Dear editor

We thank you and the reviewer for your time and valuable comments. Please find our replies displayed in blue color below each comment. A manuscript version with tracked changes is attached.

Kind regards

Mareike Ließ

Interactive comment on "Error propagation in spectrometric functions of soil organic carbon" by Ellinger et al.: Response to RC3

General comments

Most comments on technical and minor topical suggestions were addressed in the new manuscript.

Authors' reply: "The PLSR components vary largely in dependence on the pre-processing method. The information on the number of selected components, therefore, did not result very informative. Furthermore, it distracts the reader from the main message. We, therefore, refrain from including it in this publication. We have adapted the results section concerning the uncertainty scenarios. We have also extended the introduction section to elaborate on the various sources of uncertainty." I insist on giving finally selected tuning parameters for all the modeled scenarios. The reader requires such information to judge the uncertainty in relation to model complexity within different error propagation scenarios. Hyperparameter optimization through Model tuning is a key aspect of spectral modeling, and all scientifically-sound publications in this field report these results. Parameter selection results can easily be included within a model assessment table. This does not distract the reader. This information will reveal also potential over-fitting effects due to replicate spectra in the inner tuning loop in presence of multiple spectral replicates. The main concern from the last review round has been mentioned in the conclusion, but is still not resolved. Specifically, the group stratification for replicates of the same sample was not performed for the tuning procedure. The author's specifically mention this now in the conclusion. However, the author's refuse to report the finally selected PLSR parameters, but at the same time admit that suboptimal model parameters might be selected. Based on the results shown, no conclusive statement on this issue can be drawn. In order to get this manuscript eligible for final publication, the issues arising from multiple replicates in tuning during cross-validation needs to be addressed. At least, the authors should report a majority consensus value of the number of PLSR components selected in the final model. This will at least indicate whether the model errors are biased by resampling artifacts.

Reply: We decided to reply to these general comments collectively as they are all related to one another and refer to the applied nested cross-validation (CV) approach. There seems to be a misunderstanding. Model validation and tuning are both conducted with a group CV assigning replicate sample measurements and scans to the same fold. We adapted the corresponding section in the Materials and Methods and discussion section for better understandability. Please compare lines 248-283 and lines 405-408. "We agree that model complexity should always be kept in mind. However, overfitting was prevented by the applied nested group cross-validation (CV) approach. On the other hand, the mere inclusion of a table giving the number of components per model scenario would not suffice, as this would definitely require an extended discussion section, which is out of scope of this manuscript."

Specific comments

Abstract

The statement in l. 24–26 about precise protocol and measurement protocol is out of scope and needs to be removed. It isnot the main topic of this study. Such protocols and procedures are mostly well documented in soil spectroscopy literature and there are recommendations on this, consider for example Wetterlind et al., 2013.

Reply: Although there are protocols and procedures – we cite Pimstein et al. 2011 in lines 423-425 - there is still no agreement within the soil spectroscopy community on the applied protocol and procedure. As a consequence, the number of scans and replicate measurements per sample differ in each working group. As we have shown that the applied procedure has an impact on model performance, it is important to describe it in each study in order to allow for comparison between studies.We adapted the text section on Table 4 (lines 414-439) to that extent. As a consequence, the statement is in fact a major conclusion from this manuscript.

Introduction

The introduction needs a major rewrite, there are many grammatical, topical, and stylistic errors. See the technical corrections for some examples and suggestions.

Reply: Thank you, we have checked the introduction, thoroughly.

Material and Methods

Authors' reply: "Reference to soil treatment and scattering effects was made in the introduction. We refrain from referring to soil texture as we are at within-field scale and do not have a pronounced textural variability in our dataset. A reference to sample origin is included in the discussion section." Although the present study covers within-field variability, general conclusions regarding spectral error propagation need to consider soil texture as it affects scattering and averaging effects.

Reply: Soils with different properties may have a different amount of light scattering and might therefore require a differing amount of scans and replicate measurements per sample. However, this is beyond the scope of this paper. Based on the soil samples we measured we cannot make any statement in this regard.

Results and Discussion

Some paragraphs contain statements that should be in the Material and Methods section, see e.g. lines 269–271 (see also technical corrections). Table 3 and the corresponding text needs to be moved to the Material and Methods section.

Reply: We checked the Results and Discussion section, thoroughly. Table 3 and the corresponding text were moved to the Material and Methods section.

Many sections contain present tense where past tense is needed.

Reply: adapted accordingly

The section 3.1 on soil organic carbon reference values is too long, and needs some general revision. This section needs to be further summarized and presented in a more concise manner.

Reply: The text section was revised and summarised.

L. 283–284: "The plots for "A" and all samples show higher and lower SOC values than the archive data due to the fact that those data are obtained from compound samples for one plot.': It it not clear what the authors mean by compound plots..

Reply: The term "compound sample" was explained. The sentence was adapted to "The plots for "A" and all samples show higher and lower SOC values than the archive data due to the fact that those data were

obtained from compound samples , i.e. a number of distributed soil samples were taken per LTFE plot and mixed before they were subjected to soil laboratory analysis." (lines 325-326).

Authors' reply: We are not aware of any study that actually quantified the effect of spectral pre-processing on model performance and, therefore, refrain from deleting it from our study. This is simply not correct. There are many soil spectral studies addressing pre-processing with regard to model performance. Please consider e.g. Stevens et al., 2013 (see References at the end).

Reply: Yes, you are right. There are quite some publications that compare model performance in dependence on various pre-processing methods. We have actually cited some of them. The difference in model performance, due to the applied pre-processing, is usually not explicitely reported, though, but could be calculated. Still pre-processing is such an important aspect in VIS-NIR spectrometry that we find it important to report its impact on model performance in relation to the other aspects we investigated. Furthermore, the applied pre-processing changes the impact of the uncertainty propagation; in some cases the typical pattern is even reversed (Please compare lines 366-374).

The discussion around the pre-processing is still way to long.

Reply: The discussion of pre-processing only relates to its impact on model performance (now lines 375-383). We find this short paragraph of adequate length.

Table 4 on R2 values is not informative. First, it only reports R2, which is strictly not a measure of performance. Second, it is relative to the range of the measured property, which is not given. The only reference to this table is that the error conditional on the input data is different, and this information is missing in the table. Therefore, this table and discussion around it should be removed.

Reply: R^2 is still the most reported metric when comparing performance results between publications. We included the SOC range and further information and adapted the corresponding text section (lines 414-439) so that the reason for its inclusion is understandable. It relates to the information content of error values and the applied measurement protocol.

In general, the authors should stay focused on the key topics under investigation. Many sections are too long and therefore distract the reader. The Results and Discussion requires some more work to offer the audience a better flow.

Reply: Thank you. We have thoroughly revised the whole manuscript.

Conclusion

The conclusion should fit on half a page. Focus on the key findings and topics that the study addressed.

Reply: Adapted accordingly

Technical corrections

L. 12: "...the exact monitoring of..." -> "...precise monitoring of..."

Reply: changed accordingly

L. 14 "...to enhance conventional SOC analysis and has often been used to predict SOC" -> "...to complement conventional SOC analysis."

Reply: changed accordingly

L. 24–26: "We emphasize...and allow for a comparison between publications."

Reply: changed accordingly

L. 35: “production of energy”: Energy cannot be produced, rather use “production of energy crops”

Reply: changed accordingly

L. 36: remove Stenberg et al., 2010: this does not fit the context.

Reply: adapted accordingly

L. 36–37: “quality of soil” needs to be described, too generic -> soil properties, soil type...

Reply: adapted accordingly

L. 38–40: “...SOC is also interesting when it comes to the global warning issue...” -> “SOC is also relevant for the global warning issue...”

Reply: changed accordingly

L. 66: “However, the application of ....”. -> Delete “However” because there is no reference sentence.

Reply: However relates to the previous paragraph

L. 67: “...standard lab analysis” -> ”...standard laboratory analysis”

Reply: changed accordingly

L. 68: “...on the other hand side” -> either “hand” or “side” ... (see general comments for abstract)

Reply: changed accordingly

L. 100: Missing dot after “(Merbach and Schultz, 2013)”

Reply: changed accordingly

L. 131: “Elementaranalysator” -> “elemental analyser”

Reply: changed accordingly

L. 155–160: Give original publications for all pre-processing techniques; only give these and remove the other references. L. 160–182: All references that are not original method publications for pre-processing need to be removed.

Reply: Adpated accordingly.

L. 276: “the SOC distribution of “A” and “B” samples differ” -> “the SOC distribution of “A” and “B” samples differed”

Reply: changed accordingly

L. 307–308: “The model results are now compared based on their mean $RMSE_{MV}$ and their interquartile range”: this should be in Material and Methods...

Reply: adapted accordingly

L. 308: “It is not surprising that the dataset...”: style -> use “We expected that/It was expected that...” or similar

Reply: changed accordingly

L. 312: “It seems that the within sample variation concerning soil spectra con somehow compendate the within sample variability concerning SOC within the model building process, although replicate measurements do not match” -> consistently use past tense.

Reply: Thank you.We have thoroughly revised the results and discussion section.

L. 387: “cross validation” -> “cross-validation”

Reply: changed accordingly

All figures need to be in vector graphics format or need a better resolution.

Reply: When embedding vector graphics in the applied software for manuscript writing, they get automatically rasterised. Vector files will be provided for final publication.

Figure 9: The text annotation for “RMSE” must be changed from “RMSE <number>” to “RMSE = <number>”

Reply: adapted accordingly

References

Stevens, A., Nocita, M., Tóth, G., Montanarella, L., vanWesemael, B. (2013). Prediction of Soil Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy. PLoS ONE, 8(6), e66409. https://doi.org/10.1371/journal.pone.0066409

Wetterlind, J., Stenberg, B., Rossel, R. A. V. (2013). Soil Analysis Using Visible and Near Infrared Spectroscopy. In F. J. M.

Maathuis (Ed.), Plant Mineral Nutrients (pp. 95–107). https://doi.org/10.1007/978-1-62703-152-3_6

Error propagation in spectrometric functions of soil organic carbon

Monja Ellinger [a], Ines Merbach [b], Ulrike Werban[c], Mareike Ließ [a]

[a] Department Soil System Science, Helmholtz Centre for Environmental Research – UFZ, Halle (Saale), Germany

[b] Department Community Ecology, Helmholtz Centre for Environmental Research – UFZ, Bad Lauchstädt, Germany

[c] Department Monitoring & Exploration Technologies, Helmholtz Centre for Environmental Research – UFZ, Leipzig, Germany

Correspondence to: Mareike Ließ (mareike.liess@ufz.de)

Abstract

Soil organic carbon (SOC) plays a major role concerning ~~the~~ chemical, physical and biological soil properties and functions. To get a better understanding of how soil management affects the SOC content, the ~~exact~~ precise monitoring of SOC on long-term field experiments (LTFE) is needed. Visible and ~~near~~ near-infrared (Vis-NIR) reflectance spectrometry provides an inexpensive and fast opportunity~~possibility~~ to ~~enhance~~ complement conventional SOC analysis and has often been used to predict SOC. For this study, 100 soil samples were collected at an LTFE in central Germany by two different sampling designs. SOC values ranged between 1.5 and 2.9%. Regression models were built using partial least square regression (PLSR). In order to build robust models, a nested repeated 5-fold group cross-validation approach was used, that ~~comprises~~ comprised model tuning and evaluation. Various aspects that influence the obtained error measure were analysed and discussed. Four pre-processing methods were compared in order to extract information regarding SOC from the spectra. ~~Overall~~Finally, the best model performance which did not consider error propagation ~~corresponds~~ corresponded to a mean $RMSE_{MV}$ of 0.12-% SOC (R^2=0.86). This model performance ~~is~~ was impaired by $\Delta RMSE_{MV}$ = 0.04% SOC while considering input data uncertainties (ΔR^2=0.09), and by $\Delta RMSE_{MV}$ = 0.12% SOC (ΔR^2=0.17) considering an inappropriate pre-processing. The effect of the sampling design ~~amounts~~ amounted to a $\Delta RMSE_{MV}$ of 0.02% SOC (ΔR^2=0.05). Overall, w~~W~~e emphasize the necessity of a transparent and precise documentation of the measurement protocol, the model building, and validation procedure~~, including the calculation of the error measure,~~ in order to assess model performance in a comprehensive way and allow for a comparison between publications. The consideration of uncertainty propagation is essential when applying Vis-NIR spectrometry for soil monitoring.

