Geoderma 405 (2022) 115332

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

An integrated approach for the evaluation of quantitative soil maps through Taylor and solar diagrams

Alexandre M.J-C. Wadoux [a,*], Dennis J.J. Walvoort [b], Dick J. Brus [c]

[a] Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, Australia
[b] Soil, Water and Landuse group, Wageningen Environmental Research, The Netherlands
[c] Biometris, Wageningen University & Research, The Netherlands

ARTICLE INFO

Handling Editor: Kristin Piikki

Keywords:
Soil science
Pedometrics
Goodness-of-fit
Error
Digital soil mapping
Accuracy

ABSTRACT

For many decades, soil scientists have produced spatial estimates of soil properties using statistical and non-statistical mapping models. Commonly in soil mapping studies the map quality is assessed through pairwise comparison of observed and predicted values of a soil property, from which statistical indices summarizing the quality of the entire map are computed. Often these indices are based on average error and correlation statistics. In this study, we recommend a more appropriate and effective method of map evaluation by means of Taylor and solar diagrams. Taylor and solar diagrams are summary diagrams exploiting the relationship between statistical indices to visualize differentiable aspects of map quality into a single plot. An important advantage over current map quality evaluation is that map quality can be assessed from the combined effect of a few statistical quantities, not just on the basis of a single index or list of indices. We illustrate the use of common statistical indices and their combination into summary diagrams with a simulation study and two applications on soil data. In the simulation study nine maps with known statistical properties are produced and evaluated with tables and summary diagrams. In the first case study with soil data, change in the quality of a large-scale topsoil organic carbon map is tracked for a number of permutations in the mapping model parameters, whereas in the second case study several maps of topsoil organic carbon content for the same area, made by various statistical and non-statistical models, are compared and evaluated. We consider that in all cases better insights in map quality are obtained with summary diagrams, instead of using a single index or an extensive list of indices. This underpins the importance of using integrated summary graphics to communicate on quantitative map quality so as to avoid excessive trust that a single map quality index may suggest.

1. Introduction

A primary objective in digital soil mapping (DSM) is the production of soil maps by exploiting the quantitative (statistical) relationships between laboratory measured soil data and a set of environmental covariates. Recent examples of soil mapping studies depicting quantitative (continuous) soil properties are Nussbaum et al. (2018); Ramirez-Lopez et al. (2019) or Heuvelink et al. (2021). Soil maps are imperfect, so that there is always a deviation between the map with predicted values and the true spatial pattern of a soil property. The smaller this deviation, the higher the map quality.

Determining soil map quality has been an area of research for many years. Early studies include Webster and Beckett (1968), Burrough et al. (1971) and Marsman and Gruijter (1986). These studies aimed at assessing the quality of conventional (discrete or thematic) soil maps with accuracy statistics. Finke (2006) and Bishop et al. (2001) evaluated soil maps with quality indices of the user's versus producer's perspective or information theory, respectively. Lark (2000) proposed a quality index for evaluating the kriging variance that combines the prediction error and the prediction error variance, more specifically to compute the sample average of the squared prediction error divided by the associated kriging variance. Also Malone et al. (2011) proposed to evaluate a soil map on the basis of both the predictions and their uncertainties. Piikki et al. (2021) provided a list of existing statistical indices to be used for the evaluation of quantitative soil maps. Brus et al. (2011) stressed that it is important to define map quality indices at the level of the entire map, as population means, not just as sample averages, and advocate estimation of the map quality indices through probability sampling.

* Corresponding author at: Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, New South Wales, Australia
E-mail address: alexandre.wadoux@sydney.edu.au (A.M.J-C. Wadoux).

https://doi.org/10.1016/j.geoderma.2021.115332
Received 4 May 2021; Received in revised form 28 June 2021; Accepted 29 June 2021
Available online 20 September 2021

In DSM studies, map quality is usually assessed by global statistical indices of the error computed by pairwise comparison of soil map prediction and observations. These indices are global in the sense that they summarize the quality of the entire map into a single statistical index, i.e. they quantify the overall accuracy of all locations in the mapping area taken together. Often these map quality indices are based on correlation statistics (e.g. the Pearson's correlation coefficient or concordance correlation coefficient) or average error indices (e.g. the mean error or (root) mean square error). While valid and useful for the assessment of specific aspects of map quality, a number of tools developed outside soil science are available to allow one to evaluate simultaneously multiple statistical indices with summary diagrams. The aim of this paper is to introduce these tools, which provide a more integrated approach to the evaluation of quantitative soil maps, to the soil science community.

This paper is structured as follows. First, we evaluate a set of common statistical indices for estimating continuous soil map quality. Each index shows a specific aspect of map quality. Second, we suggest that multiple aspects of map quality can be evaluated simultaneously, by visualization of the Taylor and the solar diagram. The statistical description of the indices and the diagrams are illustrated with a simulation study. Finally, the use of map quality indices and their interpretation with the Taylor and solar diagrams are illustrated with two case studies on evaluation of a digital soil map of organic carbon in the Eastern Amazon rainforest.

2. Evaluation of soil maps

Consider the prediction error as the resultant of the observed minus predicted values of a soil property at the i^{th} spatial location, that is $\varepsilon(\mathbf{s}_i) = z(\mathbf{s}_i) - \widehat{z}(\mathbf{s}_i)$, in which z and $\widehat{z}$ denote the observed and predicted soil variable at location $\mathbf{s}_i (i = 1,\ldots,N; \mathbf{s}_i \in \mathscr{A})$, respectively, and N is the total number of population units in the study area $\mathscr{A}$. We consider the vectors $\mathbf{z}$ and $\widehat{\mathbf{z}}$ of equal lengths N with means $\overline{z}$ and $\overline{\widehat{z}}$ and standard deviations σ_z and $\sigma_{\widehat{z}}$. Further, we assume that $\mathbf{z}$ does not contain measurement error. In practice, ε is estimated from a probability sample of size $n << N$, or through cross-validation if collecting such a sample is not feasible.

2.1. Average error indices

Commonly reported average error indices are the mean prediction error (ME), mean absolute prediction error (MAE) and root mean squared prediction error (RMSE). They are easily interpretable because they are in the unit of the soil property, and defined as:

$$\text{ME} = \frac{1}{N}\sum_{i=1}^{N}\varepsilon(\mathbf{s}_i), \tag{1}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\varepsilon(\mathbf{s}_i)|, \tag{2}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\varepsilon(\mathbf{s}_i)^2}. \tag{3}$$

The ME can be positive or negative, while the MAE, RMSE and its square ($\text{MSE} = \text{RMSE}^2$) are nonnegative statistics with no upper-bound and an optimal value of 0. The ME (also called bias) is an indicator of whether the predicted values are systematically under- or overpredicted. The ME shows the central tendency (average error) but does not discern the magnitude of the error because of cancellation between positive and negative error. This cancellation is prevented in the MAE and RMSE by taking the absolute value of the error or by squaring the error before taking the average, respectively. The square term in Eq. (3) causes large errors to have a relatively greater importance than small errors on the total square error. From Eqs. (1)–(3) it follows that $\text{ME} \leqslant \text{MAE} \leqslant \text{RMSE}$.

Legates and McCabe (1999) have noted that the degree to which $\text{RMSE} \geqslant \text{MAE}$ is an indicator that the variance of the error is disproportionately altered by large individual errors. Several authors (e.g. Armstrong, 2001; Willmott and Matsuura, 2005) have thus warned against the use of the RMSE for the evaluation of the quality of a prediction when the distribution of the error is not symmetric.

2.2. Coefficient of determination

The coefficient of determination (r^2 or R^2) is the square of the Pearson's product-moment correlation coefficient (r), given by:

$$r = \frac{\sum_{i=1}^{N}\left(z(\mathbf{s}_i) - \overline{z}\right)\left(\widehat{z}\left(\mathbf{s}_i\right) - \overline{\widehat{z}}\right)}{\sqrt{\sum_{i=1}^{N}\left(z(\mathbf{s}_i) - \overline{z}\right)^2}\sqrt{\left(\widehat{z}\left(\mathbf{s}_i\right) - \overline{\widehat{z}}\right)^2}}. \tag{4}$$

The r^2 can be interpreted as the squared covariance between predicted and observed values (i.e. how much the two variables change together), normalized by the product of the two variances. Therefore it describes the linear correlation between the observed and predicted values and ranges from 0 to 1, with higher values indicating better correlation. The obvious limitation of the r^2 for evaluation of mapping accuracy is that the index is insensitive to additive or proportional differences between the observed and predicted values (Willmott, 1984). In other words, the r^2 quantifies the dispersion around the linear regression line between predicted and observed values $\widehat{\mathbf{z}} = \beta_0 + \beta_1\mathbf{z}$, but not the dispersion from the line of equality $\mathbf{z} = \widehat{\mathbf{z}}$, so that for any nonzero value of β_1 and any value of β_0, the squared correlation coefficient can be 1 (Legates and McCabe, 1999). A high value of r^2 can be obtained even in case of strongly biased predictions, yielding a misleading picture of the map quality. Further, it can be shown that the lower limit of the r^2 is found when $\widehat{z}(\mathbf{s}_i) = \overline{z}$, i.e. when the mean of the observations is used as predictor ($\beta_1 = 0$ and $\beta_0 = \overline{z}$). In such case, there is no correlation between the predictions and observations and $r^2 = 0$.

