A global spectral library to characterize the world's soil

R.A. Viscarra Rossel, T. Behrens, E. Ben-Dor, D.J. Brown, J.A.M. Demattê, K.D. Shepherd, Z. Shi, B. Stenberg, A. Stevens, V. Adamchuk, H. Aïchi, B.G. Barthès, H.M. Bartholomeus, A.D. Bayer, M. Bernoux, K. Böttcher, L. Brodský, C.W. Du, A. Chappell, Y. Fouad, V. Genot, C. Gomez, S. Grunwald, A. Gubler, C. Guerrero, C.B. Hedley, M. Knadel, H.J.M. Morrás, M. Nocita, L. Ramirez-Lopez, P. Roudier, E.M. Rufasto Campos, P. Sanborn, V.M. Sellitto, K.A. Sudduth, B.G. Rawlins, C. Walter, L.A. Winowiecki, S.Y. Hong, W. Ji

PII: S0012-8252(16)30011-3
DOI: doi: 10.1016/j.earscirev.2016.01.012
Reference: EARTH 2222

To appear in: *Earth Science Reviews*

Received date: 30 October 2015
Revised date: 24 January 2016
Accepted date: 25 January 2016

A global spectral library to characterize the world's soil

R.A. Viscarra Rossel[a,*], T. Behrens[b], E. Ben-Dor[c], D.J. Brown[d], J.A.M. Demattê[e], K.D. Shepherd[f], Z. Shi[g], B. Stenberg[h], A. Stevens[i], V. Adamchuk[j], H. Aïchi[k], B.G. Barthès[l], H.M. Bartholomeus[m], A.D. Bayer[n], M. Bernoux[l], K. Böttcher[o], L. Brodský[p], C.W. Du[q], A. Chappell[a], Y. Fouad[r], V. Genot[s], C. Gomez[t], S. Grunwald[u], A. Gubler[v], C. Guerrero[w], C. B. Hedley[x], M. Knadel[y], H.J.M. Morrás[z], M. Nocita[aa], L. Ramirez-Lopez[ab], P. Roudier[x], E.M. Rufasto Campos[ac], P. Sanborn[ad], V.M. Sellitto[ae], K.A. Sudduth[af], B.G. Rawlins[ag], C. Walter[r], L.A. Winowiecki[ah], S.Y. Hong[ai], W. Ji[a,g,j]

Short title: *A global spectral library*

*Correspondence: R.A. Viscarra Rossel. E-mail: raphael.viscarra-rossel@csiro.au

[a]*CSIRO Land and Water, PO Box 1666, Canberra ACT 2601, Australia,* [b]*Institute of Geography, University of Tüebingen, Germany,* [c]*The Remote Sensing and GIS laboratory Department of Geography, PO Dox 39040, Tel-Aviv University 69989, Israel,* [d]*Washington State University, USA,* [e]*Department of Soil Science, College of Agriculture Luiz de Queiroz, São Paulo University, Piracicaba, São Paulo, Brasil,* [f]*World Agroforestry Centre, ICRAF, PO Box 30677-00100, Nairobi, Kenya,* [g]*Institute of Applied Remote Sensing and Information Technology, College of Environmental and Resource Sciences, Zhejiang University, 866 Yuhangtang Road, Hangzhou, 310058, China,* [h]*Swedish University of Agricultural Sciences, Department of Soil and Environment, PO Box 234, 532 23 Skara, Sweden,* [i]*Georges Lemaître Centre for Earth and Climate Research, Earth and Life Institute, UC Louvain, Louvain-la-Neuve, Belgium,* [j]*Bioresource Engineering Department, McGill University,*