Keywords: Soil organic carbon, Laboratory Vis-NIR spectrometry, Partial least square regression, nested cross-validation

1 Introduction

Soil is at the same time one of the most important and one of the most limited natural resources. Most of all, it is needed for food production, but also for the production of energy crops and fibre, or for the provision of fresh water (Johnson, 2008; Lorenz and Lal, 2016~~; Stenberg et al., 2010~~). All these aspects depend on the quality of the soil, which is determined by its site-specific properties~~of the existing soil~~. And t~~T~~his quality ~~in turn~~, in turn, is much influenced by its SOC content since it affects chemical, physical and biological soil properties and functions (Knadel et al., 2015; Lorenz and Lal, 2016). Additionally, SOC is also ~~interesting when it comes to~~relevant in the context of global warming ~~the global warming issue~~ since the soil is the largest terrestrial reservoir of organic carbon ~~in the world~~ (Conforti et al., 2015; Johnson, 2008; McBratney et al., 2014; Stockmann et al., 2013). ~~The SOC content of soils can be increased through the sequestration of atmospheric CO_2 into long living components of soils~~SOC sequestration may lead to long-term SOC storage in relatively stable soil fractions (Lal, 2004; McBratney et al., 2014). Thus, the SOC stocks of soils could be used as a manageable sink for atmospheric carbon (Stockmann et al., 2013), achieving both, food security and a strategy against the increasing CO_2-concentration in the ~~global~~ atmosphere (Lal, 2004; Lorenz and Lal, 2016; McBratney et al., 2014). As the SOC content of soils reacts very slowly to environmental changes (Meersmans et al., 2009), long-term field experiments (LTFE) are required to understand the impact of soil management and farming systems on the rate of SOC sequestration (Lal, 2004), as well as on yield and crop quality in the long run.

The precise monitoring of SOC on an LTFE with conventional laboratory analysis is labour- and cost-intensive ~~and expensive~~ (Adamchuk and Viscarra Rossel, 2010; Loum et al., 2016) as it requires the analysis of a rather high amount of samples. Visible and ~~near~~ near-infrared (Vis-NIR) reflectance spectrometry can facilitate this procedure. It is non-destructive, fast and economical (Mouazen et al., 2010; ~~Tekin~~ Summers et al., ~~2014~~2011), and requires~~requiring~~ the conventional laboratory analysis to be conducted~~only~~ on a small number of soil samples, only, ~~and~~ as well as little sample preparation (Conforti et al., 2015). ~~In addition, no chemicals are needed and o~~The obtained~~ne~~ spectrum contains information about many different soil components (Conforti et al., 2015; Viscarra Rossel et al., 2006b);~~.~~ please compare Stenberg et al. (2010) for a review on the past and current role of Vis-NIR spectrometry in soil science. Spectral absorption features are caused by vibrational stretching and bending of structural molecule groups and electronic excitation (Ben-Dor et al., 1999; Dalal and Henry, 1986). Molecule vibrations from hydroxyl, carboxyl, and amine functional groups produce ~~soil~~ absorption features related to soil organic matter in the mid-infrared (MIR) region of the spectra (Croft et al., 2012). In comparison, Vis-NIR spectra show only broad and unclear adsorption features related to overtone vibrations from the MIR, but instruments are

less cost-intensive and available for field monitoring as well (Stenberg and Viscarra Rossel, 2010; Viscarra Rossel et al., 2006a). Furthermore, in diffuse reflectance spectroscopy, scattering properties depend on the particular wavelengths and can vary significantly over the ~~VIS~~Vis-NIR spectral range (Pilorget et al., 2016). Hence, the pre-processing of Vis-NIR spectra is necessary in order to extract soil ~~property~~ property-related information (Stenberg and Viscarra Rossel, 2010). As there is no standard pre-processing technique which works on all spectral data (Stenberg and Viscarra Rossel, 2010), it is recommended to always test various techniques and to choose the one which performs best for the respective data. Several studies, therefore, compared a rather high number of pre-processing methods (e.g. Peng et al., 2014; Nawar et al., 2016). Scattering and other effects attributed to within-sample variance can be addressed by repeated measurements of replicate samples (e.g. Pimstein et al., 2011). A~~l~~ltogether, Vis-NIR soil spectrometry has been used on many occasions to build SOC prediction models (Jiang et al., 2016; Kuang and Mouazen, 2013; Nocita et al., 2013).

However, the application of Vis-NIR soil spectrometry for SOC determination involves a couple of uncertainties. The required calibration data are determined with standard laboratory analysis, e.g. dry combustion, with associated uncertainties. On the other hand ~~side~~, the spectral measurements are affected by the sample preparation, e.g. drying, sieving, grinding (e.g. Nduwamungu et al., 2010). Furthermore, sensor noise and other spectrometer internal sources (electronic and mechanical) can affect the measurements (Schwartz et al., 2011). Finally, these two uncertain data sources are related by a regression model. And the model building procedure involves a couple of error sources itself. The development of robust models requires a resampling ~~process~~ procedure to determine the model parameters and to avoid overfitting; the applied resampling method impacts model performance (e.g. Molinaro et al., 2005, Beleites et al., 2005). Further aspects that impact model performance are: the available dataset in concordance with the applied sampling design, the handling of outliers, spectral pre-processing, and last but not least the model evaluation procedure. In most studies dealing with SOC prediction from Vis-NIR spectra, no clear statement about input data uncertainties or their handling is made. The reported prediction errors only refer to the model building procedure, while uncertainties from laboratory measurements are neglected. Commonly, only a single SOC measurement per soil sample is available, and in spectrometric laboratory measurements~~. In spectral soil sensing in lab applications~~, the general approach consists~~t~~ in averaging the multiple measured spectra of one sample to one spectrum which is then used for model building (Ge et al., 2011; Stevens et al., 2013; Viscarra Rossel et al., 2003). ~~But~~ However, the number of measurements used to gain one averaged spectrum differs between studies. Jiang et al. (2016), for example, averaged 10 measurements to receive one spectrum, while Volkan Bilgili et al. (2010) and Wang et al. (2014) used four measurements. This difference is also assumed to have an influence on the uncertainties containe~~implemente~~d in the input data.

Overall, to allow for comparis~~i~~on between studies, in terms of predictive uncertainty in % SOC, a modelling procedure is required that deals with the propagation of the input data ~~uncertainites~~uncertainties. For discussion of the general concept, please refer to Jansen (1998), for applications in soil modelling compare e.g. Heuvelink et al. (1999) and Poggio and Gimona (2014). Although~~,~~ the problem of the involved uncertainties in Vis-NIR spectrometry is well-known (e.g. Gholizadeh et al., 2013, Nduwamungu et al., 2010, Mortensen, 2014), implementations of uncertainty propagation in Vis-NIR spectrometric modelling are lacking.

2 Material and Methods

2.1 The static fertilization experiment Bad Lauchstädt

The soil samples were taken at the LTFE site "Static Fertilisation Experiment" in Bad Lauchstädt in central Germany (Körschens and Pfefferkorn, 1998). Positioned at 51°~~-~~24' N, 11~~-~~°~~-~~53' E and with an altitude of 113 m a.s.l. (Körschens and Pfefferkorn, 1998), the climate is characterized by a mean annual precipitation of 470 – 540 mm and an average mean annual temperature of 8.5 – 9.0~~–~~°C. The soil type was characterized as a haplic Chernozem developed from loess (Altermann et al., 2005) with a soil texture of 21.1 ± 1.2~~-~~% clay, 72.1 ± 1.7~~-~~% silt, and 6.9 ± 1.9~~-~~% sand (Dierke and Werban, 2013). Saturated water conductivity and air capacity are medium to high in the top~~-~~soil (Altermann et al., 2005). The Static Fertilization Experiment was initialized in 1902 by Schneidewind and Gröbler and is about 4 ha in size (Merbach and Schulz, 2013). Its objective is to investigate the impact of organic and mineral fertilization on soil fertility as well as yield and quality of crops (Körschens and Pfefferkorn, 1998; Schulz, 2017). The experiment includes eight subfields with a width from 25.2 m to 28.5 m and a length of 190 m which are each divided into 18 plots that are treated with different mineral and organic fertilizer as well as planted with different crops following a crop rotation (Körschens and Pfefferkorn, 1998). The plots of subfields 4 and 5 are additionally parted into 5 smaller subplots.

2.2 Sampling design

A total of 100 soil samples were taken at the soil surface (0-10 cm) in September 2016. The exact location of the sampling points was determined by a differential GPS GNSS LEICA Viva GS08. It was decided to sample at precise point locations instead of taking samples representative for LTFE plots to allow for a direct comparison with spectrometric field measurements for area-wide regionalisation (not included in this study). The sampling points were determined beforehand by two sampling designs. Based on the LTFE treatment factors and per-plot soil archive data including C_{org}, N_{tot}, ~~plant~~ plant-available P, ~~plant~~ plant-available K (both with DL-Method

(VDLUFA, 2012)~~)~~ and pH (Fig. 1) both designs strived to select a dataset of 50 samples representative for the soil variability of the entire LTFE. Categorical and continuous data first entered a factor analysis ~~with~~ for mixed data (FAMD) performed with R package FactoMineR (Lê et al., 2008) to allow for further joint analysis. For design 'A' the LTFE plots were then grouped by a k-means cluster analysis. R package NbClust (Charrad et al., 2014) automatically determines the optimal number of clusters making use of 30 indices. In the end, ten plots were randomly selected from each of the resulting five clusters, making a total of 50 plots to be sampled. For design 'B', the Kennard-Stone algorithm was applied with R package prospectr (Kennard and Stone, 1969; Stevens and Ramirez Lopez, 2014). 50 LTFE plots were selected involving 5 repetitions of the algorithm to reduce inter-point dependence. Finally, one sampling point was randomly selected from each of the 50 LTFE plots ~~from~~ for design A and B based on a 5 x 5 cm raster. Plot margins of 1.5 m (3 m between plots) were excluded. Fig. 2 shows the location of the so obtained 100 soil samples.

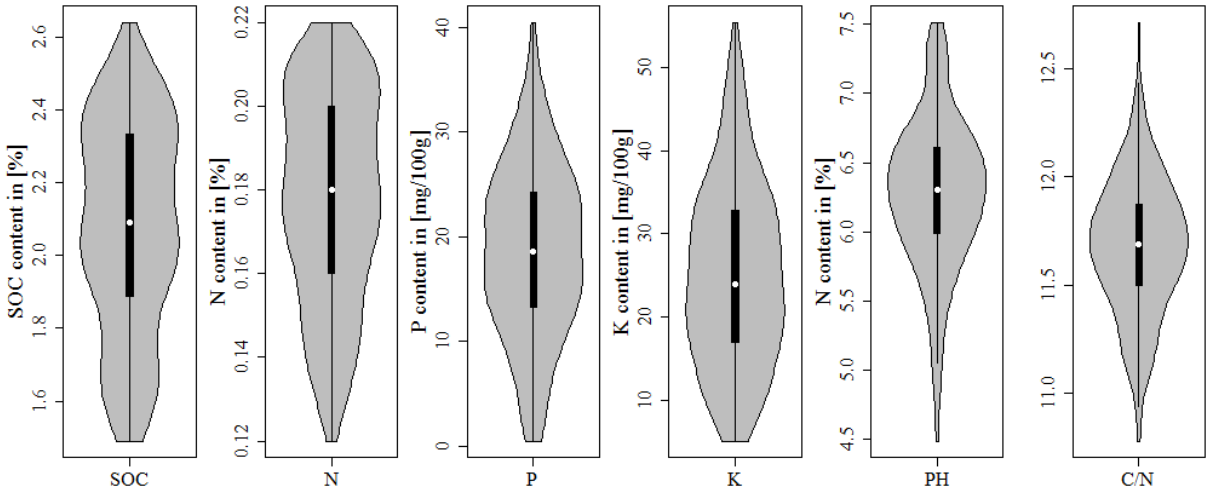
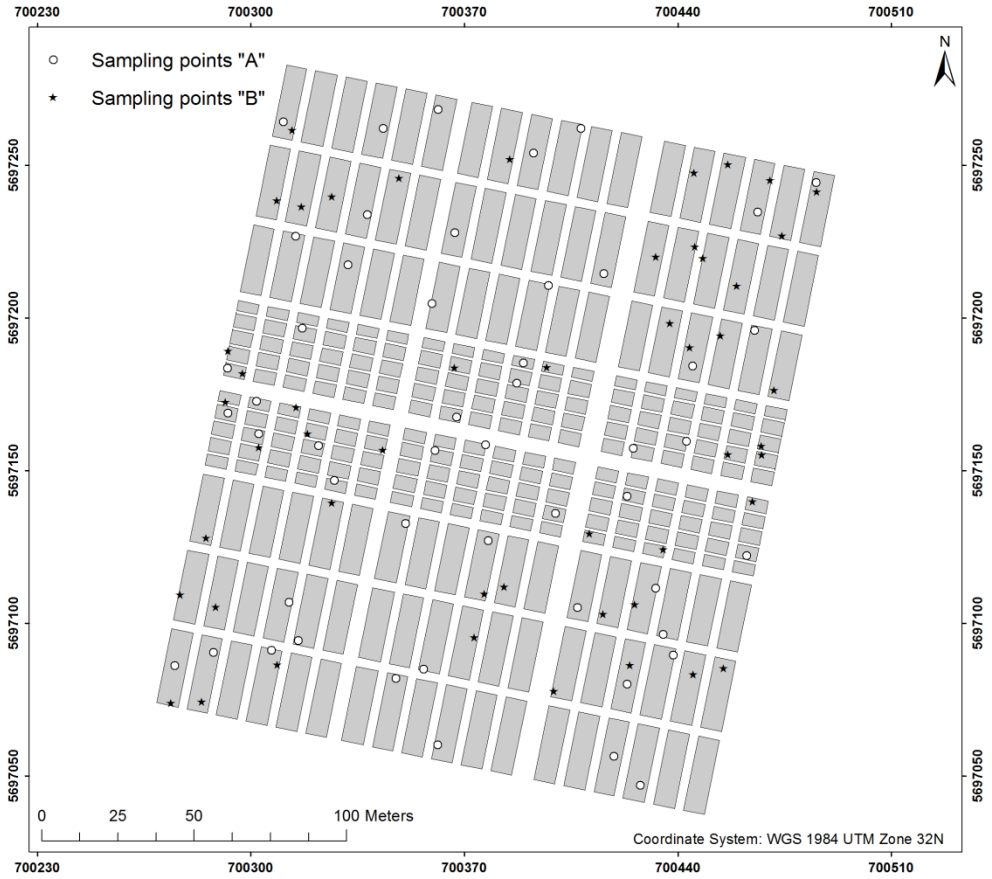


Fig. 1 Soil archive data of the LTFE measured from 2004 to 2007 (Reports of the experimental station Bad Lauschstädt 2004-2007 (unpublished).



