Layer-specific Analysis and Spatial Prediction of Soil Organic Carbon using Terrain Attributes and Results from Soil Redistribution Modelling

Abstract

High-resolution soil organic carbon (SOC) maps are a major prerequisite for many environmental studies dealing with carbon stocks and fluxes. Especially in hilly terrain, where SOC variability is most pronounced, high quality data are rare and costly to obtain.

In this study factors and processes influencing the spatial distribution of SOC in three soil layers (< 0.25, 0.25-0.5, and 0.5-0.9 m) in a sloped agricultural catchment (4.2 ha) were statistically analysed, utilising terrain parameters and results from water and tillage erosion modelling (with WaTEM/SEDEM). Significantly correlated parameters were used as covariables in regression kriging (RK) to improve SOC mapping for different input data densities (6-37.9 soil cores per hectare) compared to ordinary kriging (OK).

In general, patterns of more complex parameters representing soil moisture and soil redistribution correlated highest with measured SOC patterns, and correlation coefficients increased with soil depth. Analogously, the relative improvement of SOC maps produced by RK increased with soil depth. Moreover, an increasing relative improvement of RK was achieved with decreasing input data density. Hence, an expectable decline in interpolation quality with decreasing data density could be reduced especially for the subsoil layers incorporating soil redistribution and wetness index patterns in RK.

The optimal covariable differed between the soil layers indicating that bulk SOC mapping deduced from topsoil SOC measurements might not be appropriate in sloped agricultural landscapes. However, generally more complex covariables, especially patterns of soil redistribution, exhibit a great potential in improving subsoil SOC mapping.

Abbreviations
A, aspect; CA, catchment area; C-plan, plan curvature; C-prof, profile curvature; CV, coefficient of variation; DEM, digital elevation model; DOC, dissolved organic carbon; E_{til}, tillage erosion; E_{tot}, total erosion; E_{wat}, water erosion; GPS, global positioning system; EBLUP, empirical best linear unbiased predictor; KED, kriging with external drift; Max, maximum; ME, mean error; MEF, model-efficiency coefficient; Min, minimum; OK, ordinary kriging; RE, relative elevation; REML, residual maximum likelihood; RI, relative improvement; RK, regression kriging; RMSE, root mean square error; RUSLE, revised universal soil loss equation; R_{17}, 17 m input raster; R_{25}, 25 m input raster; R_{50}, 50 m input raster; S, slope; SC, skewness coefficient; SD, standard deviation; SEDEM, sediment delivery model; SOC, soil organic carbon; SPI, stream power index; WaTEM, water and tillage erosion model; WI, wetness index.
Introduction

Soils play a major role in the global carbon cycle. Approximately 1500 Pg C are stored in the topmost metre of soils worldwide corresponding to twice the amount of atmospheric C and triple the amount of C stored in the biosphere (Schlesinger, 2005). Nevertheless, the role of this reservoir as CO$_2$ sink or source in global climate and environmental issues is not clearly understood. To analyse the possibilities of soils of sequestering atmospheric CO$_2$ as well as for other environmental issues (e.g. analysis of soil quality and adaptation of management practices) detailed and precise maps of the distribution of soil organic carbon (SOC) are an essential prerequisite. Especially in agricultural regions the complex arrangement and combination of topography, soil and management practices and thereby controlled biological processes lead to a high spatial variability of SOC. Under rolling topography the spatial heterogeneity of SOC in agricultural fields is also affected by soil redistribution processes.

Most studies dealing with soil and SOC redistribution indicate an increase of SOC in depositional areas as compared to regions of erosion, where SOC is depleted (e.g. Ritchie et al., 2007; Mabit et al., 2008). However, there are also opposite findings published in literature. For example, Arriaga and Lowery (2005) found that the introduction of clayey subsoil material into the plough layer due to erosion of the topsoil stabilised and hence increased SOC content in the topsoil. Below the plough layer the expected decrease in SOC occurred in areas of erosion, while more or less constant SOC contents were found throughout the soil profile in regions of soil deposition (Arriaga and Lowery, 2005).

To produce accurate SOC maps, in general, different kinds of interpolation schemes are applied based on point measurements. As field measurements are costly and time-consuming, the improvement of interpolation methods while using secondary information was extensively tested (e.g. Odeh et al. 1994, 1995; Takata et al. 2007). Therefore, terrain attributes of various complexity were used as proxies for relief driven processes of pedogenesis. In most studies primary terrain parameters, which can nowadays easily be derived from digital elevation
models (DEM), such as (relative) elevation (Mueller and Pierce, 2003; Ping and Dobermann, 2006; Sumfleth and Duttmann, 2008), slope (Mueller and Pierce, 2003; Ping and Dobermann, 2006; Takata et al., 2007; Sumfleth and Duttmann, 2008), aspect (Odeh et al., 1994; 1995), and curvature (Terra et al., 2004; Takata et al., 2007) were used as secondary information.

These primary terrain parameters can also be combined to more complex secondary terrain parameters or indices comprising landscape processes more explicitly. Often the wetness (or topographic) index (Beven and Kirkby, 1979) is tested for its capability to improve the interpolation of soil organic carbon and other soil properties (e.g. Herbst et al., 2006; Takata et al., 2007; Sumfleth and Duttmann, 2008). Besides these terrain parameters also other parameters are used as covariables for interpolation schemes in literature. Takata et al. (2007), for example, use the enhanced vegetation index, whereas Chen et al. (2000) use soil colour to successfully predict the spatial distribution of SOC. Both parameters are derived from remote sensing data. Another covariable utilised effectively to improve the interpolation of SOC is the electrical conductivity of the topsoil layer (Terra et al., 2004; Simbahan et al., 2006; Ping and Dobermann, 2006).

A variety of statistical and geostatistical methods for interpolating point data with and without consideration of secondary information exist (e.g. Isaaks and Srivastava, 1989; Webster and Oliver, 2001). While more simple statistical approaches, like a (multiple) linear regression performed well under certain circumstances to interpolate SOC (e.g. Mueller and Pierce, 2003), often geostatistical kriging approaches accounting for the spatial structure of SOC as well as of that of covariables performed better. Whereas ordinary kriging utilises the spatial autocorrelation of the target variable alone, there are several geostatistical techniques that allow for the incorporation of a spatial trend caused by spatial patterns of secondary parameters in the kriging approach. Most often regression kriging (RK) or kriging with external drift (KED) are applied. In contrast to KED, which is an one-algorithm system, RK is a stepwise approach combining a regression between target and covariable with simple or
ordinary kriging of the regression residuals. Whereas the target and the covariable have to be linearly related in KED, RK also allows for the integration of more complex regression models (i.e. multiple linear or non-linear functions). KED and linear RK only differ in the computational steps used, but the resulting predictions are the same given the same input data (target and covariable) and the same regression fitting method (Hengl et al., 2007).

Odeh et al. (1994, 1995) defined three types of regression kriging, of which regression kriging model C, where the trend function is calculated using ordinary least squares, and the residuals are interpolated using ordinary kriging, was successfully used in improving the interpolation of soil organic carbon as well as that of other soil properties in many studies (e.g. Terra et al., 2004; Herbst et al., 2006; Takata et al., 2007; Sumfleth and Duttmann, 2008).

A geostatistically more sophisticated approach, which overcomes some statistical deficiencies of KED and RK in incorporating secondary parameters, is REML-EBLUP (Lark et al., 2006). In this method the trend model is estimated using residual maximum likelihood (REML), and subsequently the estimated parameters are used for the empirical best linear unbiased prediction (EBLUP). However, Minasny and McBratney (2007a; 2007b) who compared RK model C with REML-EBLUP for interpolating four different soil properties concluded that, although statistically somewhat inappropriate, RK used in many SOC studies (e.g. Terra et al., 2004; Takata et al., 2007; Sumfleth and Duttmann, 2008) has proven to be a robust technique for practical applications. In concordance to these findings (Chai et al., 2008), who analysed the effect of different covariables on the spatial interpolation of soil organic matter, concluded that REML-EBLUP performed more stable in their study, but that the improvement was not significant compared with RK.