Ste-Anne-de-Bellevue, Quebec, Canada, [k]*Higher School of Agriculture, Mograne Tunisia,* [l]*IRD, UMR Eco&Sols, SupAgro, 2 place Viala, 34060 Montpellier, France,* [m]*Laboratory of Geo-Information Science and Remote Sensing, Wageningen University, P.O. Box 47, 6700 AA Wageningen, the Netherlands,* [n]*Karlsruhe Institute of Technology (KIT), Institute of Meteorology and Climate Research, Atmospheric Environmental Research, Kreuzeckbahnstraße 19, 82467 Garmisch-Partenkirchen, Germany,* [o]*Joint Research Centre, Institute for Sustainability, Via E. Fermi 2749, 21027 Ispra, Italy; Finnish Environment Institute, Mechelininkatu 34 A, 00251 Helsinki, Finland,* [p]*Department of Soil Science and Soil Protection, Czech University of Life Sciences Prague, Czech Republic,* [q]*State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science Chinese Academy of Sciences East Beijing Road 71, Nanjing 210008, China,* [r]*UMR 1069 SAS, INRA, Agrocampus Ouest, Rennes, France,* [s]*University of Liege, Gembloux Agro-Bio Tech (Ulg?GxaBt?Belgium), Soil Science Unit, Passage des Déportés, 2, 5030 Gembloux, Belgium* [t]*IRD, UMR LISAH, INRA, SupAgro, F-34060 Montpellier, France,* [u]*Soil and Water Science Department, University of Florida, USA,* [v]*Swiss Soil Monitoring Network NABO, Agroscope, Reckenholzstr. 191, 8046 Zurich, Switzerland,* [w]*Departamento de Agroquímica y Medio Ambiente, Universidad Miguel Hernndez de Elche Avenida de la Universidad, E-03202, Elche, Spain,* [x]*Landcare Research, Private Bag 11052, Palmerston North, 4442, New Zealand,* [y]*Department of Agroecology, Aarhus University, Blichers Allé 20, PO 50, 8830 Tjele, Denmark,* [z]*Instituto de Suelos, Centro de Investigacin de Recursos Naturales (CIRN) –INTA, 1712 Castelar, Provincia de Buenos Aires, Argentina,* [aa]*Presence-Earthcollective, PO Box, 237 Patensie 6335, Eastern Cape, South Africa,* [ab]*BUCHI Labortechnik AG, Meierseggstr. 40, Switzerland,* [ac]*Departamento Académico de suelos, Facultad de Agronomía, Universidad Nacional Pedro Ruiz Gallo, Lambayeque, Perú,* [ad]*Ecosystem Science and Management Program, University of Northern British Columbia, 3333 University Way, Prince George, BC, V2N 4Z9, Canada,* [ae]*University of Molise,* *Dipartimento Agricoltura Ambiente Alimenti (DIAAA), v. De Sanctis, 86100 Campobasso (CB), Italy* [af] *USDA-ARS Cropping Systems and Water Quality Research Unit, Columbia, Missouri, USA,* [ag] *British Geological Survey, Keyworth, Nottingham, NG12 5GG, UK,* [ah] *International Center for Tropical Agriculture, CIAT, Nairobi, Kenya,* [ai] *National Institute of Agricultural Sciences, Rural Development Administration, Wanju 55365, Republic of Korea.*

Abstract

Soil provides ecosystem services, supports human health and habitation, stores carbon and regulates emissions of greenhouse gases. Unprecedented pressures on soil from degradation and urbanization are threatening agro-ecological balances and food security. It is important that we learn more about soil to sustainably manage and preserve it for future generations. To this end, we developed and analyzed a global soil visible–near infrared (vis–NIR) spectral library. It is currently the largest and most diverse database of its kind. We show that the information encoded in the spectra can describe soil composition and be associated to land cover and its global geographic distribution, which acts as a surrogate for global climate variability. We also show the usefulness of the global spectra for predicting soil attributes such as soil organic and inorganic carbon, clay, silt, sand and iron contents, cation exchange capacity, and pH. Using wavelets to treat the spectra, which were recorded in different laboratories using different spectrometers and methods, helped to improve the spectroscopic modelling. We found that modelling a diverse set of spectra with a machine learning algorithm can find the local relationships in the data to produce accurate predictions. The spectroscopic models that we derived are parsimonious and robust, and using them we derived a harmonized global soil attribute dataset, which might serve to facilitate research on soil at the global scale. This spectroscopic approach should help to deal with the shortage of data on soil to better understand it and to meet the growing demand for information to assess and monitor soil at scales ranging from regional to global. We hope that this work might reinvigorate our community's discussion towards larger, more coordinated collaborations and encourage other contributions. We also hope that use of the database will deepen our understanding of soil so that we might sustainably manage it and push the research outcomes of the soil, earth and environmental sciences towards applications that we have not yet dreamed of.