135

Fig. 2 Site of the Static Fertilisation Experiment in Bad Lauchstädt with LTFE plots and sampling points according to design A and B. Plot margins excluded from sampling are visible as 3 m wide stripes between plots.

2.3 Laboratory measurements

The soil samples were air-dried, sieved and ~~grinded~~ ground prior to ~~C~~ carbon measurements with dry combustion. A High-end ~~Elementaranalysator~~ elemental analyser ~~vario~~ Vario EL Cube-CN was used. Measurements were repeated in three replicate samples. C~~-~~arbon measurements were taken as ~~organic carbon~~SOC due to negligibly small carbonate contents (below detection limit). The Vis-NIR contact measurements were performed on air-dried and sieved (2 mm) samples in July 2017, using Veris® VIS-NIR Spectrophotometer by Veris ~~technologies~~Technologies, Inc. (hereinafter called Veris) containing an Ocean Optics USB4000 instrument (200 to 1100 nm) and a Hamamatsu Mini-spectrometer TG series (1100 to 2200 nm, resolution 6 nm). The device was warmed up for at least 20 minutes before performing measurements. All measurements were taken in a dark room to prevent daylight from affecting the outcome. The soil samples were scanned from the top. Before and between soil sample measurements, Veris was calibrated using four Avian Technologies Fluorilon™ gray~~-~~scale standards. Each soil sample was divided into three sub-samples filled into ~~petri~~ Petri dishes (Schott Duran ~~petri~~ Petri dishes;

Duran Group, Mainz, Germany). These replicate samples were not related to the three replicate samples used for ~~C/N~~ SOC measurements. For each replicate sample, six spectra were gained by measuring each replicate sample three times, rotating it by 90 degrees and then measuring it three times again. This procedure resulted in 18 spectra for each soil sample. Internally the spectrometer averaged 25 scans for each spectrometer reading (spectrometer setting).

2.4 Spectral pre-processing

Veris is equipped with two spectrometers. At the beginning and end of their respective wavelength ranges noise occurs in the measurements. Therefore, the spectra between these wavelengths (1000 to 1100 nm) had to be removed. Additionally, the spectra were cut at the beginning (402 nm) and the end (2220 nm) ~~to remove noise~~. A number of pre-processing methods were tested to enhance the information regarding SOC in the Vis-NIR spectra. The spectra were tested for outliers using R package mvoutlier (Filzmoser and Gschwandtner, 2017). For this procedure, a PCA is performed, using then the first two obtained PCs for outlier detection with function aq.plot. Out of the tested pre-processing methods, four different combinations are shown in this study in order to demonstrate their ~~different effects on the prediction model~~impact on model performance. Their application resulted in spectra with different wavelength ranges (Table 1) and different appearance (Fig. 3). These pre-processing techniques include the Savitzky-Golay algorithm (SG), the continuum removal (CR), the standard normal variate (SNV), the first derivative (d1) and the gap-segment algorithm (gapDer). ~~The application of the different~~All pre-processing methods for this study ~~was~~ere conducted using R package prospectr (Stevens and Ramirez Lopez, 2014). The SG algorithm fits a polynomial regression on the spectral data to find the derivative at a center point i of a defined smoothing window (w) (~~Rinnan et al., 2009;~~ Savitzky and Golay, 1964~~; Swarbrick, 2016~~). CR can be seen as a spectra ~~normalization~~ normalisation technique which enables to compare different absorption characteristics from a mutual baseline ~~(Kokaly, 2001; Mutanga and Skidmore, 2003)~~. ~~It identifies the local reflectance spectra maximum points and connects those points to form a convex hull (Mutanga and Skidmore, 2003; Stevens and Ramirez Lopez, 2014)~~The continuum is calculated by linear interpolation of the reflectance spectrum's maxima. We implemented CR following Stevens and Ramirez Lopez (2014) by c~~C~~alculating

$$\phi_i = \frac{x_i}{c_i} \quad (1)$$

for i = {1, … , p) with x_i and c_i being the initial and the continuum reflectance values at wavelength i of a set of p wavelengths. ϕi then gives the continuum-removed reflectance value. ~~ϕi (Stevens and Ramirez Lopez, 2014). All~~

~~other data have values between 1 and 0 (Mutanga and Skidmore, 2003; Schmidt and Skidmore, 2001). Thus the absorption peaks are enhanced (Schmidt and Skidmore, 2001).~~ SNV is a scatter-corrective pre-processing method (~~Rinnan et al., 2009~~Barnes et al., 1989). The basic formula is as follows

$$x_{corr} = \frac{x_{org} - a_0}{a_1} \quad (2)$$

with a_0 as the measured spectrum's average value which shall be corrected and a_1 being the sample spectrum's standard deviation. x_{org} ~~are~~ is the original ~~spectra~~ spectrum and x_{corr} the corrected ~~spectra~~ spectrum after applying SNV. In this study, SNV operates row-wise, ~~so~~ each observation is processed on its own (~~Rinnan et al., 2009;~~ Stevens and Ramirez Lopez, 2014). ~~d1 represents the slope of the spectrum, showing peaks where the spectrum displays its maximum slope and crossing zero where the spectrum shows peaks (Leone et al., 2012). According to Knadel et al. (2015) and Smith (2002), d1 can be used to remove baseline offsets from the spectra. The estimation of d1 is done by computing the difference between two batched~~d1 is calculated by the finite difference method, i.e. the difference between two subsequent data points ~~spectral points~~ x_i and x_{i-1} (Eq. 3)

$$x_i' = x_i - x_{i-1} \quad (3)$$

with x_i' as the value of the first derivative at the i^{th} wavelength (Rinnan et al., 2009). The downside of using derivative spectra is their tendency to ~~over-fit the calibration model (Stevens and Ramirez Lopez, 2014). Moreover, derivatives may~~ increase noise so that ~~a~~ smoothing of the data is required (~~Leone et al., 2012;~~ Stevens and Ramirez Lopez, 2014). With the gapDer, ~~a~~ smoothing is performed under a chosen segment size (s) and then a derivative follows (Stevens and Ramirez Lopez, 2014). ~~The application of the different pre-processing methods for this study was conducted using R-package prospectr (Stevens and Ramirez Lopez, 2014).~~

Table 1 Combinations of pre-processing techniques used in this study; w = window size, s = segment size.

Pre-processing methods	Wavelength range	Abbreviation~~.~~
Savitzky-Golay (w =11 nm)	432 – 2201 nm	SG
Savitzky-Golay (w=11 nm) and continuum removal	432 – 2201 nm	SGCR
Standard normal variate and 1st derivative	408 – 2186 nm	SNVd1

Gap-segment algorithm (w = 11 nm, s = 10 nm)	490 – 2163 nm	gapDer

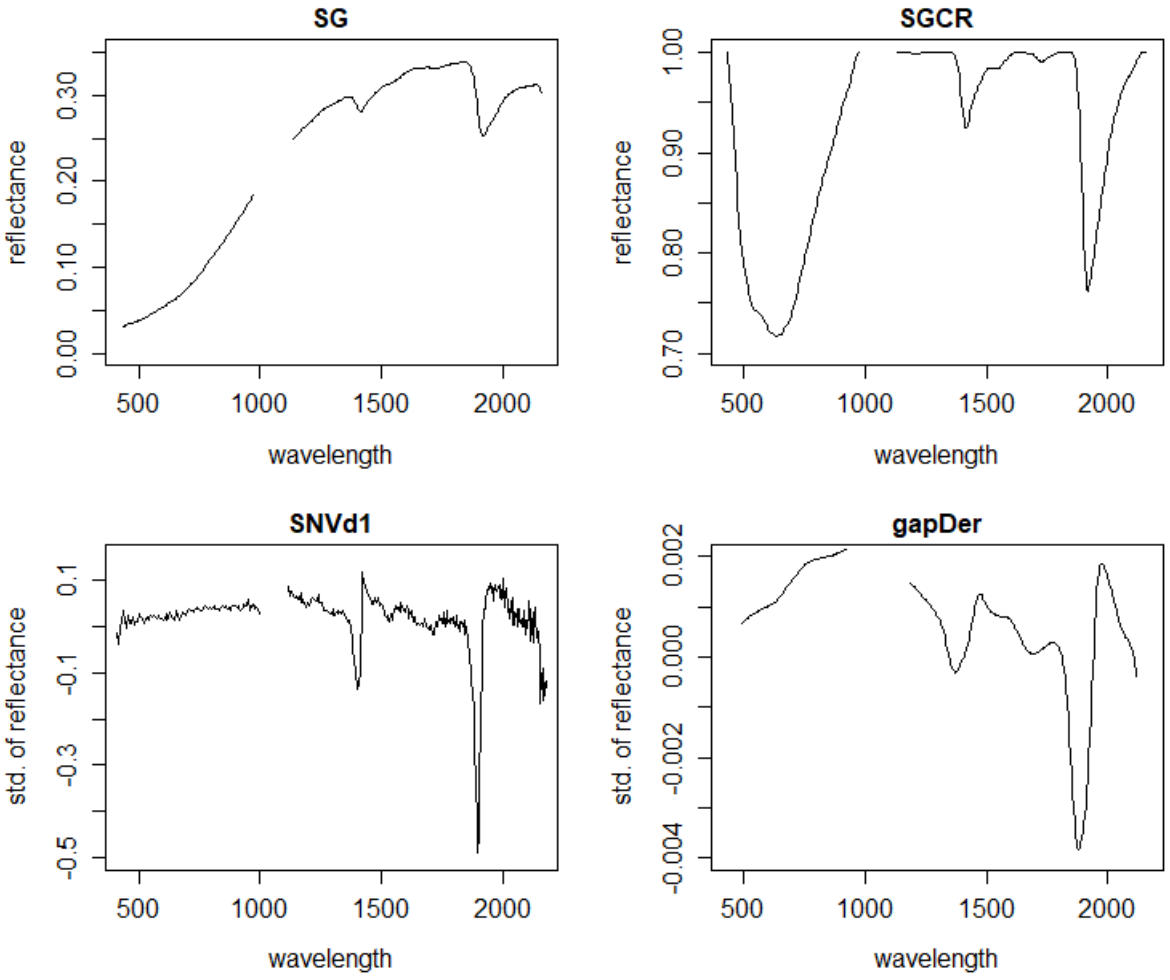


Fig. 3 Impact of different pre-processing techniques on a spectrum; SG = Savitzky-Golay, CR = continuum removal, SNV = standard normal variate, d1 = 1st derivative, gapDer = Gap-segment algorithm.