2.3. Modelling efficiency coefficient

The modelling efficiency coefficient (MEC, Janssen and Heuberger (1995)) is widely used in DSM studies. As for the r^2 its optimal value is 1 but it can be negative if the root mean square error exceeds the standard deviation of the observations. It is computed as follows:

$$\text{MEC} = 1 - \frac{\sum_{i=1}^{N}(z(\mathbf{s}_i) - \widehat{z}(\mathbf{s}_i))^2}{\sum_{i=1}^{N}(z(\mathbf{s}_i) - \overline{z})^2}, \tag{5}$$

which is equivalent to the ratio of the MSE to the variance of the observation, subtracted from unity (i.e. $\text{MEC} = 1 - \text{MSE}/\sigma_z^2$). In hydrology, the MEC is known as the Nash–Sutcliffe model efficiency coefficient (Nash and Sutcliffe, 1970). The MEC quantifies the improvement made by the model over using the mean of the observations as prediction. A value of 1 (MEC = 1) indicates a perfect match between the observed and predicted values of the soil property, whereas a value of 0 (MEC = 0) indicates that the mean of the observed values ($\overline{z}$) is as good predictor as the model. A negative value (MEC < 0) occurs when the mean of the observed values used as prediction is a better predictor than the model, i.e. when the residual variance (the numerator in Eq. (5)) is larger than the variance of the observations (the denominator in Eq. 5). Unlike the r^2, the MEC is sensitive to additive and proportional differences between observed and predicted values, that is, it considers the dispersion from the line of equality. If $\widehat{\mathbf{z}} = \beta_0 + \beta_1\mathbf{z}$ where $\beta_0 = 0$ and $\beta_1 = 1$, the MEC will decrease for change of β_0 and β_1 from 0 and 1, respectively.

Note the sensitivity of the r^2, MEC (as well as the MSE, or RMSE) to large error values. When the error distribution is strongly skewed towards large values, squaring the error severely inflates the map quality statistics which does not reflect the actual relationship between

measured and predicted values of the soil property. In such case, more robust statistics are the MAE, the (root) median square error, or adjusted statistics such as the modified MEC index proposed in Legates and McCabe (1999). The latter is computed by replacing the numerator in Eq. (5) (i.e. the MSE) by the MAE (Eq. (2)) and the denominator by the mean of absolute deviation (i.e. the mean of the absolute values of the distances from $z(\mathbf{s}_i)$ to $\overline{z}$).

3. Compact visualization of map quality

3.1. The Taylor diagram

Taylor diagrams (Taylor, 2001) quantify how closely the predicted values match the observed values in terms of correlation (r), standard deviation of the error (SDE) and standard deviation of observed (σ_z) and predicted ($\sigma_{\hat{z}}$) values. The statistics are related one to another by:

$$\text{SDE}^2 = \sigma_z^2 + \sigma_{\hat{z}}^2 - 2\sigma_z\sigma_{\hat{z}}r, \tag{6}$$

which have a geometrical relationship in the triangle through the law of cosines:

$$a^2 = b^2 + c^2 - 2bc\cos\alpha, \tag{7}$$

where a, b and c are the length of the triangle sides and α is the angle between sides b and c. A representation of this geometric relationship is shown in Fig. 1.

Eq. (6) above is the basis of the Taylor diagram as it enables to compare the individual contributions of the correlation and standard deviations to the SDE, i.e. to the standard deviation of the error ε, given by:

$$\text{SDE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\varepsilon(\mathbf{s}_i) - \overline{\varepsilon})^2}. \tag{8}$$

The Taylor diagram is shown in Fig. 2. In a polar coordinate system, the radial distance from the origin is assigned to the standardized (indicated by asterisk *) standard deviation $\sigma^* = \sigma_{\hat{z}}/\sigma_z$ and the angular position to the correlation coefficient between the map with predicted values and the map with the true values. The radial lines from the origin in Fig. 2 are determined by the inverse cosine of the correlation coefficient. A negative correlation of -1 is thus plotted 180^o away from a positive correlation value of 1.

The reference point is the point in the diagram obtained when the predictions are perfect (all prediction errors are 0). In that case the standardized standard deviation of the predictions is 1, and the correlation of the predictions with the true values is 1. This point is plotted along the abscissa at $(1, 0)$ and is shown in Fig. 2 in red. In the same figure the other point in black is called "test map" and represents the map quality indices of a test map to be evaluated.

The distance between any point in the diagram and the reference point is proportional to the SDE. Standardizing the SDE of Eq. (6) (i.e. dividing by σ_z^2) and taking the square root yields:

$$\text{SDE}^* = \sqrt{1 + \sigma^{*2} - 2\sigma^* r}. \tag{9}$$

The relative merits of different models are quantified by the distance SDE^* from the reference point, represented by the red dashed lines in Fig. 2. This is what makes the Taylor diagram useful: the position of a single point in this diagram, not only provides information about precision (semicircles around the reference point), but also about spatial (or temporal) patterns (r) and the degree of smoothing (semicircles around the origin).

In Fig. 2 the correlation coefficient between the test map predictions and the observations (i.e. the reference point) is 0.76, $\sigma_z = 1.17$ and $\sigma_{\hat{z}} = 0.90$ so that $\sigma^* = 1.30$ (i.e. the degree of smoothing). Accordingly, the test map in Fig. 2 is at a distance $\text{SDE}^* = 0.84$ from the reference point.

3.2. The solar diagram

An alternative to the Taylor diagram is the solar diagram derived from the relation (Eq. (10)) between the standard deviation of the error (SDE), mean error (ME) and RMSE. The squared SDE (in other words, the error variance) is the squared bias (ME^2, Eq. (1)) subtracted from the squared RMSE (Eq. (3)), thus verifying the quadratic relationship:

$$\text{RMSE}^2 = \text{ME}^2 + \text{SDE}^2. \tag{10}$$

A geometric representation of this relationship is provided in Fig. 3. The x-axis indicates the ME. A negative values represents systematic overestimation whereas a positive value an underestimation. The y-axis indicates the standard deviation of the error. The distance between the origin and any point is equal to the RMSE.

The solar diagram is based on Jolliff et al. (2009)'s target diagram. In the solar diagram only two quadrants are used, while Jolliff et al. (2009) used the remaining two quadrants to indicate whether the standard deviation of the observations is greater than the standard deviation of the predictions, by multiplying the SDE by $\text{sign}(\sigma_{\hat{z}} - \sigma_z)$. This way the target diagram also indicates whether the map is more smooth (positive SDE) or more rough (negative SDE) than the observations (in other words, whether the standard deviation of the predictions is larger or smaller than the standard deviation of the observations). However, maps with similar quality (in terms of bias, RMSE, and SDE) may appear as points far away from each other in the target diagram. It makes the target diagram more difficult to interpret and may lead to wrong interpretations.

A solar diagram is shown in Fig. 4. In a Cartesian coordinate system, the x-axis represents the standardized ME and the y-axis the standardized SDE. By standardizing all the quantities by the standard deviation of

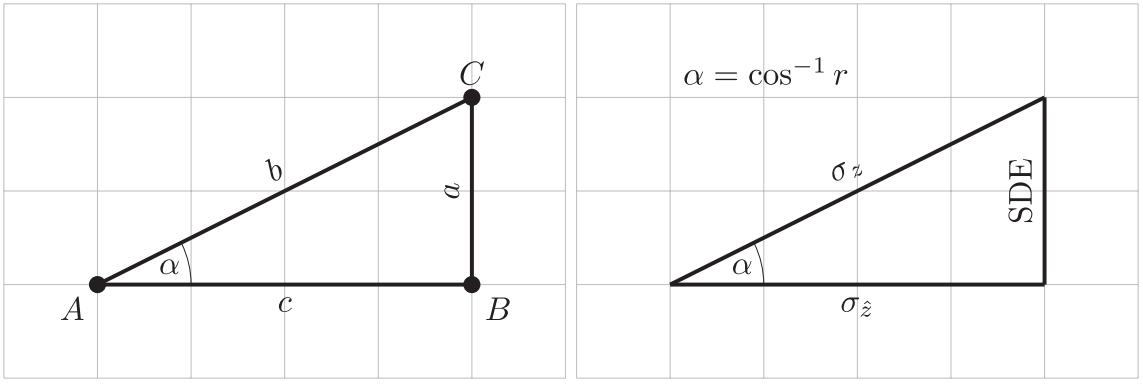

Fig. 1. Graphical representation of the relationship between the law of cosines and the map quality indices: the standard deviation of the error (SDE), variance of the prediction ($\sigma_{\hat{z}}$) and of the observation (σ_z).