Keywords : soil spectral library; global soil dataset; soil vis–NIR spectra; vis–NIR spectroscopy, multivariate statistics; machine learning; wavelets.

1 Introduction

Soil is a vital component of the Earth's critical zone. It provides ecosystem services, filters water, supplies nutrients to plants, provides us with food, fibre and energy, supports human health and habitation, stores carbon, regulates the emissions of greenhouse gases and it affects our climate. There are unprecedented pressures on soil from degradation and urbanisation, which are threatening those functions, agro-ecological balances and food security. It is important that we learn more about soil to manage it sustainably, in the context of the Sustainable Development Goals (UN Sustainable Development Knowledge Platform, 2015), and to preserve it for future generations.

To gain a better understanding of soil, its properties, processes and functions, all of which vary at different spatial and temporal scales, we need to develop effective methods to measure and monitor it. Conventional laboratory methods used to analyse soil properties are generally impractical because they are time-consuming, expensive and sometimes imprecise (e.g. Lyons et al., 2011). Often, these methods need significant amounts of sample preparation, harmful reagents and sometimes use complex apparatus that are inadequate when many measurements are needed, for example for soil mapping, monitoring and modelling.

Visible and infrared spectroscopy can effectively characterize soil. Spectroscopic measurements are rapid, precise and inexpensive. The spectra encode information on the inherent composition of soil, which comprises minerals, organic compounds and water. The minerals and the tightly bound water are traits that soil has inherited from its parent material and has acquired during its formation from that material in response to its environment and treatment by man. All of these encodings are represented in the spectra as absorptions at specific wavelengths of electromagnetic radiation, and we can use measurements of them to describe soil both qualitatively and quantitatively.

Many researchers have shown that spectra in the visible and near infrared (vis–NIR) can characterize the chemical, physical and mineralogical composition of the soil (Stoner and Baumgardner, 1981; Ben-Dor and Banin, 1994). Broad weakly expressed absorption bands at wavelengths smaller than 1000 nm can result from chromophores and iron oxides; narrow, well-defined absorption bands near 1400 and 1900 nm are due to hydroxyl bonds and water; absorptions near 2200 nm arise from clay minerals; organic matter absorbs at various wavelengths throughout the vis–NIR spectrum. Spectroscopy also provides information on soil particle size and thus information on the soil matrix. Another attractive feature of spectroscopy is that spectra can be recorded at points or by imaging, from different platforms; by proximal sensing in the field, in the laboratory using sampled material, or from remote sensing platforms with multi- and hyperspectral capabilities (Figure 1).

Figure 1 Platforms near here

Visible–near infrared spectrometers have developed considerably over the past 30 years (Figure 2). Currently, new technologies that use microelectromechanical structures (MEMSs) (Johnson, 2015), thin film filters, lasers, light emitting diodes (LED), fibre optic assemblies, and high performance detector arrays (Coates, 2014) are being used to produce miniaturised hand-held instruments that are rugged and cheap. Continual improvements in computing and statistics have helped to extract useful information from the spectra and to improve our understanding of soil. Figure 2 shows citations for some of the earlier studies that report the effects of soil water, particle size and chemical composition on vis–NIR spectra, as well as a sample of published research to date.

