A global predictive model of carbon in mangrove soils

Sunny L Jardine¹ and Juha V Siikamäki²

¹ University of Delaware, School of Marine Science and Policy, 308 Robinson Hall, Newark, DE 19716, USA
² Resources for the Future, 1616 P Street NW, Washington, DC 20036, USA

E-mail: jardine@udel.edu

Received 16 June 2014, revised 8 September 2014
Accepted for publication 10 September 2014
Published 14 October 2014

Abstract
Mangroves are among the most threatened and rapidly vanishing natural environments worldwide. They provide a wide range of ecosystem services and have recently become known for their exceptional capacity to store carbon. Research shows that mangrove conservation may be a low-cost means of reducing CO₂ emissions. Accordingly, there is growing interest in developing market mechanisms to credit mangrove conservation projects for associated CO₂ emissions reductions. These efforts depend on robust and readily applicable, but currently unavailable, localized estimates of soil carbon. Here, we use over 900 soil carbon measurements, collected in 28 countries by 61 independent studies, to develop a global predictive model for mangrove soil carbon. Using climatological and locational data as predictors, we explore several predictive modeling alternatives, including machine-learning methods. With our predictive model, we construct a global dataset of estimated soil carbon concentrations and stocks on a high-resolution grid (5 arc min). We estimate that the global mangrove soil carbon stock is 5.00 ± 0.94 Pg C (assuming a 1 meter soil depth) and find this stock is highly variable over space. The amount of carbon per hectare in the world’s most carbon-rich mangroves (approximately 703 ± 38 Mg C ha⁻¹) is roughly a 2.6 ± 0.14 times the amount of carbon per hectare in the world’s most carbon-poor mangroves (approximately 272 ± 49 Mg C ha⁻¹). Considerable within country variation in mangrove soil carbon also exists. In Indonesia, the country with the largest mangrove soil carbon stock, we estimate that the most carbon-rich mangroves contain 1.5 ± 0.12 times as much carbon per hectare as the most carbon-poor mangroves. Our results can aid in evaluating benefits from mangrove conservation and designing mangrove conservation policy. Additionally, the results can be used to project changes in mangrove soil carbon stocks based on changing climatological predictors, e.g. to assess the impacts of climate change on mangrove soil carbon stocks.

Online supplementary data available from stacks.iop.org/ERL/9/104013/mmedia

Keywords: mangroves, blue carbon, REDD, soil carbon, ecosystem services, predictive modeling

1. Introduction
Mangroves have long been recognized for the broad range of ecosystem services they provide, including serving as primary nursery habitat for many species of fish, crustaceans, birds, and marine mammals, and protecting coastal communities from coastal erosion and damage from storms and other natural hazards (Mumby et al 2004, Spalding et al 2010, Twilley et al 1996, Shepard et al 2011). More recently, mangroves have also received attention for their capacity to store large volumes of carbon (Donato et al 2011, Pendleton et al 2012, Siikamäki et al 2012). For example, on average, mangroves contain three to four times the mass of carbon typically found...
in boreal, temperate, or upland tropical forests (Donato et al. 2011). Much of this carbon storage, however, is at risk of being lost, because mangroves are among the most threatened and rapidly vanishing ecosystems globally, with habitat loss rates similar or greater to those in tropical forests (UN Food and Agricultural Organization 2007, Valiela et al. 2001).

Recent studies point to mangrove conservation as a potentially low-cost option for reducing CO\textsubscript{2} emissions (Pendleton et al. 2012, Siikamäki et al. 2012). For example, Siikamäki et al. (2012) find that in most mangrove areas of the world, protecting mangroves achieves emissions reductions at a lower cost than reducing emissions elsewhere in the economy. Accordingly, there is growing interest in developing and implementing market-based mechanisms such as carbon offsets, to credit mangrove conservation for associated emissions reductions, using a framework similar to the REDD (reduced emissions from deforestation and degradation) programs designed to protect tropical forests. The purpose of these programs is to provide market incentives to reduce emissions from deforestation by, for example, encouraging developing countries to reduce deforestation in return for compensation from developed countries committed to emission reductions (Angelsen 2008, Kindermann et al. 2008).