2.5 Error propagation

A problem occurring in every model building process is uncertainty propagation. Uncertainties of the input data and model, result in uncertainties in the output (Brown and Heuvelink, 2006). Uncertainties in the input data are caused by errors in data acquisition (e.g. measurement errors) as well as variation in the data themselves (e.g. within-sample variability) (Heuvelink, 1999). For this study, there are two different sources for errors in data acquisition: the measurement of the spectral data and the measurement of the SOC content of the soil samples. In order to investigate the influence of these errors, different datasets were built in this study. Fig. 4 gives an overview. From the measured Vis-NIR spectra, three different spectral data variants were created (Fig. 4, step 1). For the first variant, all 18 spectra were retained. The inclusion of all 18 spectra reveals the influence of the error implemented in the spectral measurements as well as the influence of the within-sample variability. For the second variant, the three measurements obtained before and after sample rotation were averaged separately resulting in 6 spectra per sample showing the influence of within-sample variability (replicate measurements). For the third data variant, all 18 spectra were averaged to 1 mean spectrum per sample, removing the influence of the measurement error as well as the within-sample variability. The different spectra obtained through this procedure can be seen in Fig. 5; only parts of the spectra are depicted in order to show their differences. The three different spectral data variants were then pre-processed with the different pre-processing methods from Table 1 (Fig. 4, step 2), resulting into 12 different spectral datasets (Fig. 4, step 3). These were then combined with single and averaged SOC values in step 4 so that altogether 24 datasets were obtained (Fig. 4, step 5). In order to compare the two sampling designs, this procedure was carried out for the 50 soil samples labelled "A" 'A' and "B" 'B' and also for the complete set of soil samples. In this way, three different soil sample sets ("A" 'A', "B" 'B' and "all samples" 'all' samples) were achievedobtained.

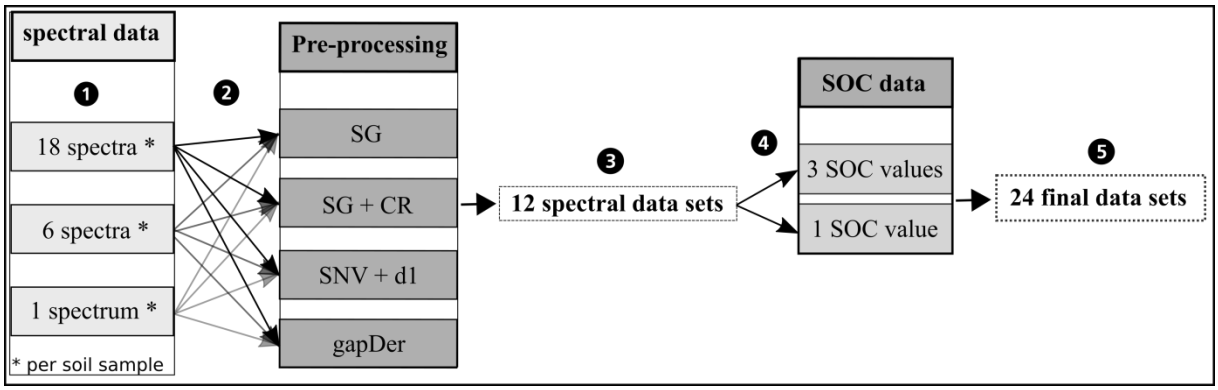


Fig. 4 Datasets to investigate the uncertainty propagation. SG = Savitzky-Golay, CR = continuum removal, SNV = standard normal variate, d1 = 1st derivative, gapDer = Gap-segment algorithm.

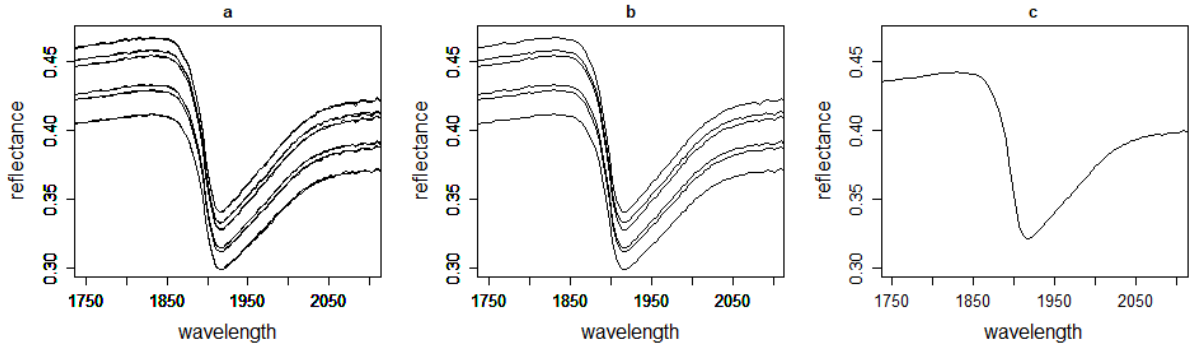


Fig. 5 Zoom-in to a sample's spectral dataset: a) 18 spectra comprised of 6 replicate sample measurements with 3 scans each, b) 6 spectra related to replicate sample measurements (average of three scans each) and c) 1 averaged spectrum.

2.6 Model building and validation

Regression models were built using partial least square regression (PLSR). Out of the many algorithms, PLSR is seen as a standard method for spectral calibration and prediction (Mouazen et al., 2010; ~~Tekin et al., 2014;~~ Viscarra Rossel et al., 2006b). For recent applications to predict SOC from Vis-NIR soil spectra, see e.g. Liu et al. (2018) and Yang et al. (2019).~~For applications to predict SOC from Vis-NIR soil spectra see (Conforti et al., 2015; Jiang et al., 2016; Kuang and Mouazen, 2013; Nocita et al., 2013).~~ PLSR is described in detail by Martens and Næs (1989) and Naes et al. (2002). It incorporates characteristics from ~~principle~~ principal component analysis (PCA) and multiple regression (Abdi, 2007). The concept behind PLSR is to seek a small number of linear combinations (components or latent factors) obtained from the measured spectral data and to use them in the regression equation to predict SOC instead of the initial values (Martens and Næs, 1989; Naes et al., 2002). These components are constructed so that they account for most of the variance in the measured spectral data (X) and the SOC content (Y), and at the same time maximize the correlation between X and Y. In other words, PLSR leads to the covariance between X and Y being maximized (Bjørsvik and Martens, 2008; ~~Summers et al., 2011; Tekin et al., 2014;~~ Wehrens, 2011).

In order to receive a robust model, it is important not to include too many components in model building as this will lead to over-fitting (Hastie et al., 2009; Kuhn and Johnson, 2013). On the other hand, the inclusion of too few components comprises the risk of building an under-fitted model which is too small to cover the variability existing in the soil spectral data (Naes et al., 2002). The selection of the optimal number of components is hereinafter referred to as *model tuning*. In order to receive a robust model, resampling is commonly applied for model ~~building and~~ validation. But resampling can also be used for *model tuning* to receive robust tuning parameters (Guio Blanco et al., 2018; Hastie et al., 2009; Kuhn and Johnson, 2013). For small datasets, k-fold cross-validation (CV) is recommended (Hastie et al., 2009).

In this study, model building, *model validation* and *model tuning* ~~was~~ were ~~done~~ implemented using a nested ~~repeated 5-fold cross-validation (~~CV~~)~~ approach (Fig. 6)~~(~~,~~ e.g. Varma and Simon, 2006, Guio Blanco et al., 2018). The CV for model validation and tuning consisted of a repeated k-fold group CV. ~~Repeated 5-fold CV can increase the precision of the prediction while maintaining a small bias (Kuhn and Johnson, 2013). Five Repetitions of 5-fold CV were conducted in this case.~~ In order to calculate reliable error measures, the subdivision of the spectral data into the folds had to account for repeated scans and replicate measurements per sample. Accordingly, all spectra for one sample were assigned to the same fold during k-fold CV, i.e. k-fold group CV. Furthermore, to allow for comparison of the models built on behalf of the 24 datasets (Figure 4), the created folds coincide for all datasets; the data of certain sample IDs were always assigned to the same fold ID. For the model validation CV, two further aspects were taken into account that were neglected for the model tuning CV. The group CV was adapted to also guarantee that neighbouring points of ≤ 5 m distance were assigned to the same fold to avoid spatial autocorrelation and too optimistic error measures. Furthermore, the response variable's density distribution was taken into account during fold creation, i.e. a stratified CV. Overall, a nested repeated k-fold group CV was applied. Five Repetitions of a 5-fold group CV were conducted in this case. Kuhn and Johnson (2013) recommend 5-fold CV as it can increase the precision of the prediction while maintaining a small bias.

Fig. 6 shows the various steps of the modelling procedure involving repeated 5-fold group CV for *model tuning* (right box) and *validation* (left box). In the process, the dataset (n = 100%) is randomly subdivided into 5 folds of equal size (step 1). One of the 5 folds is held out as a test set and the other four are used as the training set and partitioned again into 5 folds for *model tuning* (step 2). The optimal number of components (*best Ncomp*) is then determined by computing a PLSR on the resampled data, testing 1 to 30 components (step 3), and calculating the repeatedly 5-fold cross-validated RMSE of *model tuning* ($RMSE_{MT}$) corresponding to each number of components (step 4). ~~This~~ The latter was implemented with the trainControl() function ~~in~~ of R package caret (Kuhn, 2017). The optimal number of components (step 5) is then used in model building (step 6). The resulting model's test set RMSE of model validation ($RMSE_{MV}$) is determined in step 7. The whole procedure is repeated until all folds ~~in the boxes~~ have once been used as the test set to have a simple 5-fold group CV. A repeated 5-fold group CV means that the model tuning CV and model validation CV each have to be rerun according to the number of repetitions. Finally, t~~T~~he ~~model~~ performance ~~results~~ of the models built with the 24 datasets are ~~now~~ compared based on their ~~mean~~ $RMSE_{MV}$ mean and ~~and their~~ interquartile range. Table 2 displays the respective dataset size per soil sample. The resulting datasets and models were named after the following scheme: $Dataset_{x1\ x2\ x3}$ with the SOC measurement error (x1), the spectral measurement error (x2) and the within-sample variability (x3). A value of 1

indicates that the respective error is included in the model, a value of 0 shows that the error was removed beforehand by averaging the data.

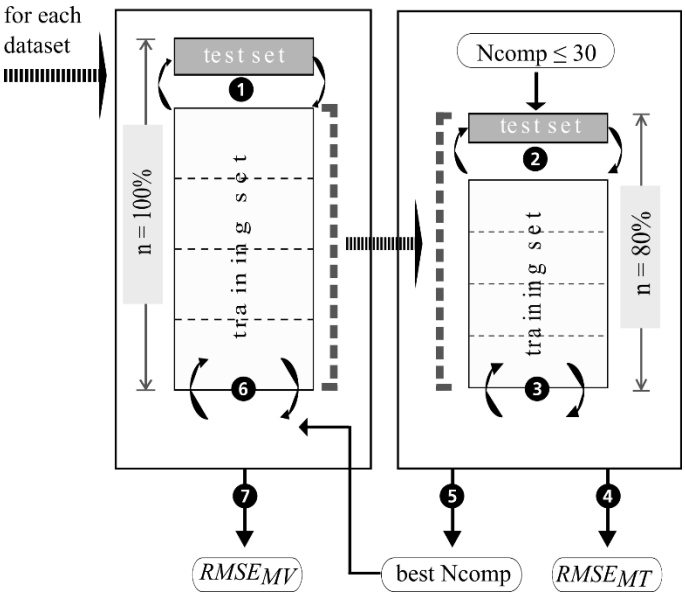


Fig. 6 M*odel tuning* and *model validation* procedure with a nested k-fold group ~~cross validation~~CV approach. The right box shows the *model tuning*, the left one the *model validation* procedure; Ncomp = number of components; adapted from Guio Blanco et al. (2018).