Figure 2 Timeline near here

Over the past few decades the exponential increase in the numbers of articles on spectroscopy in the soil science and related literature (Guerrero et al., 2010) has been out of proportion to the reporting of truly novel research. Many of these articles report little more than the outcomes of multivariate calibrations with data from small experiments in individual fields, though there have been some for larger regions and countries. They have shown that the technique can be used to predict attributes such as amounts of organic carbon, clay and water in soil and cation exchange capacity. They have also shown that the predictions of some attributes, such as the soil's pH and the contents of plant nutrients, cannot be predicted consistently (Stenberg et al., 2010). But there has been much duplication of objectives and rather few publications describing significant advances or novel applications. It seems that the adoption of vis–NIR in laboratories is, in practice, only incremental and still waiting to happen.

These are reasons for the growing interest in an international database of vis–NIR spectra linked to information on the soil's composition. The database might then be used to further the research on soil vis–NIR spectroscopy and for the prediction of the soil's attributes, condition and functions where measurements of those qualities are lacking and would be too expensive to make using conventional laboratory methods (Viscarra Rossel et al., 2006; Nocita et al., 2015b).

In 2008 we began to develop a global library of soil vis–NIR spectra as a voluntary collaborative project in response to the growing interest mentioned above. We were scientists from six countries, each representing Africa, Asia, Australia, Europe, North America and South America. Part of our aim was to bring together a community of scientists to further the research, encourage the development of new applications and the adoption of spectroscopy in the soil, earth and environmental sciences. The spectral library might then be used for applications at a range of spatial scales, in the laboratory, in the field and from the air (Figure 1). The scientists in this core group discussed how the project might proceed and produced the general guidelines, standards and protocols for the project and for the consistent measurement of spectra in the laboratory (see Appendix A, B).

Here we report on this global effort and our findings thus far. We describe the development of the global vis–NIR spectral library. We show that the information it contains can be used to characterize soil and its variability and diversity globally, and that by deriving a spectral classification we can describe the associations between spectra, soil, land cover and geography. We also show the usefulness of the global database for predicting soil attributes, such as soil organic and inorganic carbon, clay, silt, sand and iron contents, cation exchange capacity and pH. We propose that this spectroscopic approach should help to deal with the shortage of data on soil (Sanchez et al., 2009) and to meet the growing demand for information to assess and monitor soil at scales ranging from regional to global.

2 The global vis–NIR spectroscopic database

For spectra to be included in the database, we requested that they be from air dry ≤2 mm soil and in the range between 350 to 2500 nm recorded in intervals of one, two, five or 10 nm. We requested a minimum set of analytical data, geographic location and metadata, but they were not always supplied. We asked contributors that the spectra be representative of the variation in their spectroscopic databases and if possible the variation of soil in their countries. Contributors were provided with guidelines, minimum requirements and the measurement protocol for consistent measurement of spectra in the laboratory. These are shown in Appendix A and B. To date, 23 631 soil spectra have been contributed to the database by around 45 soil scientists and researchers from 35 institutions. The contributors to the database so far, by country and continent are listed in Table 1. The number of spectra by both country and continent are given in Table 2.

Tables 1 and 2 Contributors and spectra by country near here

The global database has spectra from 92 countries, representing seven continents (Table 2). It includes spectra from soil in the World Soil Information (ISRIC) collection, which were recorded by the World Agroforestry Centre (ICRAF) (ICRAF, 2015). It also includes spectra with corresponding soil attribute data from other published, multi-national, national, regional and local databases. They are listed in Table 1. The geographic locations of the spectra in the database are shown in Figure 3.

Figure 3 World near here

There are many large gaps in Figure 3. Many countries are poorly represented with only very few spectra. We would like to encourage participation from as many countries as possible, particularly, we would like contributions from counties in Central and South America, Mexico, Canada, Russia and Eastern Europe, Africa and Asia.