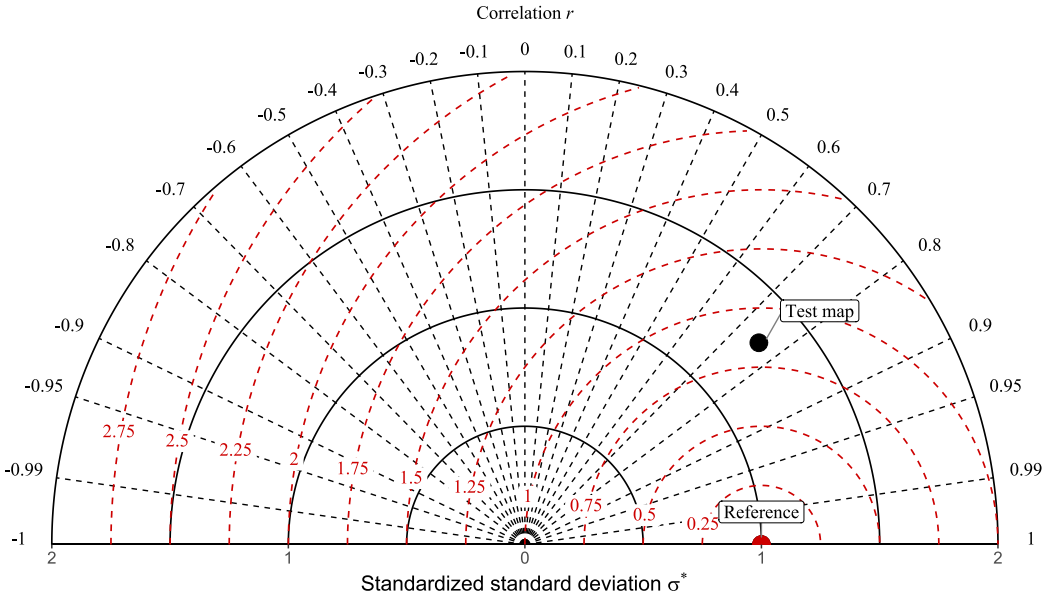

Fig. 2. Taylor diagram with a test map and a "reference" point. The red point at $(1, 0)$ is the reference point. The red dashed lines represent the distance in standardized SDE (Eq. (9)) from the reference point.

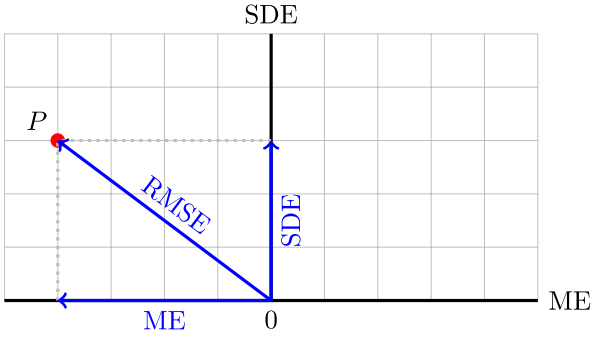

Fig. 3. Graphical representation of the geometric relationship between the map quality indices, the root mean square error (RMSE), mean error (ME) and standard deviation of the error (SDE). The distance from the origin at $(0, 0)$ to point P equals to the RMSE.

the observations, the distance in the diagram is expressed in standardized units, not in the unit of the observations. The distance from the origin to any point in the diagram is then expressed in standardized RMSE (RMSE*):

$$\text{RMSE}^{*2} = \text{ME}^{*2} + \text{SDE}^{*2}, \tag{11}$$

where $\text{ME}^{*} = \text{ME}/\sigma_z$ and SDE^{*} is given in Eq. (9).

The resulting diagram provides multiple information on map quality: whether there is a positive or negative bias in the prediction (the ME on the x-axis), information on the standard deviation of the error, and on the relative performance of different maps, quantified by the RMSE and represented by the distance to the origin.

The solar diagram takes its name from the coloured semicircles around the origin. Jolliff et al. (2009) showed the statistical relationship of the Pearson's r coefficient with the standardized SDE (SDE^{*}, Eq. (9)). Since the minimum value of SDE^{*} occurs when $\sigma^{*} = r, \forall\, r > 0$, the minimum SDE^{*} for a positive r value is expressed by $\min(\text{SDE}^{*}) = \sqrt{1 + r^2 - 2r^2}$, i.e. any point between $\min(\text{SDE}^{*})$ and the origin has a Pearson's r correlation value greater than r. Fig. 4 shows shades of yellow corresponding to classes of $\min(\text{SDE}^{*})$ values for $r > 0.95, r > 0.9, r > 0.7$ and $r > 0$. For example, a point with $r = 1, \text{SDE}^{*} = 0$ and $\text{ME}^{*} > 0$ can be anywhere on the x-axis of the solar diagram because r is always greater than the lower bounds indicated by the yellow colours. More information on the relationship between r and SDE^{*} is found in Jolliff et al. (2009, p.72).

The outer circle with standardized RMSE value of 1 also provides a convenient marker on the overall map quality. Points with a value of RMSE* greater than 1 indicate that the prediction is worse than the mean of the observations taken as prediction. The marker RMSE* = 1 is also directly related to the MEC (Eq. (5)) by $\text{MEC} = 1 - \text{RMSE}^{*2}$ and may be added to the solar diagram (the outermost black line in Fig. 4) as a reference of average map quality.

Finally, more information may be added to the solar diagram by means of a colour scale on the points to represent any supplemental information, for example, the modelling efficiency coefficient (MEC, Eq. (5)).

4. Simulation experiment and case studies

4.1. Simulation experiment

Methods On a grid of size 100 × 100 cells, we simulated a realisation of a map composed of a linear spatial trend superimposed on a Gaussian random field. The trend had an intercept of 5 and slope parameter of 0.1 for the x-axis and 0.05 for the y-axis. The Gaussian random field had a mean of zero and a covariance given by $C(h) = 5\exp\left(-\frac{h}{10}\right)$, where h is the lag distance. This map was multiplied by a factor of 0.3 to obtain the reference map. Nine modifications of the reference map were created to give nine maps with predictions:

1. a map positionally shifted by 20 cells in the x-direction and the y-direction,
2. a map where the x- and y-coordinates are reversed,
3. a map containing the mean value of the reference map at each coordinate,

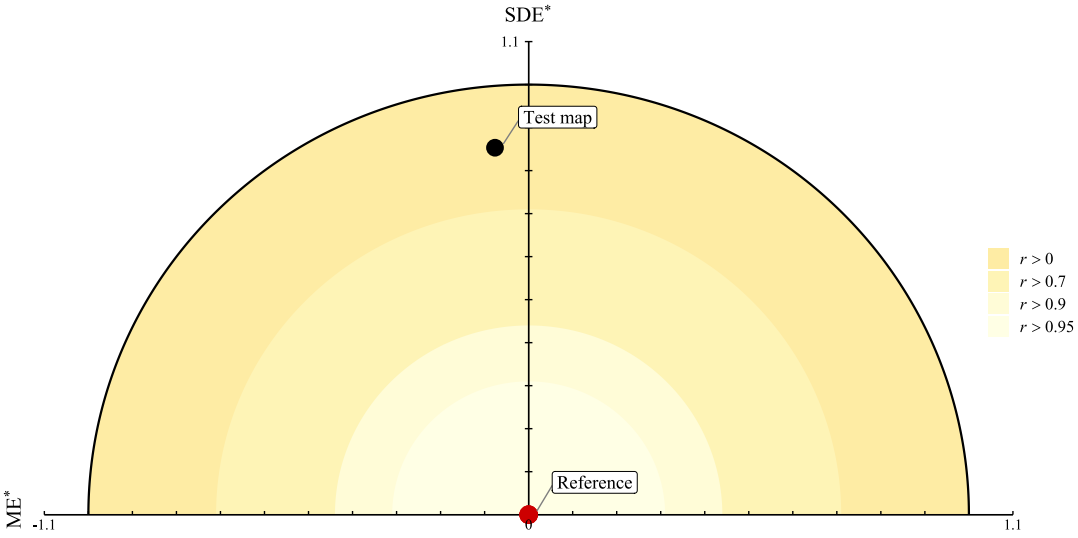

Fig. 4. Standardized solar diagram rendering the simulated maps and the reference map. The points inside the thick black line at RMSE* = 1 indicates the threshold for which the map is better than the mean of the observations taken at prediction. The yellow colour gradient represents areas where Pearson's correlation coefficient is greater than 0, 0.7, 0.9 and 0.95. The solid outer circle also delimits the area where Pearson's correlation coefficient is greater than 0 and where MEC > 0.

