Article

Large-Scale Mapping of Carbon Stocks in Riparian Forests with Self-Organizing Maps and the k-Nearest-Neighbor Algorithm

Leonhard Suchenwirth [1,2,*]**, Wolfgang Stümer** [3]**, Tobias Schmidt** [1]**, Michael Förster** [1] **and Birgit Kleinschmit** [1]

[1] Geoinformation in Environmental Planning Lab, Technische Universität Berlin, Office EB 5, Straße des 17. Juni 145, 10623 Berlin, Germany; E-Mails: leonhard.suchenwirth@tu-berlin.de (L.S.); t-schmidt@mailbox.tu-berlin.de (T.S.); michael.foerster@tu-berlin.de (M.F.); birgit.kleinschmit@tu-berlin.de (B.K.)
[2] GeoVille Information Systems GmbH, Sparkassenplatz 2, Innsbruck 6020, Austria
[3] Thünen Institute of Forest Ecosystems, Alfred-Möller-Straße 1, Eberswalde 16225, Germany; E-Mail: wolfgang.stuemer@ti.bund.de

* Author to whom correspondence should be addressed; E-Mail: lsuchenwirth@hotmail.com; Tel.: +49-152-26043271.

Abstract: Among the machine learning tools being used in recent years for environmental applications such as forestry, self-organizing maps (SOM) and the k-nearest neighbor (kNN) algorithm have been used successfully. We applied both methods for the mapping of organic carbon (C_{org}) in riparian forests due to their considerably high carbon storage capacity. Despite the importance of floodplains for carbon sequestration, a sufficient scientific foundation for creating large-scale maps showing the spatial C_{org} distribution is still missing. We estimated organic carbon in a test site in the Danube Floodplain based on RapidEye remote sensing data and additional geodata. Accordingly, carbon distribution maps of vegetation, soil, and total C_{org} stocks were derived. Results were compared and statistically evaluated with terrestrial survey data for outcomes with pure remote sensing data and for the combination with additional geodata using bias and the Root Mean Square Error (RMSE). Results show that SOM and kNN approaches enable us to reproduce spatial patterns of riparian forest C_{org} stocks. While vegetation C_{org} has very high RMSEs, outcomes for soil and total C_{org} stocks are less biased with a lower RMSE, especially when remote sensing and additional geodata are conjointly applied. SOMs show similar percentages of RMSE to kNN estimations.

1. Introduction

In recent decades, machine learning approaches have been introduced to manage the vast amount of data produced by various scientific disciplines, including environmental sciences such as forestry. One of the most intricate and specific neural networks techniques are self-organizing maps (SOM), which combine a high level of biological plausibility with applicability to numerous information processing and optimization problems. SOM allows one to reduce high dimensional information. The term "maps" refers to the low dimensionality and does not necessarily imply a spatial or geographical application; in fact, the technique emerged from neurosciences and there are a many examples from biosciences and engineering applications [1–3]. It has been described as an unsupervised learning technology [4].

A different approach to the spatial classification of data is the k-nearest neighbor (kNN) technique; this so-called instance-based, 'lazy' learning algorithm often serves as a benchmark for other methods [4]. It has been applied in a number of forest inventories, e.g., in Finland [5,6], New Zealand [7], Austria [8] or Ireland [9]. Some studies explicitly used kNN to estimate C_{org} [10–12]. The majority of studies are based on the use of Landsat data, few of them used VHSR (very high spatial resolution) satellite data.

Lek and Guégan [13] give a broad overview of applications in ecological and environmental sciences; recent applications include monitoring of river quality [14,15], urban modelling [16] and forestry applications [17,18]. Li *et al*. [19] applied an artificial neural network based approach for predicting soil matter across China. For the estimation of C_{org}, Stümer *et al*. [12] successfully applied SOM and compared it with the kNN algorithm for the assessment of biomass (and thus C_{org}) in Thuringian forests.

In the wake of the climate change discussion, it has become an essential task not only to decrease carbon emissions but also to identify natural carbon sinks in ecosystems all over the globe. Among terrestrial ecosystems, mangroves, peat lands and wetlands have especially shown an increased potential to sequester organic carbon in addition to other ecosystem services. For the case of riparian wetlands, several studies have underlined the high storage capacity [20–23].

The sequestration potential of floodplains is dependent both on vegetation (including forests, reed beds, and meadows), and soils. The important link between C_{org} stocks of forests and underlying soils has been demonstrated by a whole range of studies inside [24–26] and outside Europe [27,28].

Even though the value of riparian ecosystems has been recognized [22], the scientific underpinning for mapping large-scale carbon stocks is yet to be established. On a global scale [29], as well as on the national level [30,31], C_{org} maps have been produced and validated; however local validation of results is typically not obtainable. Various remote sensing analyses of C_{org} stocks have been utilized for non-floodplain habitats, especially forests [32–34], but most of these studies have focused either on

C_{org} stocks in soil or vegetation. To all of our knowledge, detailed C_{org} maps of floodplain areas have seldom been produced, apart from Suchenwirth *et al.* [35,36] and Güneralp *et al.* [37].

In the presented study, we estimate C_{org} above and below ground in a test site in the Danube Floodplain based on a SOM and kNN classification of VHSR RapidEye data and ancillary geodata. Both results are compared to field survey data. In contrast to Stümer's application of SOM and kNN for Thuringian forests [12], we consider the vegetation (above ground), soil (below ground) and total C_{org} in a floodplain area on a more detailed spatial resolution. Vegetation below ground (e.g., plant roots) is not separately considered, yet calculations can be done according to guidelines by the IPCC [29]. Besides remote sensing data, we introduce additional auxiliary geodata as input source data for both algorithms. In this way, we compare the outcomes for remote sensing (RS) input information with the results for RS and additional information. We decided to apply SOM and kNN for the C_{org} models, as previous methods such as the derivation of C_{org} stocks from classified vegetation types [35] or the derivation via quantiles in a classification and regression tree (CART) approach [36] had only limited success.

The specific aims of this paper are as follows:

(1) to create distribution maps of vegetation, soil, and total C_{org} stocks in a riparian forest, based on SOM and kNN algorithms and compare the results;
(2) to compare and evaluate results with previous estimation techniques;
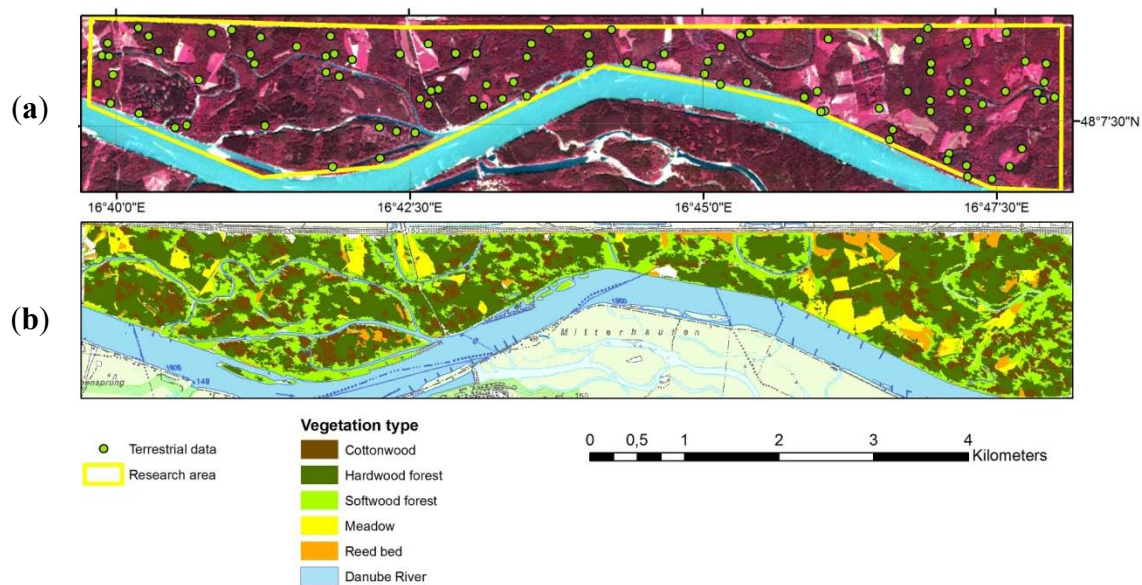(3) to evaluate the influence of additional geodata on estimation quality.

2. Material and Methods

2.1. Study Area

The research area is located inside the Danube Floodplain National Park (*Nationalpark Donauauen)* in Austria (16.66° E, 48.14° N). The area is a pristine floodplain area with few human impacts. Human activities included hunting in previous centuries, the construction of the Marchfeld dike in the 19th century, and the plantings of hybrid poplars (*Populus × canadensis*). Apart from these cottonwood plantations, the area is characterized by riparian vegetation, such as softwood forests (dominated by *Salix alba, Acer negundo*), hardwood forests (dominated by *Quercus robur, Fraxinus excelsior* and *Acer campestre*), as well as meadows and reed beds. Our study area (11.7 km^2) is limited by the Marchfeld dike (locally named Hubertusdamm dike) in the north, and the main river course towards the south. Geographic coordinates are given in Figure 1.

The area was chosen for our study due to its high protection status, a good base of geographic data, and previous research in the area [38–40]. Mean C_{org} storage in the area was estimated at 359.1 Mg ha^{-1} by Cierjacks *et al.* [41], and as 428.9 Mg ha^{-1} by Suchenwirth *et al.* [35]. Figure 1a presents a RapidEye scene of the Danube Floodplain area. Red color indicates pixels with high content of active biomass, *i.e.*, trees and bushes in comparison to bare soils and impervious areas. Figure 1b shows the distribution of existing vegetation types.

Figure 1. (**a**) Research area depicted as RapidEye Near-Infrared (NIR) composite with terrestrial survey data (green dots; above) and (**b**) vegetation classification (derived from [35]; below).

2.2. Data

We obtained a cloudless satellite image from RapidEye (acquired on August 1, 2009 in level 3A with a spatial resolution of 5.0 m [42]; Figure 1 above, Table 1). The image was provided by the German Aerospace Center, in the UTM WGS 1984 reference system. We reprojected the image into the Austrian MGI M34 projected coordinate system, as local data were mainly available in the local reference system. Atmospheric correction was not performed as we did not work with time series. RapidEye data were used as the high spatial resolution reflects the spatial heterogeneity of carbon distribution in floodplains. Notably, the RedEdge channel has already been successfully applied to improve classifications of vegetation [43].

Table 1. Available geodata and derived parameters.

Available geodata	Derived parameters	Abbreviations
RapidEye image	Blue channel (440–510 nm)	B
(1 August 2009)	Green channel (520–590 nm)	G
	Red channel (630–685 nm)	R
	Red edge channel (690–730 nm)	RE
	Near infrared channel (760–850 nm)	NIR
Digital elevation model	Elevation above river level	altitude
Ground water model	Ground water level	MGW
Topographic map 1:50,00 (ÖK 50)	Distance to river	distance
C_{org} ground survey data from 2008	Above ground carbon stocks	C_{org_veg}
and 2010	Below ground carbon stocks	C_{org_soil}
	Total carbon stocks	C_{org_tot}

A digital elevation model (DEM) derived from LiDAR data was used to compute altitude above river level; a groundwater model indicating median ground water depth was provided by the Vienna University of Technology. Distance to river (main stream) was derived from a topographic map. The topographic map is issued and updated every seven years by the Austrian Federal Office of Metrology and Surveying (*Bundesamt für Eich- und Vermessungswesen*).

A total of 104 *in situ* inventory plots (10 × 10 m) for vegetation and soil were established within two terrestrial surveys in 2008 and 2010. The point selection based on a stratified sampling design, described in detail by Cierjacks *et al*. [41] and Rieger *et al*. [44], who also measured and calculated C_{org} content of soil and vegetation for each sample point. In a nutshell, soil samples were extracted from 0–100 cm using an auger, with each soil horizon was sampled separately and carbonate concentration and the concentration of C_{org} were determined. For vegetation, height and circumference were measured at breast height of all trees >15 cm in circumference.

Based on these data, stem number per ha, mean height, and mean diameter was calculated. Total C_{org} consists of C_{org} in soil, vegetation, and dead wood on the ground.

2.3. Self-Organizing Maps (SOM)

The SOM approach is used to produce maps of C_{org} stocks in riparian forests of the Danube Floodplain. The method has been described in detail by Kohonen [45,46], and has frequently been used and described by other authors [4,12,18].

The application of SOMs is generally divided into two modes or phases: a learning (or training) phase and a classification or mapping phase. SOMs structure the neurons in the form of rectangular or hexagonal arrays or grids of nodes with n dimensions, with an associated weight vector attached to each node. The procedure of placing a vector from the high-dimensional data space into the two dimensional map space is performed by identifying the node with the closest associated distance to the presented data space vector, *i.e.*, the winner pixel or best matching unit (BMU) is selected; its position within the grid is the excitation centre. Subsequently, differences between the weight vector and the data space vector are reduced. Afterwards, vectors in the neighborhood are adapted. The distance of the feature space is defined as the Euclidean distance. The learning process of the winner selection and adaption process is iteratively repeated until no further adaption is necessary, as the initial learning rate is much smaller than in the first stage and a stable state is reached. At this moment, the learning phase is completed.

