Plant Soil (2011) 340:7–24
DOI 10.1007/s11104-010-0425-z

REGULAR ARTICLE

Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem

Martin Wiesmeier · Frauke Barthold · Benjamin Blank · Ingrid Kögel-Knabner

Received: 29 October 2009 / Accepted: 5 May 2010 / Published online: 30 May 2010

Abstract Spatial prediction of soil organic matter is a global challenge and of particular importance for regions with intensive land use and where availability of soil data is limited. This study evaluated a Digital Soil Mapping (DSM) approach to model the spatial distribution of stocks of soil organic carbon (SOC), total carbon (C_{tot}), total nitrogen (N_{tot}) and total sulphur (S_{tot}) for a data-sparse, semi-arid catchment in Inner Mongolia, Northern China. Random Forest (RF) was used as a new modeling tool for soil properties and Classification and Regression Trees (CART) as an additional method for the analysis of variable importance. At 120 locations soil profiles to 1 m depth were analyzed for soil texture, SOC, C_{tot}, N_{tot}, S_{tot}, bulk density (BD) and pH. On the basis of a digital elevation model, the catchment was divided into pixels of 90 m×90 m and for each cell, predictor variables were determined: land use unit, Reference Soil Group (RSG), geological unit and 12 topography-related variables. Prediction maps showed that the highest amounts of SOC, C_{tot}, N_{tot} and S_{tot} stocks are stored under marshland, steppes and mountain meadows. River-like structures of very high elemental stocks in valleys within the steppes are partly responsible for the high amounts of SOC for grasslands (81–84% of total catchment stocks). Analysis of variable importance showed that land use, RSG and geology are the most important variables influencing SOC storage. Prediction accuracy of the RF modeling and the generated maps was acceptable and explained variances of 42 to 62% and 66 to 75%, respectively. A decline of up to 70% in elemental stocks was calculated after conversion of steppe to arable land confirming the risk of rapid soil degradation if steppes are cultivated. Thus their suitability for agricultural use is limited.

Keywords Classification and Regression Trees (CART) · Soil organic carbon (SOC) · China · Grassland

Responsible Editor: Elizabeth (Liz) A. Stockdale.

M. Wiesmeier (✉) · I. Kögel-Knabner
Lehrstuhl für Bodenkunde,
Department für Ökologie und Ökosystemmanagement,
Wissenschaftszentrum Weihenstephan für Ernährung,
Landnutzung und Umwelt,
Technische Universität München,
85350 Freising-Weihenstephan, Germany
e-mail: wiesmeier@wzw.tum.de

F. Barthold · B. Blank
Institute for Landscape Ecology and Resources
Management, Justus-Liebig-University Giessen,
Heinrich-Buff-Ring 26,
35392 Giessen, Germany

Introduction

Detailed knowledge about the storage of soil organic matter (SOM) in soils is essential in the light of rising

demand for agricultural land, ongoing soil degradation and desertification and requirements for sequestration of atmospheric CO_2. Despite numerous local and global soil inventories (e.g. Batjes 1996), providing a global, high-resolution map of functional soil properties remains a challenge for soil scientists (Lal 2009; Sanchez et al. 2009). This is particularly true in emerging nations like China, where land use has intensified rapidly and cultivation is expanding in ecologically sensitive regions, e.g. semi-arid grasslands. Grasslands play a crucial role in the storage of SOM at a global scale, containing approximately 15% of total soil organic carbon (SOC) stocks (Anderson 1991; Lal 2004). However, land use changes, particularly overgrazing, have caused a considerable loss of SOC in semi-arid grasslands of Northern China (Xie et al. 2007; Zhou et al. 2007).

Digital Soil Mapping (DSM) provides a widely accepted framework to map the spatial patterns of soil properties (McBratney et al. 2000; Scull et al. 2003). The basic assumption of DSM is that soil development is a function of climate, organisms, topography, parent material and time (Jenny 1941). Hence, soil properties of a location can theoretically be estimated if information about those variables is available for the location. McBratney et al. (2003) introduced further variables space and soil information derived from other investigations, as soil can be predicted from its own properties in the so-called "scorpan model".

The distribution and storage of SOC has been estimated for China and partially for the semi-arid region in Northern China in several studies, but sampling density was generally low with one observation per 256–3195 km^2 (Ni 2002; Song et al. 2005; Wang et al. 2002; Wu et al. 2003; Yu et al. 2007). Few investigations have considered the spatial distribution of SOC in this ecosystem at a regional scale (Liu et al. 2006; Wei et al. 2008). In addition, recent studies have revealed a high spatial variability of soil properties at the plant and field scale in the steppes of Northern China (Steffens et al. 2009a; Wiesmeier et al. 2009). For these extensive grassland regions, a soil mapping approach is required that estimates soil properties satisfactorily, even when collection of soil data is limited.

Grunwald (2009) reviewed various DSM approaches, which are designed to correlate environmental variables derived from investigations, field sampling, remote sensing and digital elevation models (DEM) quantitatively with soil properties. Among those methods are Classification and Regression Trees (CART) and Random Forest (RF). Classification and Regression Tree analysis has been widely used for the spatial prediction of soil properties (Barthold et al. 2008; Henderson et al. 2005; McKenzie and Ryan 1999; Stoorvogel et al. 2009; Vasques et al. 2008) as well as in a broad range of other ecological sciences (De'ath and Fabricius 2000). Random Forest is a relatively new method that was developed as an extension of CART in order to improve the prediction accuracy (Breiman 2001; Liaw and Wiener 2002). Until now, it has been applied primarily in remote sensing studies (Gislason et al. 2006; Lawrence et al. 2006), but has also been used in ecology (Peters et al. 2008; Prasad et al. 2006) and genetics (Wu et al. 2009). In the field of soil science RF has been applied once for the spatial prediction of SOC concentrations and stocks of a region in Panama (Grimm et al. 2008). Prasad et al. (2006) combined CART techniques to predict species distributions under current and future climate scenarios using four different models: Regression Tree Analysis (RTA), Bagging Trees (BT), RF and Multivariate Adaptive Regression Splines (MARS). They concluded that RF has superior predictive capabilities whereas RTA, and to some degree BT, provided interpretive results.

We have used of the benefits of a combined model procedure as described in Prasad et al. (2006) and applied RF together with CART as a new tool for the prediction of SOC, total carbon (C_{tot}), total nitrogen (N_{tot}) and total sulphur (S_{tot}) stocks, subsequently referred to as elemental stocks, for a catchment in the semi-arid steppe region of Northern China. In the study area, intensification of grazing in recent decades has led to severe degradation of the grasslands (Steffens et al. 2008; Wiesmeier et al. 2009). The objectives of this study are:

(1) To evaluate a combined RF-CART approach as a method for the spatial prediction of elemental stocks in a data sparse region of Inner Mongolia.
(2) To identify the environmental variables controlling the spatial distribution of SOM in a semi-arid steppe region.
(3) To estimate elemental stocks in a semi-arid catchment and the human impact on SOM storage.

Materials and methods

Study area

The study area is a 3600 km^2 section of the Xilin River basin (43°24′ to 44°40′ N and 115°20′ to 117° 13′ E) located in the Xilingol steppe in the autonomous province of Inner Mongolia, China, approximately 450 km north of Beijing (Fig. 1). The Chinese Academy of Sciences maintains the Inner Mongolia Grassland Ecosystem Research Station (IMGERS) within the study area and located about 60 km southeast of the city of Xilinhot. Elevation in the area is between 1010 and 1609 m above sea level. The region is part of the continental semi-arid grasslands of the Central Asian steppe ecosystem, with a dry and cold mid-latitude climate (Kawamura et al. 2005). Mean annual temperature at IMGERS is 0.7°C and mean annual precipitation is around 350 mm, with the highest values in the summer from June to August. The vegetative growth period from May to September is relatively short (<150 days). The grasslands are typically used for semi-nomadic or static grazing with sheep and goats.

Sampling design

For field sampling, a design-based, stratified two-stage sampling scheme was developed with land use and topography as stratifying variables (Brus and deGruijter 1997; McKenzie and Ryan 1999). Classification of land use was based on a Landsat 7 Thematic Mapper image of August 17th, 2005. Six land use units were classified in two consecutive steps: (1) a supervised classification resulted in the delineation of four main ecological units (sand dunes, mountain meadow, marshland and steppe) and (2) a further refinement was achieved via unsupervised classification where a cluster analysis of bands one to five and seven lead to a delineation of 11 land use types. For this study, the 11 classes were lumped into six classes: (1) arable land, (2) bare soil, (3) marshland/water, (4) mountain meadow, (5) steppe and (6) sand dunes (Fig. 2, Table 1).