To our knowledge all studies explicitly dealing with the interpolation of SOC and its possible improvement by incorporating covariables in the interpolation process are focused on the topsoil layer (< 0.3 m, e.g. Mueller and Pierce, 2003; Terra et al., 2004; Simbahan et al., 2006; Ping and Dobermann, 2006; Takata et al., 2007; Sumfleth and Duttmann, 2008).
However, the spatial patterns of SOC in agricultural catchments might differ substantially in different soil depths, an aspect which should be taken into account for soil carbon balancing studies as well as for simulations of soil carbon dynamics. The objectives of this study are: (i) To evaluate the soil layer specific spatial patterns of SOC in a small agricultural catchment and to analyse their relation to spatial patterns of terrain parameters and results from soil redistribution modelling, and (ii) to evaluate if these (easily available) parameters can serve as improving covariables in a layer-specific interpolation of SOC data by regression kriging, and hence potentially allow for a reduction of SOC sampling density without loss of mapping quality.

Materials and Methods

Test site

The test site is part of the Pleiser Hügelland, a hilly landscape located about 30 km in the southeast of Cologne in North Rhine-Westphalia, Germany. It covers a small catchment of approximately 4.2 ha at an altitude of 125-154 m a.s.l. (Fig. 1) which is part of a larger agricultural field (50°43’N, 7°12’E). Slopes range from 1° in the western up to 9° in the eastern part with a relatively flat thalweg area heading to the outlet.

The mean annual air temperature was 10.0 °C and the average precipitation per year was 765 mm (1990-2006) with the highest rainfall intensities occurring from May to October (data from the German Weather Service station Bonn-Roleber, situated about 1 km to the west of the test site, 159 m a.s.l.).

Due to its fertile, loess containing silty and silty-loamy soils classified as Alfisols (USDA, 1999) and its proximity to the agglomeration of Cologne-Bonn the test site is intensively used for arable agriculture. The present crop rotation consists of sugar beet (Beta vulgaris L.), winter wheat (Triticum aestivum L.), and winter barley (Hordeum vulgare L.). Since 1980 a
no-till system was established with mustard (*Sinapis arvensis* L.) cultivated as cover crop after winter barley.

Soil sampling and SOC measurement

In order to investigate the vertical and horizontal distribution of SOC in the test site a first set of soil samples was taken in April 2006. It consisted of 92 soil cores of which 71 cores were situated in a regular 25 x 25 m raster. To account for a possible small scale spatial variability of SOC, additionally a north south transect in the eastern part of the test site with point distances of 12.5 m and two micro-plots consisting of nine sample points each in a 1 x 1 m raster were augered. In each of the micro-plots the central sample point belongs to the regular 25 x 25 m raster. Micro-plots were located in order to cover different slope positions. To densify this first sampling grid, in March 2007 a second set of soil cores (*n* = 65) was taken in a 25 x 25 m raster which was offset by 12.5 m to the north and west in relation to the 2006 raster. Additionally, three samples were taken near the outlet of the test site to account for a small colluvial area. Thus, soil samples exist on a regular 17.7 x 17.7 m raster with a density of 37.9 samples per ha (Fig. 1), with additional samples along the transect and in the micro-plots. Within each sampling campaign soil cores were extracted with a Pyrckhauer soil auger (approximately 2 cm diameter) and soil samples were taken in three depths (I: 0-0.25 m, II: 0.25-0.5 m, and III: 0.5-0.9 m). All sampling points were surveyed with a dGPS (differential Global Positioning System) with a horizontal accuracy between 0.5 and 2 m.

After oven drying at 105°C for 24 hours the samples were ground and coarse particles were seperated by 2 mm-sieving. Recognisable undecomposed organic matter particles were removed. Total C content was determined by dry combustion using a CNS elementar analyser (vario EL, Elementar, Germany). Although loess soils in the area are in most cases deeply decalcified all soil samples were checked for lime (CaCO₃) with hydrochloric acid (10 %). If any inorganic C content was recognised, its amount was determined according to the
Scheibler method (Deutsches Institut für Normung, 1996). Combining both methods if necessary, soil organic carbon (SOC) was calculated from total minus inorganic carbon.

Calculation of terrain parameters and spatial patterns of soil redistribution

Three types of parameters possibly affecting the spatial distribution of SOC were calculated:

(i) primary terrain attributes, (ii) secondary indices combining different primary terrain attributes and representing landscape processes more explicitly, and (iii) parameters representing soil redistribution patterns based on water and tillage erosion modelling. The derivation of these parameters was based on a digital elevation model (DEM) with a 6.25 x 6.25 m² grid. The DEM was derived from laser scanner data (2–3 m point distance) provided by the Landesvermessungsamt North Rhine-Westfalia using ordinary kriging within the Geostatistical Analyst of the Geographical Information System ArcGis 9.2 (ESRI Inc., USA). The grid size of 6.25 m² was chosen to assure that each sampling point is located in the centre of a grid cell.

The following primary terrain attributes were calculated using ArcGis 9.2: The relative elevation (RE), which is the vertical distance of every grid cell to the outlet of the catchment, the slope S, the aspect A, and the curvature. The curvature is the second derivative of the surface and is separated into profile curvature (C-prof; curvature in the direction of maximum slope) and plan curvature (C-plan; curvature perpendicular to the direction of maximum slope). Another primary terrain attribute used in this study is the catchment area CA calculated for each grid cell using the extension HydroTools 1.0 for ArcView 3.x (Schäuble, 2004) applying the multiple flow algorithm of Quinn et al. (1991). The catchment area takes into consideration the amount of surface water that is distributed towards each grid cell. The parameter thus is related to soil moisture and infiltration as well as erosion and deposition.

The two combined indices, wetness index (WI) and stream power index (SPI), differentiate between these two process groups more explicitly through the incorporation of the local slope
gradient. The wetness index (WI) characterises the distribution of zones of surface saturation
and soil water content in landscapes (Beven and Kirkby, 1979) and is calculated as:
\[ WI = \ln \frac{SCA}{\tan S} \]  
(1)

where SCA is the specific catchment or contributing area (m$^2$ m$^{-1}$) orthogonal to the flow
direction and is calculated as the catchment area CA divided by the grid size (6.25 m) and S is
the slope (°).

The stream power index (SPI) is the product of the specific catchment area SCA (m$^2$ m$^{-1}$) and
slope S (°) (Moore et al., 1993). It is directly proportional to stream power and can thus be
interpreted as the erosion disposition of overland flow.

\[ SPI = SCA \cdot \tan S \]  
(2)

One deficit of the SPI is that deposition is not represented. In order to more precisely consider
soil redistribution processes, namely water ($E_{\text{wat}}$), tillage ($E_{\text{til}}$) and total ($E_{\text{tot}}$) erosion and
deposition, corresponding patterns were calculated applying the long-term soil erosion and
sediment delivery model WaTEM/SEDEM version 2.1.0 (Van Oost et al., 2000; Van
Rompaey et al., 2001; Verstraeten et al., 2002).

WaTEM/SEDEM is a spatially distributed model combining WaTEM (Water and Tillage
Erosion Model) (Van Oost et al., 2000) and SEDEM (Sediment Delivery Model) (Van
Rompaey et al., 2001). WaTEM consists of a water and a tillage erosion component that can
be run separately. The water erosion component uses an adapted version of the revised
Universal Soil Loss Equation (RUSLE, Renard et al., 1996). Adaptations consist of the
substitution of slope length with the unit contributing area calculated following Desmet and
Govers (1996) and the integration of sedimentation following an approach of Govers et al.
(1993). Tillage erosion is caused by variations in tillage translocations over a landscape and
always results in a net soil displacement in the downslope direction. The net downslope flux
$Q_{\text{til}}$ (kg m$^{-1}$ a$^{-1}$) due to tillage implementations on a hillslope of infinitesimal length and unit
width is calculated with a diffusion-type equation adopted from Govers et al. (1994) and is proportional to the local slope gradient:

\[ Q_{til} = k_{til} \cdot S = -k_{til} \frac{dH}{dx} \]  

(3)

where \( k_{til} \) is the tillage transport coefficient (kg m\(^{-1}\) a\(^{-1}\)), \( S \) is the local slope gradient (\%), \( H \) is the height at a given point of the hillslope (m) and \( x \) the distance in horizontal direction (m).