Designing and evaluating market mechanisms for mangrove conservation requires several spatially explicit scientific inputs, including information on the mangrove area susceptible to deforestation, carbon in mangrove biomass and soils, annual carbon sequestration, the emissions profiles of mangroves converted to other uses, and the opportunity cost of protecting mangroves (Siikamäki et al. 2013). A growing number of researchers have recognized the need for more and better science and are calling for continued research on the potential to include mangrove conservation in climate change policy (e.g. Mcleod et al. 2011, Pendleton et al. 2012, Siikamäki et al. 2013).

A key challenge in assessing the carbon benefits from mangrove conservation is the lack of rigorous spatial estimates of mangrove soil carbon stocks. Unlike other tropical forests, for which the bulk of carbon storage is in biomass, mangrove carbon is primarily stored in the soil. For example, Donato et al. (2011) estimate that soil carbon comprises 49–98% of carbon in mangrove forests. Siikamäki et al. (2012) developed the first spatial estimates of global mangrove soil carbon, using country and regional level mean estimates of soil carbon concentrations derived from the literature. While this provides an important first step, the estimates do not capture the fine-scale variation in mangrove soil carbon concentrations and, therefore, the fine-scale variation in potential benefits from mangrove conservation in different locations.

We address this data gap by developing a predictive model of mangrove soil carbon to estimate global mangrove soil carbon concentrations (mg C cm\textsuperscript{-2}) and stocks (Pg C) on a high-resolution grid (5 arc min). Our predictive model is based on soil carbon measurements compiled in the meta-analyses by Chmura et al. (2003), Kristensen et al. (2008), and Donato et al. (2011). These three studies combined include data from over 900 samples collected in 28 countries, which contain 64.4% of global mangroves. We explore several prediction methods, including machine-learning algorithms, to assess the generalizability and predictive performance of alternative model specifications.

Our findings contribute to the science needed to accurately quantify the benefits from emissions reductions achieved with mangrove conservation. Combining results from our predictive model with information on the spatial distribution of above-ground biomass (e.g. Hutchison et al. 2013) enables improved estimation of the magnitude and spatial distribution of mangrove carbon storage, thereby aiding the design and evaluation of mangrove conservation projects. Additionally, because we find that climatological factors are important predictors of mangrove soil carbon concentrations, our predictive model can be helpful in assessing the impacts of climate change on mangrove carbon storage.

2. Methods

2.1. Data

We use mangrove soil carbon measurements from meta-analyses by Chmura et al. (2003), Kristensen et al. (2008), and Donato et al. (2011). Combining data from these sources, after removing observations where the study location is not documented, yields a total of 932 samples that were collected in 28 countries by 61 independent studies. The locations of the soil carbon samples are shown in figure 1.

Observations included in Chmura et al. (2003) and Donato et al. (2011) are measurements of the soil carbon bulk density, while Kristensen et al. (2008) record the per cent organic carbon (POC) of each soil sample. To combine the observations, we convert POC measurements into bulk density measurements following Donato et al. (2011) with the estimated relationship:

\[ \text{poc} = 3.0443 \text{bd}^{-1.313} \]  \hspace{1cm} (1)

where poc is the per cent weight of organic carbon in the soil and bd is the soil bulk density measured in grams per cubic centimeter. We then use the soil bulk density to calculate the soil carbon content in g C per cubic centimeter.

We combine the soil carbon dataset with information on several predictor variables that explain the carbon concentration in mangrove soils, including the distance of the observation’s sampling location from the equator (the absolute value of latitude coordinates to a 0.1 of a degree), several variables describing climate conditions at the sampling locations, and regional indicators. The data are available in the online supplementary materials at stacks.iop.org/ERL/9/104013/mmedia.