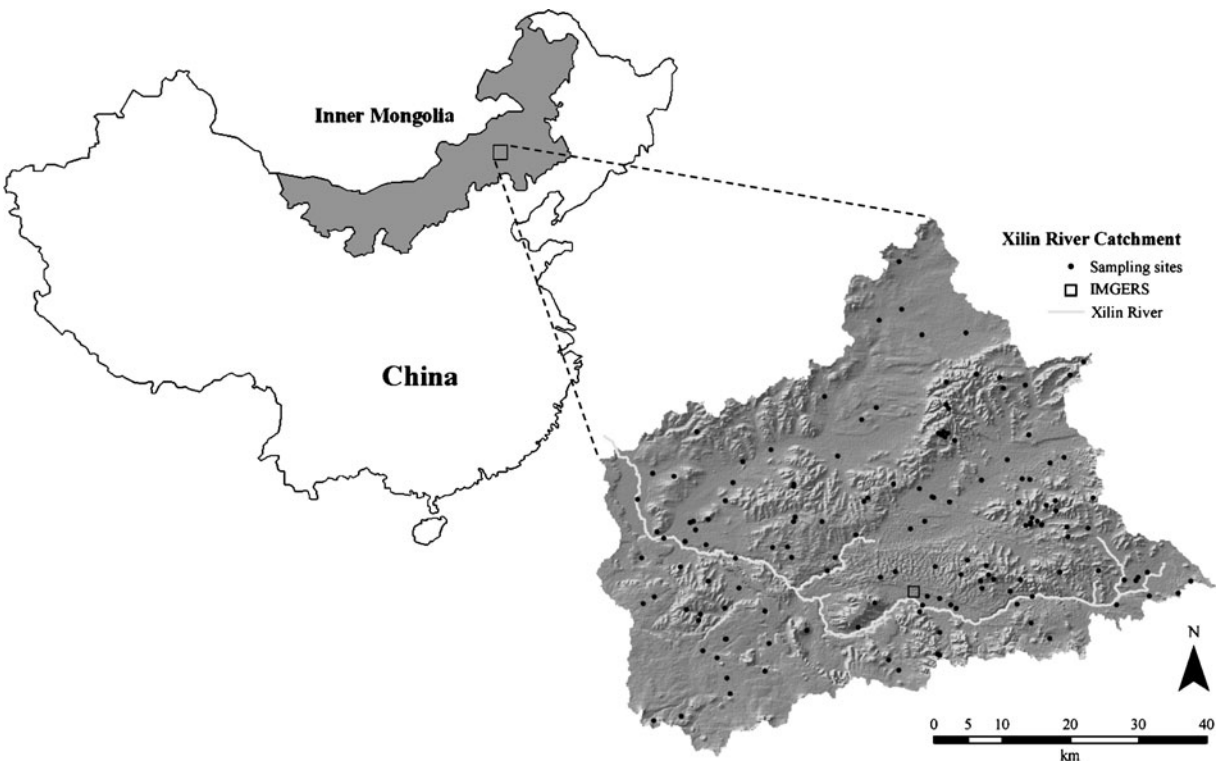

Fig. 1 Map of the Xilin River Catchment in Inner Mongolia, Northern China, with the location of the sampling sites and the Inner Mongolia Grassland Experimental Research Station (IMGERS)

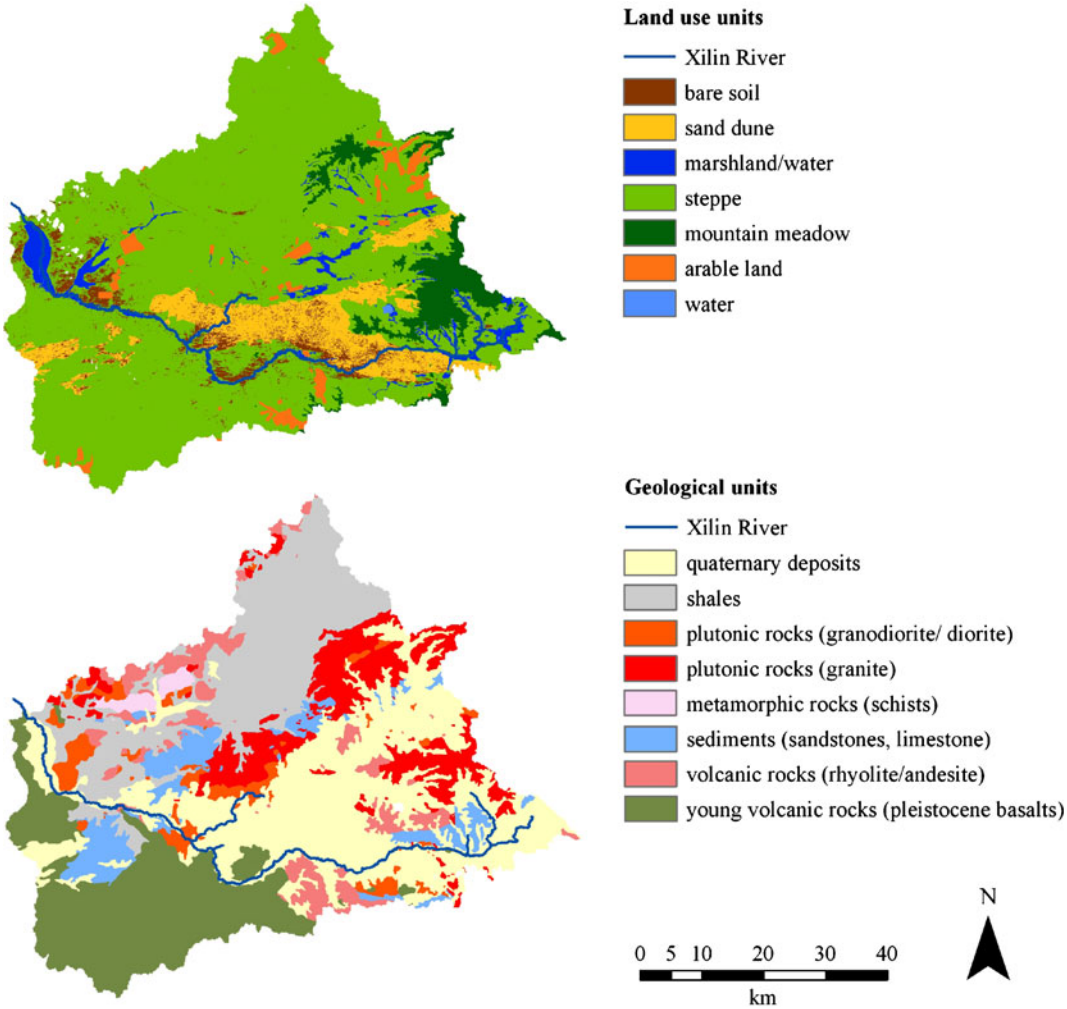

Fig. 2 Land use units and geological units of the Xilin River catchment

The Topographic Wetness Index (TWI; Sorensen et al. 2006) was classified into three equal area classes derived from a Digital Elevation Model (DEM) that was provided by the NASA Shuttle Radar Topography Mission (SRTM; Tile 60_04, data version 4.1; Jarvis et al. 2008). In combination the six land use classes and the three TWI classes resulted in 18 distinct environments, which were randomly sampled at 120 locations (Fig. 1). In each distinct environment, sampling locations were randomly determined. Due to large differences in the proportions of the land use units, the number of sampling locations for smaller land use units was reduced (Table 1) and the consequent sampling density is one sample per 30 km^2 of the catchment area.

At each location each horizon of the soil profile to 1 m depth was described according to the Food and Agriculture Organization's guidelines (FAO 2006). Undisturbed soil samples were taken from each horizon using steel cylinders with a volume of 100 cm^3 and these were analyzed for soil texture, SOC, C_{tot}, N_{tot}, S_{tot}, bulk density (BD) and pH.

Determination of soil properties

For the analyses of texture, soil samples (<2 mm) were oxidized with H_2O_2 to remove organic matter. The remaining material was dispersed with $Na_4P_2O_7$ and shaken for 16 to 24 h, followed by wet sieving to isolate sand fractions >63 μm. To determine silt and clay fractions, approximately 3 g of the <63 μm

Table 1 General information, Reference Soil Groups (IUSS Working Group WRB 2006) and mean values of chemical and physical soil properties of the land use units and the total catchment derived from 120 sampling locations (± standard deviation)

land use unit	arable land	bare soil	marshland/water	mountain meadow	steppe	sand dunes	total catchment
abbreviation	al	bs	mw	mm	s	sd	tc
area (km^2)	99	180	170	203	2605	351	3608
proportion (%)	3	5	5	6	72	10	100
sampling locations	13	12	9	26	38	22	120
Reference Soil Group (n)	*Chernozem* (9)	*Arenosol* (7)	*Arenosol* (1)	*Calcisol* (1)	*Calcisol* (2)	*Arenosol* (22)	*Arenosol* (30)
	Kastanozem (2)	*Cambisol* (3)	*Cryosol* (1)	*Chernozem* (4)	*Cambisol* (1)		*Calcisol* (3)
	Phaeozem (2)	*Chernozem* (1)	*Gleysol* (6)	*Phaeozem* (21)	*Chernozem*(13)		*Cambisol* (4)
		Kastanozem (1)	*Phaeozem* (1)		*Kastanozem* (2)		*Chernozem*(27)
					Phaeozem (20)		*Cryosol* (1)
							Gleysol (6)
							Kastanozem (5)
							Phaeozem (44)
depth[a] (cm)	76±24	48±23	43±22	73±30	70±21	30±20	60±29
SOC[b] ($kg\ m^{-2}$)	11.3±3.8	5.4±2.4	35.1±18.6	26.5±14.8	14.2±5.3	5.3±2.8	15.6±13.0
C_{tot}[b] ($kg\ m^{-2}$)	14.6±6.2	6.0±2.8	38.8±21.2	27.8±14.7	17.1±7.3	5.3±2.8	17.5±14.1
N_{tot}[b] ($kg\ m^{-2}$)	1.2±0.4	0.6±0.3	3.2±1.8	2.4±1.3	1.5±0.5	0.5±0.3	1.5±1.2
S_{tot}[b] ($kg\ m^{-2}$)	0.23±0.04	0.15±0.05	0.85±0.60	0.34±0.14	0.27±0.13	0.14±0.05	0.29±0.25
SOC ($mg\ g^{-1}$)/N_{tot} ($mg\ g^{-1}$)	9.4±1.1	9.2±1.3	10.2±1.4	10.9±0.9	9.6±0.8	10.0±1.3	9.9±1.2
BD[a] ($g\ cm^{-3}$)	1.33±0.19	1.53±0.10	1.07±0.20	1.20±0.21	1.39±0.10	1.50±0.23	1.35±0.22
pH[a] ($CaCl_2$)	6.5±0.5	6.2±0.2	6.7±0.4	6.0±0.3	6.2±0.3	6.1±0.2	6.2±0.4
sand[a] (%)	59±16	84±10	85±7	35±17	72±15	88±9	71±22
silt[a] (%)	27±11	10±6	10±5	44±13	17±8	8±6	18±15
clay[a] (%)	14±5	6±4	5±2	21±5	11±8	4±3	10±8

[a] A horizon [b] 1 m depth

fraction was suspended in deionized water with $Na_4P_2O_7$ and an ultrasonication for 3 min with 75 J ml^{-1} was conducted. Afterwards, the distribution of silt and clay fractions was obtained by measuring the X-ray absorption of the soil-water suspension during sedimentation of the soil particles with a Micromeritics Sedigraph 5100 (Micromeritics, Norcross, USA; Spörlein et al. 2004).