The local erosion or deposition rate \( E_{til} \) (kg m\(^{-2}\) a\(^{-1}\)) is then calculated as:

\[ E_{til} = -\frac{dQ_{til}}{dx} = -\frac{d^2H}{d^2x} \]  

(4)

As tillage erosion is controlled by the change of the slope gradient and not by the slope gradient itself, erosion takes place on convexities and soil is accumulated in concavities. The intensity of the process is determined by the constant \( k_{til} \) that ranges between 500 and 1000 kg m\(^{-1}\) a\(^{-1}\) in western Europe (Van Oost et al., 2000).

A second module of WaTEM/SEDEM is the calculation of sediment transport and sedimentation. The sediment flow pattern is calculated with a multiple flow algorithm. The sediment is routed along this flow pattern towards the river taking into account its possible deposition. Deposition is controlled by transport capacity computed for each grid cell. The transport capacity is the maximal amount of sediment that can pass through a grid cell and is assumed to be proportional to the potential rill (and ephemeral gully) erosion volume (Van Rompaey et al., 2001). If the local transport capacity is lower than the sediment flux, deposition is modelled.

WaTEM/SEDEM requires the input of several GIS maps as well as various constants and was implemented as follows: The 6.25 x 6.25 m\(^2\) DEM served as the basis for the calculations. Additionally, a land use map containing field boundaries and a map containing the tillage direction of the test site were derived from digital aerial photographs delivered by the Landesvermessungsamt North Rhine-Westfalia. The K factor of the RUSLE was also given as a map with values of 0.058 and 0.061 kg h m\(^{-2}\) N\(^{-1}\) in the test site. This map was deduced from...
a digital soil map (scaled 1:50000) provided by the Geological Survey of North Rhine-Westfalia. Accounting for the crop rotation and the implemented soil conservation practice in the test site the C factor was set to 0.05 (Deutsches Institut für Normung, 2005). The R factor of the USLE was calculated with a regression equation between R factor and mean daily summer precipitation developed for North Rhine-Westfalia (Deutsches Institut für Normung, 2005). Therefore precipitation data (1990-2006) of the German Weather Service station Bonn-Roleber were used, resulting in an R factor of 67 N h\(^{-1}\) a\(^{-1}\). Since no sediment yield data for model calibration were available, modelling was first performed on a 20 x 20 m\(^2\) grid, which equals the grid size in earlier, calibrated simulations under similar environmental conditions in the Belgium Loess Belt (Verstraeten et al., 2006). The results of this first simulation were used to recalibrate the transport capacity coefficients to run the model on a 6.25 x 6.25 m\(^2\) grid. All other constants necessary for running WaTEM/SEDEM were set to default, since no absolute but only relative erosion and deposition values were needed.

Statistical and geostatistical analysis

Statistical analysis

For statistical and geostatistical analysis three SOC input grids with different sampling densities were created. To achieve a dense 17.7 x 17.7 m sample raster (R\(_{17}\)) SOC data of the 2006 and 2007 sampling campaign were combined in each soil layer. For the topsoil layer it was assumed that inter-annual differences of sampling date and thus of planted crops, soil management, and climate could lead to differences of SOC concentrations gained from the two sampling campaigns. Thus, after an estimation of normal distribution by skewness coefficients, a Student’s T-test (although not optimal when used with spatially autocorrelated data) was applied to estimate the equality of means of the SOC data of the two sampling years. In the two deeper soil layers these influences were considered negligible. Here the SOC contents of the two sampling dates were simply combined to one data set. The 2006 sampling
points \((n = 92)\) arranged in a 25 m raster \((R_{25})\) served as input data set with a medium density of 16.9 samples/ha for each soil layer. To produce a low density 50 m input raster \((R_{50}; n = 44)\) every second data point of \(R_{25}\) was eliminated resulting in a density of approximately 6 samples/ha. Each raster contained the transect and the two micro-plots to incorporate the short distances in geostatistics.

To test the relation between the spatial patterns of SOC and the spatial patterns of potential covariables, Pearson correlation coefficients were calculated between all parameters and the SOC data for each soil layer and each raster width, respectively. For this correlation analysis the eight additional points of the micro-plots were excluded, since all nine sampling points of a micro-plot are located in one grid cell with one value for the relevant parameter. Parameters significantly \((p < 0.05)\) related to SOC in a soil layer were tested for their potential to improve interpolation results when used as a linear trend in regression kriging.

Geostatistical analysis

Geostatistical methods are based on the theory of regionalised variables (Matheron, 1963). For further information concerning the theoretical background of geostatistics we refer to Isaaks and Srivastava (1989) or Webster and Oliver (2001). The basic assumption is that sample points close to each other are more similar than sample points that are far away from each other. This spatial autocorrelation is quantified in the empirical semivariogram of the sampled data, where the semivariance is plotted as a function of lag distance. For a data set \(z(x_i), i = 1, 2, ...,\) the semivariance \(\gamma\) of a certain lag distance \(h\) is calculated as

\[
\gamma(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} (z(x_i) - z(x_i + h))^2 \quad (5)
\]

with \(n(h)\) being the number of pairs of data points separated by \(h\). To apply this semivariogram in the following interpolation process, known as kriging in geostatistics, a theoretical model has to be fit to the sample variogram.
Ordinary kriging (OK) that only uses the spatial autocorrelation of the target variable can be considered as the basic geostatistical interpolation method. It is a kind of weighted spatial mean, where sample point values \( x_j \) are weighted according to the semivariance as a function of distance to the prediction location \( x_0 \). The weights \( \lambda_i \) are chosen by solving the ordinary kriging system in order to minimise the kriging variance:

\[
\sum_{i=1}^{n} \lambda_i \gamma(x_i, x_j) + \varphi = \gamma(x_i, x_0)
\]

\[
\sum_{i=1}^{n} \lambda_i = 1
\]

where \( \gamma(x_i, x_j) \) is the semivariance between the sampling points \( x_i \) and \( x_j \) and \( \gamma(x_i, x_0) \) is the semivariance between the sampling point \( x_i \) and the target point \( x_0 \) and \( \varphi \) is a Lagrange-multiplier necessary for the minimisation process (Ahmed and DeMarsily, 1987).

The regression kriging used in this study follows regression kriging model C described in Odeh et al. (1995) and accounts for a possible trend in the data combining linear regression with ordinary kriging of the residuals. In a first step a linear regression function of the target variable with the covariables is used to create a spatial prediction of the target variable at the new locations. In a second step ordinary kriging is applied to the residuals of the regression resulting in a spatial prediction of the residuals. Finally, the spatially distributed regression results and the kriged residuals are added to calculate the target variable at all new locations.

As a prerequisite of geostatistics, SOC data in each soil layer and in each raster width should be normally distributed. Following Kerry and Oliver (2007a) this prerequisite can be met in geostatistical analysis if the absolute skewness coefficient (SC) is < 1. Moreover, data with an asymmetry caused by aggregated outliers need not to be transformed if the absolute SC is < 2 (Kerry and Oliver, 2007b). If this was true, SOC data were not transformed. For use in regression kriging the residuals resulting from linear regression with the significantly correlated parameters in each soil layer and in each raster width should also be normally distributed. Skewness coefficients as well as normal Q-Q plots of residuals were analysed. In
case residuals showed strong deviations from normal distribution, the corresponding parameters were transformed to logarithms and linear regression was performed again (subsequently these transformed covariables are indicated by the subscript \( tr \)). For each raster width and for each of the three soil layers SOC was interpolated using OK and RK with the selected parameters as covariables to target points spanning a 6.25 x 6.25 m raster within the test site. For the construction of omnidirectional empirical semivariograms of the original SOC data as well as of the residuals the maximum distance up to which point pairs are included was set to 200 m which is half of the maximum extent of the test site in east-west-direction. Lag increments were set to 10 m. In each approach two theoretical variogram models (exponential and spherical) and three methods for fitting the variogram model to the empirical variogram including ordinary least squares (i.e. equal weights to all semivariances) and two weighted least square methods (weighting by \( n_p = \text{number of pairs} \) and weighting by \( n_p h^{-2} \) with \( h = \text{lag distance} \) ) were applied. To evaluate the various theoretical variograms against the original data and to choose the best model a cross-validation procedure was implemented. In cross-validating each of the original data points is left out one after another and the value at that location is estimated by kriging (OK and RK) with the selected variogram model and the remaining data.