For the majority of the soil carbon data (76% of the observations), source studies report precise locational data, i.e. latitude coordinates reported to the 0.1 of a degree, which we are able to use directly. In some instances (9% of the observations), a group of soil carbon samples were taken from neighboring sites and only the boundary latitude and longitude coordinates are provided. In these cases we use the mean latitude coordinate value for each of the soil carbon estimates in that study. In studies where only imprecise
Locational data were provided, but a detailed map of the sample location was included (13% of the observations), we manually obtain a more precise location. Finally, where locational data were reported in the meta-analysis but could not be verified (e.g. unpublished data), we use the locational data reported in the meta-analysis (2% of the observations).

The latitude data are used to calculate distance from the equator, which has been linked to mangrove productivity (Twilley et al. 1992), and therefore may be an important predictor of mangrove soil carbon, a hypothesis which we test in our analysis.

Climate variables are from the WorldClim Bioclim data (Hijmans et al. 2005). Of the 19 variables included in the Bioclim data, we use mean annual temperature (Bioclim 1), mean temperature in the coldest quarter (Bioclim 11), total annual precipitation (Bioclim 12), and seasonality in precipitation (Bioclim 15). Our choice of climate variables is motivated by a large body of literature linking mangrove productivity with temperature and precipitation (Ellison 2003, Field 1995, McKee 1993) and extreme cold events (Cavanaugh et al. 2013, Snedaker 1995, Woodroffe and Grindrod 1991), implying that these climate variables should be included in our model.

Finally, we include regional indicators, where regions are defined according to the ten biogeographic regions for mangroves developed by Spalding et al. (2010) (see the SI for countries represented in the data from each region). Our geographic predictors (distance from the equator and regional indicators) control for the impacts of unobserved factors on the carbon concentration in mangrove soils, to the extent to which the unobserved factors vary over space. For example, mangrove species are more homogeneous within mangrove bioregions (Spalding et al. 2010), mangrove soil carbon concentrations may also be more homogeneous within mangrove bioregions, which we control for with regional indicators.

2.2. Predictive models

We develop two classes of statistical models to predict soil carbon concentrations, including parametric predictive models and models developed with machine learning algorithms. Constructing a robust predictive model requires balancing the ability to explain the most variation in the data sample with the model’s generalizability, i.e. its ability to predict out of sample (Babyak 2004). We evaluate the alternatives models along these two dimensions.

2.3. Parametric prediction

We first develop a parametric predictive model with regression analysis. Specifically, we estimate the following model:

\[
sc_i = \alpha + \beta \text{avlatitude}_i + \beta \text{climate}_i + \delta \text{region}_i + \epsilon_i,
\]

where \(sc_i\) is soil carbon content in sampling location \(i\), \(\text{avlatitude}_i\) is the absolute value of latitude for location \(i\) (distance from the equator), \(\text{climate}_i\) is a vector containing all Bioclim variables for site \(i\) and a squared term for average annual temperature based on evidence of a nonlinear relationship between mangrove soil carbon and temperature (e.g. Gilman 2008), and \(\epsilon_i\) is an assumed mean-zero normally distributed random error. We estimate several specifications of equation (2), where certain parameters or sets of parameters are constrained to equal zero. The model is estimated using ordinary least squares (OLS).

We compare the parametric predictive models based on R-squared statistics, signaling in-sample predictive power, and the Akaike information criterion (AIC), indicating model generalizability as it relates to model simplicity.
2.4. Machine learning prediction

Machine-learning (ML) algorithms offer an alternative to single-equation parametric predictive modeling. They involve systematic computational learning in the model-building process, which improves the model’s predictive performance and generalizability by minimizing overfitting. The ML methods used here are also flexible in modeling nonlinearities and/or interactions between the predictors. ML algorithms are especially useful in instances where the functional relationship between the outcome variable and the predictors is unknown, as is the case with mangrove soil carbon concentrations (Kristensen et al. 2008). ML methods have a long history in the field of medical biosciences (Kononenko 2001) and are rapidly gaining ground in several fields ranging from economics (Varian 2013), to conservation biology and ecology (Gomes 2009, Dietterich 2009).

We examine two ML algorithms as potential alternatives to our parametric predictive model of mangrove soil carbon: a boosted decision tree (DT) model and a bag DT model (see Hastie et al. 2009 and Breiman 1996 for a detailed description of the two methods respectively). Both are tree-based methods, which are well suited for complex ecological data (De’ath and Fabricius 2000), but the relative performance of each method depends on the dataset (Quinlan 1996).