In the mapping phase, the input vector for which the prediction is necessary is presented to the map; distances from this location to all neurons are calculated. As a result, the BMU of the map is selected, providing a representative group of data samples to which the predicted input is most similar.
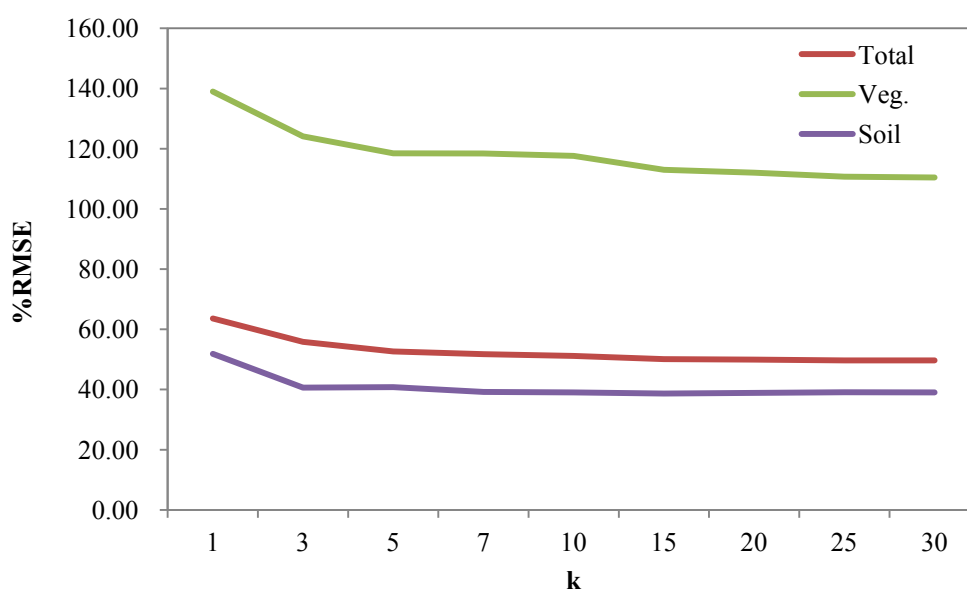
In our approach, we applied the algorithm programmed by Stümer *et al*. [12]. We use the RapidEye scene with additional geodata (see data section) for our classification as the initial layer. For the analysis, we used the following standard parameters: A feature space distance of five or eight (depending on the number of used channels/parameters), a start distance δ_{start} of 100,000 and an end distance δ_{end} of 100, and five iterations (t_{max}) were applied. It is necessary to set the start distance high in order to sufficiently consider the terrestrial samples, while at the end only the necessary neighbors shall be regarded.

2.4. k-Nearest Neighbor (kNN)

To compare the operational applicability of SOM we use the kNN method to provide spatially explicit results. It is described as the simplest, intuitively understandable and purely data-driven algorithm and is applied frequently for classification or regression tasks, or to provide a quick visualization or benchmark. It classifies a point by calculating the distances between the point and the points in the training data set. Then, the point is assigned to the class which is most common among its k-nearest neighbors (with k being an integer number). There is no learning phase, since all training examples are simply stored in the memory for further predictions. The method was described by several authors [4,47,48]. We follow the method applied by Stümer *et al*. [12].

For our kNN classification, we used standard settings to compare classifications: $k = 5$ neighbors; Euclidean distance $d_{(x1,x2)}$, of 2, and a distance weight $w_{(i),p}$ of 2. These parameter settings were often described as a compromise between a limited number of neighbors and a sufficient accuracy in other studies [9,10]. In order to prove the appropriate number of *k* neighbors, a sensitivity analysis (Figure 2) showed that RMSE strongly declined between $k = 1$ and $k = 5$, and showed a very gentle decrease for $k>5$. In order to establish a tradeoff between low error and readability and applicability of data, the *k* value of 5 was chosen.

Figure 2. Sensitivity analysis based on Root Mean Square Error (RMSE) for $k = 1$ to $k = 30$.

2.5. Validation

The reliability of C_{org} estimates obtained by the SOM- and kNN- approach is quantified by the bias and the root mean square error (RMSE). The bias is calculated as the difference between measured and estimated C_{org} stock; the RMSE includes variance of estimated C_{org} stock and the bias. The % RMSE facilitates comparisons between C_{org} measurements. In order to use terrestrial samples for both calibration and validation we used the Leave-one-out (L1o) cross-validation [49,50].

3. Results

The SOM and the kNN approach were used in the Danube Floodplain National Park. We produced two types of results: (1) spatially explicit maps of the vegetation, soil and total C_{org} stocks per unit (Mg C_{org} ha^{-1}), and (2) statistical estimates for vegetation, soil and total C_{org} stocks. The maps obtained by the SOM-approach were compared to alternative maps based on the kNN-approach. Terrestrial data were used as a basis for comparison of the statistical estimates obtained by the SOM- and kNN-approaches.

3.1. C_{org} Stock Estimations

C_{org} stock maps of vegetation, soil and total carbon, -based on the SOM method are displayed in Figure 3a,b. C_{org} maps based on the kNN method are presented in Figure 4a,b. C_{org} stocks in the maps are displayed in a color range from yellow to red where lower stocks are indicated in light yellow, higher stocks in dark red, and for total C_{org} stocks color tones with higher values are in brown tones. All figures show the same detail of the area. The C_{org} stocks is given in tons per ha (Mg C_{org} ha^{-1}).

We can see from the satellite image and the vegetation map (Figure 1a and b) that the wooded area has a dispersed distribution, with a high variation of vegetation within a small scale. This results in fragmented C_{org} stock maps. It is apparent that C_{org} stocks in soils are generally classified higher and with fewer divisions than those in vegetation. A comparison of the maps shows that forest areas are indirectly classified by both approaches due to higher concentrations of vegetation C_{org}. Stocks over 100 Mg C_{org} ha^{-1} are found mainly in areas recognizable as forests in the satellite imagery.