As a measure of spatial dependence the ratio of nugget to sill (%) was calculated reflecting the influence of the random component to the spatial variability. Following Cambardella et al. (1994) nugget-to-sill ratios between 0 and 25 % show that data are highly spatially structured with low nugget variances, whereas ratios between 25 and 75 % indicate moderate spatial dependence. Data with ratios > 75 % are weakly spatially structured with a high proportion of unexplained variability.

Validation
To validate the kriging results and to compare the different geostatistical approaches made with high-density R₁₇ as input grid, cross validation was used, since no independent validation data set for this raster width was available. No sample points should be left out for the creation of a validation data set in order to avoid a loss of information in the input data. When using the reduced sampling grids R₂₅ or R₅₀ as input data, the 2007 sampling points (n = 67) were used for validation and for comparing the different kriging approaches within each raster width.

To evaluate the goodness-of-fit of the various kriging results a set of indices was used. To account for the bias and the precision of the prediction the mean error ME and the root mean square error RMSE were calculated:

\[
ME = \frac{1}{n} \sum_{i=1}^{n} z(x_i) - \hat{z}(x_i) \tag{7}
\]

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (z(x_i) - \hat{z}(x_i))^2} \tag{8}
\]

where n constitutes the number of points in the validation sample or the number of points used for cross-validation, \(z(x_i)\) the observed and \(\hat{z}(x_i)\) the predicted values. The ME should be close to zero for unbiased predictions, and the RMSE should be as small as possible.

Additionally, the model-efficiency coefficient (MEF) by Nash and Sutcliffe (1970) was calculated.

\[
MEF = 1 - \frac{\sum_{i=1}^{n} (z(x_i) - \hat{z}(x_i))^2}{\sum_{i=1}^{n} (z(x_i) - \bar{x})^2} \tag{9}
\]

The MEF is a measure of the mean squared error to the observed variance and ranges between \(-\infty\) and 1. If the value of MEF = 1, the model or interpolation represents a perfect fit. If the error is the same magnitude as the observed variance (MEF = 0), the arithmetic mean \(\bar{x}\) of the observed values can represent the data as good as the interpolation.
The relative improvement RI (%) of prediction precision of RK with the selected covariables compared to OK was derived as:

\[ RI = \frac{\text{RMSE}_{\text{OK}} - \text{RMSE}_{\text{RK}}}{\text{RMSE}_{\text{OK}}} \times 100 \quad (10) \]

where RMSE_{\text{RK}} and RMSE_{\text{OK}} are root mean square errors for a certain regression kriging approach and for ordinary kriging, respectively.

The statistical and geostatistical analysis was carried out with GNU R version 2.6 (R Development Core Team, 2007) and the supplementary geostatistical package gstat (Pebesma, 2004).

Results and discussion

Measured horizontal and vertical SOC distribution

Since Student’s T test clearly showed that the of SOC contents of the two sampling dates in soil layer I belong to the same population, SOC contents in each soil layer were combined to one data set. After merging the data sets SOC values in soil layer I range from 0.68 to 1.67 % kg^{-1}, in soil layer II from 0.13 to 1.19 % kg^{-1} and in soil layer III from 0.04 to 1.18 % kg^{-1} (Tab. 1). Maximum values in all soil layers can be found in the flat area near the outlet of the test site (Fig. 2) indicating accumulation of SOC by depositional processes.

Another small area of relatively high SOC concentrations most pronounced in the two upper layers is located in the upper part near the southern boundary of the test site. We assume that this was caused by a former area of dung storage, but no detailed data to verify or falsify this assumption regarding its location exist. Remarkably, the SOC distribution of the mid soil layer shows more small scale variability than that of the other two soil layers.

In general, a decrease of SOC content and an increase of spatial variability expressed by the coefficient of variation (CV) with increasing soil depth can be observed (Tab. 1). The low spatial variability in soil layer I can be deduced to homogenisation caused by management
practices as well as to high turnover rates of soil organic matter. Skewness coefficients indicate that only the third soil layer was not normally distributed. This non-normality is caused by outliers aggregated in the depositional area near the outlet of the test site (Fig. 2) and was therefore not corrected for further geostatistical analysis.

Terrain parameters and patterns of soil redistribution

The spatial patterns of the calculated terrain and soil redistribution parameters are shown in Fig. 3. They show a considerable spatial variability within the test site, indicating their appropriateness for use in RK. The relative elevation RE has a clear tendency from west to east with a maximum value of 27.4 m at the western boundary of the test site and a minimum of 0 m at the outlet. The slope S shows a more complex pattern: almost two third of the test site (mid to western part) are relatively flat with slopes ranging between 1 and 2°. Steep slopes (up to approximately 9.5°) exist in the eastern part. Incorporated into this easterly part is a very small thalweg area with still higher slopes (3-5°) than the flat westerly part. The spatial distribution of the aspect A indicates the differentiation between a south facing (values > 135°) and a north facing slope (values < 45°) in the east. The flat westerly part is orientated to the east with aspects ranging from approximately 60 to 120°. Profile and plan curvature show a diffuse behaviour in the flat west, whereas the pattern of convexities and concavities in the east corresponds well to the derived slope pattern. The catchment area CA and the two indices WI and SPI are distributed in similar patterns with a concentrated area of high values near the outlet of the test site. Compared to SPI this area is smaller in north-south-direction and more elongated in east-west-direction in the patterns of CA and WI.

Comparing the distributions of tillage and water erosion (E_{til} and E_{wat}) derived from WaTEM/SEDEM clearly shows different spatial patterns of erosion and deposition resulting from these two processes, which agrees well with other studies (Govers et al., 1994; Van Oost et al., 2000). Areas with the steepest slopes have the highest water induced erosion rates
resulting in an aggregated area of high erosion rates with values between \(-1.5\) and \(-5.8\) mm a\(^{-1}\) in the test site. This aggregated area corresponds well to the areas of high values of SPI indicating that these parameters represent similar processes. The rest of the test site is dominated by only slight water induced erosion rates with values between \(-1\) and 0 mm a\(^{-1}\). No water induced deposition is calculated inside the test site, since WaTEM/SEDEM is not capable to model the backwater effect induced by landuse change at the outlet of the test site.

Tillage induced erosion generally occurs on convexities and on the downslope side of field boundaries, whereas deposition occurs on concavities and on the upslope side of field boundaries (Govers et al., 1994; Van Oost et al., 2000). High tillage induced deposition rates with values ranging from 2 up to 15 mm a\(^{-1}\) occur in the thalweg area near the outlet of the test site, whereas highest erosion rates (\(-0.5\) to \(-3\) mm a\(^{-1}\)) occur on the shoulders of the north-south-facing slope in the easterly part. The most pronounced difference between water and tillage erosion patterns can be found along the thalweg: here deposition by tillage counteracts with water induced erosion. The pattern of total erosion (\(E_{\text{tot}}\)) combines the two soil redistribution patterns. Most grid cells experiencing tillage induced deposition in the thalweg area are still depositing sites in the total erosion pattern.

Relation between SOC and secondary parameters

Among the primary terrain attributes, C-prof, C-plan and CA show significant linear relationships to SOC in all soil layers and in all input rasters (Tab. 3). Correlation coefficients with C-prof and CA are always positive, whereas correlations with C-plan are always negative. Additionally, RE shows negative correlations with SOC in soil layer III for all raster widths. The SPI and the soil redistribution patterns based on water and tillage erosion modelling all significantly correlate with SOC in all soil layers and in all raster widths, whereas the WI is only significantly correlated with SOC in the two subsoil layers. Correlations between SOC and \(E_{\text{til}}\) and \(E_{\text{tot}}\), respectively, are positive in each soil layer and in
each raster width indicating an accumulation of SOC in depositional sites and a loss of SOC on eroding sites. In contrast and unexpectedly, the water induced erosion pattern expressed by $E_{\text{wat}}$ and SPI results in a different picture: Here high erosion rates correspond to high SOC concentrations in each soil layer. This might be due to the special characteristics of the test site, where high SOC concentrations in each soil layer exist in an area around the small thalweg near the outlet of the test site comprising some parts with relatively steep slopes. Additionally, the backwater effect induced by landuse change leading to deposition of soil transported by water at the outlet is not accounted for in both parameters.