Table 2. Data basis per soil sample

	Number of SOC values per sample	Number of measured spectra per sample	The resulting size of the dataset per sample
Dataset$_{111}$	3	18	54
Dataset$_{101}$	3	6	18
Dataset$_{100}$	3	1	3
Dataset$_{011}$	1	18	18
Dataset$_{001}$	1	6	6
Dataset$_{000}$	1	1	1

~~In this study, the subdivsision of the spectral data into the folds for 5-fold CV had to account for repeated scans and replicate measurements per sample. All spectra for one sample were assigned to the same fold during k-fold CV, i.e. k-fold group CV. The folds contained always the same sample IDs for the various data variants described in Figure 4. For model validation, folds were created following a repeated stratified group CV approach (5-fold). The data were primarily subdivided into 5 equal probability strata based on their density function. The data from each of the strata were then randomly assigned to the 5 folds. Neighbouring points of ≤ 5 m distance were assigned to the same fold to avoid spatial autocorrelation and too optimistic error measures.~~

3 Results and Discussion

3.1 Soil organic carbon content

Fig. 7 ~~shows~~ compares the distribution of the SOC content of the three soil sample sets~~, consisting of 50 soil samples labelled "A" and "B" and 100 soil samples referred to as "all samples"~~to the LTFE archive data (Fig. 1). ~~The measurement error existing in the SOC measurements, here defined as the difference between replicate measurements, ranges from 0.003 to 0.229 % SOC with a mean of 0.048 % SOC. The aim of this study was not to compare the two different sampling designs among each other, but to test whether they are representative of the SOC values existing on the LTFE. For this purpose, per plot soil archive data from the years 2004 to 2007 are also displayed.~~ A Mann–Whitney U test was applied. The statistics of the data are given in Table ~~2~~3. ~~In order to compare distributions between the archive SOC data and "A", "B" and "all samples", a Mann–Whitney U test was applied to the data, testing the "A", "B" and "all samples" against the archive data, respectively.~~ In all cases, no significant difference between the ~~different~~ respective ~~sample~~ datasets and the archive data could be found. This shows that all soil sample sets used in this study ~~are~~ were representative ~~for~~ of the SOC ~~values existing~~variability ~~o~~in the LTFE. Nevertheless, the SOC distribution of ~~"A"~~'A' and ~~"B"~~'B' samples differed~~, with the "A" samples resembling the distribution of all 100 samples more than the "B" samples.~~ . The ~~"A"~~'A' samples contained more samples representing higher SOC values, whereas the ~~"B"~~'B' samples showed a higher representation for lower SOC values. ~~This difference in the distribution of SOC values may have an influence on the prediction results of the models built with "A" and "B" samples. "A" models may be better in predicting higher SOC values, while simultaneously failing to estimate lower SOC values in an appropriate way. To the contrary, "B" models may predict lower SOC values more accurate than higher SOC values.~~ The violin plots of all three ~~soil sample~~data sets do not resemble the archive violin plot very much. The plots for ~~"A"~~'A' ~~and all~~ samples show higher and lower SOC values than the archive data, 'B' samples share the same minimum value with the archive data but display slightly higher SOC values. This difference is likely due to the fact that the archive~~hose~~ data ~~a~~were obtained from compound samples ~~for one plot~~, i.e. a number of distributed soil samples were taken per LTFE plot and mixed before they were subjected to soil laboratory analysis. ~~The "B" samples share the same minimum value with the archive data, but display slightly higher SOC values. This indicates that the choice of the sampling design might have an influence on the model outcome, even if both designs represent the SOC values on the experimental field in an appropriate way.~~

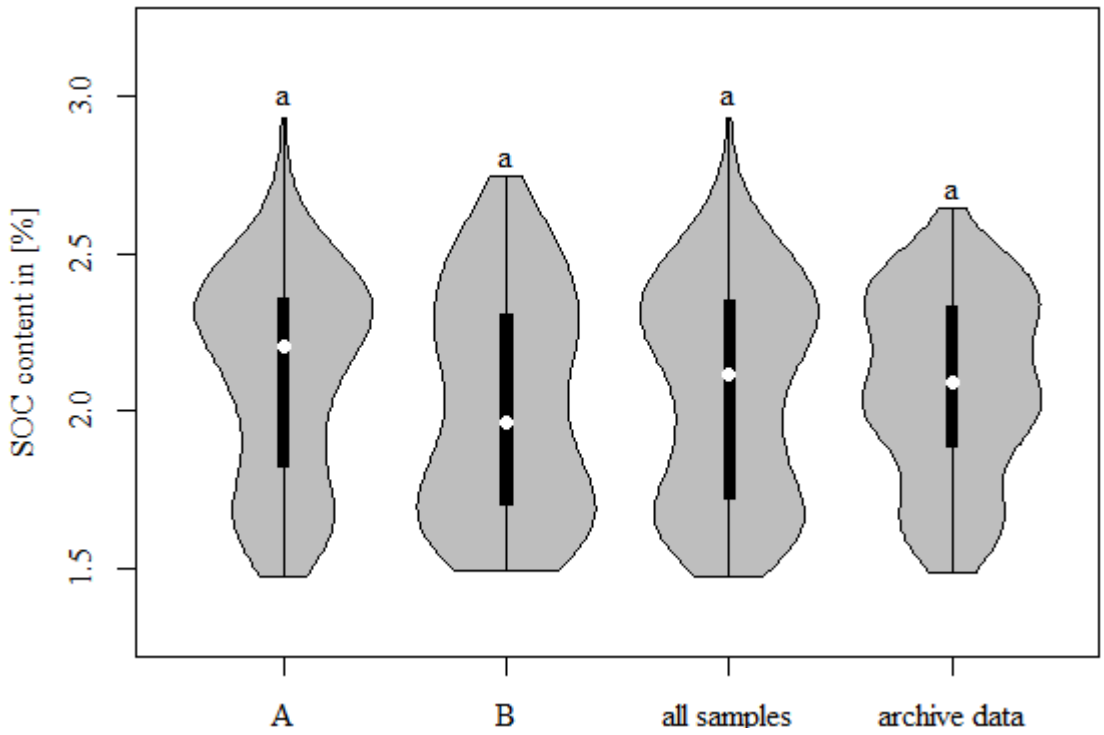


Fig. 7 Soil organic carbon (SOC) content of the three soil sample sets ~~"A"~~'A' (left), ~~"B"~~'B' (middle), and all (middle) and of archive data measured from 2004 to 2007 (right); The thin line shows the 95 % confidence interval, the bar the interquartile range and the dot the median; Mann–Whitney U test was used to compare ~~"A"~~'A', ~~"B"~~'B', and all samples to the archive data; the three soil sample sets were not compared among each other.

Table ~~2~~ 3. Statistics of soil organic carbon in [%] for the three different soil sample sets and the per-plot soil archive data.

Samples	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
~~"A"~~'A'	1.47	1.82	2.21	2.11	2.36	2.93
~~"B"~~'B'	1.49	1.70	1.97	2.02	2.31	2.74
'all'	1.47	1.72	2.12	2.01	2.35	2.93
Archive data	1.49	1.89	2.09	2.08	2.33	2.64

3.2 Comparison of datasets and pre-processing methods

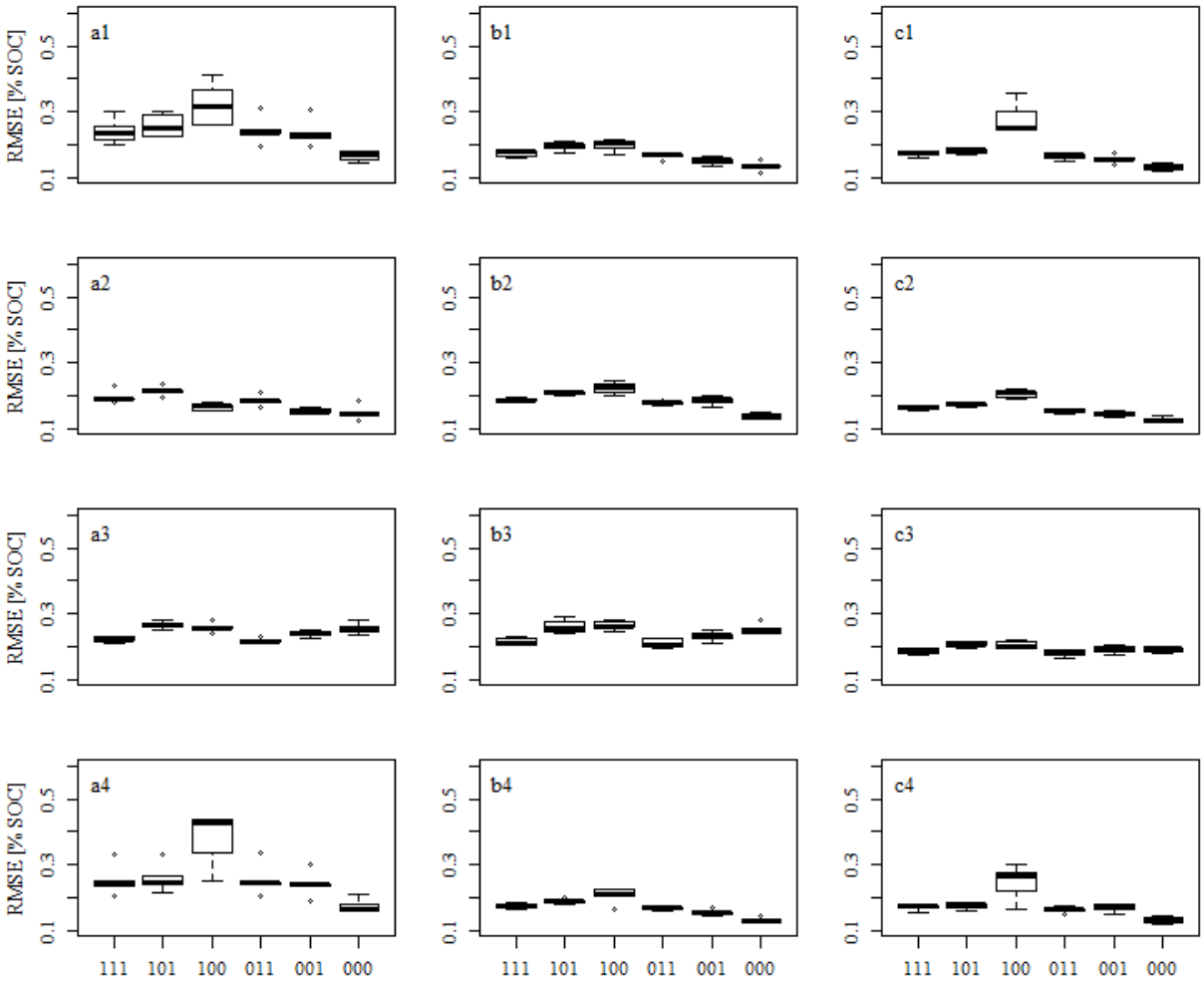
~~According to Filzmoser (2005), the Mahalanobis distance of normally distributed data follows a chi square distribution. Observations which lay beyond a certain quantile of this distribution are marked as outliers and removed from the data (Filzmoser and Gschwandtner, 2017). In this study, no outliers were detected.~~

~~Table 3 displays the different combinations of soil spectra and SOC values per soil sample. The resulting models are then named after the following scheme: Dataset$_{x1\ x2\ x3}$ with the SOC measurement error (x1), the spectral measurement error (x2) and the within-sample variability (x3). The number 1 indicates that the respective error is included into the model, the number 0 shows that the error was removed beforehand by averaging the data.~~

~~**Table 3 Data basis per soil sample**~~

	~~**Number of SOC values per sample**~~	~~**Number of measured spectra per sample**~~	~~**Resulting size of the dataset per sample**~~
~~**Dataset$_{111}$**~~	~~3~~	~~18~~	~~54~~
~~**Dataset$_{101}$**~~	~~3~~	~~6~~	~~18~~
~~**Dataset$_{100}$**~~	~~3~~	~~1~~	~~3~~
~~**Dataset$_{011}$**~~	~~1~~	~~18~~	~~18~~
~~**Dataset$_{001}$**~~	~~1~~	~~6~~	~~6~~
~~**Dataset$_{000}$**~~	~~1~~	~~1~~	~~1~~

Fig. 8 shows the boxplots of the $RMSE_{MV}$. The results of the six datasets corresponding to different information concerning SOC values and spectra (Table 2) are displayed in one plot, the results according to the various pre-processing methods (compare Fig. 4) are displayed in figure lines 1 to 4 ~~anel~~, and t~~. T~~he results of the models ~~using~~ built from the data corresponding to ~~"A"~~'A' samples, 'B' samples and 'all' samples are shown in ~~the 1st~~ figure columns a, b and c~~(a), "B" samples in the 2nd column (b) and all samples in the 3rd column (c); Figure lines 1 to 4 refer to the used pre-processing method~~. As 5-fold CV with five repetitions was performed, five $RMSE_{MV}$ are shown in each boxplot.



355 **Fig. 8 Boxplots of testset RMSE$_{MV}$ ~~(5 repetetions)~~ obtained with the various datasets; Figure columns refer to datasets using a) 'A' samples, b) 'B' samples and c) 'all' samples; figure rows refer to the applied pre-processing, 1 = SG, 2 = SGCR, 3 = SNVd1, 4 = gapDer.**

~~It is not surprising that~~As expected, the dataset of 3 SOC replicate measurements with 1 averaged spectrum (Dataset$_{100}$) ~~results~~ resulted in low model performance, as the within-sample variance concerning SOC ~~cannot~~ 360 could not be explained by the contained predictor information; the input data uncertainty propagate~~d~~s through the model building process. This model performance wa~~i~~s impaired in some cases by Dataset$_{101}$ which ~~combines~~ combined the 3 SOC measurements with 6 replicate spectral measurements (Figures 8b$_1$, a$_2$, a$_3$, b$_3$, c$_3$) ~~but not always~~. It seems that the ~~within~~ within-sample variation concerning soil spectra ~~can~~ was somehow able to compensate for the ~~within~~ within-sample variability concerning SOC ~~with~~in the model building process, although 365 replicate measurements ~~do~~ did not match. Considering the dataset with 18 spectra and 3 SOC measurements (Dataset$_{111}$), model performance ~~improves~~ improved even further ~~(Dataset$_{111}$)~~. In contrast to this, we f~~i~~ound the expected pattern while only 1 SOC measurement wa~~i~~s considered: model performance results display an increase

of RMSE values from $Dataset_{000}$ to $Dataset_{001}$ to $Dataset_{011}$ due to the fact that more spectral variance wa~~i~~s related to the same target information concerning SOC. This applies for three of the four spectral pre-processing variants (SG, SGCR, gapDer), while SNVd1 pre-processing displays an unexpected pattern with datasets including replicate measurements and multiple scans even outperforming those with averaged data. Overall SGCR resulted in the best model performance for data 'A' (Figure $8a_2$) and 'all' samples (Figure $8c_2$), while SG pre-processing resulted best for data 'B' (Figure $8b_1$). However, the latter does not apply for $Dataset_{000}$ where gapDer pre-processing resulted in the best model performance with $RMSE_{MV}$= 0.13.