Comparing outcomes from SOM and kNN, we can identify a more distinct spatial pattern in SOM classifications. This is evident in the classification of soil C_{org}, where, while kNN classifications show a highly homogeneous surface with tiny differences, SOM classifications exhibit clear differences between forested areas and meadows and reed beds.
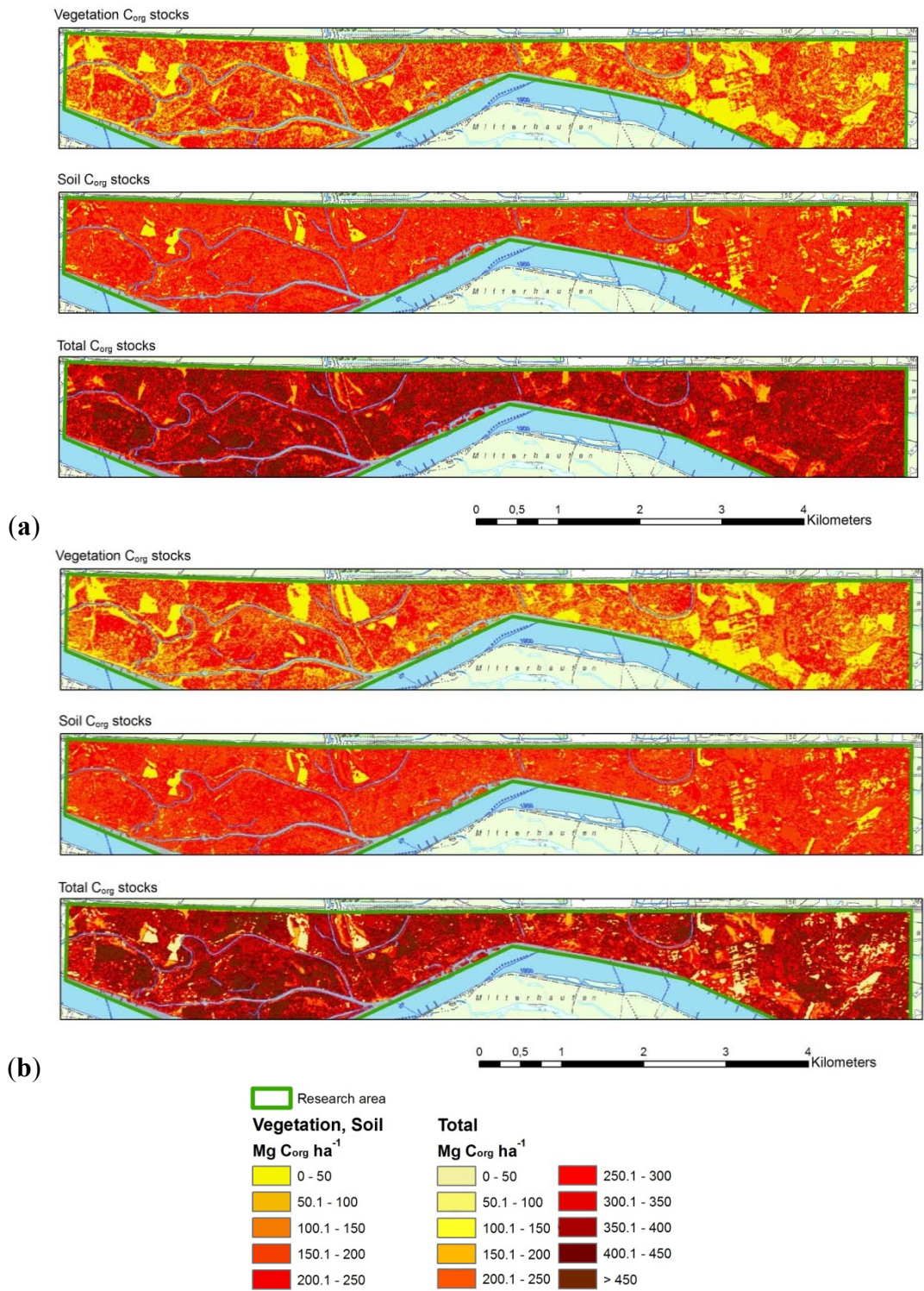
Comparing the maps generated by pure remote sensing data and the combination of remote sensing and auxiliary data, we can observe greater details for classifications with combined data, which is especially visible for classifications of total C_{org} stocks, where the range of possible values is much more highlighted.

The review of the SOM- and kNN approach was complemented by a comparison of statistical estimates for the test area. Table 2 shows the results of C_{org} stock provided by the SOM- and kNN-approaches. The results presented are based on the entire set of point estimates used for producing the test area maps. The differences between SOM- and kNN-based estimates range between 3.87 Mg ha^{-1} (soil and total C_{org} stocks) and 46.14 Mg ha^{-1} (total C_{org} stocks).

The kNN approach with the RapidEye dataset provides generally higher values in comparison to the SOM approach. The differences between the two approaches including additional data do not indicate a one-sided bias structure. Estimations of vegetation, soil and total C_{org} stocks are independent from each other, so vegetation and soil C_{org} stocks do not necessarily add up to total C_{org} stocks. In order to analyze the accuracy in comparison with the field data, we have to consider the error estimates. In general, values are slightly lower than values of previous results [35,41].
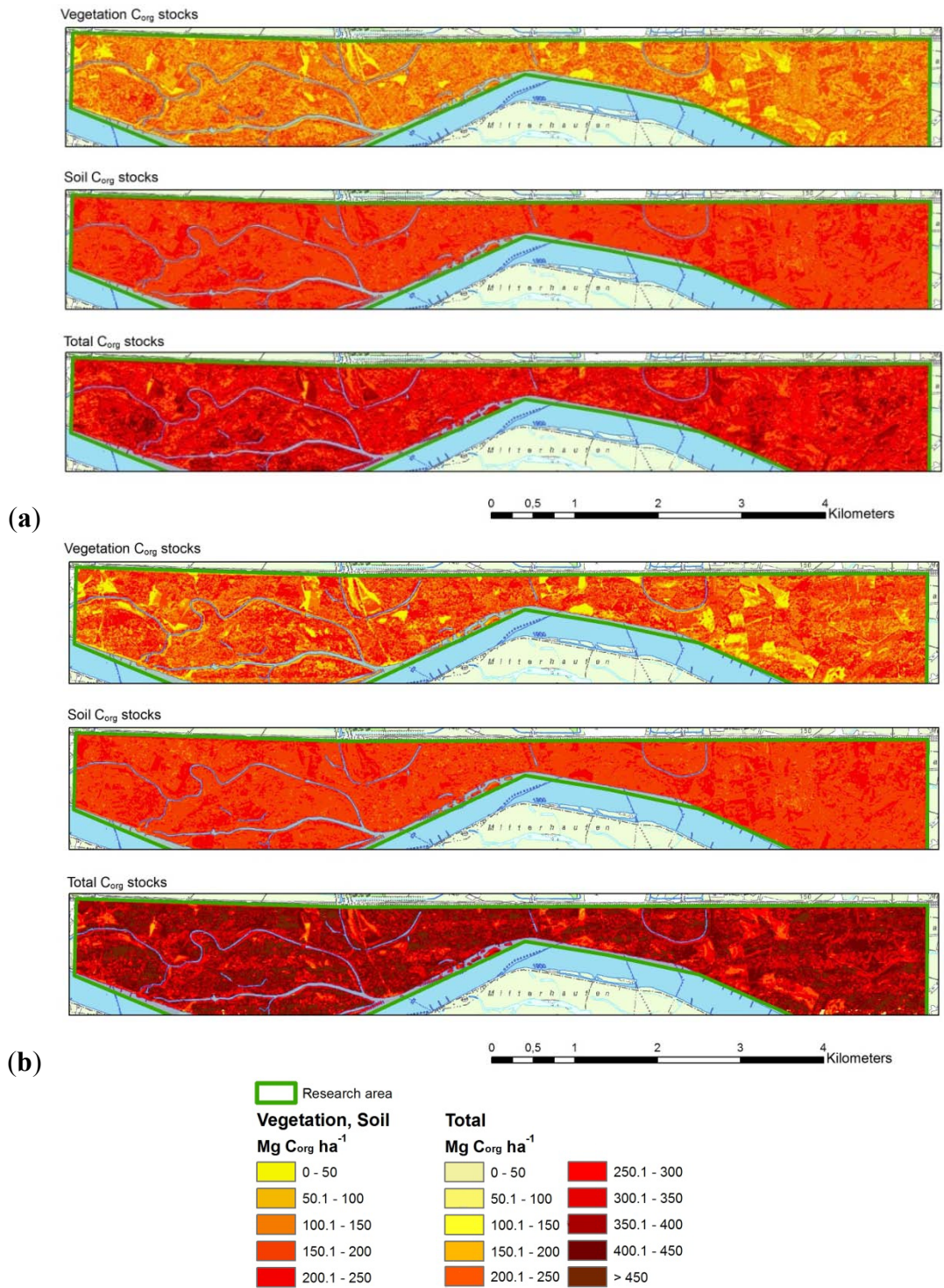
Figure 3. (**a**) C_{org} stocks in vegetation (above), soil (middle) and total (below), calculated by SOM method based on RapidEye; (**b**) C_{org} stocks in vegetation (above), soil (middle) and total (below), calculated by SOM method based on RapidEye and additional data.