In general, the linear relationship between SOC and the two indices as well as between SOC and the erosion/deposition patterns increases with increasing soil depth within each raster width. The same is true for the relationship between SOC and CA. This indicates (i) that relief driven processes play a less significant role in the topsoil layer where periodic management practices homogenise soil properties in agricultural areas and (ii) that more process-related terrain attributes such as CA, the two indices WI and SPI, and the patterns of soil redistribution play a more important role in the spatial distribution of soil organic carbon in the deeper soil layers. The correlation between SOC and water erosion ($S_{\text{PI}}$ and $E_{\text{wat}}$) as well as between SOC and tillage erosion ($E_{\text{til}}$) indicates the importance of erosion and deposition in the deeper soil layers. The increasing correlations of SOC with CA and WI with increasing soil depth indicate that also processes affecting soil moisture and infiltration influence the SOC patterns in these soil layers. The WI represents areas where water accumulates, and zones with higher WI values tend to have higher biomass production, lower SOC mineralisation, and higher sediment deposition compared to zones of low WI (Terra et al., 2004).

In some respect our results disagree with other results where correlations between SOC and various primary terrain attributes could be found. Mueller and Pierce (2003) for example derived the highest correlation coefficients between SOC and elevation in three different
raster widths for the topsoil layer. Other authors (e.g. Mueller and Pierce, 2003; Terra et al.,
2004; Takata et al., 2007) also found significant correlations with slope. The positive
correlations to CA and WI correspond with findings of other authors (e.g. Terra et al. 2004;
Sumfleth and Duttmann, 2007).

SOC kriging results

High density SOC input data

In the three soil layers different combinations of theoretical variogram models and weighting
methods performed best for the original SOC data. Theoretical variogram parameters (Tab. 4)
show that the original SOC data of R17 are moderately to highly spatially structured for all
three soil layers with low nugget-to-sill ratios. Ranges are much larger than the raster width
with a maximum value of 216 m for the SOC data in soil layer I, indicating that the sampling
scheme used here accounts for most of the spatial variation of SOC in the three soil layers.
The nugget variances comprising small scale variability as well as measurement errors are
close to zero in all soil layers. Mean errors calculated from cross-validation for OK in each
soil layer are close to zero indicating unbiased predictions (Tab. 4). Root mean square errors
resulting from OK are 0.12, 0.20 and 0.15 % kg kg\(^{-1}\) SOC for soil layers I, II, and III,
respectively, corresponding to approximately 10, 28 and 44 % of the mean SOC values in the
different soil layers (Tab. 1). This indicates a loss of precision with increasing soil depth. In
contrast, model efficiency (MEF) is highest in soil layer III (MEF = 0.53) and lowest in soil
layer II (MEF = 0.23). The SOC maps derived from OK (Fig. 4) represent well the spatial
distributions of SOC in each soil layer which were already visible in the patterns of the
measured SOC values at the sampling points (Fig. 2).

Regarding the theoretical variogram parameters of the residuals resulting from linear
regression with the different significantly related covariables in the three soil layers (Tab. 4),
the same conclusions as for the original SOC data in each soil layer can be drawn. The
residuals are moderately or even highly spatially structured, and ranges are larger than the raster width. The sill of the different residual variograms is reduced compared to the sill of the raw data in all soil layers reflecting the success of regression fitting (Hengl et al., 2004; Terra et al., 2004). Nugget variances are all close to zero.

In all three soil layers the geostatistical interpolation of SOC could be improved incorporating covariables in RK (Tab. 4). For soil layer I this was only one covariable, namely C-prof. For soil layer II C-prof, C-plan, CA, WI, and the three soil redistribution patterns derived from modelling were able to ameliorate interpolation results, and in soil layer III improvements were achieved by using C-plan, CA_tr, SPI_tr, WI_tr, E_til and E_tot as covariables in RK. Mean errors were still close to zero for all kriging approaches in all soil layers indicating unbiased predictions. Due to the high spatial density of the original SOC data relative improvements of the described RK approaches compared to OK were only low to moderate in all three soil layers. In soil layer II and III the integration of the more complex covariables outperformed that of the primary terrain parameters (Tab. 4). In general, spatial distributions resulting from the best RK approach in each soil layer (Fig. 4) are similar to those derived from OK but show more small scale variability.

Medium to low density SOC input data

Although a minimum number of at least 50 better 100-150 sampling points is recommended for geostatistical analysis (Webster and Oliver, 2001), the theoretical semivariogram parameters of the SOC data and the values describing the goodness-of-fit for OK of the two reduced input rasters R_{25} (n = 92) and R_{50} (n = 44) still show reasonable results in each soil layer (Tab. 5 and Tab. 6). As for the high density sampling grid (R_{17}) different combinations of theoretical variogram models and weighting methods performed best for the original SOC data. Nugget-to-sill ratios show that the primary SOC data in the two subsoil layers are highly spatially structured in both rasters, and SOC data in the topsoil are moderately spatially
dependent. This indicates that the low density sampling schemes are still suitable to resolve
the spatial continuity of the original SOC data. For R_{25} the ranges are larger than the raster
width, only in soil layer II in R_{50} this is not the case. But since the short distances formed by
the transect and the two micro-plots were kept in each input raster, we assume that the results
of the R_{50} interpolation are still reasonable. Nugget and sill variances of the original SOC data
tend to be in the same order of magnitude than in R_{17} for each soil layer. Mean errors resulting
from OK are still relatively low indicating unbiasedness, and relations of the root mean square
errors to the mean values of the original SOC data also remain similar compared with the
relations in R_{17} for each soil layer. Model efficiency through OK is 0.23 (R_{25}) and 0.14 (R_{50})
in soil layer I and 0.01 (R_{25}) and 0.15 (R_{50}) in soil layer II. Higher values for OK are again
reached in the deepest soil layer where a MEF of 0.34 (R_{25}) and 0.39 (R_{50}) can be found.

The interpolated SOC distributions resulting from OK with the medium and low input data
sets (Fig. 4) are smoothed compared to those using the high density input data set in each soil
layer. But even with coarse sampling (R_{50}), there is still a pronounced area with high SOC
concentrations in the east in all soil layers. The second region with high SOC values (southern
edge and centre), however, is no longer detectable in the R_{50}-interpolation results for the
deepest soil layer.

As was the case in R_{17}, nugget-to-sill ratios and the ranges of the various residuals for the
different soil layers show a moderate to high spatial structure. Sills are also lower than for the
original SOC data, and nuggets are close to zero.

In contrast to the use of the high resolution sampling grid R_{17} as input data no improvements
compared to OK were achieved in soil layer I by RK when using R_{25} and R_{50} (Tab. 5 and Tab.
6). In soil layer II RK including total erosion improved predictions best with R_{25} and R_{50}
(RI = 8.4 and 6.2%, respectively). Relative improvements in soil layer III were even higher in
the medium and low density raster than for soil layer II. In R_{25} the spatial pattern of tillage
erosion and in $R_{50}$ the wetness index WI performed best in improving RK results (Tab. 5 and 6).

In general, relative improvements in soil layer II and III referring to RK vs. OK were more pronounced in case of medium and low density compared to the high density input data. Moreover, SOC maps produced by RK in soil layer II and III with $R_{25}$ and $R_{50}$ (Fig. 4) show considerably more detail and compare more favourably to the spatial patterns produced with $R_{17}$ input data. Although a direct comparison of the interpolation results with the different input rasters is not possible due to a missing independent data set, it has to be recognised that a reduction of input data density seems to slightly decrease MEF and increase RMSE with decreasing data density.