4. a smoothed map obtained by a moving average window of size 5 × 5 units,
5. a negatively biased map obtained by adding a value of 1 to the reference map,
6. a map predicted by a random forest model with a single tree and default parameters from the R package ranger (Wright and Ziegler, 2017) using the reference map as calibration data and the x- and y-coordinates as predictors,
7. a map predicted by a single regression tree with default parameters from the R package rpart (Therneau et al., 2019) using the reference map as calibration data and the x- and y-coordinates as predictors,
8. a map of the upper quartile, made by assigning the mean of the reference map to the values lower than the upper quartile and value of the reference map otherwise,
9. a map of the lower quartile, made by assigning the mean of the reference map to the values higher than the lower quartile and the value of the reference map otherwise.

Results Fig. 5 shows the reference map that represents a simulated map of a soil property and the nine variants of this map. The map quality indices described in Section 2 are shown in Table 1.

Table 1 shows that the MAE and RMSE are always greater or equal to the ME. There is no clear relationship between the ME and the MAE,

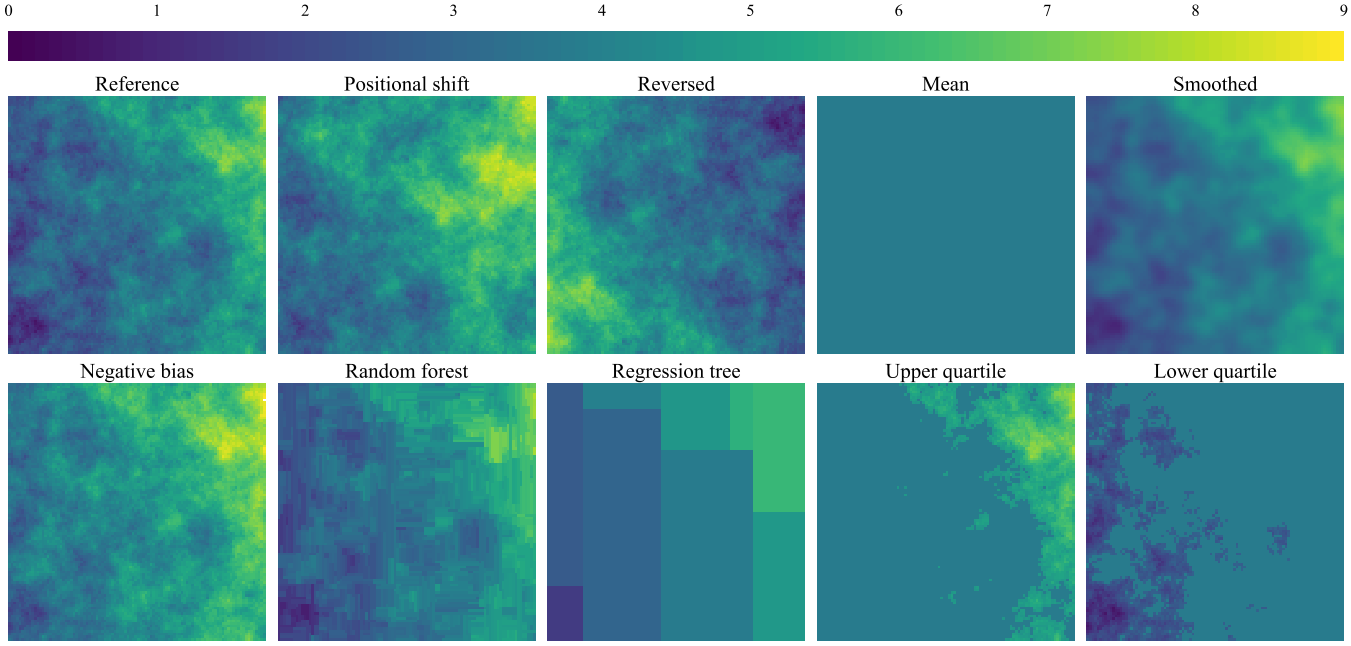

Fig. 5. Simulated reference map and modifications: (1) positionaly shifted, (2) reversed, (3) map of the mean, (4) smoothed, (5) with negative bias, (6) predicted by random forest, (7) predicted by regression tree, (8) upper quartile or mean, and (9) lower quartile or mean.

Table 1
Statistics described in Section 2 of the quality for the simulated and reference maps.

	ME	MAE	RMSE	r	r^2	MEC
Reversed	0	1.88	2.31	−0.7	0.48	−2.39
Positional error	−0.87	1.05	1.31	0.72	0.51	−0.1
Mean	0	1.01	1.25	0	0	0
Smoothed	0	0.19	0.24	0.98	0.97	0.96
Negative bias	−1	1	1	1	1	0.36
Random forest	0	0.2	0.29	0.97	0.95	0.95
Regression tree	0	0.49	0.61	0.87	0.76	0.76
Upper quartile	−0.43	0.58	0.84	0.84	0.7	0.55
Lower quartile	0.36	0.65	1	0.69	0.48	0.37

except for the negative bias maps, in which the MAE is positive (as is by definition) while the ME is negative. The RMSE is affected by the scale of the error. For the lower quartile map (i.e. when assigning the mean to the values higher than the lower quartile) the RMSE is 54% larger than the MAE. Large value of the ratio RMSE/MAE points to the presence of large error values affecting the RMSE (e.g. in the lower quartile and random forest maps). The characteristics of the RMSE and its decomposition was further discussed in Section 3. Note also that the mean map and the negative bias map have the same MAE, even-though the negative bias map is more plausible as judged by the eye.

Values of r and its square are nearly one for the random forest, biased and smoothed maps, indicating a strong correlation between the reference and predicted pattern, yet the values for these two maps are substantially dissimilar from those of the reference map. At the other extreme, it is possible to obtain an r value and its square of 0 when the map is a constant value but that no variation occurs in the predicted map, as shown by the r value of 0 for the mean map.

Table 1 shows that the values of r^2 and MEC are equal for the map predicted by random forest and the map predicted by a regression tree. This is because in both cases the predicted values are unbiased (ME = 0) and follow the line of equality $\mathbf{z} = \widehat{\mathbf{z}}$. This is confirmed by fitting a linear model with intercept between the predicted values by random forest and regression tree and the reference map values. For both predicted maps, the intercept is nearly zero and the slope nearly one, indicating that there is no substantial deviation from the 1:1 line. In this situation, both r^2 and MEC are the same. For the case of the reverse map with ME = 0, the MEC is negative because the MSE is larger than the variance of the true z values ($\sigma_z^2 = 1.57$).

The nine simulated maps are plotted in the Taylor diagram in Fig. 6. The reversed map is the farthest from the reference point, at a distance $\text{SDE}^* = \sqrt{1 + 1^2 - 2 \times 1 \times -0.69} = 1.84$. The points that come closer to the reference point are the ones with the smallest unbiased RMSE values (negatively biased map, smoothed map and map predicted by random forest). Indeed the smoothed map and the random forest map (Fig. 5) seem the most similar to the reference map, but the location of the negatively biased map close to the reference map shows the major shortcoming of the Taylor diagram, that is the absence of information on bias. Points close to the reference point can have large bias, as is the case for the negatively biased map in which the RMSE is only composed of bias (RMSE = 1, ME = 1, $\sigma^* = 1$). Including the bias is possible by adding a colour scale associated to the point, but the Taylor diagram may nonetheless be misinterpreted. This is further discussed in Section 5.