2.1 Metadata

Spectra were recorded with Fieldspec®, Agrispec®, Terraspec® or Labspec® instruments (PANalytical Inc., formerly Analytical Spectral Devices–ASD, Boulder, CO), with a spectral range of 350–2500 nm and spectral resolution of 3 nm at 700 and 10 nm at 1400 and 2100 nm, and mostly with a contact probe® or muglite® lightsource also from PANalytical Inc.

The spectral resolution varies somewhat depending on type of spectrometer. However, like Knadel et al. (2013), we also found that different instrumental resolutions have no noticeable influence on the spectroscopic modelling. The most common material used to calibrate the instruments was a Spectralon® white reference panel, although a different standard (Pimstein et al., 2011) was also used in some cases to assess instrumental drift. The frequency of calibration with a reference panel ranged between once every measurement and once every 50 soil measurements, with a median of once every 10 measurements. The number of readings averaged during calibration and measurement ranged between 10 and 50 readings. The median number of readings averaged during instrument calibration was 30 and during measurements it was 10. The number of replicates per sample ranged between no replication and six replicates; in most instances, however, there was no replication.

Approximately 84 % of the spectra have coordinates recoded in WGS84 latitude and longitudes, which belong to 12 509 unique sites (Figure 2). Eleven percent of the spectra have no depth information recorded. Of those with a record, around 60 % of the spectra originate from within the 0–30 cm soil layer, 30 % from within the 30–100 cm layer and 10 % of the spectra originate from samples collected at depths greater than 1 m. Around 15 % of the spectra have information on the soil horizons from where they originate, 95 % are assigned a soil classification using the FAO–WRB system (IUSS Working Group WRB, 2006). Land cover is recorded for around 80 % of the samples. The number of samples in the database by WRB major soil groups and land cover type are listed in Table 3.

Table 3 WRB and land cover near here

Around 80 % of the samples have measurements of soil organic carbon and clay content, 65 % have pH measured in water, around 50 % have measurements of silt and sand contents, 30 % have pH measured in calcium chloride and cation exchange capacity (CEC) and 20 % have measurements of inorganic carbon and extractable iron contents (Table 4). Around 25 % of the measurements have records of the laboratory method used in the analyses.

Table 4 Laboratory analysis near here

Table 5 lists by continent and overall, the number of data, the soil attributes and their statistics.

Table 5 Summary of soil data near here

The mean of the organic C data in the global database is 2.16%, the distribution is positively skewed and the median is 1.00% (Table 5). The distributions of the data on inorganic C and extractable Fe content were also positively skewed. The median inorganic C content is 2.10% and the median extractable Fe content is 1.00% (Table 5). The average CEC of the samples is 17.5 $\text{cmol}_c\,\text{kg}^{-1}$, average pH_{Water} is 6.57 and on average the samples in the database have more sand than clay and silt (Table 5). The ranges of the soil attribute distributions are wide and their coefficients of variation large, underlining the varied and diverse origin of the samples (Table 2).

3 Methods

3.1 A spectral classification for characterizing soil globally

To test whether spectra can be used to characterize soil and its variation globally, we averaged the reflectance spectra from several depths (when recorded) to obtain a single spectrum of the soil at each site. In doing so, we used the spectra from only the top metre of soil; that gave us 12 509 individual units to analyse. We used spectra in the range from 350 to 2500 nm, re-sampled at 10-nm intervals, so that we had 216 wavelengths for this analysis.

3.1.1 Continuum removal

The lower curves in Figure 4 show the average reflectance spectrum of each continent and their corresponding standard deviation. All the spectra have a similar general form with reflectance increasing with increasing wavelength in the visible range

(400–700 nm) and a broad region within which are sharp absorption bands in the infrared (700–2500 nm). We removed the general form of the reflectance spectra by fitting a convex hull to each spectrum and computing the deviations from the hull (Clark and Roush, 1984). The upper curves in Figure 4 show the form of the average continuum removed (CR) spectra by continent. These CR spectra can be used to isolate and identify characteristic absorptions of minerals, organic compounds and water.