Except for the high density input data, our results for the topsoil layer are in correspondence with Terra et al. (2004) who found that OK predicted SOC best compared to cokriging, regression kriging and multiple regression for the uppermost 30 cm and for three different densities of input data (8, 32 and 64 samples/ha). In contrast, other authors (e.g. Mueller and Pierce, 2003; Simbahan et al., 2006; Sumfleth and Duttmann, 2008) could improve the prediction of SOC in the topsoil layer by using (relative) elevation and electrical conductivity, respectively, as covariables in regression kriging and/or kriging with external drift. Their studies show that the sampling density plays an important role for improving the performance of geostatistics when incorporating covariables. E.g. Mueller and Pierce (2003) also used three different input rasters in their test site. For their high resolution input raster with a density of 10.7 samples/ha Mueller and Pierce (2003) also found only modest differences between the applied interpolation techniques, but for their two reduced rasters (2.7 and 1 sample/ha) different interpolation methods incorporating covariables could outperform OK. This was also true for the three test sites of Simbahan et al. (2006) with sampling densities of 2.5 to 4.2 samples/ha. Our result of no or only slight improvements in the first soil layer might
be caused by (i) our high sampling densities (37.9, 16.9, and 6 samples/ha), (ii) homogenisation effects of management, and (iii) the area of high SOC concentrations at the southern boundary of the test site which is most pronounced in the topsoil. This area cannot be deduced to relief driven processes, and in combination with homogenisation is thus leading to relatively low correlations between SOC and the various parameters in the topsoil layer.

In contrast, considerable improvements of RK over OK in our study were achieved in the two subsoil layers. In soil layer II this improvement was highest when using the patterns of tillage or total erosion as covariable in RK. This indicates that especially tillage induced erosion and deposition processes affect the SOC distribution in this layer. This makes it necessary to not only consider water induced soil redistribution processes, which are already represented in other primary and secondary terrain attributes (CA and SPI) used here and in other studies. Relative patterns of tillage erosion and deposition can easily be derived with well tested and relatively simple erosion and sediment delivery models like the WaTEM/SEDEM model. To implement the tillage erosion component only a DEM and an estimation of the tillage transport coefficient are required (Van Oost et al., 2000).

Although the tillage and total erosion pattern could also significantly improve SOC prediction in the deepest soil layer in all three raster widths, comparable and in some instances even better results were produced by RK with CA and WI. This indicates that not only soil redistribution processes affect the spatial distribution of SOC in the deepest soil layer but also processes concerning the spatial distribution of infiltration and soil moisture. Both processes may increase SOC contents in the thalweg area due to (i) infiltration and absorption of dissolved organic carbon (DOC) and (ii) limited mineralisation of SOC in case of high soil moisture contents.

In contrast to the topsoil layer in the two subsoil layers improved SOC interpolations can actually be obtained when using a high density of input data for RK. This is possibly caused
by higher spatial variations of SOC in these soil layers, expressed as coefficient of variation (Tab. 1).

Our results indicate that SOC patterns in different soil layers can be linked to different processes. Whereas the topsoil pattern is homogenised by tillage operations, the patterns of the subsoil layers are more pronounced and driven by soil redistribution and moisture/infiltration differences. Patterns of topsoil SOC distribution might be dissimilar to subsoil layers particularly in hilly agriculturally used areas. Thus, estimating total SOC pools from topsoil SOC, for instance by applying remote sensing techniques (e.g. Stevens et al., 2008), might not be appropriate.

Conclusions

Factors and processes affecting the spatial distribution of soil layer specific SOC in a hilly agricultural catchment were analysed, while correlating measured SOC data with primary terrain parameters, combined indices as well as spatial patterns of soil redistribution derived from modelling. In general, Pearson correlation coefficients showed that the linear relationship between SOC and the more process-related indices and erosion/deposition patterns was higher than between SOC and relatively simple terrain parameters. Correlation coefficients increased with increasing soil depth indicating that relief driven processes in the small catchment play a less significant role in the topsoil layer, where periodic agricultural management practices homogenise soil properties.

To produce detailed and precise maps of the SOC distribution in the three soil layers, the performance of OK and RK with the significantly correlating parameters was tested using three input rasters with decreasing sampling density. Results showed that especially in the subsoil layers the geostatistical interpolation of SOC could be improved, when covariables were incorporated. In the mid soil layer (0.25-0.5 m) the best result was produced by RK with the patterns of tillage and total erosion, indicating the importance of soil redistribution
(especially the inclusion of tillage erosion) for the spatial distribution of SOC in agricultural areas. In the third soil layer (0.5-0.9 m) tillage and total erosion as well as the wetness index and partly the catchment area performed best. This indicates that besides soil redistribution also processes concerning the distribution of soil moisture affect the spatial pattern of SOC in the deepest soil layer. In general, relative improvements of RK vs. OK increased with increasing soil depth and with reduced sampling density. Hence, the expectable decline in interpolation quality with decreasing data density can be reduced for the subsoil layers integrating results from soil redistribution modelling and spatial patterns of soil moisture indices as covariables in RK.

In general, it could be shown that (especially) an integration of patterns in soil redistribution in kriging approaches, can substantially improve SOC interpolation of subsoil data in hilly arable landscapes. This is an important finding insofar as high resolution subsoil SOC data are rare and most promising data improvements due to new remote sensing techniques are limited to topsoil SOC.


Govers, G., T.A. Quine and D.E. Walling. 1993. The effect of water erosion and tillage movement on hillslope profile development: a comparison of field observation and


Tab. 1: Statistics of SOC content [% kg kg⁻¹] for the 2006, 2007 and the merged (merg.) dataset in three soil depths (I: 0-0.25 m; II: 0.25-0.5 m; III: 0.5-0.9 m).

<table>
<thead>
<tr>
<th>Soil layer</th>
<th>Data</th>
<th>n</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>CV</th>
<th>Min</th>
<th>Max</th>
<th>SC††</th>
</tr>
</thead>
<tbody>
<tr>
<td>I 2006</td>
<td>92</td>
<td>1.16</td>
<td>1.14</td>
<td>0.18</td>
<td>15.13</td>
<td>0.68</td>
<td>1.68</td>
<td>0.75</td>
<td></td>
</tr>
<tr>
<td>II 2006</td>
<td>92</td>
<td>0.67</td>
<td>0.67</td>
<td>0.22</td>
<td>32.62</td>
<td>0.13</td>
<td>1.19</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td>III 2006</td>
<td>92</td>
<td>0.32</td>
<td>0.24</td>
<td>0.18</td>
<td>62.13</td>
<td>0.05</td>
<td>0.90</td>
<td>1.51</td>
<td></td>
</tr>
<tr>
<td>I 2007</td>
<td>67</td>
<td>1.11</td>
<td>1.08</td>
<td>0.16</td>
<td>12.23</td>
<td>0.85</td>
<td>1.43</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td>II 2007</td>
<td>68</td>
<td>0.75</td>
<td>0.77</td>
<td>0.23</td>
<td>30.44</td>
<td>0.18</td>
<td>1.18</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>III 2007</td>
<td>68</td>
<td>0.36</td>
<td>0.33</td>
<td>0.22</td>
<td>62.55</td>
<td>0.04</td>
<td>1.18</td>
<td>1.57</td>
<td></td>
</tr>
<tr>
<td>I merg.</td>
<td>159</td>
<td>1.14</td>
<td>1.12</td>
<td>0.17</td>
<td>14.91</td>
<td>0.68</td>
<td>1.68</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>II merg.</td>
<td>160</td>
<td>0.71</td>
<td>0.71</td>
<td>0.22</td>
<td>30.99</td>
<td>0.13</td>
<td>1.19</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>III merg</td>
<td>160</td>
<td>0.34</td>
<td>0.27</td>
<td>0.21</td>
<td>62.61</td>
<td>0.04</td>
<td>1.18</td>
<td>1.50</td>
<td></td>
</tr>
</tbody>
</table>