The nine simulated maps are plotted in the solar diagram in Fig. 7. Note that all quantities are standardized by the standard deviation of the values on the reference map (true values) to remove the unit of measurement. The standardized mean error along the x-axis shows that several maps are biased, either positively (lower quartile map) or negatively (positional error map, negative bias map). Few maps are unbiased, see, for example, the reversed, mean or random forest maps. These unbiased maps are on the y-axis, which means that the difference between the reference and predicted maps is due to variance. Two maps are outside the outermost circle at a RMSE* distance of 1. Using these maps do not represent an improvement over using the mean of the reference map as predictor of the soil property. This is confirmed by the mean map falling at the exact RMSE* distance of 1. All points between the circle a RMSE* distance of 1 and the origin are positively correlated with the reference map. The two other markers at a distance of 0.71 and 0.44 and 0.31 establish that all points falling between the markers and the origin have a correlation value greater or equal than 0.7, 0.9 and 0.95, respectively. This is illustrated with the yellow colour scale showing the Pearson's r correlation coefficient. The brighter the colour,

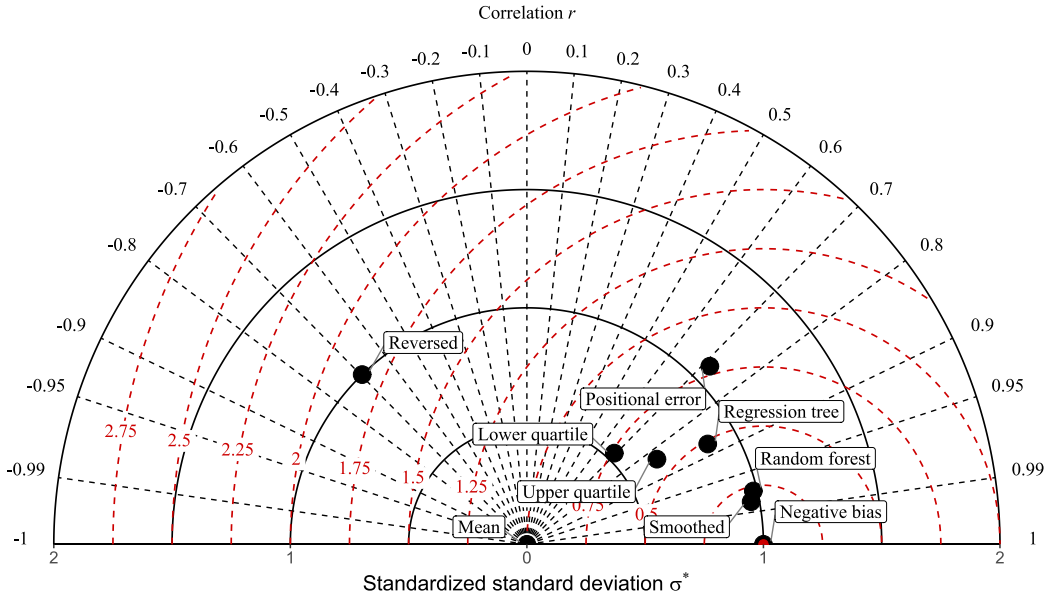

Fig. 6. Taylor diagram rendering the simulated maps and the reference map. The red point at (1,0) is the reference map. The red dashed lines represent the distance from the reference point (the standardized SDE, Eq. (9)).

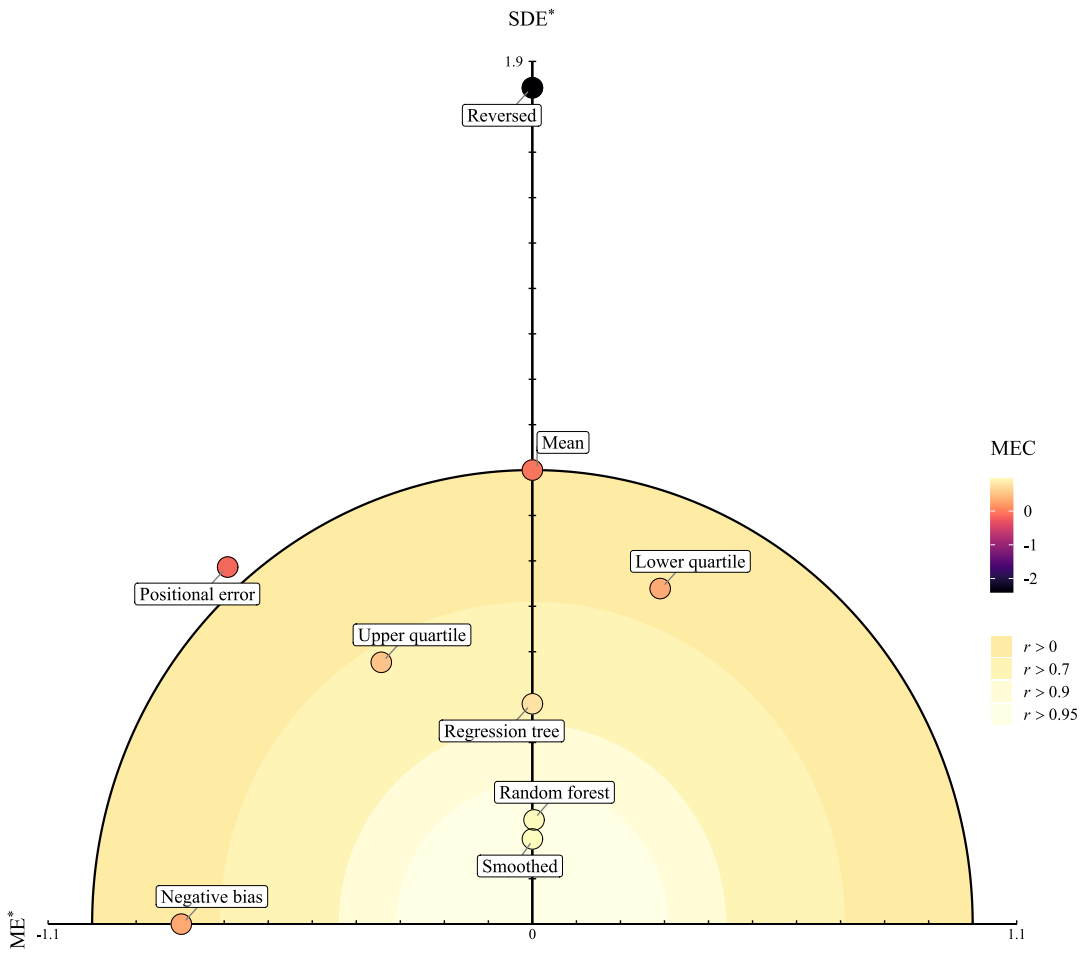

Fig. 7. Standardized solar diagram rendering the simulated maps and the reference map. The points inside the thick black line at RMSE* = 1 indicates the threshold for which the map is better than the mean of the reference map taken at prediction. The outer circle and the inner circles enclose areas where *r* is greater than 0, 0.7, 0.9 and 0.95. The colour scaling on the points represents the modelling efficiency coefficient (MEC) between the reference and simulated map.

the higher the Pearson's *r* correlation value.

4.2. Two applications studies

We selected a large rectangular area (1,523,776 km^2) in Eastern Amazonia with values of topsoil organic carbon density (SOC, in dg/kg) from the SoilGrids global dataset (Poggio et al., 2021), which we aggregated at a spatial resolution with grid cells of 1 km × 1 km. We collected an additional set of 28 environmental covariates representing average long-term climatic condition (i.e. temperature, precipitation, solar radiation, evapotranspiration), climate seasonality (temperature, precipitation and solar radiation seasonality), extreme climatic condition (temperature of driest and warmest quarter, precipitation of wettest and driest month, aridity index), topography (elevation, slope), soil (topsoil clay and sand content), vegetation (above-ground live woody biomass, Shannon enhanced vegetation index, Terra net primary production) and average long-term MODIS bands (red, green, NIR, and SWIR 1–2). These covariates were obtained from publicly available sources such as SoilGrids (Poggio et al., 2021), WorldClim (Fick and Hijmans, 2017), and Google Earth Engine. All covariates were harmonized to the spatial resolution and extent of the SOC map. We used the SOC map as proxy for the "true" SOC, and selected a sample of size 500 from this map using simple random sampling. The sample was used for model calibration and prediction in two applications, whereas map accuracy statistics were obtained from the entire population, i.e. by predicting on 1,308,838 grid cells. In the first application various parameters combinations of a mapping model were tested, and the effect of these parameters on the resulting map quality was tracked with summary diagrams. In the second application, we compared five maps made by five different models and showed how the summary diagrams can be used to convey information on map quality more efficiently than with a summary statistics table.

4.2.1. Tracking effect of model parameters on map quality

Methods We built a generalized boosted regression tree (GBM) model for mapping the topsoil organic carbon content using the set of 28 environmental covariates as predictor. We repeated the calibration of the GBM model for different sets of the shrinkage and interaction.depth parameters of the GBM implementation in the gbm R package (Greenwell et al., 2020). The shrinkage parameter corresponds to the learning rate and the interaction.depth to the maximum depth of each tree. For more information on GBM, we refer to Hastie et al. (2009). We tested 10 combinations of parameters using Latin hypercube sampling of the shrinkage parameter between 0.00001 and 1 and of the interaction.depth parameter between 1 and 10. Other model parameters were held at their default value. The map quality indices resulting from the calibration of 10 mapping models were visualized in the Taylor and solar diagrams.

Results The diagrams in Figs. 8 and 9 show the map quality for different values of the shrinkage and interaction.depth parameters of the GBM mapping model. In Fig. 9, points in the Taylor diagram are clustered between SDE* values of 0.5 and 0.75, between correlation coefficient values of 0.7 and 0.9, and between standardized standard

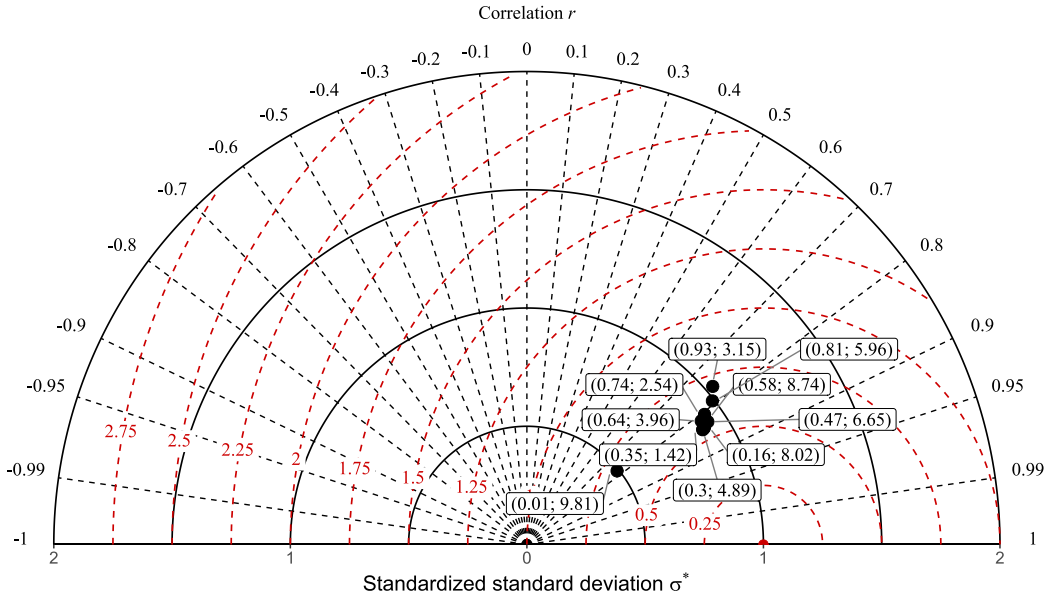

Fig. 8. Taylor diagram rendering the 10 simulated maps predicted with different sets of shrinkage and interaction.depth mapping model parameters and compared to the reference point at $(1, 0)$. The red dashed lines represent the distance from the reference point (the standardized SDE, Eq. (9)). The labels indicate the parameter values of the GBM mapping model (shrinkage; interaction.depth).

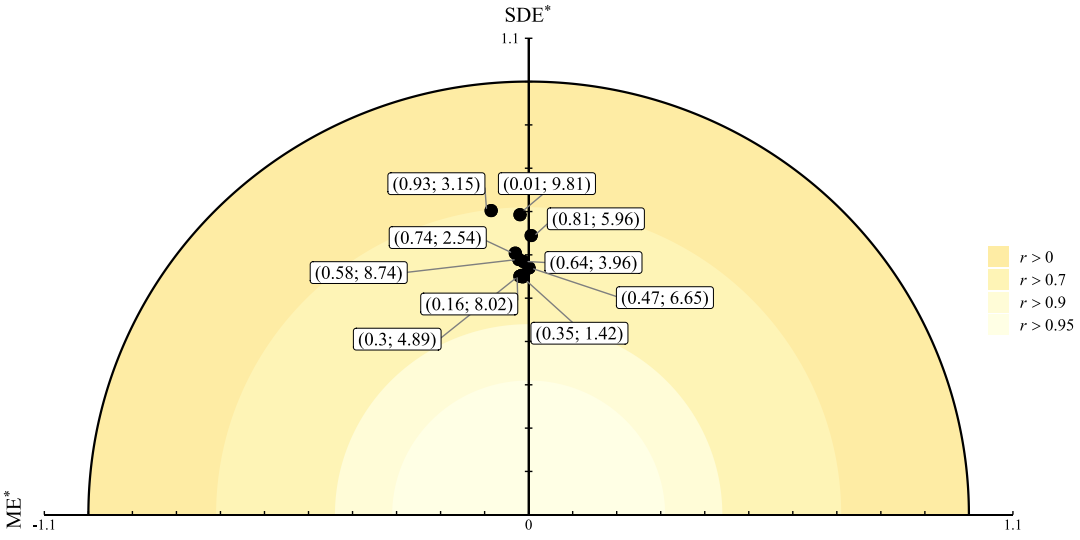

Fig. 9. Solar diagram of the 10 maps obtained with different sets of shrinkage and interaction.depth mapping model parameters. The labels indicate the parameter values of the GBM mapping model (shrinkage; interaction.depth).

deviation values of 0.5 and 1.3. The labels show that a small value of the learning rate is preferable in combination with a moderately high value of the depth parameter. Interaction of the learning rate and depth parameters also shows that the worse result is found for the combination of very small learning rate and large depth parameter.

In both diagrams, the same point (i.e. for shrinkage = 0.16 and interaction.depth = 8.02) comes closest to the reference point (Taylor diagram) and origin (solar diagram) and so have a similar SDE* and RMSE* value, which indicates that the predictions for this combination of parameters are relatively unbiased. This is confirmed in the solar diagram, where this point is close the y-axis: predictions are nearly unbiased so that the primary contribution to the RMSE is the SDE (Eq. (3)). In the absence of bias, both diagrams lead to the same conclusion on the optimal parameters combination of the mapping model. A different pattern between the plots is observed for the point with parameters shrinkage = 0.93 and interaction.depth = 3.15 and the point with parameters shrinkage = 0.01 and interaction.depth = 9.81. In the Taylor diagram, this former point appear better than the latter point, because closer to the standardized standard deviation value of 1. In the solar diagram, an opposite pattern is shown: the point with parameters

shrinkage = 0.01 and interaction.depth = 9.81 would be preferred because it has less bias for a similar value of RMSE*.

4.2.2. Comparing soil maps made by different models

Methods We built six common mapping models for mapping the topsoil organic carbon using the set of 28 environmental covariates as predictor. The parameters of models were chosen manually. Note that the aim of this case study is to compare soil maps, not to compare and evaluate the performance of models, so no conclusion about the supposed superiority of a model should be drawn from this study case. We used the R programming language (R Core Team, 2020) to implement and compare maps created by:

1. Linear regression, using the default implementation from the stats package.
2. Regression tree with the maximum depth of the tree set to 30, a complexity parameter of 10^{-9} and the maximum number of observations in a split set to 2. We used the implementation of the rpart package (Therneau et al., 2019).
3. Cubist with 10 committee models and the implementation of the Cubist package (Kuhn et al., 2020).
4. Random forest with 500 trees and other parameters set to their default value as in the ranger package (Wright and Ziegler, 2017).
5. Neural networks with a single-hidden layer, 5 units and decay parameter value of 2, using the implementation of the nnet function in the caret package (Kuhn, 2008).
6. Support vector machine with the default implementation of the e1071 package (Meyer et al., 2020).

For each map the quality indices were reported on the Taylor and solar diagrams.

Results The map quality indices are presented in Table 2 and plotted in the diagrams in Figs. 10 and 11. Table 2 shows that most models are unbiased, except cubist, neural networks and support vector machine. The largest ME values are obtained by the map made by support vector machine (ME = −7.35) and neural networks (ME = 3.52). Map quality indices presented in Table 2 show that it is generally difficult to decide on the best map because it varies from one criterion to another. For example, values of the MEC, r and its square indicate that the map made by the random forest model is the most accurate. Plotting the statistics in the diagrams in Figs. 10 and 11 facilitates the choice of a mapping model. According to Figs. 10 and 11, the map made by the regression tree is the farthest away from the reference point, whereas the map made by cubist and random forest are the second closest and closest points to the reference point, respectively. While values of SDE* (Taylor diagram) are nearly similar for both models, inclusion of the bias in the values of RMSE* as is done in the solar diagram is slightly exaggerating the advantage of the map made by random forest over that made by cubist. In both diagrams, the map made by random forest appears to be the closest from reality.

