SOIL
Discussions

Open Access

Interactive
Comment

Interactive comment on "Comparison of spatial association approaches for landscape mapping of soil organic carbon stocks" *by* B. A. Miller et al.

Anonymous Referee #2

Summary and general comments

This paper applies the Cubist modelling software to predict the spatial distribution of SOC stock across a study area in Germany. Two approaches are compared. In the direct approach, SOC stock is first calculated for the 117 profiles and subsequently modelled and predicted at unsampled locations. In the indirect approach, the SOC %, bulk density, percent of particles >2mm, and depth are first modelled and predicted at unsampled locations, before the predictions are combined to calculate SOC stocks; the uncertainty of the component predictions is also propagated in this approach. Maps of predictions and their estimated uncertainties are presented and there is a detailed discussion of the selected predictors for both approaches.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

I think that the paper provides a worthy contribution to the body of research on SOC modelling and prediction. However, I have a few questions and comments that I would like to see addressed before publication.

Specific comments

I wonder about the effect of the 10 closely-paired samples. . .in addition to ensuring the estimated errors capture random variation, could it induce a bias towards these sample points?

I am a little bit worried about the large number of potential predictors in the pool compared to the number of data on the target variable (117). I appreciate that the authors use the discussion to suggest explanations of why particular predictors were selected, and thus partly validate their selection, but am still not totally convinced that the same could not be done even with junk data and this many potential predictors. I wonder if some acknowledgement of the potential of data-mining software to overfit should be included and commented on. I don't think Cubist does anything to deal with the size of predictor pool (in a multiple hypothesis testing kind of way). . .a comment on this issue could be useful.

As you state in the methods section, for the propagation of error in the indirect approach, the variances and covariances should be those of the residuals from the fitted models, not of the data themselves. It seems that this is what was done, but lines 5 and 6 of page 770 made me wonder if the variances and covariances of the raw data had been used. Could you clarify this, as this could be an alternative explanation of the larger uncertainties resulting from the indirect approach?

Define f in Equations 2 and 3, and explain exactly what |f| is.

In the cross validation (Table 5), I am not sure why the results for predictions of SOC stock by the indirect approach are omitted. I think the table should include these.

I think it would also be good to provide some validation of the uncertainties. . .I appre-

ciate the difficulties of validating with a small, clustered dataset such as this, but I think it would be worth including some measure of the adequacy of uncertainty assessment in the cross validation. One possibility is the mean of the theta-statistic, which should be close to 1 (see e.g. Lark, RM. 2000. A comparison of some robust estimators of the variogram for use in soil survey. European Journal of Soil Science, 51, 137-157).

It is quite interesting that although the ME for all subsoil component variables was <1, the resulting predictions of SOC stock gave a ME of 1.67. This is worth commenting on in Section 3.1.1.

I think that the residuals for all variables are assumed normal. . . however, depending on the dataset, it may be more appropriate to model log SOC % as normally distributed. Some comment about this, and about the effect that this could have on predictions and uncertainties in the indirect approach, could be useful.

Is a conservative estimate of the spatial distribution the best thing? The most conservative would be to use the mean across the entire study area, but this would not be very useful. I am not sure whether the paper is recommending that the more conservative approach should be used, or just saying that the direct approach is more conservative than the indirect approach.

What exactly is meant by the 'spatial association approach'?

Were all soil profiles deeper than 2 m?

Page 767, sentence starting on line 27: 'correlation . . . of R2 = 0.59'. Correlation should be measured by r, not R2. . .reword this sentence.

Page 768, line 6: direct R2 = 0.14, but in Table 4 is 0.19. . .is this correct?

Figures 2, 3 and 4: I am not sure that the hillshade effect helps. I found it difficult to distinguish between the effect of the hillshade and the SOC stock differences. I would suggest removing this effect.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

Interactive comment on SOIL Discuss., 1, 757, 2014.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper