Answers to comments by anonymous referee #1 (RC C2487) of 31 January 2014

M. Nussbaum, A. Papritz, A. Baltensweiler, L. Walthert

March 18, 2014

Thank you for your positive feedback. We comment on your suggestions for changing our article in the subsequent text. Your main criticism relates to the fact that we used only one statistical approach in our analysis and did not include any methodological comparison of results obtained by our and other statistical approaches (regression kriging [RK] and machine learning method [ML] methods, in particular).

We agree that such a comparison would be of some interest as there are only few published studies comparing ML with geostatistical methods to date. However, our main intention when writing the paper was to document our new robust external-drift kriging (EDK) approach in detail to make it fully transparent to the reader how we calculated mean soil organic carbon (SOC) stocks. Hence, we see no easy way to shorten the text to compensate for the inclusion of new material. Adding additional material would therefore likely extend the paper beyond acceptable length.

When selecting an approach for the analysis of our data we abstained from using RK and ML methods for the following reasons:

RK is an ad hoc approach and a (theoretical) sub-optimal variant of EDK. Regression coefficients are estimated by ordinary least squares (OLS) that does not — as opposed to generalized least squares (GLS), which is used in EDK — take into account the correlation structure in the residuals when estimating the regression coefficients.

The differences of spatial predictions by RK and EDK are expected to be small (e.g. Minasny and McBratney, 2007). However, the quantification of prediction accuracy — a main focus of our paper — remains likely too optimistic with RK. Two sources of uncertainty contribute to the standard errors of plug-in EDK predictions (i.e. when the uncertainty of the variogram parameters is ignored): (1) The uncertainty about the estimated regression coefficients and (2) the uncertainty arising from spatial interpolation of the regression residuals (simple kriging variance). RK accounts only for the second component (often in an inconsistent way as mostly ordinary [OK] instead of simple kriging is used). The RK prediction standard errors are therefore very likely too small.

One could combine the OK variances with the covariance matrix of the regression coefficients as returned by the OLS fit in the same way as EDK does. However uncertainties of regression coefficients are notoriously underestimated by OLS in case of spatially correlated residuals. Furthermore, why should one want to compensate in an ad hoc manner for a flaw of RK when EDK provides a consistent solution to the problem?

ML methods do not — unlike kriging — provide consistent approaches to handle change-of-support. Therefore, the quantification of the uncertainty of regional and national SOC stock estimates with ML remains difficult. To report standard errors of such estimates was a main objective of our work. Their calculation requires the covariance matrix of the point prediction errors (see P7092 L1-20, section 2.3.4). So far, it remains unclear how this covariance matrix could be computed for Random Forest or boosted classification and regression trees (BRT). Martin *et al.* (2011) who used BRT for mapping SOC stocks across France did not provide a correct answer to this problem.

The specific comments are very welcome as the suggested changes clearly improve the clarity of the text. We accepted most of the suggestions, and we comment below only on those points where we (partly) disagree and did not change the text as proposed.

P7085 L8 Rename the title as Soil bulk density To our understanding "bulk density" refers to the density of whole soil including rock fragments (particles with diameter $>$ 2 mm). As mentioned on P7085 L9 we considered the density of soil fraction with particle size $\leq$ 2 mm (corresponding to the term "Feinerde" in the Swiss and German soil taxonomy). The use of the term "bulk density" would therefore be misleading.

P7086 L9–16 Change of formula for calculating SOC stocks We use the formula to explicitly state how the correction of the soil density for rock fragment content was done. In your suggestion this is mentioned by using "ρ_0 the soil bulk density corrected for rock fragments" and does not include details. However, we changed the wording of the paragraph and avoid now the term "volume of horizon per unit area" and use "thickness of horizon" (denoted by symbol D_i) instead.

P7095 L11–24 Shift sub-section to Material and methods Here we disagree. This subsection describes the structure of the fitted models, and this is clearly a result of our work.

Table 1 Covariates were selected with the aim to optimize the predictive power of the models. The reason for different terrain covariates being chosen for top- (0–30 cm) or bulk soil (0–100 cm) is simply because they were better predictors in either model. In general, we strove in model building for parsimonious models containing as few covariates as possible, and we took care to not include any covariate where the sign of its coefficient would contradict the dependency expected by a soil scientist. Thus, we do not believe that we can causally interpret the results of the regression analysis — this is notoriously difficult for models fitted to observational data — and see therefore no need to extend the respective part of the discussion.

Table 2 R^2 (including its robust variant) is a relative measure to describe the strength of the linear dependence of observed and predicted values. It does not depend on the variance of the data, unlike the root mean squared error (RMSE), which is an absolute measure for the precision of the predictions and thus depends on the variance of the observations. In our analysis, we used the *relative* RMSE, obtained by standardizing the prediction errors by the observations. Hence, we expect the relative RMSE to depend on the coefficient of variation (CV) of the data.

Considering the CVs of SOC stocks in the validation set (Table S1 in Supplement) we can see that a larger CV for the compartment 0–100 cm (0.63) than for 0–30 cm (0.51) is in accordance with larger relative RMSE for 0–100 cm (0.56) than for 0–30 cm depth (0.49, Table 2). The ratio of relative RMSEs for 0–100 cm and 0–30 cm (1.14) is somewhat smaller than the respective ratio of CVs (1.23) which is what we qualitatively expect if the R^2 is larger for 0–100 cm than for 0–30 cm soil depth.

References

Martin, M. P., Wattenbach, M., Smith, P., Meersmans, J., Jolivet, C., Boulonne, L., and Arrouays, D. (2011). Spatial distribution of soil organic carbon stocks in France. *Biogeosciences*, **8**(5), 1053–1065.

Minasny, B. and McBratney, A. B. (2007). Spatial prediction of soil properties using eblup with the matérn covariance function. *Geoderma*, **140**(4), 324–336.