Besides description of the diagrams in terms of their individual components (i.e. the SDE in the Taylor and RMSE in the solar diagram), indices can be considered together with respect to the aim of the map. For instance, random forests and cubist models have similar RMSE, SDE and r values, but differ in bias. If the aim is to use the map as input to a process model, then one probably would prefer the cubist model as the additional bias is judged to be less of a problem than underestimating spatial variation. However, if the aim is not only to produce a high quality map, but also to use models that are easily interpretable, one may focus on either a linear regression or a regression tree without substantial decrease in SDE and RMSE. If again one needs a map without too much smoothing, then the regression tree may be the better choice, although linear regression is slightly more accurate and precise and slightly better reproduces spatial patterns.

5. Discussion and conclusions

The literature (e.g Armstrong, 2001; Willmott and Matsuura, 2005; Jolliff et al., 2009) showed that no single statistical quantity can represent all aspects of map quality. Correlation-based indices, for example the Pearson's r and r^2, can yield a misleading picture of the map accuracy. A more useful, albeit related measure is the MEC which quantifies the improvement made by the map over using the mean of the observations as a predictor, and is sensitive to deviation from the line of equality $\mathbf{z} = \widehat{\mathbf{z}}$. In addition to relative map accuracy indices, it is strongly recommended to report an index of the overall bias by the ME, as well as average error indices such as MAE and RMSE. They are easy to interpret because they are in the unit of the soil property, but the RMSE is disproportionately affected by large error values. Overall, it is strongly recommended to use multiple and complementary map quality indices for map quality evaluation, not just a single index or a list of comparable indices.

By exploiting the statistical relationships between indices, it is sensible to plot these different indices into a single plot. Taylor and solar diagrams are designed for this purpose. The Taylor diagram explicitly conveys information on correlation, standard deviations of the predicted and observed values, and standard deviation of the error. In the solar diagram, also other aspects are included, such as bias and overall RMSE, providing a broader overview on the map quality than the Taylor diagram. It is possible to convey information on overall RMSE and bias in the Taylor diagram via the addition of a colour scale. Alternatively, considering the SDE as a vector from the point representing the map to evaluate to the reference point, Taylor (2001) proposed to visualize bias by a vector starting from the point representing the map to evaluate. This vector is perpendicular to the SDE-vector, and the vector sum of the SDE-vector and bias-vector is equal to the RMSE (i.e., it follows the geometric relationship of Eq. (10) as is done in the solar diagram). A graphical representation is provided in Fig. 12. However, this solution is possible only in the case of a limited number of maps to evaluate due to overplotting. Another way to include the bias is to extend the Taylor diagram to 3 dimensions, in which the distance between any point in the 3-D space and the reference point is equal to the RMSE.

Overall, both Taylor and solar diagrams provide an integrated and visually attractive solution to evaluate quantitative soil maps. In the examples it was shown in a variety of ways that map quality is better evaluated through the combined effect of statistical indices. In the first example we created different soil maps and compared them to a reference map. We consider that Taylor and solar diagrams provided better

Table 2
Statistics of the map quality indices described in Section 2, with SDE* and RMSE*. The SDE* measures the distance to the reference point in the Taylor diagram while the RMSE* measures the distance to the origin in the solar diagram.

	SDE*	RMSE*	ME	MAE	RMSE	r	r^2	MEC
Linear regression	0.67	0.67	−0.01	42.30	59.55	0.74	0.40	0.55
Regression tree	0.74	0.74	0.22	42.60	65.84	0.72	0.34	0.45
Cubist	0.53	0.54	−1.21	30.77	47.50	0.85	0.52	0.71
Random forest	0.52	0.52	0.14	30.36	46.19	0.86	0.58	0.73
Neural networks	0.72	0.72	3.52	43.68	64.04	0.74	0.33	0.48
Support vector machine	0.60	0.60	−7.35	33.34	53.68	0.82	0.57	0.63

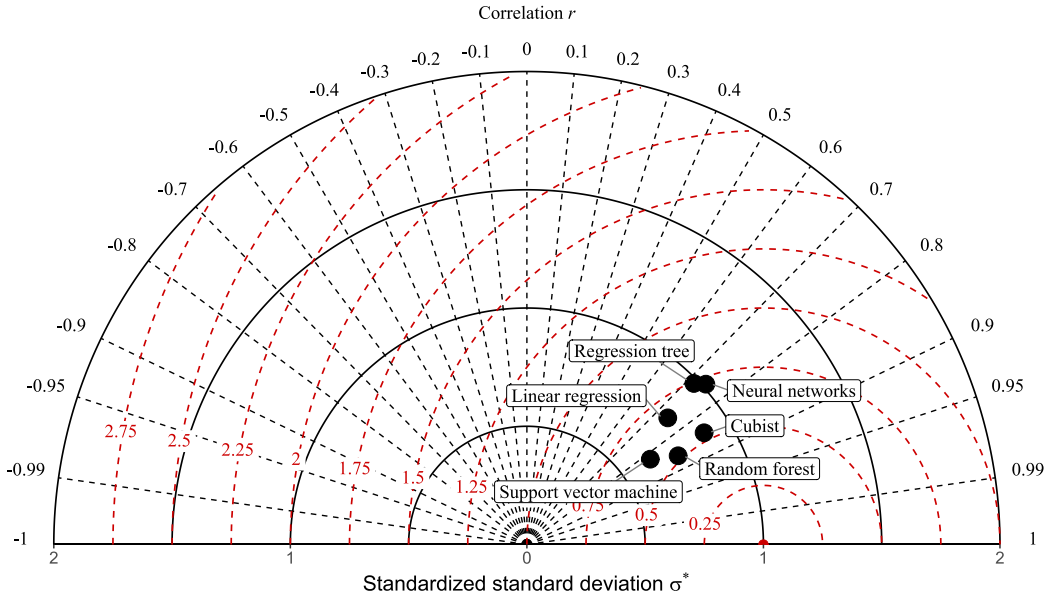

Fig. 10. Taylor diagram of the 6 common mapping models for mapping topsoil organic carbon, and compared to the reference point at $(1,\ 0)$. The red dashed lines represent the distance from the reference point (the standardized SDE, Eq. (9)).

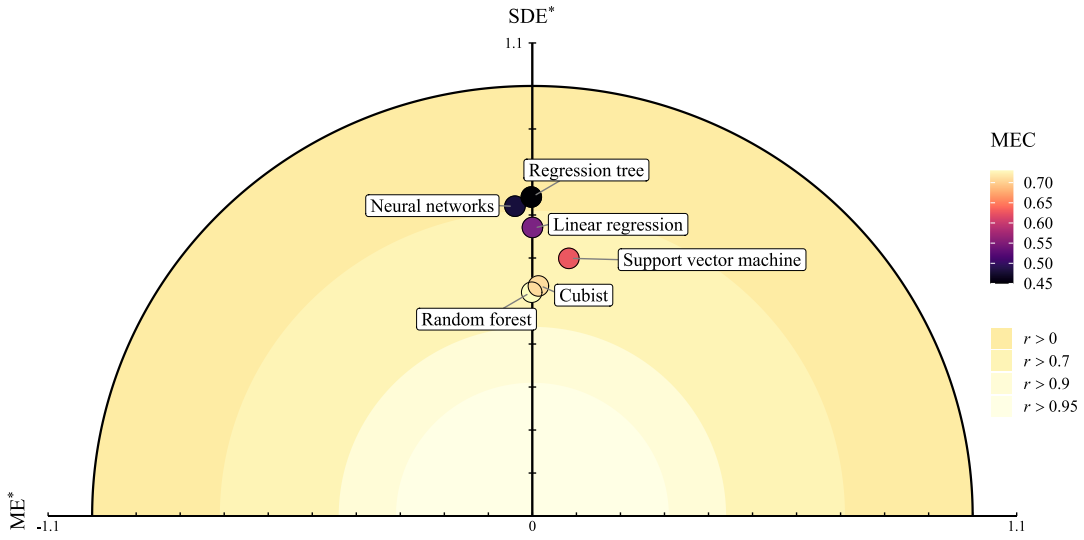

Fig. 11. Solar diagram of the 6 common mapping models for mapping the topsoil organic carbon, and compared to the reference point at coordinates $(0, 0)$. The colour scaling is the modelling efficiency coefficient (MEC) computed from the observed and predicted values of topsoil organic carbon.

insight into map quality, instead of using a list of indices. This was also the case for the second example: summary diagrams made simple the decision of the best map, via understanding of how much of the disagreement is cause by bias, variance and lack of correlation. In our example of choosing the mapping model parameters among 10 possible choices, choice of the optimal mapping model based on the r^2, as is often done in practice, yielded a different map than when using the diagrams. The diagrams also showed that for some parameter choice, using the mean of the observations was better than using the model.