The overall best pre-processing methods in this study were the combination of SG and CR as well as SG alone. SG was used successfully by many authors before for spectral pre-processing. CR was used by e.g. Viscarra Rossel et al. (2016) or Loum et al. (2016) with acceptable success. The combination of SG and CR could not be found in literature, though. SNV was applied before by other authors in order to remove baseline effects (Knadel et al., 2015; Minasny et al., 2011; Viscarra Rossel et al., 2006a). The pre-processing technique d1 was found to lead to poorer model results and rather unexpected performance patterns in this study. The former may have its cause in the tendency of d1 to increase noise (Leone et al., 2012; Stevens and Ramirez Lopez, 2014). We do not have an explanation for the latter, though. Leone et al. (2012) suggested the usage of SG in combination with d1 to solve the problem. For the usage of gapDer no comparison could be found in the literature.

Comparing the mean $RMSE_{MV}$, the ~~for~~ models built on samples ~~'A' and~~ 'B'~~, models build on samples 'B'~~ resulted in better model performance than those built on samples 'A' w~~h~~ith the exception of $Dataset_{100}$. ~~The locations of the 'B' samples were determined using the Kennard Stone algorithm, those of the 'A' samples with k means clustering algorithm. Fig. 7 allows an assessment of the data collected by those two sampling designs and shows no clear resemblance between the violin plots of 'A', 'B' , all samples, and the archive violin plot. But T~~the Mann–Whitney U test did not show a significant difference ~~between the archive data and the sample sets used in this study~~. 'A' samples ~~as well as 'B' samples~~, as well as 'B' samples, seem to represent the ~~LTFE SOC~~archive data in an adequate way. ~~As already mentioned above~~Nevertheless, the difference in the distribution of SOC values of 'A' and 'B' samples ~~might~~ may have led to ~~a~~ the observed different predictive capability in certain SOC value ranges. However, whether ~~If~~ this difference is the reason for the better performance of the 'B' models cannot be stated with certainty.

Comparing the results of $Dataset_{111}$ with those of $Dataset_{000}$, shows how the inclusion of all input data uncertainties ~~affect~~impaired~~s~~ ~~the model results~~model performance. It can be seen that a model without error propagation ($Dataset_{000}$) ~~reaches~~ achieved a mean $RMSE_{MV}$ of 0.12 % SOC and a mean R^2 of 0.86 using the pre-processing

method which delivered the best results. A model with error propagation ($Dataset_{111}$), on the other hand, ~~reaches~~ reached a mean $RMSE_{MV}$ of 0.16 % SOC and an R^2 of 0.77. This is further illustrated in Fig. 9 and could be expected, as $Dataset_{000}$ ~~contains~~ contained no input data uncertainties~~, .~~ ~~t~~The $RMSE_{MV}$ values, therefore, only correspond to the model building process. Overall, the best model performance which ~~does~~ did not consider error propagation ~~corresponds~~ corresponded to a mean $RMSE_{MV}$ of 0.12 % SOC (R^2=0.86). This model performance ~~is~~ was impaired by $\Delta RMSE_{MV}$ = 0.04% SOC while considering input data uncertainties (ΔR^2=0.09), and by $\Delta RMSE_{MV}$ = 0.12 (ΔR^2=0.17) considering an inappropriate pre-processing. The effect of the sampling design ~~amounts~~ amounted to a $\Delta RMSE_{MV}$ of 0.02% SOC (ΔR^2=0.05). Overall, the additional accounting of neighbouring sample locations during fold division not only for model validation CV but also for model tuning CV might still improve the performance of all models. This is currently not implemented in the applied R package caret. We will, therefore, opt for other implementations in future studies.

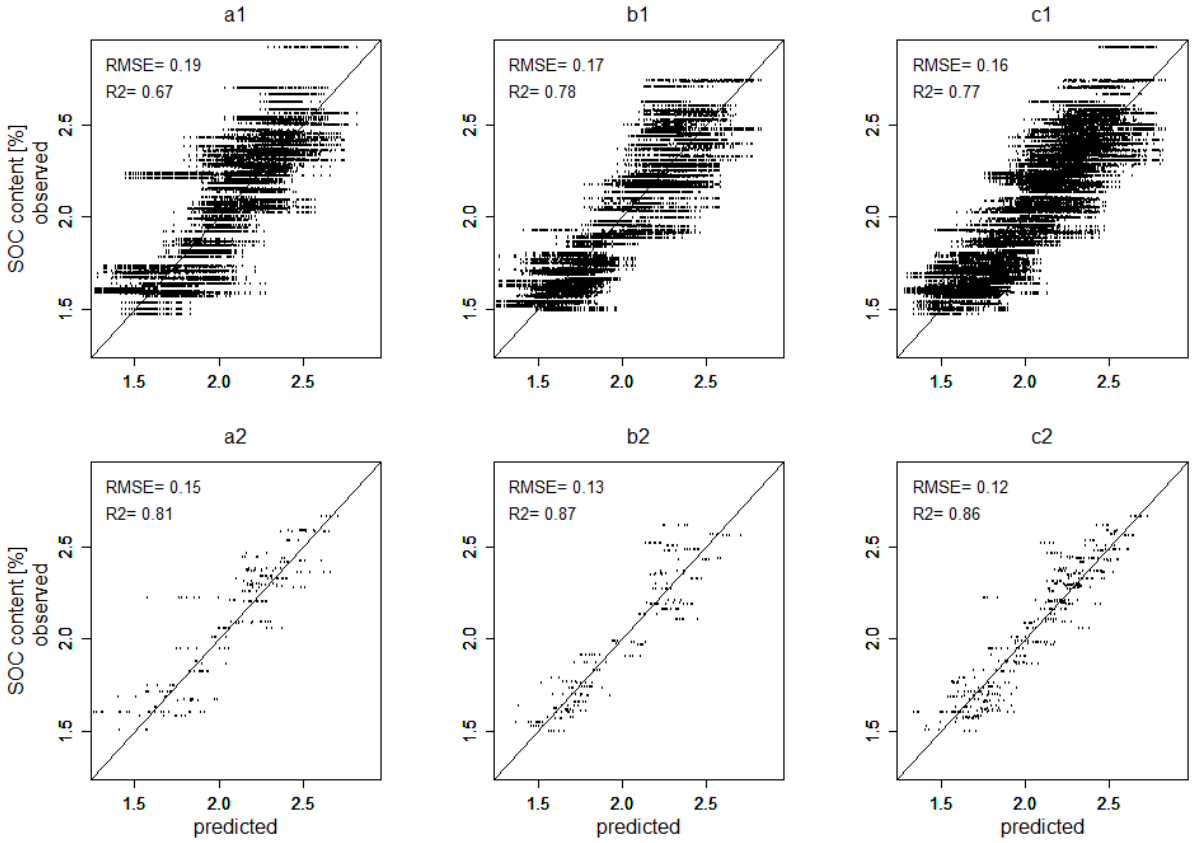


Fig. 9 Comparison of predicted and observed soil organic carbon (SOC) values for $Dataset_{111}$ (a1 to c1) and $Dataset_{000}$ (a2 to c2) for five repetitions with the corresponding best pre-processing (SGCR for data 'A' and 'all' ~~ADATA~~data, SG for data 'B'); a) shows results for 'A' samples, b) for 'B' samples and c) for all samples. The depicted RMSE and R^2 values refer to the mean of 5 repetitions.

Model performance values between studies that use VIS-NIR spectral information to predict soil properties are often compared to one another without mentioning the underlying range of the target variable, the variability of the measured soils, the applied sampling design, measurement protocol, validation approach, or applied instrumentation. Often, this information is not even provided by the respective studies. However, all of this has an impact on the calculated measures of model performance. Table 4 provides some of this information but not all details of the applied measurement protocol. ~~Examples of other studies using Vis-NIR spectra to predict SOC are listed in Table 4. Most studies~~The listed studies used a different number of scans and replicate samples to ~~get~~ calculate an averaged spectrum to predict SOC~~,~~. ~~so~~ Often, it is not specified whether the measurements refer to instrument internal scans, repeated external scans, or replicate measurements. As a consequence, the error implemented in the respective spectral input data ~~is~~ must be assumed to be different. Pimstein et al. (2011) propose~~r~~d a number of 3-5 replicate measurements as standard protocol for measu~~r~~ering Vis-NIR spectra of soil samples under laboratory conditions. Figure 5b indicated the high impact of ~~within~~ within-sample variance determined by the measurements of replicate samples, whereas the effect of the repeated scans per replicate ~~is~~ was comparatively small (compare Figures 5a and b). We dried and sieved the samples before spectral measurements but did not grind them to a fine powder. The latter might reduce the spectral variance in replicate measurements, but the benefit of Vis-NIR spectroscopy as a fast and inexpensive method ~~is~~ would be reduced. One might argue that samples ~~have~~ had to be ~~grinded~~ ground for SOC analysis, anyway. However, this requires a tiny fraction of the large amount that would have to be ~~grinded~~ ground for Vis-NIR measurements. In addition, comparison to measurements under field conditions would be~~is~~ further distorted while grinding the samples for laboratory measurements. ~~As shown in Figure 8, the input data error has a major influence on the model outcome.~~ In none of the~~se~~ studies listed in Table 4, the error in SOC measurements ~~is~~ was mentioned to be considered during model building. Also, in most studies the available dataset ~~i~~was randomly parted into calibration and validation set, using different ~~percentages~~ data proportions ~~of the data~~ for the two sets. Jeong et al.~~,~~ (~~-~~2017~~-~~) and Beleites et al. (2005) showed that different validation strategies le~~a~~d to different error ~~measures~~values. As shown in Figure 8, the input data uncertainty had a major influence on model performance. Accordingly, the applied measurement protocol should be reported in all details.

Table 4 R^2 model performance values from Vis-NIR applications to predict SOC~~from literature for soil organic carbon prediction models.~~

Authors	SOC [%]	Soil variability	instrumentation	~~number of samples~~ n	~~averaged scans per sample~~ m	calibration and validation set	R^2
~~(~~Reeves and Smith~~,~~ (2009)	0.04 – 34.2	various soil types (USA)	Digilab FTS 7000 Fourier tranform	720	64 (internal scans)	a) ~~Cross-validation with all samples~~LOO CV b) ~~Independent validation set~~cluster analysis based partition (½, ½)	a) 0.534 b) 0.335
Liu et al., 2018	0.10 - 3.40	various soil types (China)	ASD FieldSpec Pro	515	10 scans	Stratified systematic partition (75%, 25%)	0.74 – 0.83
~~(~~Islam et al.~~,~~ (2003)	0.06 - 4.95	various soil types (Australia)	Cary 500 (UV-VIS-NIR), Labsphere DRA CA-50D	161	-	~~Randomly selected from dataset~~random partition (~~121 / 40~~3/4, 1/4)	0.76
~~(Wang et al., 2014)~~	~~156~~		~~4~~	~~Randomly selected from dataset (116 / 40)~~		~~0.67 – 0.88~~	
~~(~~Volkan Bilgili et al.~~,~~ (2010)	0.39 - 0.69 (SOM)	various soil types (Turkey)	ASD FieldSpec Pro	512	~~100~~ 100 (2 replicates, 5 scans, 10	~~Randomly selected from dataset~~random partition (70~~-~~%~~,~~ ~~and~~ 30~~-~~%)	0.80

					internal scans)		
(Kuang and Mouazen, 2013)	174		10	60 % and 40 %	-		
(Jiang et al., (2016)	0.24 - 2.62	various soil types (China)	ASD FieldSpec 3	98	10	Dataset parted into calibration and validation setstratified partition (2/3, 1/3)	0.58 - 0.85
(Confort i et al., (2015)	0.3 - 6.5	various soil types (Italy)	ASD FieldSpec Pro	201	30	CV (unspecified)	-
Yang et al., 2019	0.24 - 6.05 (SOM)	single soil type, texture range (China)	ASD FieldSpec Pro	523	3 replicates, 10 internal scans	partition (2/3, 1/3)	0.81
(Leone et al., (2012)	0.04 - 21.56	various soil types (Italy)	ASD FiledSpec Pro	374	4 scans	Randomly selected from datasetrandom partition (2/3, and 1/3)	0.84 - 0.92

n = number of samples, m = averaged spectral measurements per sample, LOO CV = leave-one-out cross-validation