Bulk density was quantified from the mass of the oven-dry soil (105°C) divided by the core volume (Hartge and Horn 1989). Soil pH values were measured in 0.01 M $CaCl_2$ at a soil/solution ratio of 1-to−2.5 at room temperature. Total carbon, N_{tot} and S_{tot} were determined in duplicate by dry combustion on a Vario Max CNS elemental analyser (Elementar, Hanau, Germany). The measured C_{tot} concentrations of the samples that were free of carbonate represent SOC concentrations. Samples that contained $CaCO_3$ were heated to 500°C for 4 h to remove organic carbon and the concentration of inorganic C of the residual material was determined by dry combustion. The SOC content was calculated by subtracting the content of inorganic C from the C_{tot} concentration of the untreated material. Stocks of C_{tot}, SOC, N_{tot} and S_{tot} for each horizon were calculated using the respective elemental concentrations and bulk densities:

$$ES = EC \times BD \times (1 - vR) \times T \times 10^{-2} \quad (1)$$

ES is the elemental stock (kg m^{-2}), EC is the elemental concentration (mg g^{-1}), BD is the bulk density (g cm^{-3}), vR is the volumetric fraction of rock fragments >2 mm (%) and T is the thickness of the horizon (cm). The elemental stocks of all horizons from a location were summed to obtain comparable stocks to 1 m depth.

Environmental predictor variables

Land use units

Six land use units were characterized as described above (Fig. 2, Table 1) and were further characterized by field surveys as follows:

(1) arable land (al): cultivated areas are located predominantly in the vicinity of villages, single farms or streets and are cultivated with rapeseed, potato, maize or wheat. Both small cultivated sites without irrigation or fertilization as well as large areas of intensive vegetable production with irrigation systems can be found.

(2) bare soil (bs): degraded areas with marginal vegetation are primarily found along the Xilin River and around villages where the grazing intensity is high and the access to water is unrestricted. Eroded areas are also found in the surroundings of sand dunes and along country roads in the steppe areas.

(3) marshland/water (mw): this land use type stretches along the Xilin River system and its branches as well as in episodic water-bearing valleys and groundwater dominated areas. The marshland can either be continuously affected by water or just occasionally flooded. Typical plant species are *Phragmites australis* and *Carex appendiculata*.

(4) mountain meadow (mm): in higher elevation areas in the east and northeast of the basin, the typical steppe merges gradually to mountain meadows. Due to their remote location, these meadows are only slightly grazed and contain a number of characteristic plant species, e.g. *Agrostis gigantean*, *Carex pediformis* and *Stipa baicalensis*.

(5) steppe (s): there are two main grassland types: *Leymus chinensis* dominated steppe communities at areas with relatively wet soil conditions and *Stipa grandis* dominated communities in drier regions (Chen et al. 2005; Tong et al. 2004). Degraded steppe areas are indicated by the occurrence of *Cleistogenes squarrosa* and *Artemisia frigida*. These grasslands are almost totally used for livestock grazing.

(6) sand dune (sd): a broad belt of sand dunes stretches from east to west through the catchment. On north-facing slopes, dense vegetation can be found that comprises several tree genera such as *Ulmus*, *Betulus*, *Malus*, *Prunus* and *Populus.* In contrast, south-facing slopes are only slightly covered with shrubs and grassland species.

Geological units

A geological map at the scale of 1:200,000 from the Inner Mongolian Bureau of Geology was digitized

and georeferenced. The original 29 map units were summarized into nine geological units with similar ages or formation processes (Fig. 2). Shales from the Tertiary and Cretaceous period dominate the north and northwest of the basin. Older plutonic rocks including granite, granodiorite and diorite from the Jurassic and Carboniferous periods can be found in the center of the catchment and form the mountains in the east and northeast. Smaller areas of sedimentary rocks including sandstones, conglomerates and limestone from the Permian and Carboniferous periods stretch from west to east through the catchment. In the northwest, there are small regions of metamorphic rocks (schists). Volcanic rocks (rhyolite, andesite) from the Jurassic and Permian periods are located in the northwest and southeast. A wide plateau of Quaternary basaltic lavas covers the southwest of the basin. The geological unit with the largest extension is comprised of Quaternary deposits (loess) that are distributed in the river valley. Further information on the geology of the study area is presented by Steffens (2009).

Reference Soil Groups

At 145 locations, soils were sampled by horizon to 1 m depth and classified according to the World Reference Base (WRB) for Soil Resources (IUSS Working Group WRB 2006) (Table 1). Reference Soil Groups (RSG) were then mapped for the entire catchment using a DSM approach that included RF and CART as statistical models to infer relationships between RSGs and land use, geology and topography. In general, 9 RSGs were identified; qualifiers, which further specify the soil classification at individual locations, were not incorporated in order to obtain a consistent soil map. The dominant soils in the study area are the typical steppe soils *Phaeozems* and *Chernozems*. *Phaeozems* are the most extensive soils covering 51% of the entire catchment. They primarily occur in mountain regions in the east and northeast as well as in other elevated areas throughout the basin. *Chernozems* can generally be found in the transition zones between *Gleysols* and *Phaeozems* in lower elevated areas. They occupy with 15% of the catchment approximately as much area as the *Gleysols* (14%). *Gleysols* occur in close vicinity of the Xilin River system, in valleys and in groundwater-dominated areas. *Kastanozems* cover 1% of the catchment and are mainly located at the basalt plateau in the southwest. *Arenosols* account for 16% of the study area and their occurrence is almost congruent with sand dunes. However, they can also be found in arable land. *Calcisols* and *Regosols* also take over some parts of the area but cover only small patches (0.3 and 0.2%, respectively). They are distributed northwards and southwards of the river in the west of the basin. Very small sites of *Cryosols* (0.1%) can be found in marshlands of the Xilin River. *Cambisols* occur only at a few locations and do not cover a significant area of the catchment.

Topography

A 90 m resolution Digital Elevation Model (DEM) was provided from the NASA Shuttle Radar Topography Mission (SRTM; Jarvis et al. 2008) and used to calculate a set of 12 primary and secondary terrain attributes (Wilson and Gallant 2000). Primary terrain attributes are elevation, slope, aspect, profile curvature (profcurv), plan curvature (plancurv) and mean curvature (meancurv). Secondary terrain attributes are total upslope length (tlen), longest upslope length (plen), contributing area (ca), topographic wetness index (twi), transport capacity index (tci) and stream power index (spi).

Digital Soil Mapping (DSM)

A combined RF-CART approach was applied to identify the environmental variables that control the spatial patterns of SOM, quantify their relationships and spatially predict elemental stocks. Random Forest was used for the generation of the prediction maps and CART was applied as an additional, supportive tool to improve the interpretability of the involved variables. The underlying assumption for the applicability of this combined model procedure is that the differences between the CART tree, i.e. the splitting rules, and the trees in RF are small.

Classification and Regression Trees (CART)

Classification and Regression Tree analysis is a non-parametric data mining technique that can handle non-linear and non-additive relationships because it uses recursive partitioning of the dataset to explore relationships between a response variable and predictor

variables (Breiman et al. 1984; De'Ath 2002; Myles et al. 2004). The response as well as the predictor variables are either categorical (classification trees) or numeric (regression trees). The dataset of the response variable is split in a tree like manner into successively smaller groups on the basis of the predictor variable that maximizes the homogeneity of the groups (De'ath and Fabricius 2000). The splits are based on threshold values of the predictor variables that decrease node impurity optimally. Each of the generated leafs represents a final group with a distinct categorical or numeric prediction for the response variable. The most advantageous feature of CART is that the results are easy to interpret. Major disadvantages are that there is a high sensitivity to the selection of the dataset with respect to the resulting tree structure and that the data is often overfitted (Breiman 2001). Regression trees were generated for SOC, C_{tot}, N_{tot} and S_{tot}. The trees indicate the importance of the predictor variables and associated threshold values for the spatial prediction of elemental stocks and help to gain insight into the splitting process.

