Biogeosciences Discuss., 9, C6994–C6997, 2013
www.biogeosciences-discuss.net/9/C6994/2013/
© Author(s) 2013. This work is distributed under
the Creative Commons Attribute 3.0 License.

Biogeosciences
Discussions

Interactive comment on "Modeling soil bulk density at the landscape scale and its contributions to C stock uncertainty" *by* K. P. Taalab et al.

T. Wutzler (Referee)

twutz@bgc-jena.mpg.de

Received and published: 3 January 2013

The paper shows how topsoil bulk density can be inferred independent of organic carbon concentration from other climate and landscape information, and how this might change predicted soil carbon stocks and their uncertainty. This is a significant contribution to the soil carbon stock inventory community. Before publishing the manuscript, however, a few points should be put more clearly:

1) What results are based on the training and what on the validation dataset?

2) Comparison against calculations of unweighted average density across the entire study area is unfair. Please, concentrate on the soilscape comparisons. For the total

study area, a better solution is to compare against the area-weighted average of the carbon stocks based on average soilscape bulk density.

3) Error propagation of single pixels to areas is not described sufficiently.

1 General comments

To point 1):

The methods section describes how the data is split into training and validation dataset. However in the results and discussion it is not clear to me what results pertain to the training data and which to the validation data. It even seems to me that all the presented results pertain only to the training dataset. At the current presentation it is hard to get an impression of the best fit and the danger of overfitting.

To Point 2):

Calculating C-stocks with just one unweighted average bulk density over a broad heterogeneous area of course will lead to biased results, and I would reject such a study. One must stratify the bulk density measures or carbon stocks to areas of similar properties (e.g. soilscapes) and then aggregate the results be area weighted averages (explained e.g. with eq 1 and 2 in Wutzler et al 2006). My guess is that the carbon stocks then will be of comparable magnitude.

To point 3)

Eq. 5 describes error propagation for a single measurement point or a single pixel. How were the errors propagated to the estimates of soilscapes or total area? One cannot assume independence of the single pixels as they are predictions of the same model with uncertain parameters. One approach how to deal with this with an ordinary mixed nonlinear regression model in a different conext see e.g. appendix A2 of Wutzler et al. 2008). How was covariance between OC and D_b in eq. 5 derived?

2 Specific comments

P18836 L12: Why two numbers of landscape subdivisions? Which one was used – both? Table 1: What are $Greatgroup$, $AT0_{annual}$, FCD_{med}? A table with a short description of all predictors would be helpful (maybe include two columns for their ranking in the RF-A and RF-S models).

P 18838 L4ff. It did not become clear who derived the soilscapes I miss discussion of covariance between predictions. Especially for the soil classifications of high predictor importance, please discuss their derivation. It is only based on texture or are there topographic arguments, vegetation etc, which factors do they covary with?

I also miss a discussion about good models by chance. How many predictors/models did you compare? If you compare enough models, several will be good despite they are actually not related to the observed patterns.

Fig 3b and 3c: In order to appreciate the information contained in the mean residual, can you plot a distribution of residuals?

Table 2: I would prefer quantiles in absolute values instead of percentages.

Is is possible to compare results to studies that make use of neighborhood of sampling points, e.g. Kriging? Can this somehow be combined with the presented machine learning approaches?

3 Technical comments

18849 L11ff: The locations of table 3 have not been introduced yet and I got confused. Maybe move this section further down.

Fig 4 Panels b and c are redundant to Table 3, Shapes of Fig 4a might be included in Fig 2. Hence, Fig 4 is not necessary.

Fig 3: Adding labels directly to the fig instead of the caption would be helpful (A horizon, subsoil, RF model, ANN model)

4 Cited literature

Wutzler, T.; Köstner, B. Bernhofer, C. (2006) Spatially explicit assessment of carbon stocks of a managed forest area in Eastern Germany. European Journal of Forest Research, 126, 371-381

Wutzler, T.; Wirth, C. Schumacher, J. (2008) Generic biomass functions for Common beech (Fagus sylvatica L.) in Central Europe - predictions and components of uncertainty. Canadian Journal of Forest Research, 38, 1661–1675

Interactive comment on Biogeosciences Discuss., 9, 18831, 2012.