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.

Received: 1 April 2014; in revised form: 15 June 2014 / Accepted: 2 July 2014 / Published: 11 July 2014

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.
Keywords: organic matter; machine learning algorithm; neural network; Danube Floodplain; RapidEye; additional geodata; Austria

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, peatlands 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<sub>org</sub> stocks in soil or vegetation. To all of our knowledge, detailed C<sub>org</sub> maps of floodplain areas have seldom been produced, apart from Suchenwirth <i>et al.</i> [35,36] and Güneralp <i>et al.</i> [37].

In the presented study, we estimate C<sub>org</sub> 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 Stümser’s application of SOM and kNN for Thuringian forests [12], we consider the vegetation (above ground), soil (below ground) and total C<sub>org</sub> 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<sub>org</sub> models, as previous methods such as the derivation of C<sub>org</sub> 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<sub>org</sub> 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 (<i>Populus × canadensis</i>). Apart from these cottonwood plantations, the area is characterized by riparian vegetation, such as softwood forests (dominated by <i>Salix alba</i>, <i>Acer negundo</i>), hardwood forests (dominated by <i>Quercus robur</i>, <i>Fraxinus excelsior</i> and <i>Acer campestre</i>), as well as meadows and reed beds. Our study area (11.7 km²) 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<sub>org</sub> storage in the area was estimated at 359.1 Mg ha<sup>−1</sup> by Cierjacjks <i>et al.</i> [41], and as 428.9 Mg ha<sup>−1</sup> by Suchenwirth <i>et al.</i> [35]. Figure 1a presents a RapidEye scene of the Danube Floodplain area. Red color indicates pixels with high content of active biomass, <i>i.e.</i>, 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.

<table>
<thead>
<tr>
<th>Available geodata</th>
<th>Derived parameters</th>
<th>Abbreviations</th>
</tr>
</thead>
<tbody>
<tr>
<td>RapidEye image (1 August 2009)</td>
<td>Blue channel (440–510 nm)</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>Green channel (520–590 nm)</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>Red channel (630–685 nm)</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td>Red edge channel (690–730 nm)</td>
<td>RE</td>
</tr>
<tr>
<td></td>
<td>Near infrared channel (760–850 nm)</td>
<td>NIR</td>
</tr>
<tr>
<td>Digital elevation model</td>
<td>Elevation above river level</td>
<td>altitude</td>
</tr>
<tr>
<td>Ground water model</td>
<td>Ground water level</td>
<td>MGW</td>
</tr>
<tr>
<td>Topographic map 1:50,00 (ÖK 50)</td>
<td>Distance to river</td>
<td>distance</td>
</tr>
<tr>
<td>( C_{org} ) ground survey data from 2008 and 2010</td>
<td>Above ground carbon stocks</td>
<td>( C_{org_veg} )</td>
</tr>
<tr>
<td></td>
<td>Below ground carbon stocks</td>
<td>( C_{org_soil} )</td>
</tr>
<tr>
<td></td>
<td>Total carbon stocks</td>
<td>( C_{org_tot} )</td>
</tr>
</tbody>
</table>
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 $d_{(x_1, x_2)_p}$ 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 $C_{\text{org}}$ stocks per unit ($\text{Mg C}_{\text{org}} \text{ ha}^{-1}$), and (2) statistical estimates for vegetation, soil and total $C_{\text{org}}$ 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. $C_{\text{org}}$ Stock Estimations

$C_{\text{org}}$ stock maps of vegetation, soil and total carbon, -based on the SOM method are displayed in Figure 3a,b. $C_{\text{org}}$ maps based on the kNN method are presented in Figure 4a,b. $C_{\text{org}}$ 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 $C_{\text{org}}$ stocks color tones with higher values are in brown tones. All figures show the same detail of the area. The $C_{\text{org}}$ stocks is given in tons per ha ($\text{Mg C}_{\text{org}} \text{ ha}^{-1}$).

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 $C_{\text{org}}$ stock maps. It is apparent that $C_{\text{org}}$ 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 $C_{\text{org}}$. Stocks over 100 Mg $C_{\text{org}} \text{ ha}^{-1}$ 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 $C_{\text{org}}$, 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 $C_{\text{org}}$ 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 $C_{\text{org}}$ 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 $C_{\text{org}}$ stocks) and 46.14 Mg ha$^{-1}$ (total $C_{\text{org}}$ 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 $C_{\text{org}}$ stocks are independent from each other, so vegetation and soil $C_{\text{org}}$ stocks do not necessarily add up to total $C_{\text{org}}$ 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\textsubscript{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.
**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 $\delta_{\text{start}}$ of 100,000 and an end distance $\delta_{\text{end}}$ of 100, and have five iterations ($t_{\text{max}}$); kNN $k$: 5; Euclidean distance $d_{(x1,x2)}$: 2; distance weight $w_{(i,p)}$: 2).

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Approach</th>
<th>Vegetation C$_{org}$ stocks (average 149.65 Mg C ha$^{-1}$)</th>
<th>Soil C$_{org}$ stocks (average 192.1 Mg C ha$^{-1}$)</th>
<th>Total C$_{org}$ stocks (average 361.52 Mg C ha$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RapidEye</td>
<td>SOM</td>
<td>Bias 229.12 RMSE 146.99 % RMSE 3.01</td>
<td>Bias 113.26 RMSE 58.99 % RMSE 7.08</td>
<td>Bias 262.98 RMSE 70.85 % RMSE 7.08</td>
</tr>
<tr>
<td></td>
<td>kNN</td>
<td>39.52 RMSE 158.32 % RMSE 18.22</td>
<td>85.34 RMSE 48.27 % RMSE 7.08</td>
<td>81.26 RMSE 40.79 % RMSE 7.08</td>
</tr>
<tr>
<td>RapidEye + altitude</td>
<td>SOM</td>
<td>11.41 RMSE 143.29 % RMSE 0.28</td>
<td>108.22 RMSE 56.42 % RMSE 3.15</td>
<td>226.18 RMSE 63.11 % RMSE 3.15</td>
</tr>
<tr>
<td>+ MGW + distance</td>
<td>kNN</td>
<td>$-0.94$ RMSE 118.46 % RMSE $-4.28$</td>
<td>81.26 RMSE 40.79 % RMSE $-8.23$</td>
<td>196.66 RMSE 52.67 % RMSE $-8.23$</td>
</tr>
</tbody>
</table>

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_{\text{org}}, the bias is smaller for predictions using additional geodata.

4. Discussion

SOM and kNN have been applied for spatially explicit estimates of C_{\text{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_{\text{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_{\text{org}} distribution, while the kNN method results in rather average, homogeneous patterns.

In comparison with existing estimations, values for total C_{\text{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_{\text{org}} stocks [35].

In detail, estimations of vegetation C_{\text{org}} stocks (both kNN and SOM, based on satellite sensors and additional data) have an apparently higher RMSE than estimations of soil C_{\text{org}} stocks and total C_{\text{org}} stocks. This is not the case for bias, which is in several cases higher for the estimation of total C_{\text{org}} stocks. The RMSE of our estimations of soil (ranging from 40.79%–58.99%) and total C_{\text{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_{\text{org}}, where values range between 44.85% and 70.49%. RMSE for vegetation C_{\text{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 RMSE estimates for soil C_{\text{org}} and for estimations of vegetation and total C_{\text{org}} stocks based on RapidEye and additional data, yet has higher RMSE estimates for estimations of vegetation and total C_{\text{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 $C_{\text{org}}$, 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 OBIA's 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/$C_{\text{org}}$ 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 $C_{\text{org}}$). 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 $C_{\text{org}}$ 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 $C_{\text{org}}$ 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 $C_{\text{org}}$ stocks in
Another improvement for future research on the estimation of $C_{\text{org}}$ 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**

This study was funded by the German Research Foundation (DFG; project number KL 2215/2-1 and KL 2215/2-2). We acknowledge the DLR for RapidEye images as part of the RapidEye Science Archive (proposal 454). We would like to thank the staff of the Danube Floodplain National Park as well as Friederike Lang, Arne Cierjacks and Isaak Rieger for the provision of data. We would also like to thank Kyle Pipkins for checking our English.

**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**


© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).