Figure 4. (**a**) C_{org} stocks in vegetation (above), soil (middle) and total (below), calculated by kNN method based on RapidEye; (**b**) Corg stocks in vegetation (above), soil (middle) and total (below), calculated by kNN method based on RapidEye and additional data.

Table 2. SOM- and kNN-based estimates for vegetation, soil and total C_{org} stocks in the Danube Floodplain.

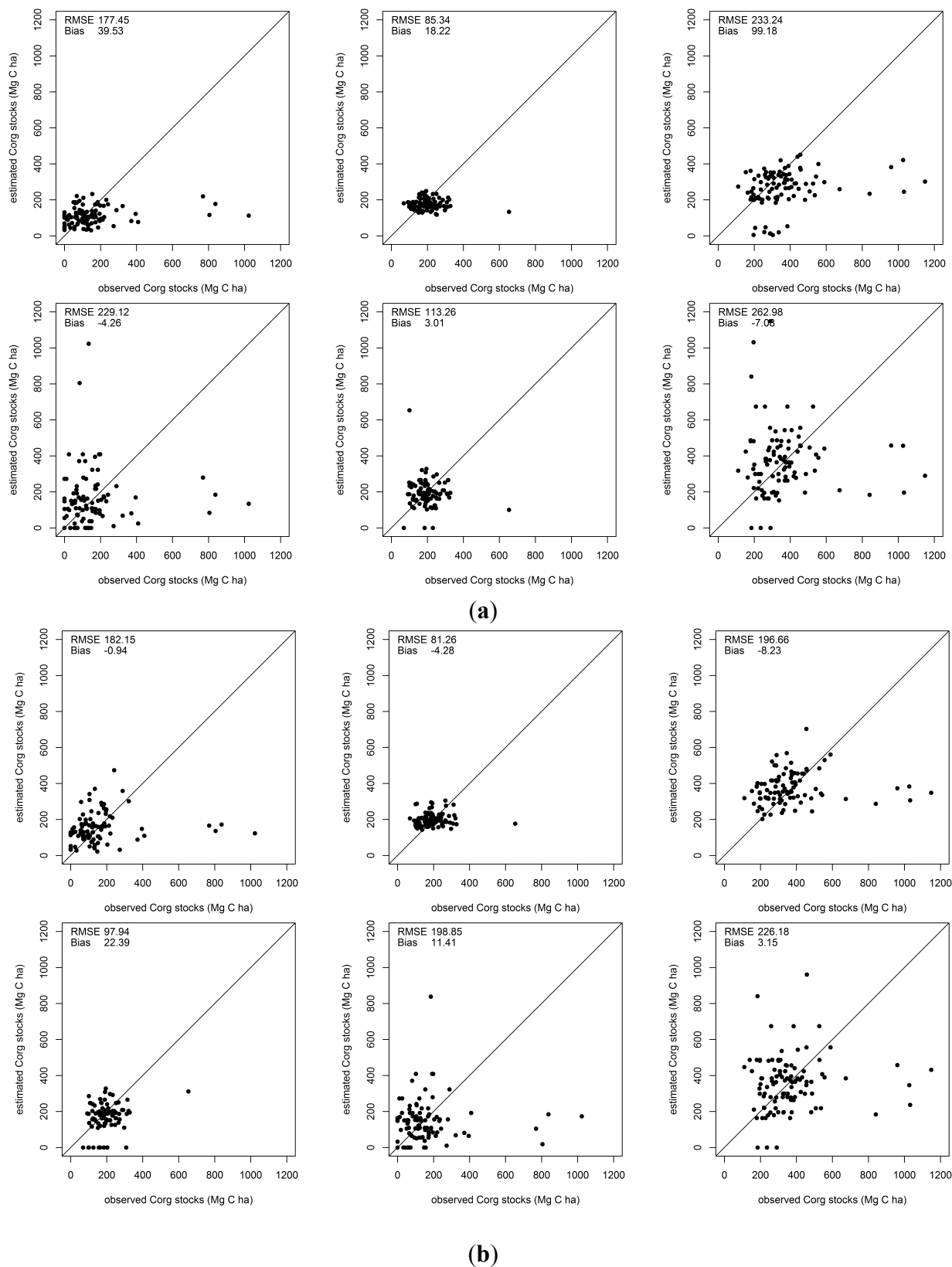| Dataset	Approach	Vegetation C_{org}: Mg C_{org} in total study area (Mg C ha^{-1})	Soil C_{org}: Mg C_{org} in total study area (Mg C ha^{-1})	Total C_{org}: Mg C_{org} in total study area (Mg C ha^{-1})
RapidEye	SOM	**144043.49** (127.47)	**198390.17** (175.57)	**393735.41** (348.44)
	kNN	**158791.28** (140.52)	**238362.66** (210.94)	**398114.52** (352.31)
RapidEye + altitude + MGW + distance	SOM	**168056.05** (148.72)	**198635.46** (175.78)	**389228.63** (344.45)
	kNN	**122856.37** (108.72)	**203001.62** (179.65)	**337092.95** (298.31)

3.2. Error Estimates

In order to evaluate their performance, SOM and kNN point estimates that coincided with terrestrial survey plots were each used to carry out an error analysis using the Leave-one-out (L1o) cross–validation [49,50] for the estimation of the average growing stock per unit area (Table 3). The values assessed on the field plots served as control values. In the research area, in total 104 terrestrial plots were available for calculating the bias and RMSE, with normalized values for bias, RMSE and % RMSE. Additionally, for each run of SOM and kNN scatterplots of estimated and observed survey plots are shown (Figure 5a,b). For vegetation C_{org} measurements, the approaches had positive and negative biases (SOM: −4.26; 11.41; kNN: 39.52; −0.94). In soil C_{org} assessments, SOM approaches yielded positive biases (3.01; 0.28), while kNN yielded positive bias for RapidEye estimation only (18.22; −4.28).

Table 3. Error estimates from SOM and kNN for vegetation, soil and total C_{org} stocks in the Danube Floodplain (SOM: start distance δ_{start} of 100,000 and an end distance δ_{end} of 100, and have five iterations (t_{max}); kNN k: 5; Euclidean distance $d_{(x1,x2)}$: 2; distance weight $w_{(i),p}$: 2).