† n: number of sample points
‡ SD: standard deviation
§ CV: coefficient of variation
¶ Min: Minimum
# Max: Maximum
†† SC: skewness coefficient
Tab. 2: Statistics of terrain attributes and soil redistribution parameters within the test site (n = 1030); considered are: relative elevation RE, slope S, aspect A, profile and plan curvature (C-prof and C-plan), catchment area CA, wetness index WI, stream power index SPI, and patterns of tillage ($E_{\text{til}}$), water ($E_{\text{wat}}$) and total ($E_{\text{tot}}$) erosion.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Median</th>
<th>SD†</th>
<th>CV‡</th>
<th>Min§</th>
<th>Max¶</th>
<th>SC††</th>
</tr>
</thead>
<tbody>
<tr>
<td>RE [m]</td>
<td>15.82</td>
<td>16.44</td>
<td>5.97</td>
<td>----</td>
<td>0.00</td>
<td>27.42</td>
<td>-0.31</td>
</tr>
<tr>
<td>S [°]</td>
<td>3.93</td>
<td>3.21</td>
<td>1.87</td>
<td>47.58</td>
<td>1.56</td>
<td>9.46</td>
<td>1.05</td>
</tr>
<tr>
<td>A [°]</td>
<td>87.33</td>
<td>78.82</td>
<td>28.17</td>
<td>40.37</td>
<td>166.38</td>
<td>1.23</td>
<td></td>
</tr>
<tr>
<td>C-prof [0.01 m]</td>
<td>-0.03</td>
<td>-0.02</td>
<td>0.20</td>
<td>---</td>
<td>-1.08</td>
<td>0.95</td>
<td>-0.10</td>
</tr>
<tr>
<td>C-plan [0.01 m]</td>
<td>-0.04</td>
<td>-0.01</td>
<td>0.26</td>
<td>---</td>
<td>-1.27</td>
<td>0.72</td>
<td>-1.05</td>
</tr>
<tr>
<td>CA [m²]</td>
<td>1568.91</td>
<td>868.15</td>
<td>2634.86</td>
<td>167.94</td>
<td>105.17</td>
<td>25612.13</td>
<td>5.45</td>
</tr>
<tr>
<td>WI</td>
<td>7.79</td>
<td>7.90</td>
<td>0.90</td>
<td>---</td>
<td>5.73</td>
<td>11.00</td>
<td>-0.03</td>
</tr>
<tr>
<td>SPI</td>
<td>20.67</td>
<td>7.15</td>
<td>41.19</td>
<td>199.27</td>
<td>0.68</td>
<td>325.99</td>
<td>4.28</td>
</tr>
<tr>
<td>$E_{\text{til}}$ [mm a⁻¹]</td>
<td>0.02</td>
<td>-0.16</td>
<td>1.49</td>
<td>---</td>
<td>-5.28</td>
<td>15.00</td>
<td>2.96</td>
</tr>
<tr>
<td>$E_{\text{wat}}$ [mm a⁻¹]</td>
<td>-0.50</td>
<td>-0.23</td>
<td>0.66</td>
<td>---</td>
<td>-5.81</td>
<td>-0.02</td>
<td>-3.58</td>
</tr>
<tr>
<td>$E_{\text{tot}}$ [mm a⁻¹]</td>
<td>-0.47</td>
<td>-0.44</td>
<td>1.30</td>
<td>---</td>
<td>-5.39</td>
<td>13.19</td>
<td>1.82</td>
</tr>
</tbody>
</table>

† SD: standard deviation
‡ CV: coefficient of variation. The CV cannot be calculated for variables containing negative values or possessing a negative skewness coefficient (Isaaks and Srivastava, 1989).
§ Min: Minimum
¶ Max: Maximum
†† SC: skewness coefficient
Tab. 3: Quality of correlation between SOC content [% kg kg\(^{-1}\)] and all calculated parameters in the three soil layers (I: 0-0.25 m; II: 0.25-0.5 m; III: 0.5-0.9 m) expressed as Pearson correlation coefficients; results are given for the three different raster widths (R\(_{17}\), R\(_{25}\), R\(_{50}\)) used as input for geostatistics; for abbreviations of parameters refer to Tab. 2.

<table>
<thead>
<tr>
<th></th>
<th>SOC (R_{17}) ((n_1 = 143), (n_{II,III} = 144))</th>
<th>SOC (R_{25}) ((n = 76))</th>
<th>SOC (R_{50}) ((n = 28))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>II</td>
<td>III</td>
</tr>
<tr>
<td>RE [m]</td>
<td>-0.04</td>
<td>-0.03</td>
<td>-0.28**</td>
</tr>
<tr>
<td>S [°]</td>
<td>0.13</td>
<td>0.03</td>
<td>0.14</td>
</tr>
<tr>
<td>A [°]</td>
<td>0.12</td>
<td>-0.01</td>
<td>0.08</td>
</tr>
<tr>
<td>C-prof [0.01 m]</td>
<td>0.37**</td>
<td>0.44**</td>
<td>0.39**</td>
</tr>
<tr>
<td>C-plan [0.01 m]</td>
<td>-0.28**</td>
<td>-0.36**</td>
<td>-0.56**</td>
</tr>
<tr>
<td>CA [m²]</td>
<td>0.19*</td>
<td>0.27**</td>
<td>0.67**</td>
</tr>
<tr>
<td>WI</td>
<td>0.14</td>
<td>0.35**</td>
<td>0.53**</td>
</tr>
<tr>
<td>SPI</td>
<td>0.25**</td>
<td>0.29**</td>
<td>0.67**</td>
</tr>
<tr>
<td>(E_{sat}) [mm a(^{-1})]</td>
<td>0.36**</td>
<td>0.45**</td>
<td>0.67**</td>
</tr>
<tr>
<td>(E_{wat}) [mm a(^{-1})]</td>
<td>-0.22**</td>
<td>-0.25**</td>
<td>-0.53**</td>
</tr>
<tr>
<td>(E_{tot}) [mm a(^{-1})]</td>
<td>0.33**</td>
<td>0.42**</td>
<td>0.55**</td>
</tr>
</tbody>
</table>

* Significant at 95 %, ** significant at 99 %.
Tab. 4: Theoretical semivariogram parameters of original SOC data and residuals resulting from linear regression with different covariables as well as cross-validation results from ordinary (OK) and regression kriging (RK) of SOC content [% kg kg⁻¹] in three soil layers (I: 0-0.25 m; II: 0.25-0.5 m; III: 0.5-0.9 m) using the 17.7 m raster data set (R₁₇) (ν₁ = 159; νᵢ = 160); RK results are included only when improving the prediction compared to OK; no covariable indicates OK; for exponential models the practical range is given; goodness-of-fit was tested using mean error (ME), root mean square error (RMSE), model efficiency (MEF), and relative improvement (RI). The transcript tr means that covariables were transformed to logarithms so that linear regression residuals meet normal distribution.

<table>
<thead>
<tr>
<th>Soil layer</th>
<th>Covariable</th>
<th>Theoretical semivariogram parameters</th>
<th>Kriging results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Model</td>
<td>Weights</td>
</tr>
<tr>
<td>I</td>
<td>---</td>
<td>exponential</td>
<td>equal</td>
</tr>
<tr>
<td></td>
<td>C-prof</td>
<td>exponential</td>
<td>equal</td>
</tr>
<tr>
<td>II</td>
<td>---</td>
<td>exponential</td>
<td>h⁻²</td>
</tr>
<tr>
<td></td>
<td>C-prof</td>
<td>exponential</td>
<td>h⁻²</td>
</tr>
<tr>
<td></td>
<td>C-plan</td>
<td>exponential</td>
<td>h⁻²</td>
</tr>
<tr>
<td></td>
<td>CA</td>
<td>exponential</td>
<td>h⁻²</td>
</tr>
<tr>
<td></td>
<td>WI</td>
<td>exponential</td>
<td>h⁻²</td>
</tr>
<tr>
<td></td>
<td>E_R</td>
<td>exponential</td>
<td>h⁻²</td>
</tr>
<tr>
<td></td>
<td>E_max</td>
<td>exponential</td>
<td>h⁻²</td>
</tr>
<tr>
<td>III</td>
<td>---</td>
<td>spherical</td>
<td>nᵢ</td>
</tr>
<tr>
<td></td>
<td>C-plan</td>
<td>spherical</td>
<td>nᵢ</td>
</tr>
<tr>
<td></td>
<td>CAₚ</td>
<td>spherical</td>
<td>nᵢ</td>
</tr>
<tr>
<td></td>
<td>WIₚ</td>
<td>spherical</td>
<td>nᵢ</td>
</tr>
<tr>
<td></td>
<td>SPIₚ</td>
<td>spherical</td>
<td>nᵢ</td>
</tr>
<tr>
<td></td>
<td>E_R</td>
<td>exponential</td>
<td>equal</td>
</tr>
<tr>
<td></td>
<td>E_max</td>
<td>spherical</td>
<td>nᵢ</td>
</tr>
</tbody>
</table>