Random Forest (RF)

Random Forest is an ensemble method that was developed as an extension of CART to improve the prediction performance of the model (Breiman 2001). The model building process is the same as in CART with the difference that many trees are built resulting in a "forest of models". For each tree, only a subset of the predictor variables is used. The number of predictors to be used in each tree-building process (m_{try}) and the number of trees to be built in the forest (n_{tree}) can be varied depending on the data set. Each tree is built from a bootstrap sample of the original data set which allows for robust error estimation with the remaining test set, the so-called Out-Of-Bag (OOB) sample. The excluded OOB samples are predicted from the bootstrap samples and by aggregating the OOB predictions from all trees the Mean Square Error (MSE_{OOB}) is calculated (Liaw and Wiener 2002):

$$\mathrm{MSE_{OOB}} = n^{-1}\sum_{i=1}^{n}\left(z_i - \hat{z}_i^{OOB}\right)^2 \qquad (2)$$

where $\hat{z}_i^{OOB}$ is the average of all OOB predictions. The MSE_{OOB} is normalized as it depends on the unit of the response variable and the percentage of explained variance (Var_{ex}) is calculated:

$$\mathrm{Var_{ex}} = 1 - \frac{MSE_{OOB}}{Var_z} \qquad (3)$$

where Var_z is the total variance of the response variable. The result of a random forest is one single prediction which is of the average over all aggregations. The advantages that arise from this procedure are (a) higher prediction performance, (b) no overfitting, (c) low correlation of individual trees since the diversity of the forest is increased through the usage of a limited number of predictors, (d) low bias and low variance due to averaging over a large number of trees and (e) robust error estimates by using the OOB data. One major disadvantage of random forest is its "black box" nature that prohibits easy interpretation of the relationships between the response and predictor variables, since it is impractical to investigate the structure of all trees in the forest (Prasad et al. 2006). In order to overcome this shortcoming, the procedure allows estimation of the variable importance measured by the mean decrease in prediction accuracy before and after permuting a variable.

Combined DSM approach

To take advantage of both modeling approaches, we used RF to establish relationships between the environmental covariates and the soil properties in order to spatially predict elemental stocks. The variable importance measures and the spatial visualizations of the RF model helped to get a first impression of the controlling variables. CART was applied to further quantify the relationships between the environmental covariates and the elemental stocks in detail. We used the "rpart" package of R (RDevelopment Core Team 2008) for generating regression trees and the "randomForest" package of R (Liaw and Wiener 2002) for all RF computations. These software packages not only allow the computations of the models but also provide built-in functions to spatially predict the variable of interest based on the RF model.

Spatial prediction of SOM

For the spatial prediction of elemental stocks of the Xilin River catchment, 40 sampling points were

randomly excluded to provide an independent validation data set. Basic descriptive statistics were used to compare the validation set with remaining 80 sampling points that is used as the training set to build the RF models. For each tree building process, five randomly selected predictor variables ($m_{try}=5$) were used and a total of 1000 trees ($m_{tree}=1000$) was calculated (Liaw and Wiener 2002). With this approach, maps for SOC, C_{tot}, N_{tot} and S_{tot} were generated with a resolution of 90 m×90 m pixels. The accuracy of the maps was evaluated with the validation set in order to obtain an independent error estimate. For the 40 validation points, differences between measured and predicted elemental stocks were calculated with the Mean Error (ME) and the Root Mean Square Error (RMSE):

$$\mathrm{ME} = n^{-1}\sum_{i=1}^{n}\left(z_i - z_i^{*}\right) \tag{4}$$

$$\mathrm{RMSE} = \sqrt{n^{-1}\sum_{i=1}^{n}\left(z_i - z_i^{*}\right)^2} \tag{5}$$

where z_i^{*} is the predicted elemental stock. The ME indicates if the prediction is biased whereas the RMSE measures the accuracy of the predictions. Additionally, the MSE and the percentage of explained variance (according to Eqs. 2, 3 and 4) as well as the coefficient of correlation (R^2) were calculated for the validation set.

To assess the degree to which land use impacted the spatial distribution of SOM, stocks of SOC, C_{tot}, N_{tot} and S_{tot} were calculated for every land use unit by spatially overlaying prediction maps with the land use map in ArcGIS (ESRI, Redlands, USA). Plots of the variable importance measures together with the prediction maps and the CART-derived regression trees allowed an interpretation of the importance of the predictor variables for the spatial distribution of SOM stocks in the Xilin River basin.

Results

Soil properties

The classification of the studied soils on the RSG level, mean values for the depth of the A horizons and mean values for chemical and physical soil properties for all land use units are given in Table 1. The thicknesses of the A horizons from steppes, mountain meadows and arable land were comparable with depths of 70–76 cm. Bare soil, marshland/water and sand dunes had considerably shallower A horizons of 30–48 cm. Generally, marshland/water showed the highest elemental stocks for 1 m depth with amounts of 35.1 kg m^{-2} for SOC, 38.8 kg m^{-2} for C_{tot}, 3.2 kg m^{-2} for N_{tot} and 0.85 kg m^{-2} for S_{tot}. Mountain meadows had 23–28% lower values for SOC, C_{tot} and N_{tot} and 60% lower S_{tot} stocks. Steppe regions showed 52–68% lower elemental values compared to marshland/water. Slightly lower amounts (61–73%) were determined for arable land. Bare soil and sand dunes showed the lowest amounts for all land use units with up to 5 to 7 times lower stocks of SOC, C_{tot}, N_{tot} and S_{tot} compared to marshland/water. Ratios of SOC concentrations to N_{tot} concentrations were in a range of 9.2–10.9 and showed no considerable differences between land use units. A negative correlation was found between all determined elemental concentrations and BD$\left(R^2 = 0.42 - 0.44\right)$ from all land use units at all horizons. Values for BD from the topsoil (A horizons) were lowest for mountain meadows with 1.07 gcm^{-3} and highest for bare soil with 1.53 gcm^{-3}. pH values between 6.0 and 6.7 were measured for the A horizons of all land use units. The soil texture of A horizons was comparable among bare soil, marshland/water, steppe and sand dunes with high sand contents of 72–88%, silt contents of 8–17% and clay contents of 4–11%. Arable land and mountain meadows showed considerably lower contents of sand (59% and 35% respectively), and accordingly higher contents for silt and clay.

RF model performance and validation of the prediction maps

The performances of the RF models indicated by MSE_{OOB} and percentages of explained variance are shown in Table 2. The MSE_{OOB} showed values between 0.05 and 104.1 for the investigated elemental stocks. Explained variances ranged between 42 and 62% for the RF modeling. Descriptive statistics revealed similar ranges, mean values and standard deviations for the validation and the training set (Table 3). The ME ranged between −0.02 and −2.09 and was negative for all elemental stocks (Table 4).

Table 2 Mean Square Error of the Out-Of-Bag sample (MSE_{OOB}) and percentage of the explained variance (Var_{ex}) for the RF prediction of elemental stocks

	SOC (kg m^{-2})	C_{tot} (kg m^{-2})	N_{tot} (kg m^{-2})	S_{tot} (kg m^{-2})
MSE_{OOB}	72.64	104.13	0.64	0.049
Var_{ex}	61.9	53.4	57.0	42.4

Table 4 Mean Error (ME), Mean Square Error (MSE), Root Mean Square Error (RMSE), percentage of explained variance (Var_{ex}) and coefficient of variation (R^2) of the validation set

	ME	MSE	RMSE	Var_{ex}	R^2
SOC (kg m^{-2})	−2.09	29.86	5.46	70.3	0.74
C_{tot} (kg m^{-2})	−2.16	32.25	5.68	68.7	0.75
N_{tot} (kg m^{-2})	−0.14	0.21	0.46	74.7	0.78
S_{tot} (kg m^{-2})	−0.02	0.008	0.09	65.6	0.67

For RMSE, values between 0.09 and 5.68 were calculated. Mean Square Errors of the validation set were considerably lower compared to MSE_{OOB} with values between 0.01 and 32.3. Calculated percentages of explained variances were generally higher and ranged between 65.6 and 74.7% for the prediction maps. Additionally, high R^2 values of 0.67 to 0.78 indicated a strong correlation between measured and predicted elemental stocks.

Importance of predictor variables

Importance measures of the investigated predictor variables by RF indicated that land use unit is the most important variable in explaining the spatial distributions of SOC, C_{tot}, N_{tot} and S_{tot} (Fig. 3). Less important but still of major dominance are RSGs, while the geological unit is third most important variable that impacts the spatial pattern of SOM. The importance of the remaining 12 environmental covariates differs largely between the different elemental stocks.

Using CART, land use unit is the variable upon which the first split in the decision trees of all analyzed parameters is based (Fig. 4). This first split divides arable land, bare soil, mountain meadows and sand dunes from marshland/water and steppe, except for S_{tot}, where marshland/water alone is separated from all other land use units. The second split is again based on land use for C_{tot} and N_{tot}. However, for SOC and S_{tot}, RSGs were responsible for further splitting up the data sets. For the trees of SOC and S_{tot}, RSG separated the *Arenosols*, *Cambisols* and *Kastanozems* from *Calcisols*, *Chernozems*, *Cryosols*, *Gleysols* and *Phaeozems*. For C_{tot} and N_{tot}, arable land and sand dunes were separated from bare soil and mountain meadows. Further splits were based upon the contributing area in all trees except for $S_{tot.}$ At the third split of the left branches, geological unit was the predictor variable for C_{tot} and S_{tot} and contributing area for SOC and N_{tot}. In terms of geology, Quaternary deposits, young volcanic rocks, volcanic rocks, sediments and for S_{tot} additionally granodiorite/diorite were separated from shales, metamorphic rocks and granite.