Dataset	Approach	Vegetation C_{org} stocks (average 149.65 Mg C ha^{-1})			Soil C_{org} stocks (average 192.1 Mg C ha^{-1})			Total C_{org} stocks (average 361.52 Mg C ha^{-1})		
		Bias	RMSE	% RMSE	Bias	RMSE	% RMSE	Bias	RMSE	% RMSE
RapidEye	SOM	−4.26	229.12	146.99	3.01	113.26	58.99	−7.08	262.98	70.85
	kNN	39.52	177.45	158.32	18.22	85.34	48.27	73.92	210.45	72.52
RapidEye + altitude + MGW + distance	SOM	11.41	198.85	143.29	0.28	108.22	56.42	3.15	226.18	63.11
	kNN	−0.94	182.15	118.46	−4.28	81.26	40.79	−8.23	196.66	52.67

Both approaches had a positive and a negative bias for estimations of total C_{org} stocks (SOM: −7.08; 3.15), kNN was positive and negative (73.92; −8.23). The positive biases are higher than the negative biases. In most cases, apart from vegetation C_{org} with additional data, the kNN approach is more biased than the SOM results. Furthermore, the kNN approach exhibits a trend of slight overfitting when using additional geodata but strong underfitting when using RapidEye only. The SOM approach shows an overfitting performance when additional geodata were utilized. If only the spectral information of RapidEye is included, SOM has no clear direction of estimation.

Figure 5. (**a**) Scatterplots of kNN (above) and SOM (below) estimated and observed survey plots for C_{org} stocks in vegetation (left), soil (central) and total (right) based on RapidEye only; (**b**) Scatterplots of kNN (above) and SOM (below) estimated and observed survey plots for C_{org} stocks in vegetation (left), soil (central) and total (right) based on RapidEye and additional data.

Discerning between SOM and kNN, the RMSE does not show a clear tendency. In some estimations, SOM has lower RMSE, in other cases kNN estimation is more accurate. The %RMSE of the SOM-approach ranged between 56.42% and 146.99% and had a smaller range than for the kNN approach (40.79%–158.32%). Biases are smaller for SOM estimations.

Regarding the use of additional geodata, there is a lower RMSE for the estimations based on additional geodata, than for estimations based on pure RapidEye datasets. Especially for kNN estimations, the error is notably lower (8%–40%), whereas for SOM estimations, errors are only slightly lower (2%–6%). Apart from the SOM approach on vegetation C_{org}, the bias is smaller for predictions using additional geodata.

4. Discussion

SOM and kNN have been applied for spatially explicit estimates of C_{org} stocks above and below ground in riparian forest zones. Terrestrial measurements and satellite data as well as additional geodata served as input data to carry out the learning and training process of a neural network. Results show that both methods, SOM and kNN, are able to mimic spatial patterns of vegetation, soil and total C_{org} stocks. Both provide spatially detailed estimates, only limited by the spatial resolution of the used imagery. However, the SOM approach supplies a far more distinct spatial pattern of the C_{org} distribution, while the kNN method results in rather average, homogeneous patterns.

In comparison with existing estimations, values for total C_{org} stocks are comparable to the results of Cierjacks *et al.* [41], but are considerably lower than results classified by Suchenwirth *et al.* [35]. This would support the assumption that the SOM and kNN methods can substitute a field-based calculation [41] better than a mere classification of vegetation types to estimate C_{org} stocks [35].

In detail, estimations of vegetation C_{org} stocks (both kNN and SOM, based on satellite sensors and additional data) have an apparently higher RMSE than estimations of soil C_{org} stocks and total C_{org} stocks. This is not the case for bias, which is in several cases higher for the estimation of total C_{org} stocks. The RMSE of our estimations of soil (ranging from 40.79%–58.99%) and total C_{org} (ranging between 52.67% and 72.52%) are in line with results of other studies using SOM [12,19,51] or kNN [9] to classify C_{org}, where values range between 44.85% and 70.49%. RMSE for vegetation C_{org} is higher (118.46%–158.32%) in our estimation.

The reason for the higher RMSE within the vegetation classification can be explained by the more complex natural structure and the resulting diversity inside riparian forest vegetation and the national park area in comparison to the structure of conventional working forests and timberland monocultures.

Comparing the results of kNN and SOM-based estimations, we can find that both provide similar results. kNN has smaller RSME estimates for soil C_{org} and for estimations of vegetation and total C_{org} stocks based on RapidEye and additional data, yet has higher RMSE estimates for estimations of vegetation and total C_{org} based solely on RapidEye. In general, we can state that kNN have a better performance regarding RMSE than SOM estimates, which is also consistent with Stümer *et al.* [12]. Contrarily, the kNN results are much more biased than the results of the SOM. Moreover, the visual impressions of the SOM-generated maps are more distinct; this distinguishes our results from Stümer's results who found a smaller bias and a higher level of detail of structures such as roads, planting rows and stand boundaries for kNN results. In conclusion, we can state that in our study, kNN provides on

average better estimates for C_{org}, but only within the restricted range of values within the test area. For a possible transfer of the method to other regions, the less biased SOM approach might be the preferable algorithm.

Both presented approaches provide greater spatial detail than comparable classifications based on object-based image analysis (OBIA) of the area [35,36,39]. Even though OBIAs have the advantage of working with distinct image objects, the process of segmentation can be challenging and even misleading for continuous objects, such as natural vegetation or ecosystems in an intricate floodplain area, and may thus be a source of error, as stated by Rocchini *et al.* [52]. In comparison to other remote sensing techniques, such as Principal Component Analysis (PCA), SOM has shown demonstrably better performance [14,53].

However, some issues may yet occur when using SOMs, as they are not self-explanatory and are generally treated as a "black box" due to unknown weights and the non-linearity of the activation functions. While Hsu and Halgamuge [54] mention the obliqueness of rectangular lattices as major sources of topographic errors, Klobucar and Subasic [53] count the repeatability of the method among the problems of SOM. The time needed to calibrate and validate neural networks should not be underrated and the decision about the termination of the learning process may be difficult.

The application of kNN did not impose greater issues, and their applicability to forest and biomass/C_{org} inventions has often been proven, even though the majority of studies have worked with Landsat data which have a lower spatial resolution and thus provide coarser imagery; the application of VHSR data is not so common yet, while the combination with auxiliary geodata has barely been used. Among the commonly mentioned disadvantages of kNN are excessive validations of each distance, and the sensitivity towards irrelevant or noisy attributes as well as towards unbalanced datasets. For our application, however, it has served as a valuable alternative to the application of SOMs.

The use of additional geodata improved the performance of both algorithms, in that all RMSE values improved, as well as the bias (with the exception of SOM assessment on vegetation C_{org}). Especially for kNN, the notable improvement of RMSE underlines the importance of combined data approaches. This also confirms previous findings of Suchenwirth *et al.* [35].

Comparing the study's method with previous methods to quantify C_{org} in the Danube floodplain, our study uses a "direct remote sensing" approach including machine learning [55], while Suchenwirth *et al.* [35] used a "stratify and multiply" approach, and Suchenwirth *et al.* [36] used a "combine and assign" approach [55].

5. Conclusions

In general, we see our study as a contribution to high-detailed C_{org} analyses and large scale maps of intricate ecosystems such as riparian forests or similar wetland areas with interfering aquatic and terrestrial environments, as they impede ground survey measurements through their restricted accessibility and require advanced methods to estimate biomass and organic carbon, such as remote sensing or machine learning.