Inputs to the ML models are identical to those in the parametric model. To compare the two ML algorithms with our parametric model, we test the out-of-sample predictive power of each model. In the out-of-sample prediction tests, we first draw 1,000 datasets by taking random samples of 30 studies (and all observations from those studies) from our original dataset, without replacement. Next, we estimate the machine learning and parametric predictive models for each of the 1,000 datasets. Finally, we use the estimated models to predict soil carbon concentrations for observations in the remaining 31 studies for each of the 1,000 datasets and calculate prediction errors. The magnitude of prediction errors indicates how sensitive the model is to the studies included in our analysis. All models were fitted in MATLAB (2013) version 8.1.0.604, using the fitensemble package in the statistics toolbox.

3. Results

3.1. Parametric prediction

Regression results are reported in table 1. We begin with a simple constant-only model showing that the mean soil carbon content in our sample is 32.14 mg cm$^{-3}$ (min. 13.48 mg cm$^{-3}$, and max. 115 mg cm$^{-3}$), or approximately an average of 321 Mg (tons) C ha$^{-1}$ in the top meter of soil. We gradually add controls to the model and find that climatic variables are significant predictors of mangrove soil carbon as are the regional controls. The regional controls contribute substantially to the explanatory power of the model. For example, the R-squared for Model 3 with climatic variables but no regional controls (fixed effects) is 0.103, but after adding the regional controls the R-squared is 0.912 (Model 4). Adding regional controls also lowers the AIC, suggesting that overfitting is not driving the increased R-squared statistic.

Figure 2 (left panel) shows the mean soil carbon by region, and the 95% confidence intervals around the means, with and without controls for climate and distance from the equator. The data show that mangroves in North and Central America contain some of the most carbon-rich soils whereas mangroves in East Asia are among the most carbon-poor soils. Soils in South East Asia, where a large fraction of the world’s mangroves are located (approx. 32.8%), have considerably greater carbon content than mangroves soils in East Asia but substantially less carbon content than mangrove soils in North and Central America. However, regional constants are no longer statistically different from each other after controlling for climate and latitude (right panel of figure 2).

Figure 2 (left panel) shows the mean soil carbon by region, and the 95% confidence intervals around the means, with and without controls for climate and distance from the equator. Uncontrolled regional means are obtained by regressing soil carbon concentrations on the full set of regional indicator variables without any other controls. The data show that mangroves in North and Central America contain some of the most carbon-rich soils whereas mangroves in East Asia are among the most carbon-poor soils. Soils in South East Asia, where a large fraction of the world’s mangroves are located (approx. 32.8%), have considerably greater carbon content than mangroves soils in East Asia but substantially less carbon content than mangrove soils in North and Central America. However, regional constants are no longer statistically different from each other after controlling for climate and latitude (right panel of figure 2).

Model 4, which controls for latitude, climatic conditions, and sample region, outperforms all other parametric predictions explored here in terms of attaining the highest R-squared and lowest AIC. Therefore, we select it as our preferred parametric specification and compare it to the two ML models.
3.2. Machine learning predictions

Table 2 contains results from our out-of-sample predictive test, where we compare the performance of our preferred parametric prediction model (Model 4) and two competing ML algorithms (the boosted decision tree and the bag decision tree). The average sum of squared errors (SSE) and mean percentage errors (MPE) for each model are summarized in the table.

The results indicate that both ML methods outperform our preferred parametric predictive model by achieving smaller average out-of-sample SSE and MPE. The improved ability of ML models to predict out of sample suggests that they offer distinct advantages in modeling mangrove soil carbon. Of the two ML methods, the bag DT performs better than the boosted DT, so we select it as our preferred ML model.