The overall best pre-processing methods in this study were the combination of SG and CR as well as SG alone. SG was used successfully by many authors before for spectral pre-processing (Bogrekci and Lee, 2006; Nocita et al., 2013; Stevens et al., 2013; Viscarra Rossel et al., 2006a). CR was used by e.g. Viscarra Rossel et al. (2016) or Loum et al. (2016) with acceptable success. The combination of SG and CR could not be found in literature, though. SNV was applied before by other authors in order to remove baseline effects (Knadel et al., 2015; Minasny et al., 2011; Viscarra Rossel et al., 2006a). The pre-processing technique d1 was found to lead to poorer model

~~results and rather unexpected performance patterns in this study. The former may have its cause in the tendency of d1 to over fitting and the increasing of noise in the data as reported by some authors before (Leone et al., 2012; Stevens and Ramirez Lopez, 2014). We do not have an explanation for the latter, though. Leone et al. (2012) suggests the usage of SG in combination with d1 to solve the problem. For the usage of gapDer no comparison could be found in literature. As there is no standard pre processing technique which works on all spectral data (Stenberg and Viscarra Rossel, 2010), it is recommended to always test various techniques and to choose the one which performs best for the respective data.~~

4 Conclusions

This study ~~aimed to investigate~~addressed the ~~influence~~ impact of various data and modelling~~different~~ aspects on ~~have on the model building process and the calculation of error measures~~model performance with a focus on the propagation of input data uncertainties.~~.~~ Overall, the best model performance which did not consider uncertainty propagation corresponded to a mean $RMSE_{MV}$ of 0.12% SOC (R^2=0.86). This model performance was impaired by $\Delta RMSE_{MV}$ = 0.04% SOC considering input data uncertainties (ΔR^2=0.09), and by $\Delta RMSE_{MV}$ = 0.12% SOC (ΔR^2=0.17) considering an inappropriate pre-processing. The effect of the sampling design amounted to a $\Delta RMSE_{MV}$ of 0.02% SOC (ΔR^2=0.05). ~~Those aspects included the input data uncertainties, the number of measurements per sample and the chosen pre processing method. Furthermore, the effect of sampling design, model tuning and validation procedure was discussed.~~ Overall, t~~T~~he applied nested k-fold group ~~cross validation~~CV approach ~~that includes resampling in model tuning as well as evaluation~~ can be recommended in general. ~~The fact that replicate measurements and scans as well as geographically near samples have to be assigned to the same fold (k-fold group CV) in order to obtain unbiased error measures is often neglected, and, therefore, has to be emphasized. Overall the best model performance which does not consider error propagation corresponds to a mean $RMSE_{MV}$ of 0.12 % SOC (R^2=0.86). This model performance is impaired by $\Delta RMSE_{MV}$ = 0.04% SOC while considering input data uncertainties (ΔR^2=0.09), and by $\Delta RMSE_{MV}$ = 0.12 (ΔR^2=0.17) considering an inappropriate pre processing. The effect of the sampling design amounts to a $\Delta RMSE_{MV}$ of 0.02% SOC (ΔR^2=0.05). We are aware that the consideration of stratified group CV for model evaluation but only partly for tuning (spectral replicate measurements and scans per sample were always assigned to the same fold) might impair model performance as suboptimal model parameters might be selected. This will be adapted in future studies.~~

~~The rather high within-sample variance of spectral replicate measurements of field-scale soil samples of very similar mineral composition requires a reconsideration of the number of replicate measurements per sample. 3-5~~

~~replicates as suggested for Vis-NIR soil measurements might simply not be enough. In general, this within-sample variability depends on the soil treatment and possibly also on the origin of the samples (e.g. agricultural or forest soils). Overall~~Furthermore, this study showed that it is of ~~important~~ uttermost importance to clarify which information is contained in the reported error ~~measure~~ values~~contains~~. ~~We are aware that the consideration of stratified group CV for model evaluation but only partly for tuning (spectral replicate measurements and scans per sample were always assigned to the same fold) might impair model performance as suboptimal model parameters might be selected. This will be adapted in future studies.~~ We, therefore, emphasize the necessity of a transparent and precise documentation of the measurement protocol, the model building and validation proce~~s~~dure, including the calculation of the error measure, in order to assess model ~~preformance~~ performance in a comprehensive way and allow for comparison between publications. Particularly, when Vis-NIR spectrometry is used for soil monitoring, the aspect of uncertainty propagation in the involved modelling procedure becomes essential.

Acknowledgements

We are grateful to the support of our colleagues from the UFZ departments Soil System Science, Community Ecology, Monitoring & Exploration Technologies, and Computational Landscape Ecology for support during the field campaign and the consequent sample preparation and laboratory analysis.

References

Abdi, H.: Partial Least Square Regression - PLS-Regression, in Encyclopedia of Measurement and Statistics, edited by N. Salkind, ThousandOaks (CA): Sage., 2007.

Adamchuk, V. I. and Viscarra Rossel, R. A.: Development of On-the-Go Proximal Soil Sensor Systems, in Proximal Soil Sensing. Progress in Soil Science, edited by R. A. Viscarra Rossel, A. McBratney, and B. Minasny, pp. 15–28, Springer, Dordrecht., 2010.

Altermann, M., Rinklebe, J., Merbach, I., Körschens, M., Langer, U. and Hofmann, B.: Chernozem — Soil of the Year 2005, J. Plant Nutr. Soil Sci., 168, 725–740, doi:10.1002/jpln.200521814, 2005.

Barnes, R. J., Dhanoa, M. S., and Lister, S. J.: Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra, Applied Spectroscopy, 43(5), 772-777, 1989.

Ben-Dor, E., Irons, J. A. and Epema, A.: Soil Spectroscopy, in Manual of Remote Sensing, edited by A. Rencz, p. 111−188, J. Wiley & Sons, Inc., NewYork., 1999.

Ben Dor, E., Ong, C. and Lau, I. C.: Reflectance measurements of soils in the laboratory: Standards and protocols, Geoderma, 245-246, 112–124, doi:10.1016/j.geoderma.2015.01.002, 2015.

Beleites, C., Baumgartner, R., Bowman, C., Somorjai, R., Steiner, G., Salzer, R., & Sowa, M. G.: Variance reduction in estimating classification error using sparse datasets. Chemometrics and Intelligent Laboratory Systems, 79(1–2), 91–100, https://doi.org/10.1016/j.chemolab.2005.04.008, 2005.

Bird, M., Brookes, P. C., Chenu, C., Jastrow, J. D., Lal, R., Lehmann, J., Donnell, A. G. O., Parton, W. J., Whitehead, D. and Zimmermann, M.: The knowns , known unknowns and unknowns of sequestration of soil organic carbon, Agric. Ecosyst. Environ., 164, 80–99, 2013.

Bjørsvik, H.-R. and Martens, H.: Data Analysis: Calibration of NIR Instruments by PLS Regression, in Handbook of Near-Infrared Analysis, edited by D. A. Burns and E. W. Ciurczak, pp. 189 –205., 2008.

Bogrekci, I. and Lee, W. S.: Effects of soil moisture content on absorbance spectra of sandy soils in sensing phosphorous cencentrations using UV-VIS-NIR spectroscopy ~~EFFECTS OF SOIL MOISTURE CONTENT ON ABSORBANCE SPECTRA OF SANDY SOILS IN SENSING PHOSPHORUS CONCENTRATIONS USING UV-VIS-NIR SPECTROSCOPY~~, Trans. Asabe, 49(4), 1175–1180, 2006.

Brown, J. D. and Heuvelink, G. B. M.: Assessing Uncertainty Propagation through Physically Based Models of Soil Water Flow and Solute Transport, in Encyclopedia of Hydrological Sciences, edited by M. G. Anderson, pp. 1181–1195, Wiley, Chicester, UK., 2006.

Charrad, M., Ghazzali, N., Boiteau, V. and Niknafs, A.: NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set, J. Stat. Softw., 61(6), 1–36 [online] Available from: http://www.jstatsoft.org/v61/i06/, 2014.

Clark, R. N., and Roush, T. L.: Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications, Journal of Geophysical Research: Solid Earth, 89(B7), 6329-6340, 1984.

Conforti, M., Castrignanò, A., Robustelli, G., Scarciglia, F., Stelluti, M. and Buttafuoco, G.: Laboratory-based Vis-NIR spectroscopy and partial least square regression with spatially correlated errors for predicting spatial variation of soil organic matter content, Catena, 124, 60–67, doi:10.1016/j.catena.2014.09.004, 2015.

Croft, H., Kuhn, N. J. and Anderson, K.: On the use of remote sensing techniques for monitoring spatio-temporal soil organic carbon dynamics in agricultural systems, Catena, 94, 64–74, doi:10.1016/j.catena.2012.01.001, 2012.

Dalal, R. C. and Henry, R. J.: Simultaneous Determination of Moisture, Organic Carbon, and Total Nitrogen by Near Infrared Reflectance Spectrophotometry1, Soil Sci. Soc. Am. J., 50(1), 120, doi:10.2136/sssaj1986.03615995005000010023x, 1986.

Dierke, C. and Werban, U.: Geoderma Relationships between gamma-ray data and soil properties at an agricultural

test site, Geoderma, 199, 90–98, doi:10.1016/j.geoderma.2012.10.017, 2013.

Filzmoser, P.: Identification of multivariate outliers: A performance study, Austrian J. Stat., 34(2), 127–138 [online] Available from: http://www.stat.tugraz.at/AJS/ausg052/052Filzmoser.pdf, 2005.

Filzmoser, P. and Gschwandtner, M.: mvoutlier: Multivariate Outlier Detection Based on Robust Methods, 2017.

Ge, Y., Morgan, C. L. S., Grunwald, S., Brown, D. J. and Sarkhot, D. V.: Comparison of soil reflectance spectra and calibration models obtained using multiple spectrometers, Geoderma, 161(3-4), 202–211, doi:10.1016/j.geoderma.2010.12.020, 2011.

Gholizadeh, A., Boruvka, L., Sbaerioon, M., Vasat, R.: Visible, near-infrared, and mid-infrared spectroscopy applications for soil assessment with emphasis on soil organic matter content and quality: State-of-the-art and key issues. Applied Spectroscopy, 67/12, 1349-1362, 2013.

Guio Blanco, C. M., Brito Gomez, V. M., Crespo, P. and Ließ, M.: Spatial prediction of soil water retention in a Páramo landscape: Methodological insight into machine learning using random forest, Geoderma, 316(November 2017), 100–114, doi:10.1016/j.geoderma.2017.12.002, 2018.

Hastie, T., Tibshirani, R. and Friedman, J. H.: The Elements of Statistical Learning, 2nd ed., Springer, New York., 2009.

Heuvelink, G. B. M.: Propagation of error in spatial modelling with GIS, in Geographical Information Systems, edited by P. A. Longley, M. F. Goodchild, D. J. Maguire, and D. W. Rhind, pp. 207–217, New York, John Wiley & Sons., 1999.

Islam, K., Singh, B. and McBratney, A. B.: Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared reflectance spectroscopy, Aust. J. Soil Res., 41(6), 1101–1114, doi:10.1071/SR02137, 2003.

Jansen, M.: Prediction error through modelling concepts and uncertainty from basic data. Nutrient Cycling in Agroecosystems, 50(1), 247–253. https://doi.org/10.1023/A:1009748529970, 1998.

Jeong, G., Choi, K., Spohn, M., Park, S. J., Huwe, B. and Ließ, M.: Environmental drivers of spatial patterns of topsoil nitrogen and phosphorus under monsoon conditions in a complex terrain of South Korea, PLoS One, 12(8), 1–19, doi:10.1371/journal.pone.0183205, 2017.

Jiang, Q., Chen, Y., Guo, L., Fei, T. and Qi, K.: Estimating Soil Organic Carbon of Cropland Soil at Different Levels of Soil Moisture Using VIS-NIR Spectroscopy, Remote Sens., 8(9), 755, doi:10.3390/rs8090755, 2016.

Johnson, M. G.: Soil carbon sequestration: Quantifying this ecosystem service, Present. Oregon Soc. Soil Sci. Annu. Meet. Newport, OR, Febr. 28 - 29, 2008.

Kennard, R. W. and Stone, L. A.: Computer Aided Design of Experiment, Technometrics, 11(1), 137–148, 1969.

Knadel, M., Thomsen, A., Schelde, K. and Greve, M. H.: Soil organic carbon and particle sizes mapping using vis-NIR, EC and temperature mobile sensor platform, Comput. Electron. Agric., 114(April), 134–144, doi:10.1016/j.compag.2015.03.013, 2015.

~~Kokaly, R. F.: Investigating a physical basis for spectroscopic estimates of leaf nitrogen concentration, Remote Sens. Environ., 75(2), 153–161, doi:10.1016/S0034-4257(00)00163-2, 2001.~~

Körschens, M. and Pfefferkorn, A.: Bad Lauchstädt - The Static Fertilization Experiment and other Long-Term Field Experiments, edited by UFZ – Umweltforschungszentrum Leipzig-Halle GmbH., 1998.