Spatial prediction of elemental stocks under different land uses

The RF modeling resulted in prediction maps for stocks of SOC, C_{tot}, N_{tot} and S_{tot} for the studied catchment (Fig. 5), and for each land use unit mean values and total amounts of elemental stocks, as well as proportions of the total catchment, were calculated (Table 5). Marshland/water showed the highest accumulation of SOM with stocks of 33.5 kg m^{-2} for SOC, 37.4 kg m^{-2} for C_{tot}, 3.2 kg m^{-2} for N_{tot} and

Table 3 Descriptive statistics including Minimum (Min), Maximum (Max), Mean values (Mean) and Standard deviation (SD) of the training and the validation data set

	Training set (n=80)				Validation set (n=40)			
	Min	Max	Mean	SD	Min	Max	Mean	SD
SOC (kg m^{-2})	0.15	52.76	16.13	13.40	1.99	58.57	14.60	12.24
C_{tot} (kg m^{-2})	0.15	70.20	18.42	14.75	1.99	58.57	15.70	12.53
N_{tot} (kg m^{-2})	0.02	4.99	1.53	1.15	0.23	6.24	1.43	1.20
S_{tot} (kg m^{-2})	0.00	1.52	0.28	0.23	0.04	1.98	0.29	0.30

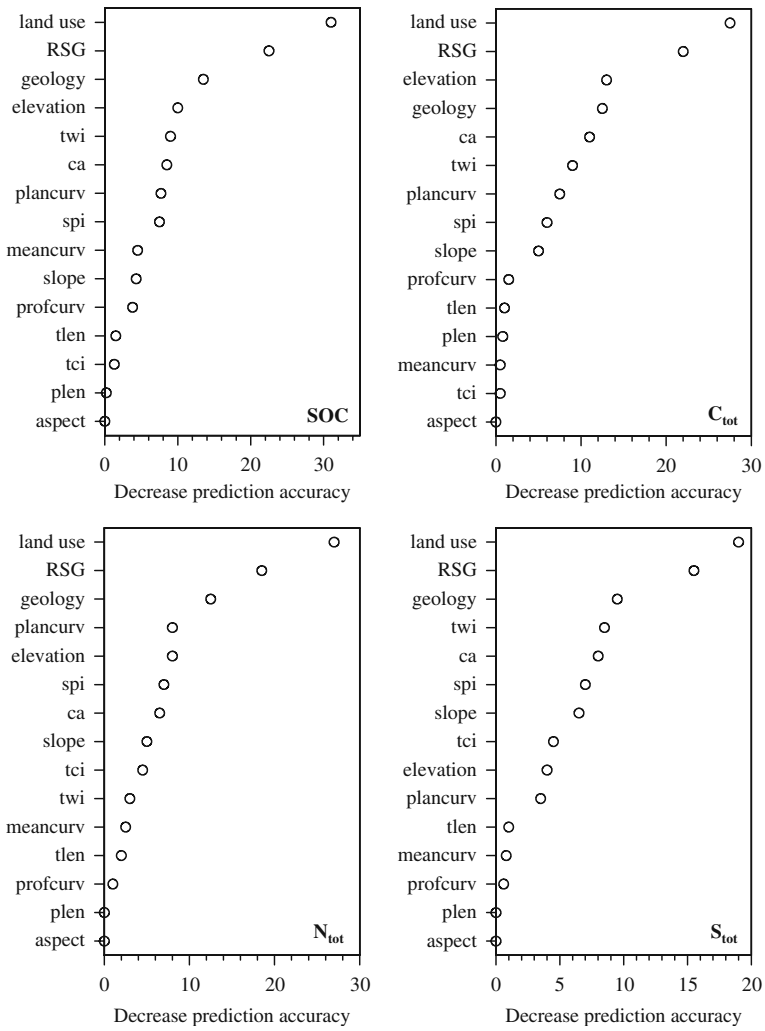

Fig. 3 Variable importances derived from RF models for SOC, C_{tot}, N_{tot} and S_{tot} (RSG = Reference Soil Group, profcurv = profile curvature, plancurv = plan curvature, meancurv = mean curvature, tlen = total upslope length, plen = longest upslope length, ca = contributing area, twi = topographic wetness index, tci = transport capacity index, spi = stream power index)

0.6 kg m^{-2} for S_{tot}. However, this land use unit only accounted for 6–7% of total elemental stocks for the basin due to its small proportion of the total catchment area. For steppes, also high average stocks were calculated with 27.3 kg m^{-2} for SOC, 29.9 kg m^{-2} for C_{tot}, 2.5 kg m^{-2} for N_{tot} and 0.6 kg m^{-2} for S_{tot}. As steppe is the dominating vegetation type in the basin, it accumulates 81–84% of total stocks. With 3–5% of total stocks, mountain meadows accumulated less SOM compared to marshland/water although their spatial abundance is slightly higher. The lowest amounts of SOM were found for land use classes in the order bare soil > arable land > sand dunes with values of 7.2–14.1 kg m^{-2} for SOC, 7.8–16.7 kg m^{-2} for C_{tot}, 0.7–1.5 kg m^{-2} for N_{tot} and 0.2–0.3 kg m^{-2} for S_{tot}. These three land use units together take over 18% of the total catchment area but only account for 1–3% of the total elemental stocks of the basin.

Based on the proportion of the land use units, soils in the Xilin River catchment accumulate on average 24.1 kg m^{-2} of SOC, 26.5 kg m^{-2} of C_{tot}, 2.2 kg m^{-2} of N_{tot} and 0.5 kg m^{-2} of S_{tot}. Totally, 86.8 Tg of SOC, 95.7 Tg of C_{tot}, 8.0 Tg of N_{tot} and 1.9 Tg of S_{tot} are stored in the soils of the studied catchment.

Discussion

Evaluation of the DSM approach

The combined RF-CART approach provided a promising framework for the spatial prediction of soil

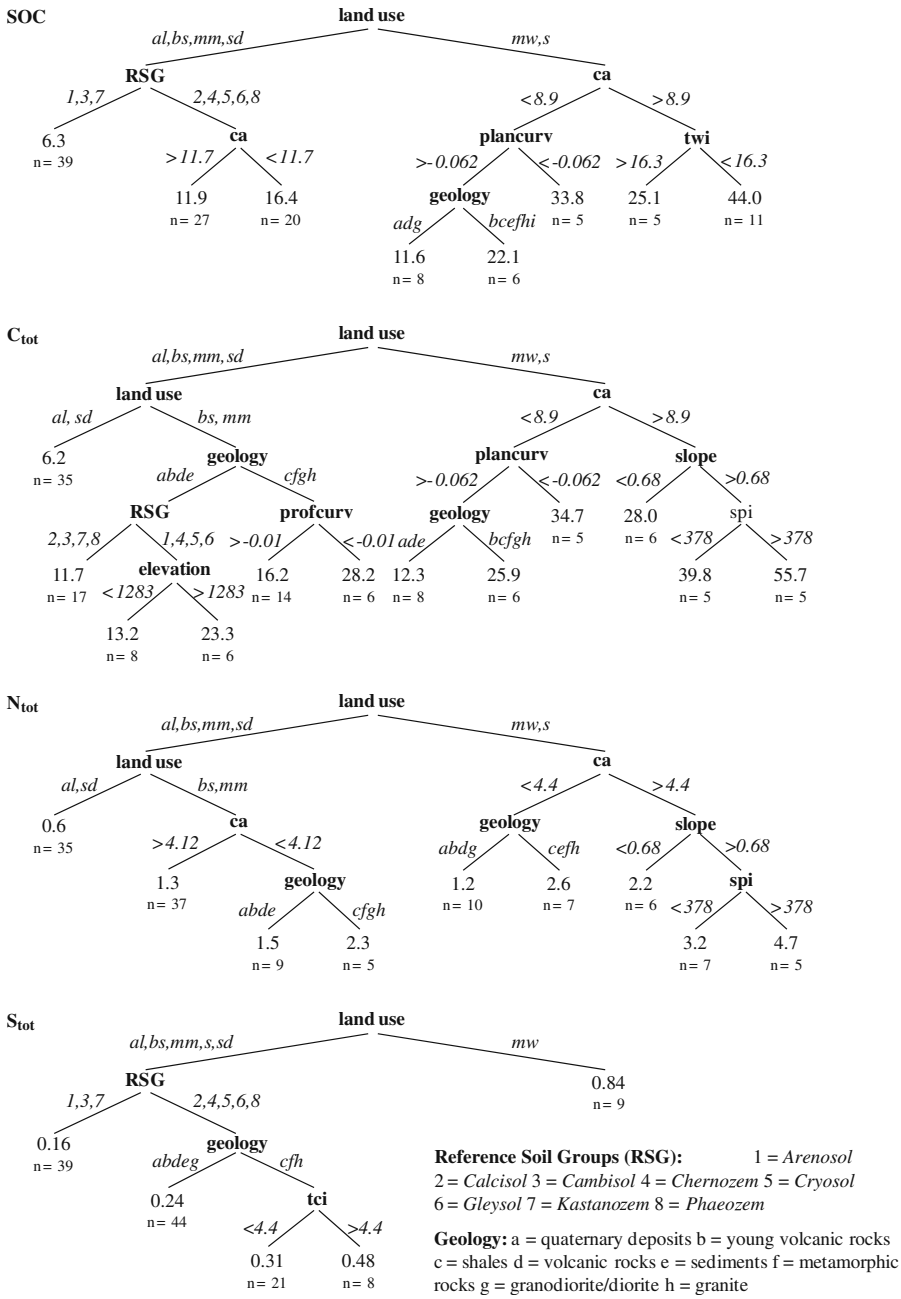

Fig. 4 Regression trees for SOC, C_{tot}, N_{tot} and S_{tot} stocks (kg m^{-2}) (Predictor variables are bold, threshold values are italic; units of predictor variables: plancurv, profcurv (radians m^{-1}), slope (°), elevation (m), ca (m^2), spi, tci and twi non-dimensional indices)

properties as the prediction accuracy of the model performance was acceptable with explained variances of 42–62% for SOC, C_{tot}, N_{tot} and S_{tot} in the model building process (Table 2). For the spatial prediction of SOC concentrations and stocks of Barro Colorado Island in Panama, Grimm et al. (2008) also applied RF but achieved much lower explained variances of 6–23% (indicated by normalized MSE_{OOB} of 0.94 to 0.77) for different soil depths. A DSM approach for the prediction of SOC contents and other soil properties of an extensive area in Kenya revealed an explained variance of 13% for SOC (Mora-Vallejo et al. 2008). As descriptive statistics for the training and validation data sets revealed similar ranges, mean values and standard deviations for both data sets (Table 3), analyses of map accuracies by the excluded validation points are representative. Mean Errors relatively close to zero indicate unbiased predictions

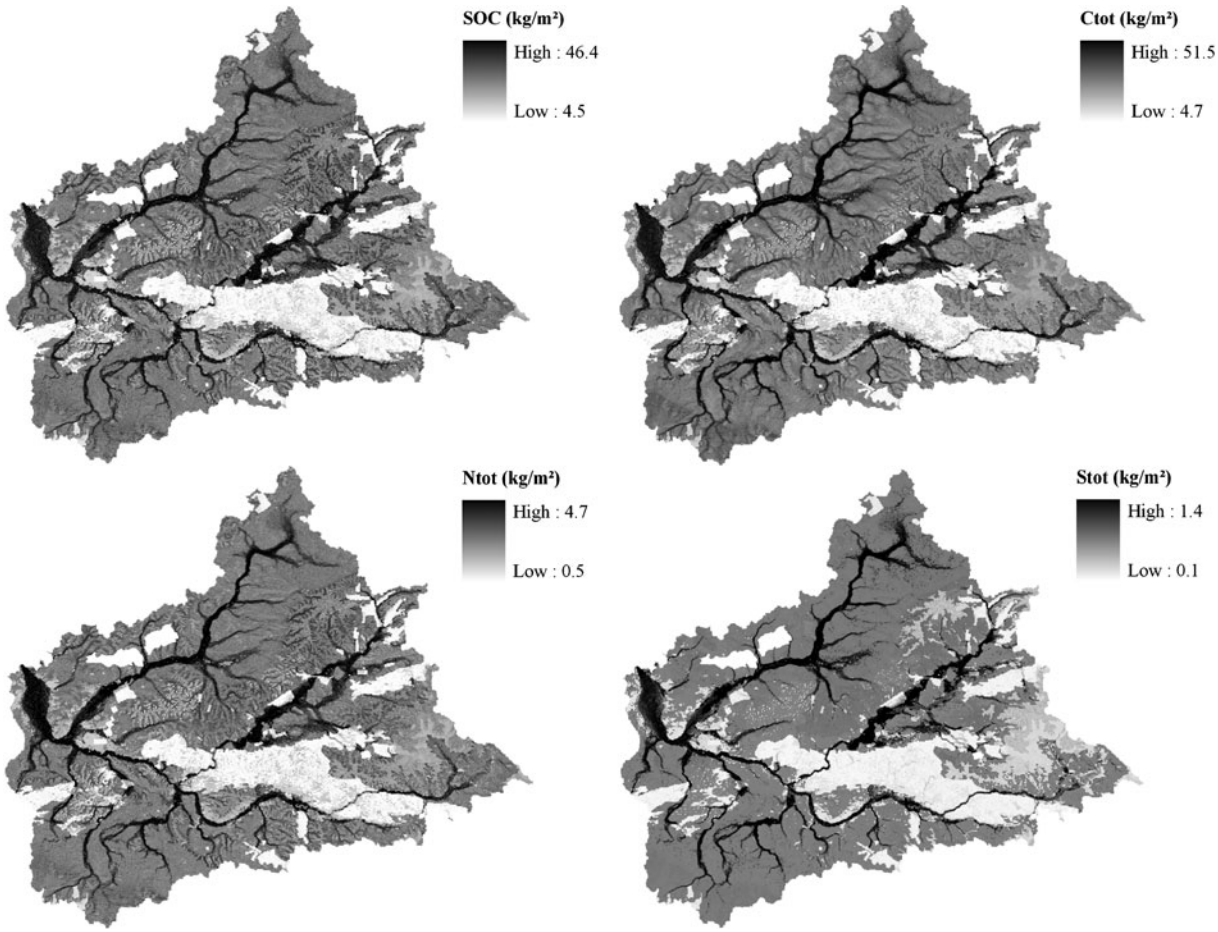

Fig. 5 Spatial prediction of SOC, C_{tot}, N_{tot} and S_{tot} stocks for the Xilin River Catchment

Table 5 Predicted mean values, total amounts and relative proportions of SOC, C_{tot}, N_{tot} and S_{tot} stocks for 1 m depth for the land use units and the total catchment derived from the predictions for the pixels of the DEM model (± standard deviation)

Land use unit	Arable land	Bare soil	Marshland/water	Mountain meadow	Steppe	Sand dunes	Total catchment
SOC (kg m^{-2})	8.3±2.9	14.1±3.0	33.5±6.5	20.3±3.1	27.3±6.0	7.2±2.2	24.1±9.1
total amount (Tg)	0.81	2.52	5.68	4.13	71.10	2.52	86.76
proportion (%)	1	3	6	5	82	3	100
C_{tot} (kg m^{-2})	9.6±3.9	16.7±4.4	37.4±7.8	23.6±2.5	29.9±6.6	7.8±2.8	26.5±9.9
total amount (Tg)	0.95	3.00	6.35	4.79	77.87	2.74	95.70
proportion (%)	1	3	7	5	81	3	100
N_{tot} (kg m^{-2})	0.8±0.3	1.5±0.3	3.2±0.8	1.9±0.3	2.5±0.6	0.7±0.2	2.2±0.8
total amount (Tg)	0.08	0.26	0.54	0.39	6.44	0.25	7.96
proportion (%)	1	3	7	5	81	3	100
S_{tot} (kg m^{-2})	0.23±0.11	0.27±0.15	0.64±0.28	0.30±0.05	0.61±0.14	0.18±0.09	0.53±0.22
total amount (Tg)	0.02	0.05	0.11	0.06	1.59	0.06	1.89
proportion (%)	1	3	6	3	84	3	100

for elemental stocks by RF, although generally negative values for all response variables point towards a slight overestimation. The accuracy of the prediction maps is relatively high as explained variances of 66 to 75% were calculated (Table 4). This was also indicated by strong correlations between measured and predicted elemental stocks. Remarkably, the MSE was much lower and explained variances were considerably higher for the map validation compared to the analysis of model performance by the OOB sample. Thus, we confirm that the application of an independent validation data set is necessary to analyse map accuracy separately for DSM approaches (Kempen et al. 2009; Mueller and Pierce 2003; Stoorvogel et al. 2009; Thompson et al. 2001).

Prediction accuracy can further be improved if the spatial correlation structure of the model residuals is analyzed. Hengl et al. (2004) provide a framework for such an analysis. They show that if the model residuals indicate spatial correlation then predictions can further be improved by geostatistical interpolation. However, while such an analysis was beyond the scope of our study, prediction accuracy with the RF model could be further enhanced by model residual analysis.

Simple comparison of the variable importance measures of both methods, RF and CART, confirms that the tree generated with CART, i.e. its splitting rules, is similar to the trees in the RF model. However, Prasad et al. (2006) suggest a more qualitative approach where similarities between the CART and the RF trees are examined by applying a BT model. The BT model averages over 50 trees. A calculation of a statistical summary of the deviances as well as the variations in the variable importances among the 50 trees may provide a less subjective measure of differences or similarities between the trees than direct comparison of the importance measures.

Environmental variables controlling the distribution of SOM

The results of the modeling approaches, RF and CART, both indicate that the most important predictor variable is land use unit. This confirms that land use, defined as physical and biological cover of the earth´s surface influenced by human activities (according to IPCC 2000) is the major factor that controls the input, decomposition and stabilization of organic matter into the soil and thus the amount and distribution of SOM. For other semi-arid regions in Northern China, Wei et al. (2008) and Liu et al. (2008) also highlighted the importance of land use in driving the spatial variability of soil properties. CART-derived decision trees showed that the six land use units were separated into several groups, which are associated with different amounts of SOM.

RSG was the second important variable associated with the storage of SOM and divided the dataset into a section of soils with very low elemental stocks and a group with low to medium stocks. The separation of these two groups demonstrates that in the studied catchment, RSGs are associated with and characterized by different amounts of SOM. The variable geological unit separated a section of different parent materials associated with low elemental stocks from a group of geological units characterized by higher stocks. Although topography was only of minor importance in the RF models, the contributing area was always one of the higher ranked variables among the topographic variables. Elemental stocks are generally higher where contributing area is smaller. Small contributing areas are associated with upstream locations at higher elevations. Thus, higher elemental stocks in these relatively undisturbed areas can be explained by an increased organic matter input and slower decomposition and turnover rates due to a colder climate.