Communicating and further using ML model results can be challenging, as the model cannot be collapsed into an equation. However, in contrast to parametric models we can calculate the relative importance of each predictor variable in the ML model (table 3) to gain insight into which predictors explain the greatest amount of variation in the data. We find that annual precipitation is the most important predictor, followed by distance from the equator and geographic region. Precipitation seasonality, mean temperature of the coldest quarter, and annual mean temperature explain less of the variation in mangrove soil carbon.

| Method       | Average SSE | Average MPE |
|--------------|-------------|-------------|
| OLS (Model 4)| 5.69 e05    | 53.84       |
| Boost DT     | 1.27 e05    | 36.99       |
| Bag DT       | 0.844 e05   | 32.26       |

Table 2. Comparison of OLS and ML models using the out-of-sample sum of squared errors (SSE) and mean percentage error (MPE), on average, from 1 000 random divisions of independent studies in the data.
Table 3. Predictor importance in the bag DT ML model. The importance score is calculated as the change in mean squared error due to splits on each predictor and then dividing the sum by the number of branch nodes. The relative importance normalizes the importance scores by the sum of importance scores over all predictors.

| Variable name | Variable description | Relative importance |
|---------------|----------------------|---------------------|
| Bioclim 12    | Annual precipitation | 0.2092              |
| AV Latitude   | Absolute value of latitude | 0.1967          |
| Region        | Geographical region  | 0.1732              |
| Bioclim 15    | Precipitation seasonality | 0.1532         |
| Bioclim 11    | Coldest quarter mean temperature | 0.1347   |
| Bioclim 1     | Annual mean temperature | 0.1331          |

Using our ML predictive model along with mangrove land cover data from Giri et al. (2011), we construct a global dataset (available online from the authors) of estimated mangrove soil carbon concentrations (mg C cm⁻²) and stocks (Pg C) on a high-resolution grid (5 arc min). Standard errors for all of our estimates are calculated from the standard deviation of predictions from 10,000 ML models constructed from 10,000 random samples (bootstrap samples) of our primary soil carbon dataset. Figure 3 maps the predicted soil carbon concentrations for the world’s mangroves and figure 4 maps mangrove soil carbon concentrations in Indonesia (and neighboring countries), a country that contains roughly 19.5% of the world’s mangroves. The figures illustrate considerable spatial variation in mangrove soil carbon concentrations.

To estimate global carbon stocks we assume a carbon-rich soil depth of 1 meter, as common in the literature (e.g. Donato et al. 2011, Pendleton et al. 2012, Siikamäki et al. 2012). Table 4 lists the estimated global and country-level stock for the top-20 countries. Globally, we estimate that mangrove soils contain 5.00 ± 0.94 Pg C and that about 80.5% of the pool is contained in 20 countries, which is roughly proportional to the per cent of the world’s mangroves these countries contain (81.3%).

Our results document considerable geographic variation in soil carbon. We estimate that global mangrove soils contain 369 ± 6.8 Mg C ha⁻¹ on average (in the top meter). However, we estimate that the amount of carbon per hectare in the world’s most carbon-rich mangroves (the highest grid cell prediction is 703 ± 38 Mg C ha⁻¹) is roughly 2.6 ± 0.14 times the amount of carbon per hectare in the world’s most carbon-poor mangroves (the lowest grid cell prediction is 272 ± 49 Mg C ha⁻¹). We also find substantial within-country variation in mangrove soil carbon. For example, in Indonesia, the country with globally the largest mangrove soil carbon stock, we estimate that the most carbon-rich mangroves contain 1.5 ± 0.12 times as much carbon per hectare as the most carbon-poor mangroves.

When examining country averages, we find that the country with the highest average soil carbon concentration has roughly twice the amount of soil carbon, per hectare, as the country with the lowest estimated soil carbon concentration. Interestingly, none of the top 20 countries, when ranked by soil carbon concentrations, overlap with the 20 countries with the largest carbon pools (table 4). In fact, the 20 countries with highest average soil carbon concentrations contain only 1.2% of the world’s mangroves, but as a consequence of their relatively high soil carbon, these countries account for 1.5% of global mangrove soil carbon.

4. Discussion

Here we develop a predictive model and global dataset of soil carbon concentrations, which documents and provides information on the spatial distribution of mangrove soil carbon. Our results indicate that the variation in soil carbon is systematically determined by several climatic variables but that locational variables are also significant predictors of mangrove soil carbon.