Kuang, B. and Mouazen, A. M.: Non-biased prediction of soil organic carbon and total nitrogen with vis e NIR spectroscopy , as affected by soil moisture content and texture, Biosyst. Eng., 114(3), 249–258, doi:10.1016/j.biosystemseng.2013.01.005, 2013.

Kuhn, M.: caret: Classification and Regression Training, 2017.

Kuhn, M. and Johnson, K.: Applied Predictive Modeling [Hardcover]., 2013.

Lal, R.: Soil Carbon Sequestration Impacts on Global Climate Change and Food Security, Science (80-.)., 304, 1623 – 1627, doi:DOI: 10.1126/science.1097396, 2004.

Lê, S., Josse, J. and Husson, F.: FactoMineR: An R Package for Multivariate Analysis, J. Stat. Softw., 25(1), 1–18, doi:10.1016/j.envint.2008.06.007, 2008.

Leone, A. P., Viscarra Rossel, R. A., Amenta, P. and Buondonno, A.: Prediction of Soil Properties with PLSR and vis- NIR Spectroscopy : Application to Mediterranean Soils from Southern Italy, Curr. Anal. Chem., 8(February 2014), 283 – 299, doi:10.2174/157341112800392571, 2012.

Liu, Y., Zhou, S., Zhang, G., Chen, Y., Li, S., Hong, Y., Shi, T., Wang, J. and Liu, Y.: Application of spectrally derived soil type as ancillary data to improve the estimation of soil organic carbon by using the Chinease soil Vis-NIR spectral library, Remote Sensing, 10 (1747), 1-16, doi:10.3390/rs10111747, 2018.

Lorenz, K. and Lal, R.: Soil Organic Carbon- An Appropriate Indicator to Monitor Trends of Land and Soil Degradation within the SDG Framework?, edited by S. M. Starke and K. Ehlers, Umweltbundesamt., 2016.

Loum, M., Diack, M., Ndour, N. Y. B. and Masse, D.: Effect of the Continuum Removal in Predicting Soil Organic Carbon with Near Infrared Spectroscopy (NIRS) in the Senegal Sahelian Soils, Open J. Soil Sci., 06(09), 135–148, doi:10.4236/ojss.2016.69014, 2016.

Martens, H. and Næs, T.: Multivariate Calibration, JohnWiley & Sons, Chichester, United Kingdom, UK., 1989.

McBratney, A. B., Stockmann, U., Angers, D. A., Minasny, B. and Field, D. J.: Challenges for Soil Organic Carbon Research, in Soil Carbon. Progress in Soil Science, edited by A. E. Hartemink and K. McSweeney, p. 57, Springer International Publishing Switzerland., 2014.

Meersmans, J., Wesemael, B. Van and Molle, M. Van: Determining soil organic carbon for agricultural soils : a comparison between the Walkley & Black and the dry combustion methods (north Belgium), Soil Use Manag., 25, 346–353, doi:10.1111/j.1475-2743.2009.00242.x, 2009.

Merbach, I. and Schulz, E.: Long-term fertilization effects on crop yields , soil fertility and sustainability in the Static Fertilization Experiment Bad Lauchstädt under climatic conditions 2001 – 2010, Arch. Agron. Soil Sci., 59(8), 1041 – 1057, doi:10.1080/03650340.2012.702895, 2013.

Minasny, B., McBratney, A. B., Bellon-Maurel, V., Roger, J.-M., Gobrecht, A., Ferrand, L. and Joalland, S.: Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon, Geoderma, 167-168, 118–124, doi:10.1016/j.geoderma.2011.09.008, 2011.

Molinaro, A.M., Simon, R., Pfeiffer, R.M.: Prediction error estimation: a comparison of resampling methods. Bioinformatics, 21(15), 3301–3307, 2005

Mortensen, P.: Myth: A partial least squares calibration model can never be more precise than the reference method..., NIR news, 25/ 3, 20-22, 2014.

Mouazen, A. M., Kuang, B., Baerdemaeker, J. De and Ramon, H.: Comparison among principal component , partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy, Geoderma, 158(1-2), 23–31, doi:10.1016/j.geoderma.2010.03.001, 2010.

~~Mutanga, O. and Skidmore, A. K.: Continuum-removed absorption features estimate tropical savanna grass quality in situ, 3rd EARSEL Work. Imaging Spectrosc. Herrsching, 542–558, 2003.~~

Naes, T., IsakssonT., Fearn, T. and Davies, T.: A User Friendly Guide to Multivariate Calibration and Classification, 2002.

Nieder, R. and Benbi, D. K.: Carbon and Nitrogen in the Terrestrial Environment, Springer Netherlands., 2008.

Nduwamungu, C., Ziadi, N., Parent, L.-E., Tremblay, G.F., Thuriès, T.: Opportunities for, and Limitations of, Near Infrared Reflectance Spectroscopy Applications in Soil Analysis: A Review, Can. J. Soil Sci.. 89(5): 531-541, 2009.

Nocita, M., Stevens, A., Noon, C. and Van Wesemael, B.: Prediction of soil organic carbon for different levels of soil moisture using Vis-NIR spectroscopy, Geoderma, 199, 37–42, doi:10.1016/j.geoderma.2012.07.020, 2013.

Pilorget, C., Fernando, J., Ehlmann B. et al.: Wavelength dependence of scattering properties in the VIS–NIR and links with grain-scale physical and compositional properties, Icarus, 267, 296-314, 2016.

Pimstein, A., Notesco, G. and Ben-Dor, E.: Performance of Three Identical Spectrometers in Retrieving Soil Reflectance under Laboratory Conditions, Soil Sci. Soc. Am. J., 75(2), 746, doi:10.2136/sssaj2010.0174, 2011.

Poggio, L. and Gimona, A.: National scale 3D modelling of soil organic carbon stocks with uncertainty propagation - An example from Scotland, Geoderma, 232-234, 284-299, 2014.

Reeves, J. B. and Smith, D. B.: The potential of mid- and near-infrared diffuse reflectance spectroscopy for determining major- and trace-element concentrations in soils from a geochemical survey of North America, Appl. Geochemistry, 24(8), 1472–1481, doi:10.1016/j.apgeochem.2009.04.017, 2009.

Rinnan, Å., Berg, F. van den and Engelsen, S. B.: Review of the most common pre-processing techniques for near-infrared spectra, TrAC - Trends Anal. Chem., 28(10), 1201–1222, doi:10.1016/j.trac.2009.07.007, 2009.

Savitzky, A. and Golay, M. J. E.: Smoothing and Differentiation of Data by Simplified Least Squares Procedures., Anal. Chem., 36(8), 1627–1639, doi:10.1021/ac60214a047, 1964.

~~Schmidt, K. S. and Skidmore, A. K.: Exploring spectral discrimination of grass species in African rangelands, Int. J. Remote Sens., 22(17), 3421–3434, doi:10.1080/01431160152609245, 2001.~~

Schulz, E.: Static Fertilization Experiment Bad Lauchstädt‘, , www.ufz.de/index.php?en=37010 , 2017.

Schwartz, G., Eshel, G., Ben-Dor, E.: Reflectance Spectroscopy as a Tool for Monitoring Contaminated Soils. In: S. Pascucci, editor. Soil Contamination. New York: InTech, Pp. 67-90. doi: 10.5772/23661, 2011.

~~Smith, B. C.: Quantitative spectroscopy : theory and practice, Academic Press. [online] Available from: https://www.sciencedirect.com/science/book/9780126503586 (Accessed 18 January 2018), 2002.~~

Stenberg, B. and Viscarra Rossel, R. A.: Diffuse Reflectance Spectroscopy for High-Resolution Soil Sensing, in Proximal Soil Sensing. Progress in Soil Science, edited by Viscarra Rossel R. A., McBratney A., and Minasny B, pp. 29–47, Springer, Dordrecht., 2010.

Stenberg, B., Viscarra Rossel, R. A., Mouazen, A. M. and Wetterlind, J.: Visible and Near Infrared Spectroscopy in Soil Science, Adv. Agron. Vol 107, 107(10), 163–215, doi:10.1016/s0065-2113(10)07005-7, 2010.

Stevens, A. and Ramirez Lopez, L.: An introduction to the prospectr package, , 1–22 [online] Available from: http://cran.r-project.org/web/packages/prospectr/vignettes/prospectr-intro.pdf, 2014.

Stevens, A., Nocita, M., Tóth, G., Montanarella, L. and van Wesemael, B.: Prediction of Soil Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy, PLoS One, 8(6), doi:10.1371/journal.pone.0066409, 2013.

Stockmann, U., Adams, M. A., Crawford, J. W., Field, D. J., Henakaarchchi, N., Jenkins, M., Minasny, B., Mcbratney, A. B., Remy, V. De, Courcelles, D., Singh, K., Wheeler, I., Abbott, L., Angers, D. A., Baldock, J., Summers, D., Lewis, M., Ostendorf, B. and Chittleborough, D.: Visible near-infrared reflectance spectroscopy as a predictive indicator of soil properties, Ecol. Indic., 11, 123–131, doi:10.1016/j.ecolind.2009.05.001, 2011.

Swarbrick, B.: Near-infrared spectroscopy and its role in scientific and engineering applications, in Handbook of

Measurement in Science and Engineering, edited by M. Kutz, John Wiley & Sons, Inc., Hoboken, NJ, USA., 2016.

Tekin, Y., Ulusoy, Y., Tümsavas, Z. and Mouazen, A. M.: Online Measurement of Soil Organic Carbon as Correlated with Wheat Normalised Difference Vegetation Index in a Vertisol Field, Sci. World J., 2014.

VDLUFA: Methodenbuch Band I Die Untersuchung von Böden, in Das VDLUFA Methodenbuch, VDLUFA-Verlag, Darmstadt., 2012.

Varma, S. and Simon, R.: Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics, 7, 91. https://doi.org/10.1186/1471-2105-7-91, 2006

Viscarra Rossel, R. A., Walter, C. and Fouad, Y.: Assessment of two reflectance techniques for the quantification of the within-field spatial variability of soil organic carbon, Precis. Agric., (January), 697–703 [online] Available from: <Go to ISI>://WOS:000186384100105, 2003.

Viscarra Rossel, R. A., McGlynn, R. N. and McBratney, A. B.: Determining the composition of mineral-organic mixes using UV-vis-NIR diffuse reflectance spectroscopy, Geoderma, 137(1-2), 70–82, doi:10.1016/j.geoderma.2006.07.004, 2006a.

Viscarra Rossel, R. A., Walvoort, D. J. J., McBratney, A. B., Janik, L. J. and Skjemstad, J. O.: Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties, Geoderma, 131(1-2), 59–75, doi:10.1016/j.geoderma.2005.03.007, 2006b.

Viscarra Rossel, R. A., Behrens, T., Ben-Dor, E., Brown, D. J., Demattê, J. A. M., Shepherd, K. D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aichi, H., Barthès, B. G., Bartholomeus, H. M., Bayer, A. D., Bernoux, M., Böttcher, K., Brodský, L., Du, C. W., Chappell, A., Fouad, Y., Genot, V., Gomez, C., Grunwald, S., Gubler, A., Guerrero, C., Hedley, C. B., Knadel, M., Morrás, H. J. M., Nocita, M., Ramirez-Lopez, L., Roudier, P., Campos, E. M. R., Sanborn, P., Sellitto, V. M., Sudduth, K. A., Rawlins, B. G., Walter, C., Winowiecki, L. A., Hong, S. Y. and Ji, W.: A global spectral library to characterize the world's soil, Earth-Science Rev., 155(February), 198–230, doi:10.1016/j.earscirev.2016.01.012, 2016.

Volkan Bilgili, A., van Es, H. M., Akbas, F., Durak, A. and Hively, W. D.: Visible-near infrared reflectance spectroscopy for assessment of soil properties in a semi-arid area of Turkey, J. Arid Environ., 74(2), 229–238, doi:10.1016/j.jaridenv.2009.08.011, 2010.

Wang, Y., Lu, C., Wang, L., Song, L., Wang, R. and Ge, Y.: Prediction of Soil Organic Matter Content Using VIS / NIR Soil Sensor, Sensors & Transducers, 168(4), 113–119, 2014.

Wehrens, R.: Chemometrics with R - Multivariate Data Analysis in the Natural Sciences and Life Sciences, edited by G. P. Robert Gentleman, Kurt Hornik, Springer-Verlag Berlin Heidelberg., 2011.

Wickham, H. and Francois, R.: dplyr: A Grammar of Data Manipulation, 2017.

Yang, M., Xu, D., Chen, S., Li, H., Shi, Z.: Evaluation of machine learning approaches to predict soil organic matter and pH using Vis-NIR spectra, Sensors, 19 (263), 1-14, doi:10.3390/s19020263, 2019.

695