The spatial distribution of elemental stocks in the Xilin River catchment was primarily influenced by land use unit and RSG. Geological unit and topographic variables are relatively weak predictors in comparison with these dominating predictor variables. This might be due to the fact that both are incorporated into the definition of RSG mapping units. Moreover, the uncertainty that adheres to the RSG and land use maps propagates into the models and spatial predictions of elemental stocks. The minor importance of topography-related variables for SOM storage might be also explained by the coarse scale at which analyses were carried out. Barthold et al. (2008) concluded that a 25 m resolution DEM was already too coarse to capture transport processes that are responsible for the spatial distribution of soil potassium and calcium. A 90 m resolution of the applied DEM is probably too coarse to capture some key topographic processes. The incorporated environmental predictor

variables probably also interact and can not be regarded as independent factors controlling the distribution of SOM.

Land use related elemental stocks

Spatial overlaying of prediction maps for SOC, C_{tot}, N_{tot} and S_{tot} stocks with the land use map allowed the calculation of elemental stocks for different land use units in the Xilin River Catchment (Fig. 5, Table 5). The high amounts of SOM in marshland/water are expected for these *Gleysol*-dominated areas, as anaerobic conditions inhibit the mineralization of organic matter. Predicted marshland SOC stocks (33.5 kg m^{-2}) are slightly higher compared to the estimations for bog soils of northeast and total China and for worldwide *Gleysols* of 20.4–29.4 kg m^{-2} (Batjes 1996; Wang et al. 2002; Wu et al. 2003).

Phaeozems dominated the mountain meadows and contained no carbonates probably because of a carbonate-free parent material (plutonic rocks) and a lower groundwater level associated with a limited capillary rise of dissolved carbonates in cooler, elevated areas. The high amounts of SOM of mountain meadows can be explained by a reduced mineralization due to cooler climatic conditions and an increased amount of stabilized, mineral-associated SOM in fine-textured soils as silt and clay contents were relatively high (44% and 21% respectively). Also, grazing intensity is probably lower in remote mountain regions with limited accessibility resulting in a higher organic matter input. Modelled SOC stocks for *Phaeozem*-dominated mountain regions of the basin with 20.3 kg m^{-2} are higher compared to the predictions for Chinese and worldwide *Phaeozems* of 12.8–14.6 kg m^{-2} (Batjes 1996; Wu et al. 2003). However, marshland/water and mountain meadows are of minor importance for total catchment stocks as they take over only a small proportion of the basin.

In contrast, steppe soils classified as *Chernozems* and *Phaeozems* are most important in terms of total catchment SOM storage as they cover the largest part of the basin and contain relatively high stocks of SOC, C_{tot}, N_{tot} and S_{tot}. The predicted high SOC stocks of 27.3 kg m^{-2} can be attributed on the one hand to a high input of organic matter as semi-arid grasslands in the Xilin River catchment belong to the most productive regions of Inner Mongolia (Chen et al. 2008). On the other hand, river-like structures of accumulated SOM contribute to the high predictions of elemental stocks (Fig. 5). These structures may indicate older, dry-fallen branches of the river system, periodically flooded valleys or simply groundwater-dominated areas where the mineralization of organic matter is delayed. These regions store more than twice the amount of SOM compared to the remaining steppe area. In contrast, other studies showed a considerable decrease of SOC at continuously grazed grasslands located in the vicinity of the Xilin River (Steffens et al. 2008; Wiesmeier et al. 2009). Mean total catchment stocks are probably higher due to a lower grazing intensity in remote grassland areas in the north and south of the basin. In other studies, much lower amounts of 8.2–23.8 kg m^{-2} for SOC were estimated for steppe soils from Northern China and other semi-arid regions of the world (Wang et al. 2002; Wu et al. 2003; Yu et al. 2007).

Arable soils were mainly classified as *Chernozems* and showed relatively high silt and clay contents (27% and 14%, respectively) although all agricultural sites were turned in the vicinity of sand dominated steppe areas. One may assume that only the most fertile steppe areas with fine-textured soils were converted to arable land during the recent intensification of agriculture in China. Nonetheless it is remarkable that cultivation of these productive steppe regions has resulted in a considerable decline of elemental stocks of up to 70%. This decline is higher than the mean decrease of SOC stocks of 59% after conversion of grassland to crops found in a meta analysis by Guo and Gifford (2002), who noted that the highest losses were found in regions with low precipitation. Loss of SOM during arable cultivation results from a low input of crop residues into the soil, mineralization of formerly protected organic matter after disruption of aggregates due to tillage and enhanced erosion of C- and N-rich particles of the topsoil (Balesdent et al. 2000; Hoffmann et al. 2008a; Steffens et al. 2009b).

For degraded bare soil areas, low amounts of SOM are likely to result from intensive soil erosion due to high grazing intensities particularly along the Xilin River and in the vicinity of villages. As these soils are very sandy (84%), they were mainly classified as *Arenosols*. Wind erosion primarily removes particles finer than sand (Hoffmann et al. 2008b) and consequently, the relative amount of sand in the topsoil increases. It should be noted that despite low

elemental stocks, bare soil still reveals almost twice the amounts of SOM compared with arable land indicating an even greater impact of tillage on soil erosion and aggregate destruction compared to intensive grazing.

The lowest predicted elemental stocks were found for *Arenosols* of less productive sand dunes where little soil development was observed. However, on north-facing slopes, shallow A horizons developed under a dense vegetation of several tree species. Sand dunes make up only a small proportion of total catchment stocks despite their considerable extent. Nevertheless, the modelled SOC stock (7.2 kg m^{-2}) is higher than estimated amounts of SOC of 2.4–3.1 kg m^{-2} for *Arenosols* from China and other regions of the world (Batjes 1996; Wang et al. 2002; Wu et al. 2003). This may be due to the dense tree vegetation and associated humus-rich topsoils on north-facing slopes.

Conclusions

The applied DSM approach including RF as a relatively new tool in the field of soil science for the prediction of elemental stocks and CART as an additional method for the analysis of predictor variables yielded promising results as the accuracy of the model and the generated prediction maps were acceptable. RF in combination with CART is a suitable approach for the spatial prediction of soil properties at the landscape scale and can also be applied in regions with limited amount of samples and data. As well as land use, RSG and geology as key factors determining the amount of SOM, incorporation of vegetation and soil moisture as predictor variables is likely to improve further the prediction of SOM storage in semi-arid environments. We recommend that topography is explicitly included in the sampling design of future soil inventories to enhance the prediction accuracy of SOM storage by accounting for SOC accumulation areas in water-dominated valleys and depressions. As high elemental stocks were found for the dominating grasslands we conclude that soils of semi-arid steppes in Northern China currently have a high C storage. However, these grasslands are susceptible to rapid SOM loss on cultivation and their suitability for agricultural use is therefore limited and will require no-till farming, windbreaks and other soil conservation techniques.

Acknowledgements The authors would like to thank Thomas Lendvaczky, Michael Schnabel, Sarah Schroeder, Hermann Autengruber, Peter Schad, Markus Steffens, Angelika Kölbl, Elfriede Schörk and Livia Wissing for their efforts in field sampling, laboratory work, helpful suggestions and for providing their knowledge. Qimei Lin and Yuandi Zhu are acknowledged for logistic handling. We also thank Xingguo Han, Yongfei Bai and the Institute of Botany (Chinese Academy of Sciences) for the opportunity to work at IMGERS. We are grateful to the Deutsche Forschungsgemeinschaft (DFG) for funding the MAGIM project (KO 1035/26-3, Forschergruppe 536 MAGIM—Matter fluxes in grasslands of Inner Mongolia as influenced by stocking rate).

References

Anderson JM (1991) The effects of climate change on decomposition processes in grassland and coniferous forests. Ecol Appl 1:326–347

Balesdent J, Chenu C, Balabane M (2000) Relationship of soil organic matter dynamics to physical protection and tillage. Soil Tillage Res 53:215–230

Barthold FK, Stallard RF, Elsenbeer H (2008) Soil nutrient-landscape relationships in a lowland tropical rainforest in Panama. For Ecol Manag 255:1135–1148

Batjes NH (1996) Total carbon and nitrogen in the soils of the world. Eur J Soil Sci 47:151–163

Breiman L (2001) Random forests. Mach Learn 45:5–32

Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and regression trees. Chapman and Hall, New York

Brus DJ, deGruijter JJ (1997) Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). Geoderma 80:1–44

Chen SP, Bai YF, Zhang LX, Han XG (2005) Comparing physiological responses of two dominant grass species to nitrogen addition in Xilin River Basin of China. Environ Exp Bot 53:65–75

Chen J, Hori Y, Yamamura Y, Shiyomi M, Huang DM (2008) Spatial heterogeneity and diversity analysis of macro-vegetation in the Xilingol region, Inner Mongolia, China, using the beta distribution. J Arid Environ 72:1110–1119

De'Ath G (2002) Multivariate regression trees: a new technique for modeling species-environment relationships. Ecology 83:1105–1117

De'ath G, Fabricius KE (2000) Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology 81:3178–3192

FAO (2006) Guidelines for soil description. Food and Agriculture Organization of the United Nations, Rome

Gislason PO, Benediktsson JA, Sveinsson JR (2006) Random Forests for land cover classification. Pattern Recognit Lett 27:294–300