For prospective applications, we envisage comparable studies with extensions of start distances and numbers of iterations, as the focus of this study lies in the comparative estimations of C_{org} stocks in

vegetation, soil, and total, with varying parameters and with two methods, and not with different settings of SOM and kNN.

Another improvement for future research on the estimation of C_{org} with remote sensing data may be to include imagery with an even higher spatial resolution, as e.g., provided by the commercial sensors Ikonos (1 m), QuickBird 2 (0.64 m), or Worldview (0.5 m). The inclusion of further datasets such as surface models including tree height, e.g., based on LiDAR, and other auxiliary data is able to additionally improve the performance.

Acknowledgments

Author Contributions

All authors contributed extensively to this work. Wolfgang Stümer supported the SOM and kNN methods by programming and provided the tool and interpreting results, while Tobias Schmidt supported the validation and review process. Michael Förster was intensively involved in interpretations and discussions of the manuscript, and Birgit Kleinschmit supervised and supported the research project as head of the geoinformation lab. Leonhard Suchenwirth carried out the data processing, analyses, and drafted and revised the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Breijo, E.G.; Pinatti, C.O.; Peris, R.M.; Fillol, M.A.; Martinez-Manez, R.; Camino, J.S. Tnt detection using a voltammetric electronic tongue based on neural networks. *Sens. Actuator A-Phys.* **2013**, *192*, 1–8.
2. Wang, Z.S.; Bian, S.R.; Liu, Y.; Liu, Z.H. The load characteristics classification and synthesis of substations in large area power grid. *Int. J. Electr. Power Energy Syst.* **2013**, *48*, 71–82.
3. Xuan, S.Y.; Wu, Y.B.; Chen, X.F.; Liu, J.; Yan, A.X. Prediction of bioactivity of hiv-1 integrase st inhibitors by multilinear regression analysis and support vector machine. *Bioorg. Med. Chem. Lett.* **2013**, *23*, 1648–1655.
4. Kanevski, M.; Timonin, V.; Pozdnukhov, A. *Machine Learning Algorithms for Spatial Data Analysis and Modelling*; EFPL Press: Lausanne, Switzerland, 2009; p. 377.
5. Tomppo, E. Satellite image-based national forest inventory of finland. *Int. Arch. Photogramm. Remote Sens.* **1991**, *28*, 419–424.

6. Tomppo, E.; Halme, M. Using coarse scale forest variables as ancillary information and weighting of variables in k-NN estimation: A genetic algorithm approach. *Remote Sens. Environ.* **2004**, *92*, 1–20.

7. Tomppo, E.; Goulding, C.; Katila, M. Adapting finnish multi-source forest inventory techniques to the new zealand preharvest inventory. *Scand. J. For. Res.* **1999**, *14*, 182–192.

8. Koukal, T.; Suppan, F.; Schneider, W. The impact of relative radiometric calibration on the accuracy of knn-predictions of forest attributes. *Remote Sens. Environ.* **2007**, *110*, 431–437.

9. McInerney, D.O.; Nieuwenhuis, M. A comparative analysis of knn and decision tree methods for the irish national forest inventory. *Int. J. Remote Sens.* **2009**, *30*, 4937–4955.

10. Fuchs, H.; Magdon, P.; Kleinn, C.; Flessa, H. Estimating aboveground carbon in a catchment of the siberian forest tundra: Combining satellite imagery and field inventory. *Remote Sens. Environ.* **2009**, *113*, 518–531.

11. Magnussen, S.; McRoberts, R.E.; Tomppo, E.O. Model-based mean square error estimators for k-nearest neighbour predictions and applications using remotely sensed data for forest inventories. *Remote Sens. Environ.* **2009**, *113*, 476–488.

12. Stümer, W.; Kenter, B.; Köhl, M. Spatial interpolation of in situ data by self-organizing map algorithms (neural networks) for the assessment of carbon stocks in european forests. *For. Ecol. Manag.* **2010**, *260*, 287–293.

13. Lek, S.; Guégan, J.F. Artificial neural networks as a tool in ecological modelling, an introduction. *Ecol. Modell.* **1999**, *120*, 65–73.

14. Astel, A.; Tsakovski, S.; Barbieri, P.; Simeonov, V. Comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets. *Water Res.* **2007**, *41*, 4566–4578.

15. Shanmuganathan, S.; Sallis, P.; Buckeridge, J. Self-organising map methods in integrated modelling of environmental and economic systems. *Environ. Modell. Softw.* **2006**, *21*, 1247–1256.

16. Arribas-Bel, D.; Nijkamp, P.; Scholten, H. Multidimensional urban sprawl in europe: A self-organizing map approach. *Comput. Environ. Urban Syst.* **2011**, *35*, 263–275.

17. Adamczyk, J.J.; Kurzac, M.; Park, Y.S.; Kruk, A. Application of a kohonen's self-organizing map for evaluation of long-term changes in forest vegetation. *J. Veg. Sci.* **2013**, *24*, 405–414.

18. Giraudel, J.L.; Lek, S. A comparison of self-organizing map algorithm and some conventional statistical methods for ecological community ordination. *Ecol. Modell.* **2001**, *146*, 329–339.

19. Li, Q.; Yue, T.; Wang, C.; Zhang, W.; Yu, Y.; Li, B.; Yang, J.; Bai, G. Spatially distributed modeling of soil organic matter across China: An application of artificial neural network approach. *Catena* **2013**, *104*, 210–218.

20. Hoffmann, T.; Glatzel, S.; Dikau, R. A carbon storage perspective on alluvial sediment storage in the rhine catchment. *Geomorphology* **2009**, *108*, 127–137.

21. Anonymous. *IPCC Special Report on Land Use, Land-Use Change and Forestry*; Cambridge University Press: Cambridge, UK, 2000; p. 24.

22. Mitra, S.; Wassmann, R.; Vlek, P.L.G. An appraisal of global wetland area and its organic carbon stock. *Anglais* **2005**, *88*, 25–35.

23. Cierjacks, A.; Kleinschmit, B.; Kowarik, I.; Graf, M.; Lang, F. Organic matter distribution in floodplain can be predicted using spatial and vegetation structure data. *River Res. Appl.* **2011**, *27*, 1048–1057.

24. Baritz, R.; Seufert, G.; Montanarella, L.; van Ranst, E. Carbon concentrations and stocks in forest soils of europe. *For. Ecol. Manag.* **2010**, *260*, 262–277.

25. Harrison, A.F.; Howard, P.J.A.; Howard, D.M.; Howard, D.C.; Hornung, M. Carbon storage in forest soils. *Forestry* **1995**, *68*, 335–348.

26. Hofmann, G.; Anders, S. Waldökosysteme als Quellen und Senken für Kohlenstoff-Fallstudie ostdeutsche Länder. *Beitr. Forstwirtsch. Landsch.* **1996**, *30*, 9–16.

27. Kooch, Y.; Hosseini, S.M.; Zaccone, C.; Jalilvand, H.; Hojjati, S.M. Soil organic carbon sequestration as affected by afforestation: The darab kola forest (north of Iran) case study. *J. Environ. Monit.* **2012**, *14*, 2438–2446.