Some of our results call for added discussion. First, although Chmura et al. (2003) find that soil carbon decreases with temperature, in our analysis, which includes and expands beyond the data in Chmura et al. (2003), the (parametric) results show that, on average, carbon content increases with temperature, albeit at a decreasing rate. Our results are consistent with studies that find mangrove productivity increases with temperature up to a threshold (Ellison 2003, Field 1995).

Second, our findings are consistent with several studies that find that extreme cold events have a significant impact on mangrove productivity (Cavanaugh et al. 2013, Snedaker 1995, Woodroffe and Grindrod 1991). For example, our (parametric) results show that a 1 degree C increase in temperature during the coldest quarter leads to an increase of 5.4 mg cm⁻³ of soil C, all else constant, or a 16.8% increase relative to the mean value of mangrove soil carbon concentrations in our sample.

Third, we find that ML algorithms perform substantially better than simple parametric predictions in predicting out of sample. Therefore, they may offer substantial gains in accurately estimating mangrove soil carbon concentrations based on available data. On the other hand, the parametric model is exceedingly practical for predictions and, as such it offers a useful first-order approximation.

Finally, to obtain the highest-quality predictive model given available data, we explored several predictive modeling techniques and compared their performance along several dimensions. Nonetheless, as with any predictive model, the quality of the predictions depends on the quality of the underlying data. For example, if the secondary observations used in this analysis targeted sampling locations based on characteristics that are unobservable to us but correlated with soil carbon concentrations (e.g. oversampling of pristine mangrove forests), then the underlying data would not be representative of the population and any predictive model would contain selection bias. The possibility that soil carbon measurements oversample pristine mangrove locations is speculated about in Kristensen et al. (2008) and Hutchinson et al. (2013). Although there is no actual evidence of sampling bias, it cannot be completely ruled out.
5. Conclusions

This analysis adds to the science necessary to design and evaluate mangrove conservation options. We develop a model to predict mangrove soil carbon, explaining substantial spatial variation in the carbon concentrations of global mangrove soils. Our predictive model is based on a rich dataset of mangrove soil carbon measurements, including over 900 observations collected in 28 countries throughout the world, which represent the majority of global mangroves. Using
the model predictions, we produce a high-resolution and spatially explicit global dataset of mangrove soil carbon concentrations. These data can help examine current mangrove conservation projects and direct future mangrove conservation efforts, thus providing an important scientific input to mangrove conservation assessments.

Because mangrove soil concentrations are determined by climate conditions, our predictive model can also help assess the impacts of a changing climate on carbon in mangrove soils, i.e. our model can be used to predict changes in mangrove soil carbon concentrations that result from changing climate conditions. This will allow for a more complete understanding of the impacts of climate change on mangroves.

Mangrove carbon storage varies substantially over space; therefore, the benefits from mangrove conservation depend critically on the location of the mangroves conserved. In principle, our results enable the targeting of mangrove conservation to maximize benefits from avoided carbon emissions. We note, however, that a more meaningfully defined conservation strategy should consider the full range of benefits from mangrove conservation, not only avoided carbon emissions. Moreover, it is not clear a priori how similar or different a more multi-objective targeting strategy would be relative to a carbon-focused targeting. Although previous research (Siikamäki et al 2012) suggests that, in general, carbon-focused mangrove conservation will target areas that are also high in biodiversity, the relationship between carbon and biodiversity may vary at a finer spatial scale than has been considered thus far in the literature. Additionally, it is unclear whether the many other benefits, such as shoreline protection or the provision of nursery habitat for fish, from mangrove conservation are strongly and positively correlated with the potential for carbon offsets in mangroves. Therefore, to develop a more comprehensive understanding of mangrove conservation, future work is needed to evaluate the full array of ecosystem services that mangroves provide. Our results can be an important input into these future comprehensive assessments.

Acknowledgements

We thank Daniel Donato and Erik Kristensen for their assistance with the data. We also thank James Sanchirico for his support.

References

Angelsen A 2008 Moving Ahead with REDD: Issues, Options and Implications (Bogor, Indonesia: CIFOR)