¹ For abbreviations of covariables refer to Tab. 2
² Weighting of the semivariogram model is done by ordinary least squares (i.e. equal weights to all semivariances) and two weighted least square methods (weighting by nᵢ = number of pairs and weighting by nᵢh⁻² with h = lag distance [m]).
Tab. 5: Theoretical semivariogram parameters of original SOC data and residuals resulting from linear regression with different covariables as well as results from ordinary (OK) and regression kriging (RK) with SOC content [% kg kg\(^{-1}\)] in three soil layers (I: 0-0.25 m; II: 0.25-0.5 m; III: 0.5-0.9 m) using the 25 m raster data set \(R_{25}\) \((n = 92)\); the values describing the goodness-of-fit result from the comparison with a validation data set \((n = 67)\); RK results are included only when improving the prediction compared to OK; no covariable indicates OK; for exponential models the practical range is given; goodness-of-fit was tested using mean error (ME), root mean square error (RMSE), model efficiency (MEF), and relative improvement (RI). The transcript tr means that covariables were transformed to logarithms so that linear regression residuals meet normal distribution.

<table>
<thead>
<tr>
<th>Soil layer</th>
<th>Covariable</th>
<th>Theoretical semivariogram parameters</th>
<th>Kriging results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Model</td>
<td>Weights(^1)</td>
</tr>
<tr>
<td>I</td>
<td>---</td>
<td>exponential</td>
<td>equal</td>
</tr>
<tr>
<td>II</td>
<td>---</td>
<td>exponential</td>
<td>(n_p h^{-2})</td>
</tr>
<tr>
<td></td>
<td>WI</td>
<td>exponential</td>
<td>(n_p h^{-2})</td>
</tr>
<tr>
<td></td>
<td>E$_{sil}$</td>
<td>spherical</td>
<td>(n_p)</td>
</tr>
<tr>
<td></td>
<td>E$_{sat}$</td>
<td>exponential</td>
<td>(n_p h^{-2})</td>
</tr>
<tr>
<td></td>
<td>E$_{tot}$</td>
<td>exponential</td>
<td>(n_p h^{-2})</td>
</tr>
<tr>
<td>III</td>
<td>---</td>
<td>spherical</td>
<td>equal</td>
</tr>
<tr>
<td></td>
<td>C-plan</td>
<td>spherical</td>
<td>(n_p h^{-2})</td>
</tr>
<tr>
<td></td>
<td>CA$_{tr}$</td>
<td>spherical</td>
<td>equal</td>
</tr>
<tr>
<td></td>
<td>WI$_{tr}$</td>
<td>spherical</td>
<td>equal</td>
</tr>
<tr>
<td></td>
<td>SPI$_{tr}$</td>
<td>spherical</td>
<td>equal</td>
</tr>
<tr>
<td></td>
<td>E$_{sil}$</td>
<td>exponential</td>
<td>(n_p h^{-2})</td>
</tr>
<tr>
<td></td>
<td>E$_{sat}$</td>
<td>spherical</td>
<td>equal</td>
</tr>
<tr>
<td></td>
<td>E$_{tot}$</td>
<td>spherical</td>
<td>equal</td>
</tr>
</tbody>
</table>

\(^1\) For abbreviations of covariables refer to Tab. 2

\(^2\) Weighting of the semivariogram model is done by ordinary least squares (i.e. equal weights to all semivariances) and two weighted least square methods (weighting by \(n_p\) = number of pairs and weighting by \(n_p h^{-2}\) with \(h =\) lag distance [m]).
Tab. 6: Theoretical semivariogram parameters of original SOC data and residuals resulting from linear regression with different covariables as well as results from ordinary (OK) and regression kriging (RK) of SOC content [\% kg kg\(^{-1}\)] in three soil layers (I: 0-0.25 m; II: 0.25-0.5 m; III: 0.5-0.9 m) using the 50 m raster data set (R\(_{50}\)) (n = 44); the values describing the goodness-of-fit result from the comparison with a validation data set (n = 67); RK results are included only when improving the prediction compared to OK; no covariable indicates OK; for exponential models the practical range is given; goodness-of-fit was tested using mean error (ME), root mean square error (RMSE), model efficiency (MEF), and relative improvement (RI). The transcript tr means that covariables were transformed to logarithms so that linear regression residuals meet normal distribution.

<table>
<thead>
<tr>
<th>Soil layer</th>
<th>Covariables</th>
<th>Theoretical semivariogram parameters</th>
<th>Kriging results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Model</td>
<td>Weights(^{\dagger})</td>
</tr>
<tr>
<td>I</td>
<td>---</td>
<td>spherical</td>
<td>(n_p)</td>
</tr>
<tr>
<td>II</td>
<td>---</td>
<td>exponential</td>
<td>(n_p h^2)</td>
</tr>
<tr>
<td></td>
<td>C-prof</td>
<td>exponential</td>
<td>(n_p h^2)</td>
</tr>
<tr>
<td></td>
<td>E(_{nil})</td>
<td>spherical</td>
<td>(n_p)</td>
</tr>
<tr>
<td></td>
<td>E(_{tot})</td>
<td>spherical</td>
<td>(n_p)</td>
</tr>
<tr>
<td>III</td>
<td>---</td>
<td>spherical</td>
<td>(n_p)</td>
</tr>
<tr>
<td></td>
<td>W1</td>
<td>spherical</td>
<td>(n_p)</td>
</tr>
<tr>
<td></td>
<td>E(_{nil})</td>
<td>spherical</td>
<td>(n_p)</td>
</tr>
<tr>
<td></td>
<td>E(_{tot})</td>
<td>spherical</td>
<td>equal</td>
</tr>
</tbody>
</table>

\(^{\dagger}\) For abbreviations of covariables refer to Tab. 2

\(^{\dagger}\) Weighting of the semivariogram model is done by ordinary least squares (i.e. equal weights to all semivariances) and two weighted least square methods (weighting by \(n_p\) = number of pairs and weighting by \(n_p h^2\) with \(h\) = lag distance [m]).
Fig. 1: Test site with location of soil sampling points; each of the two micro-plots (MP) consists of nine sample points arranged in a 1 x 1 m grid; flow direction is from east to west.

Fig. 2: Measured SOC contents [% kg kg\(^{-1}\)] at the 17.7 x 17.7 m raster sampling points for soil layers I (0-0.25 m), II (0.25-0.5 m), and III (0.5-0.9 m).

Fig. 3: Maps of terrain attributes and patterns of soil redistribution derived from WaTEM/SEDEM; for abbreviations of parameters refer to Tab. 2.; a positive curvature (C-prof and C-plan) indicates that the surface is upwardly convex, and a negative value indicates that the surface is upwardly concave; regarding the erosion patterns (E\(_{\text{til}}\), E\(_{\text{wat}}\), and E\(_{\text{tot}}\)) negative values represent erosion, while positive ones represent deposition.

Fig. 4: Maps of SOC content [% kg kg\(^{-1}\)] for three soil layers (I: 0-0.25 m; II: 0.25-0.5 m; III: 0.5-0.9 m) resulting from ordinary (OK) and the best regression kriging (RK) approach using three different raster widths \(R_{17}\) (17.7 x 17.7 m), \(R_{25}\) (25 x 25 m) and \(R_{50}\) (50 x 50 m) as input; covariables of the RK approaches are given above each map; for abbreviations of covariables refer to Tab. 2. The transcript tr means that covariables were transformed to logarithms so that linear regression residuals meet normal distribution.
- sampling points 2006
- sampling points 2007
- 1 m contour lines