28. Lal, R. Forest soils and carbon sequestration. *For. Ecol. Manag.* **2005**, *220*, 242–258.

29. 2006 IPCC Guidelines for National Greenhouse Gas Inventories. *National Greenhouse Gas Inventories Programme*; Eggleston, H.S., Buendia, L., Miwa, K., Ngara, T., Tanabe, K., Eds.; IPCC National Greenhouse Gas Inventories Programme Technical Support Unit: Hayama, Kanagawa, Japan, 2006.

30. Beets, P.N.; Brandon, A.M.; Goulding, C.J.; Kimberley, M.O.; Paul, T.S.H.; Searles, N. The national inventory of carbon stock in New Zealand's pre-1990 planted forest using a LiDAR incomplete-transect approach. *For. Ecol. Manag.* **2013**, *280*, 187–197.

31. Smith, J.E.; Heath, L.S.; Hoover, C.M. Carbon factors and models for forest carbon estimates for the 2005–2011 National Greenhouse Gas Inventories of the United States. *For. Ecol. Manag.* **2013**, *307*, 7–19.

32. Olofsson, P.; Lagergren, F.; Lindroth, A.; Lindström, J.; Klemedtsson, L.; Kutsch, W.; Eklundh, L. Towards operational remote sensing of forest carbon balance across northern europe. *Biogeosciences* **2008**, *5*, 817–832.

33. Patenaude, G.; Milne, R.; Dawson, T.P. Synthesis of remote sensing approaches for forest carbon estimation: Reporting to the kyoto protocol. *Environ. Sci. Policy* **2005**, *8*, 161–178.

34. Gallaun, H.; Zanchi, G.; Nabuurs, G.J.; Hengeveld, G.; Schardt, M.; Verkerk, P.J. EU-wide maps of growing stock and above-ground biomass in forests based on remote-sensing and field measurements. *For. Ecol. Manag.* **2010**, *260*, 252–261.

35. Suchenwirth, L.; Förster, M.; Cierjacks, A.; Lang, F.; Kleinschmit, B. Knowledge-based classification of remote sensing data for the estimation of below- and above-ground organic carbon stocks in riparian forests. *Wetl. Ecol. Manag.* **2012**, *20*, 151–163.

36. Suchenwirth, L.; Förster, M.; Lang, F.; Kleinschmit, B. Estimation and mapping of carbon stocks in riparian forests by using a machine learning approach with multiple geodata. *Photogramm. Fernerkund. Geoinforma.* **2013**, *4*, 333–349.

37. Güneralp, I.; Filippi, A.M.; Randall, J. Estimation of floodplain aboveground biomass using multispectral remote sensing and nonparametric modeling. *Int. J. Appl. Earth Obs. Geoinforma.* **2014**, *33*, 119–126.

38. Lair, G.J.; Zehetner, F.; Fiebig, M.; Gerzabek, M.H.; van Gestel, C.A.M.; Hein, T.; Hohensinner, S.; Hsu, P.; Jones, K.C.; Jordan, G.; *et al.* How do long-term development and periodical changes of river-floodplain systems affect the fate of contaminants? Results from european rivers. *Environ. Pollut.* **2009**, *157*, 3336–3346.

39. Wagner-Lücker, I.; Lanz, E.; Förster, M.; Janauer, G.A.; Reiter, K. Knowledge-based framework for delineation and classification of ephemeral plant communities in riverine landscapes to support ec habitat directive assessment. *Ecol. Inf.* **2013**, *14*, 44–47.

40. Zehetner, F.; Lair, G.J.; Gerzabek, M.H. Rapid carbon accretion and organic matter pool stabilization in riverine floodplain soils. *Glob. Biogeochem. Cycles* **2009**, *23*, GB4004.

41. Cierjacks, A.; Kleinschmit, B.; Babinsky, M.; Kleinschroth, F.; Markert, A.; Menzel, M.; Ziechmann, U.; Schiller, T.; Graf, M.; Lang, F. Carbon stocks of soil and vegetation on danubian floodplains. *J. Plant Nutr. Soil Sci.* **2010**, *173*, 644–653.

42. Sandau, R. Status and trends of small satellite missions for earth observation. *Acta Astronaut.* **2010**, *66*, 1–12.

43. Schuster, C.; Förster, M.; Kleinschmit, B. Testing the red edge channel for improving land-use classifications based on high-resolution multi-spectral satellite data. *Int. J. Remote Sens.* **2012**, *33*, 5583–5599.

44. Rieger, I.; Lang, F.; Kleinschmit, B.; Kowarik, I.; Cierjacks, A. Fine root and aboveground carbon stocks in riparian forests: The role of diking, environmental gradients and dominant tree species. *Plant Soil* **2013**, *2*, 1–13.

45. Kohonen, T. Self-organized formation of topologically correct feature maps. *Biol. Cybern.* **1982**, *43*, 59–69.

46. Kohonen, T. *Self-organizing Maps*, 3rd ed.; Springer: Berlin/Heidelberg, Germany; New York, NY, USA, 2001; Volume 30.

47. Hall, P.; Park, B.U.; Samworth, R.J. Choice of neighbor order in nearest-neighbor classification. *Ann. Stat.* **2008**, *20*, 1236–1265.

48. Kanevski, M.; Maignan, M. *Analysis and Modelling of Spatial Environmental Data*; EFPL Press: Lausanne, Switzerland, 2004; p. 288.

49. Richter, K.; Atzberger, C.; Hank, T.B.; Mauser, W. Derivation of biophysical variables from earth observation data: Validation and statistical measures. *J. Appl. Remote Sens.* **2012**, *6*, 63557–63551.

50. Hastie, T.; Tibshirani, R.; Friedman, J.H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer: New York, NY, USA, 2009.

51. Tuominen, S.; Pekkarinen, A. Performance of different spectral and textural aerial photograph features in multi-source forest inventory. *Remote Sens. Environ.* **2005**, *94*, 256–268.

52. Rocchini, D.; Foody, G.M.; Nagendra, H.; Ricotta, C.; Anand, M.; He, K.S.; Amici, V.; Kleinschmit, B.; Förster, M.; Schmidtlein, S.; *et al.* Uncertainty in ecosystem mapping by remote sensing. *Comput. Geosci.* **2013**, *50*, 128–135.

53. Klobucar, D.; Subasic, M. Using self-organizing maps in the visualization and analysis of forest inventory. *J. Biogeosci. For.* **2012**, *5*, 216–223.

54. Hsu, A.L.; Halgamuge, S.K. Enhancement of topology preservation and hierarchical dynamic self-organising maps for data visualisation. *Int. J. Approx. Reason.* **2003**, *32*, 259–279.

55. Goetz, S.; Baccini, A.; Laporte, N.; Johns, T.; Walker, W.; Kellndorfer, J.; Houghton, R.; Sun, M. Mapping and monitoring carbon stocks with satellite observations: A comparison of methods. *Carbon Balanc. Manag.* **2009**, *4*, 2.