SOIL Discuss.,
https://doi.org/10.5194/soil-2017-30-SC1, 2018

Interactive comment on "Hot regions of labile and stable soil organic carbon in Germany – Spatial variability and driving factors" *by* Cora Vos et al.

L. Cécillon

lauric.cecillon@irstea.fr

Received and published: 15 January 2018

Comment on "Hot regions of labile and stable soil organic carbon in Germany - Spatial variability and driving factors" by Vos et al. (SOIL Discuss.)

Lauric Cécillon1,2, Bernard Barthès3, Suzanne Lutfalla2, Laure Soucémarianadin2, Pierre Barré2

1 Université Grenoble Alpes, Irstea, UR LESSEM, 2 rue de la Papeterie, 38402 St-Martin-d'Hères, France 2 Laboratoire de Géologie, PSL Research University, CNRS-ENS UMR 8538, Ecole Normale Supérieure, 24 rue Lhomond, 75231 Paris CEDEX 5, France 3 IRD, UMR Eco&Sols, Montpellier SupAgro, 2 place Viala, 34060 Montpellier Cedex 2, France

In their interesting and stimulating paper, Vos and colleagues aimed at studying the distribution of labile and stable soil organic carbon (SOC) fractions in a large set of agricultural (cropland and grassland) topsoils (0–10 cm depth) from the German Agricultural Soil Inventory (2900 sites).

They define the labile SOC fraction as low density (d < 1.8) particulate organic matter (POM) carbon, while the stable SOC fraction is defined as high density (d > 1.8) mineral-associated organic matter (MOM) carbon. Both labile and stable SOC fractions were isolated by a soil organic matter density fractionation scheme on a subset of 145 samples.

They then calibrated using this subset of 145 samples a multivariate regression model relating the absolute content (g/kg) and proportion (%) of SOC in the POM and in the MOM fractions to soil absorbance in the near-infrared spectral domain (predictor variables; 1300–3300 nm). They used a leave-one-out cross-validation to validate their model, and briefly reported the cross-validated predictive performance of the model in the Material & methods section of the paper and in the Supplementary Figure S1. The authors mentioned that more details of the multivariate regression model based on near-infrared spectroscopy can be seen in a submitted paper (Jaconi et al., submitted) with no reference to the journal and/or submission tracking number of this manuscript.

The authors then used their multivariate regression model based on near-infrared spectroscopy to predict the absolute content (g/kg) and proportion (%) of SOC in the POM and in the MOM fractions of the remaining topsoils of the German Agricultural Soil Inventory (n = 2755).

Finally, Vos and colleagues used a random forest algorithm to investigate the relative importance of 75 potential drivers of differences in carbon proportions in the labile and stable pools using both the calibration set (n = 145) and whole dataset (n= 2900).

We have a concern regarding the use of the cross-validated regression model based on near-infrared spectroscopy to predict the size of SOC labile and stable pools in “new” samples of the German Agricultural Soil Inventory.

We regret the use a regression model that has not been published yet, impeding us from a clear understanding of the actual predictive performance of the model on “new” topsoil samples. Here, the details provided by the authors regarding the predictive performance of the multivariate regression model (see Material & methods section 2.4 at lines 189–194 and Supplementary Figure S1) do not demonstrate its ability to accurately predict the absolute content (g/kg) and proportion (%) of SOC in the POM and in the MOM fractions of the 2755 “new” samples.

Specifically, the authors have only assessed the predictive performance of their model using a leave-one-out cross-validation. Leave-one-out cross-validation is not the optimal method to validate a partial least-squares (PLS) regression model when 145 samples with reference measurements are available. It may be recommended for smaller datasets when a proper validation procedure (see below) cannot be done.

An acceptable procedure for validating this PLS regression model would be adding an independent validation step to the current validation scheme:

i/ first run a leave-one-out or k-fold cross-validation on a subset of ca. 110 samples with reference measurements, that would provide a Q^2 (= coefficient of determination of the model in cross-validation, not a R^2), and a first assessment of the mean error of prediction of the PLS regression model in cross-validation (RMSECV).

ii/ use this cross-validated PLS model to predict the values of the absolute content (g/kg) and proportion (%) of SOC in the POM and in the MOM fractions of the ca. 35 independent samples with reference measurements not used for cross-validation (and independent from the ca. 110 samples used for cross-validation). The coefficient of determination (actual coefficient of determination of the model in validation, R^2) and mean error of prediction of the PLS regression model in validation (RMSEP) would provide acceptable criteria for the reliable (independent) assessment of the actual predictive performance of the model for prediction on “new” topsoil samples.

iii/ if the R^2 and RMSEP (or RPD) of the PLS regression model obtained on the 35 independent validation samples were judged acceptable, then the model may be used to predict the values of the absolute content (g/kg) and proportion (%) of SOC in the POM and in the MOM fractions of the 2755 remaining topsoils of the German Agricultural Soil Inventory.

We therefore argue that the PLS regression model based on near-infrared spectroscopy presented by the authors cannot be used in its current form to predict labile and stable SOC fractions on "new" topsoil samples of the German Agricultural Soil Inventory.

At this stage (i.e. unreliable assessment of the predictive performance of the PLS regression model), the authors can only use the reference data (n = 145) of the absolute content (g/kg) and proportion (%) of SOC in the POM and in the MOM fractions to investigate the potential drivers of the distribution of SOC kinetic pools on this limited dataset. This would already be a significant piece of work.

Furthermore, Vos and colleagues used the particulate organic matter (POM) fraction to represent the labile SOC kinetic pool. However, the POM fraction could contain substantial (and variable) amounts of pyrogenic carbon with residence time in soils higher than the mean residence time of total SOC. This limitation of the SOC density fractionation scheme should be mentioned and discussed in the text, as it is not possible to guaranty that the POM fraction truly represents the actual labile SOC pool for all investigated samples.

We are looking forward to seeing an improved version of this work in SOIL.

Interactive comment on SOIL Discuss., https://doi.org/10.5194/soil-2017-30, 2017.