Taylor (2001) and Jolliff et al. (2009) described improvement of Taylor and solar diagrams by means of skill scores instead of RMSE-based metrics. RMSE-based metrics are potentially deficient in that for a given value of the correlation (r less than unity), the minimum value of RMSE is found when the standardized standard deviation σ^* equals the correlation (i.e. $\sigma^* = r$, see also Section 3.2), instead of the ideal value $\sigma^* = 1$. Take the case of the positional error and lower quartile map in Fig. 6. Both maps have correlation nearly equal $r = 0.7$, and for this correlation value the minimum SDE* occurs at $\sigma^* = 0.7$. Consequently the lower quartile map appears more accurate because closer to $\sigma^* = 0.7$, but in reality the positional error map could be preferred because closer to the ideal value $\sigma^* = 1$. In other words, RMSE-based metrics may be inappropriate if the ultimate goal is to move both correlation and standardized standard deviation to an optimal value of 1. For these reasons, Taylor (2001) and Jolliff et al. (2009) proposed skill diagrams using skill scores for which reduction in the RMSE or SDE are not necessarily made at the expense of the standardized standard deviation. Integrating the skill scores in Taylor and solar diagrams would certainly be a valuable contribution to future research.

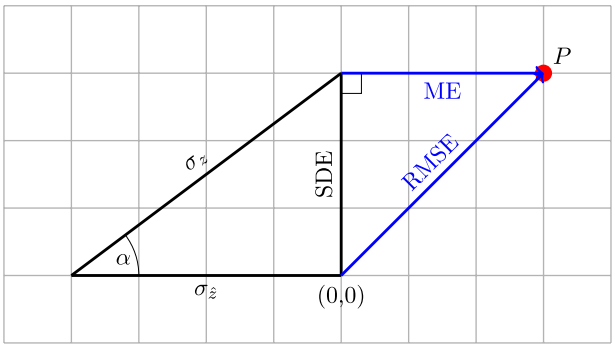

Fig. 12. Graphical representation between the law of cosines and the map quality indices in the Taylor diagram, with the inclusion of the ME and RMSE. The distance from the origin at $(0, 0)$ to point P is the RMSE, calculated by the vector-sum of the SDE vector and ME vector. The ME vector is perpendicular to the SDE vector.

Finally, it is in principle possible to account for measurement error of the validation data. Estimation of the measurement error is beyond the scope of this paper, but in estimating soil property errors may arise from the laboratory analysis method and in the type of inference, for example, when soil data are inferred with infrared spectroscopy techniques or pedotransfer functions. When explicit assessment of the error associated with soil data is made, it can be included in the diagrams in various ways. Taylor (2001) simply plotted both the reference point and the reference point resulting from measurement error. The distance between the two is a visual representation of measurement error. Jolliff et al. (2009) represented this error with a circle centered at the origin, the radius of which equals measurement error. Points falling between this circle and the origin are within the range or soil data measurement error. In this range, further improvement in mapping accuracy is unlikely to be meaningful. Both diagrams could also potentially be extended in three dimensions. Further details for this extension is found in Taylor (2001).

6. Code availability

Code for computing the map quality indices, Taylor and solar diagrams in the R programming language are available via an open access link at https://github.com/AlexandreWadoux/MapQualityEvaluation, along with a reproducible example using simulated data.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Armstrong, J.S., 2001. Evaluating forecasting methods. In: Armstrong, J.S. (Ed.), Principles of Forecasting. Springer, Boston, USA, pp. 443–472.
Bishop, T.F.A., McBratney, A.B., Whelan, B.M., 2001. Measuring the quality of digital soil maps using information criteria. Geoderma 103, 95–111.
Brus, D.J., Kempen, B., Heuvelink, G.B.M., 2011. Sampling for validation of digital soil maps. European Journal of Soil Science 62, 394–407.
Burrough, P.A., Beckett, P.H.T., Jarvis, M.G., 1971. The relation between cost and utility in soil survey (I–III). Journal of Soil Science 22, 359–394.
Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology 37, 4302–4315.
Finke, P.A., 2006. Quality assessment of digital soil maps: producers and users perspectives. In: Lagacherie, P., McBratney, A.B., Voltz, M. (Eds.), Developments in Soil Science, vol. 31. Elsevier, Amsterdam, NL, pp. 523–631.
Greenwell, B., Boehmke, B., Cunningham, J., GBM Developers 2020. gbm. url:https://CRAN.R-project.org/package=gbm R package version 2.1.8. Accessed 21.02.2021.
Hastie, T., Tibshirani, R., Friedman, J.H., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York, USA.
Heuvelink, G.B.M., Angelini, M.E., Poggio, L., Bai, Z., Batjes, N.H., van den Bosch, H., Sanderman, J., 2021. Machine learning in space and time for modelling soil organic carbon change. European Journal of Soil Science 72, 1607–1623.
Janssen, P.H.M., Heuberger, P.S.C., 1995. Calibration of process-oriented models. Ecological Modelling 83, 55–66.
Jolliff, J.K., Kindle, J.C., Shulman, I., Penta, B., Friedrichs, M.A.M., Helber, R., Arnone, R.A., 2009. Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment. Journal of Marine Systems 76, 64–82.
Kuhn, M., 2008. Building predictive models in R using the caret package. Journal of Statistical Software 28, 1–26.
Kuhn, M., Weston, S., Keefer, C., Coulter, N., 2020. Cubist. url:https://CRAN.R-project.org/package=Cubist R package version 0.2.3. Accessed 21.02.2021.
Lark, R.M., 2000. A comparison of some robust estimators of the variogram for use in soil survey. European Journal of Soil Science 51, 137–157.
Legates, D.R., McCabe Jr, G.J., 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resources Research 35, 233–241.
Malone, B.P., de Gruijter, J.J., McBratney, A.B., Minasny, B., Brus, D.J., 2011. Using additional criteria for measuring the quality of predictions and their uncertainties in a digital soil mapping framework. Soil Science Society of America Journal 75, 1032–1043.
Marsman, B.A., Gruijter, J.J., 1986. Quality of Soil Maps: a Comparison of Survey Methods in a Sandy Area Soil Survey Papers. Technical Report Netherland Soil Survey Institute. Wageningen, the Netherlands.
Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C.-C., Lin, C.-C., 2020. e1071. url:https://CRAN.R-project.org/package=e1071 R package version 1.7.4. Accessed 23.02.2021.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I-A discussion of principles. Journal of Hydrology 10, 282–290.
Nussbaum, M., Spiess, K., Baltensweiler, A., Grob, U., Keller, A., Greiner, L., Schaepman, M.E., Papritz, A.J., 2018. Evaluation of digital soil mapping approaches with large sets of environmental covariates. SOIL 4, 1–22.
Piikki, K., Wetterlind, J., Söderström, M., Stenberg, B., 2021. Perspectives on validation in digital soil mapping of continuous attributes–A review. Soil Use and Management 37, 7–21.
Poggio, L., de Sousa, L.M., Batjes, N.H., Heuvelink, G.B.M., Kempen, B., Riberio, E., Rossiter, D., 2021. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. SOIL 7, 217–240.
Ramirez-Lopez, L., Wadoux, A.M.J.-C., Franceschini, M.H.D., Terra, F.S., Marques, K.P.P., Sayão, V.M., Demattê, J.A.M., 2019. Robust soil mapping at the farm scale with vis–NIR spectroscopy. European Journal of Soil Science 70, 378–393.
R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing Vienna, Austria. url:https://www.R-project.org/.
Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research: Atmospheres 106, 7183–7192.
Therneau, T., Atkinson, B., Ripley, B., 2019. rpart. url:https://CRAN.R-project.org/package=rpart R package version 4.1-15. Accessed 21.02.2021.
Webster, R., Beckett, P.H.T., 1968. Quality and usefulness of soil maps. Nature 219, 680–682.
Willmott, C.J., 1984. On the evaluation of model performance in physical geography. In: Gaile, G.L., Willmott, C.J. (Eds.), Spatial Statistics and Models. Springer, Dordrecht, NL, pp. 443–460.
Willmott, C.J., Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research 30, 79–82.
Wright, M.N., Ziegler, A., 2017. ranger: A fast implementation of random forests for high dimensional data in C++ and R. Journal of Statistical Software 77, 1–17.