stat. Sci. 16(3), 199-215. http://dx.doi.org/10.1214/ss/1009213726, 2001.

Fourcade, Y., Besnard, A.G. and Secondi, J. Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. Glob. Ecol. Biogeogr. 27(2), 245-256. http://dx.doi.org/10.1111/geb.12684, 2018.

Friedman, J.H. Greedy function approximation: A gradient boosting machine. Ann. Stat. 29(5), 1189-1232. http://dx.doi.org/10.1214/aos/1013203451, 2001.

Geurts, P., Ernst, D. and Wehenkel, L. Extremely randomized trees. Mach. Learn. 63(1), 3-42. http://dx.doi.org/10.1007/s10994-006-6226-1, 2006.

Ghosal, R., Maity, A., Clark, T. and Longo, S.B. Variable Selection in Functional Linear Concurrent Regression. arXiv preprint arXiv:1904.08507. 2019.

Greve, M.H., Greve, M.B., Kheir, R.B., Bøcher, P.K., Larsen, R. and McCloy, K. Comparing Decision Tree Modeling and Indicator Kriging for Mapping the Extent of Organic Soils in Denmark, in: Boettinger, J.L., Howell, D.W., Moore, A.C., Hartemink, A.E. and Kienast-Brown, S. (Eds.), Digital Soil Mapping: Bridging Research, Environmental Application, and Operation. Springer Netherlands, Dordrecht, 267-280. http://dx.doi.org/10.1007/978-90-481-8863-5_22, 2010.

Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B. and Gräler, B. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 6, e5518. http://dx.doi.org/10.7717/peerj.5518, 2018.

Holmes, K.W., Griffin, E.A. and Odgers, N.P. Large-area spatial disaggregation of a mosaic of conventional soil maps: Evaluation over Western Australia. Soil Res. 53(8), 865. http://dx.doi.org/10.1071/sr14270, 2015.

Keskin, H., Grunwald, S. and Harris, W.G. Digital mapping of soil carbon fractions with machine learning. Geoderma 339, 40-58. http://dx.doi.org/10.1016/j.geoderma.2018.12.037, 2019.

Koch, J., Berger, H., Henriksen, H.J. and Sonnenborg, T.O. Modelling of the shallow water table at high spatial resolution using random forests. Hydrol. Earth Syst. Sci. 23(11), 4603-4619. http://dx.doi.org/10.5194/hess-23-4603-2019, 2019.

Lagacherie, P., Arrouays, D., Bourennane, H., Gomez, C., Martin, M. and Saby, N.P.A. How far can the uncertainty on a Digital Soil Map be known?: A numerical experiment using pseudo values of clay content obtained from Vis-SWIR hyperspectral imagery. Geoderma 337, 1320-1328. http://dx.doi.org/10.1016/j.geoderma.2018.08.024, 2019.

McBratney, A.B., Mendonça Santos, M.L. and Minasny, B. On digital soil mapping. Geoderma 117(1-2), 3-52. http://dx.doi.org/10.1016/s0016-7061(03)00223-4, 2003.

Møller, A.B., Beucher, A.M., Pouladi, N. and Greve, M.H. Oblique geographic coordinates as covariates for digital soil mapping. SOIL Discuss. http://dx.doi.org/10.5194/soil-2019-83, 2019.

Nussbaum, M., Spiess, K., Baltensweiler, A., Grob, U., Keller, A., Greiner, L., Schaepman, M.E. and Papritz, A. Evaluation of digital soil mapping approaches with large sets of environmental covariates. SOIL 4(1), 1-22. http://dx.doi.org/10.5194/soil-4-1-2018, 2018.

Poggio, L. and Gimona, A. National scale 3D modelling of soil organic carbon stocks with uncertainty propagation - An example from Scotland. Geoderma 232-234, 284-299. http://dx.doi.org/10.1016/j.geoderma.2014.05.004, 2014.

Pouladi, N., Møller, A.B., Tabatabai, S. and Greve, M.H. Mapping soil organic matter contents at field level with Cubist, Random Forest and kriging. Geoderma 342, 85-92. http://dx.doi.org/10.1016/j.geoderma.2019.02.019, 2019.

Rudiyanto, Minasny, B., Setiawan, B.I., Saptomo, S.K. and McBratney, A.B. Open digital mapping as a cost-effective method for mapping peat thickness and assessing the carbon stock of tropical peatlands. Geoderma 313, 25-40. http://dx.doi.org/10.1016/j.geoderma.2017.10.018, 2018.

Sandri, M. and Zuccolotto, P. A bias correction algorithm for the Gini variable importance measure in classification trees. J. Comput. Graph. Stat. 17(3), 611-628. http://dx.doi.org/10.1198/106186008x344522, 2008.

Sandri, M. and Zuccolotto, P. Analysis and correction of bias in Total Decrease in Node Impurity measures for tree-based algorithms. Stat. Comput. 20(4), 393-407. http://dx.doi.org/10.1007/s11222-009-9132-0, 2009.

Wadoux, A. Interactive comment on "Oblique geographic coordinates as covariates for digital soil mapping" by Anders Bjørn Møller et al. SOIL Discuss. http://dx.doi.org/10.5194/soil-2019-83-SC1, 2019.

Wadoux, A.M.J.C., Samuel-Rosa, A., Poggio, L. and Mulder, V.L. A note on knowledge discovery and machine learning in digital soil mapping. Eur. J. Soil Sci. http://dx.doi.org/10.1111/ejss.12909, 2019.

Wu, Y.J., Boos, D.D. and Stefanski, L.A. Controlling variable selection by the addition of pseudovariables. J. Am. Stat. Assoc. 102(477), 235-243. http://dx.doi.org/10.1198/016214506000000843, 2007.

Zhang, X., Lin, F., Jiang, Y., Wang, K. and Wong, M.T. Assessing soil Cu content and anthropogenic influences using decision tree analysis. Environ. Pollut. 156(3), 1260-1267. http://dx.doi.org/10.1016/j.envpol.2008.03.009, 2008.