Grimm R, Behrens T, Marker M, Elsenbeer H (2008) Soil organic carbon concentrations and stocks on Barro Colorado Island—digital soil mapping using Random Forests analysis. Geoderma 146:102–113

Grunwald S (2009) Multi-criteria characterization of recent digital soil mapping and modeling approaches. Geoderma 152:195–207

Guo LB, Gifford RM (2002) Soil carbon stocks and land use change: a meta analysis. Glob Chang Biol 8:345–360

Hartge KH, Horn R (1989) Die physikalische Untersuchung von Böden. Enke Verlag, Stuttgart

Henderson BL, Bui EN, Moran CJ, Simon DAP (2005) Australia-wide predictions of soil properties using decision trees. Geoderma 124:383–398

Hengl T, Heuvelink GBM, Stein A (2004) A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma 120:75–93

Hoffmann C, Funk R, Li Y, Sommer M (2008a) Effect of grazing on wind driven carbon and nitrogen ratios in the grasslands of Inner Mongolia. Catena 75:182–190

Hoffmann C, Funk R, Wieland R, Li Y, Sommer M (2008b) Effects of grazing and topography on dust flux and deposition in the Xilingele grassland, Inner Mongolia. J Arid Environ 72:792–807

IPCC (2000) IPCC Special Report Land use, Land-use change, and Forestry. Intergovernmental Panel on Climate Change (IPCC), Geneva

IUSS Working Group (2006) World reference base for soil resources 2006. World Soil Resources Reports No. 103, Food and Agriculture Organization (FAO), Rome

Jarvis A, Reuter HI, Nelson A, Guevara E (2008) Hole-filled seamless SRTM data V4.1, International Center for Tropical Agriculture (CIAT), http://srtm.csi.cgiar.org

Jenny H (1941) Factors of soil formation—a system of quantitative pedology. McGraw-Hill, New York

Kawamura K, Akiyama T, Yokota H, Tsutsumi M, Yasuda T, Watanabe O, Wang SP (2005) Quantifying grazing intensities using geographic information systems and satellite remote sensing in the Xilingol steppe region, Inner Mongolia, China. Agric Ecosyst Environ 107:83–93

Kempen B, Brus DJ, Heuvelink GBM, Stoorvogel JJ (2009) Updating the 1:50, 000 Dutch soil map using legacy soil data: a multinomial logistic regression approach. Geoderma 151:311–326

Lal R (2004) Carbon sequestration in dryland ecosystems. Environ Manag 33:528–544

Lal R (2009) Challenges and opportunities in soil organic matter research. Eur J Soil Sci 60:158–169

Lawrence RL, Wood SD, Sheley RL (2006) Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest). Remote Sens Environ 100:356–362

Liaw A, Wiener M (2002) Classification and regression by randomForest. R News 2(3):18–22

Liu DW, Wang ZM, Zhang B, Song KS, Li XY, Li JP, Li F, Duan HT (2006) Spatial distribution of soil organic carbon and analysis of related factors in croplands of the black soil region, Northeast China. Agric Ecosyst Environ 113:73–81

Liu HY, Yin Y, Tian YH, Ren J, Wang HY (2008) Climatic and anthropogenic controls of topsoil features in the semi-arid East Asian steppe. Geophys Res Lett 35, doi:10.1029/2007GL032980

McBratney AB, Odeh IOA, Bishop TFA, Dunbar MS, Shatar TM (2000) An overview of pedometric techniques for use in soil survey. Geoderma 97:293–327

McBratney AB, Santos MLM, Minasny B (2003) On digital soil mapping. Geoderma 117:3–52

McKenzie NJ, Ryan PJ (1999) Spatial prediction of soil properties using environmental correlation. Geoderma 89:67–94

Mora-Vallejo A, Claessens L, Stoorvogel J, Heuvelink GBM (2008) Small scale digital soil mapping in Southeastern Kenya. Catena 76:44–53

Mueller TG, Pierce FJ (2003) Soil carbon maps: enhancing spatial estimates with simple terrain attributes at multiple scales. Soil Sci Soc Am J 67:258–267

Myles AJ, Feudale RN, Liu Y, Woody NA, Brown SD (2004) An introduction to decision tree modeling. J Chemom 18:275–285

Ni J (2002) Carbon storage in grasslands of China. J Arid Environ 50:205–218

Peters J, Verhoest NEC, Samson R, Boeckx P, De Baets B (2008) Wetland vegetation distribution modelling for the identification of constraining environmental variables. Landsc Ecol 23:1049–1065

Prasad AM, Iverson LR, Liaw A (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9:181–199

RDevelopment Core Team (2008) An introduction to R. http://www.r-project.org/

Sanchez PA, Ahamed S, Carre F, Hartemink AE, Hempel J, Huising J, Lagacherie P, McBratney AB, McKenzie NJ, Mendonca-Santos MD, Minasny B, Montanarella L, Okoth P, Palm CA, Sachs JD, Shepherd KD, Vagen TG, Vanlauwe B, Walsh MG, Winowiecki LA, Zhang GL (2009) Digital soil map of the world. Science 325:680–681

Scull P, Franklin J, Chadwick OA, McArthur D (2003) Predictive soil mapping: a review. Prog Phys Geogr 27:171–197

Song GH, Li LQ, Pan GX, Zhang Q (2005) Topsoil organic carbon storage of China and its loss by cultivation. Biogeochemistry 74:47–62

Sorensen R, Zinko U, Seibert J (2006) On the calculation of the topographic wetness index: evaluation of different methods based on field observations. Hydrol Earth Syst Sci 10:101–112

Spörlein P, Dilling J, Joneck M (2004) Pilot study to test the equivalence or comparability of soil-particle-size analysis according to E DIN ISO 11277: 06.94 (pipette method) and by the use of the sedigraph. Journal of Plant Nutrition and Soil Science-Zeitschrift für Pflanzenernährung und Bodenkunde. J Plant Nutr Soil Sc 167:649–656

Steffens M (2009) Soils of a semiarid shortgrass steppe in Inner Mongolia: organic matter composition and distribution as affected by sheep grazing. Dissertation, Lehrstuhl für Bodenkunde, Technische Universität München

Steffens M, Kölbl A, Totsche KU, Kögel-Knabner I (2008) Grazing effects on soil chemical and physical properties in a semiarid steppe of Inner Mongolia (PR China). Geoderma 143:63–72

Steffens M, Kölbl A, Giese KM, Hoffmann C, Totsche KU, Breuer L, Kögel-Knabner I (2009a) Spatial variability of topsoil and vegetation in a grazed steppe ecosystem in Inner Mongolia (PR China). Journal of Plant Nutrition and Soil Science-Zeitschrift für Pflanzenernährung und Bodenkunde. J Plant Nutr Soil Sc 172:78–90

Steffens M, Kölbl A, Kögel-Knabner I (2009b) Alteration of soil organic matter pools and aggregation in semi-arid steppe topsoils as driven by organic matter input. Eur J Soil Sci 60:198–212

Stoorvogel JJ, Kempen B, Heuvelink GBM, de Bruin S (2009) Implementation and evaluation of existing knowledge for digital soil mapping in Senegal. Geoderma 149:161–170

Thompson JA, Bell JC, Butler CA (2001) Digital elevation model resolution: effects on terrain attribute calculation and quantitative soil-landscape modeling. Geoderma 100:67–89

Tong C, Wu J, Yong S, Yang J, Yong W (2004) A landscape-scale assessment of steppe degradation in the Xilin River Basin, Inner Mongolia, China. J Arid Environ 59:133–149

Vasques GM, Grunwald S, Sickman JO (2008) Comparison of multivariate methods for inferential modeling of soil carbon using visible/near-infrared spectra. Geoderma 146:14–25

Wang SQ, Zhou CH, Liu JY, Tian HQ, Li KA, Yang XM (2002) Carbon storage in northeast China as estimated from vegetation and soil inventories. Environ Pollut 116:S157–S165

Wei JB, Xiao DN, Zeng H, Fu YK (2008) Spatial variability of soil properties in relation to land use and topography in a typical small watershed of the black soil region, northeastern China. Environ Geol 53:1663–1672

Wiesmeier M, Steffens M, Kölbl A, Kögel-Knabner I (2009) Degradation and small-scale spatial homogenization of topsoils in intensively grazed steppes of Northern China. Soil Tillage Res 104:299–310

Wilson J, Gallant J (2000) Terrain analysis: principles and applications. Wiley, New York

Wu HB, Guo ZT, Peng CH (2003) Distribution and storage of soil organic carbon in China. Global Biogeochem Cycles 17:1048. doi:10.1029/2001GB001844

Wu JS, Liu HD, Duan XY, Ding Y, Wu HT, Bai YF, Sun X (2009) Prediction of DNA-binding residues in proteins from amino acid sequences using a random forest model with a hybrid feature. Bioinformatics 25:30–35

Xie ZB, Zhu JG, Liu G, Cadisch G, Hasegawa T, Chen CM, Sun HF, Tang HY, Zeng Q (2007) Soil organic carbon stocks in China and changes from 1980s to 2000s. Glob Chang Biol 13:1989–2007

Yu DS, Shi XZ, Wang H, Sun WX, Chen JM, Liu QH, Zhao YC (2007) Regional patterns of soil organic carbon stocks in China. J Environ Manag 85:680–689

Zhou ZY, Sun OJ, Huang JH, Li LH, Liu P, Han XG (2007) Soil carbon and nitrogen stores and storage potential as affected by land-use in an agro-pastoral ecotone of northern China. Biogeochemistry 82:127–138