Figure 4 Spectra by continent

3.1.2 Principal component analyses

The CR spectra were centred and analysed by principal component analysis (PCA) with the iterative NIPALS algorithm (Martens and Næs, 1989). The algorithm avoids the computation of the covariance matrix, which when analysing large sets of data with many variates, can be computationally inefficient. We did not standardize the data to unit variance because all of our wavelengths are in the same units and the differences in variation between them are inherently important. We used both the scores and eigenvectors of the PCA to help interpret the global data. Table 6 shows the results from the PCA. The three leading principal components of the spectra described 86 % of the information. The remaining components represent only small proportions of the variance in the spectra and so they were not used subsequently.

Table 6 PCA

3.1.3 Fuzzy c-means classification

To provide a more general description of the 12 509 spectra, the first three principal component scores (Table 6) were classified by the fuzzy-*c*-means algorithm (Bezdek et al., 1984). We used fuzzy classification instead of a 'crisp' method of multivariate classification, such as the *k*-means technique, because the fuzzy approach provides information on class overlaps, which helped to account for the continuous and complex nature of the information in the spectra.

The fuzzy-*c*-means technique subdivides a set of multivariate data, in our case the PCA scores, into *c* classes so that the pooled within-class variance is minimized, and provides for individual units a fuzzy membership in each class centre, or centroid. The membership functions as they are called, are continuous and range from 0 to 1. As memberships approach zero the degree of similarity between the unit and the particular class decreases, and as they approach 1, similarity increases. A description of the fuzzy-*c*-means algorithm we used, based on Euclidean distances between individual scores and class centroids, can be found in Bezdek et al. (1984), and we direct the reader there for a full description.

To determine the optimal number of classes we used two validity indices. The partition coefficient (C_p) and partition entropy (E_p) (Bezdek et al., 1984). The indices focus on the within-class variance (or compactness) and the separation between the classes (or isolation), respectively. The optimal, most compact partition with the largest separation is obtained by maximization of the C_p and minimization of the E_p.

To assess the classification and the relative associations between the memberships, we also calculated a confusion index for each unit as the ratio of the second most dominant membership and the most dominant one. We plot and interpreted these in the Results below.

3.1.4 Correspondence analysis

We used a correspondence analysis to investigate the associations between soil type, land cover, and geography, and the spectral classes defined above. Correspondence analysis (CA) is conceptually similar to PCA in that it enables one to reduce the dimensions of the data into a few orthogonal components that explain most of the variation in the data. It is more general than PCA in that it can be applied to categories in a contingency table and not only to continuous data. Our main purpose in using it is to summarize the associations between the units in the rows (*i.e.* the spectral classes) and the variables in the columns, (*i.e.* the categories of soil type, land cover and geography) of the contingency table as a small number of principal coordinates. The technique transforms both units and variables into the same set of dimensions, so that one can visualize both the units and variables on the same space. The algorithm is described in Greenacre (2007).

We plotted the scores for each spectral class and soil type, land cover, continents and countries, jointly in ordination graphs to get insights into the associations between them and to provide general descriptions of the soil, as represented by the spectral classes, in the various countries and continents from which the spectra originate and the soil types and land cover class in which they occur.

3.2 Global prediction of soil attributes

For the spectroscopic modelling we used data from different depth layers, not the averaged spectra for each site, like for the analysis above.

3.2.1 Data and spectral screening, preprocessing and transformations

Not all 23 631 spectra have corresponding soil analytical data, and the analytical methods used to measure the attributes were not recorded for all soil samples. The records also show that different analytical methods were used to measure individual soil attributes (Table 4).

We did not attempt to harmonise the soil analytical data for each attribute because of the largely incomplete metadata. However, for each soil attribute, when discrepancies in the units used to report the data were apparent, we converted the units to a common form. Then, we identified outliers in both the spectra and the soil attributes, both visually and using the Mahalanobis distances on the correlations between the spectra and the soil attributes. These were then removed from the data set before the statistical analysis. Some attributes had strongly positively skewed distributions, and to stabilize their variances for the spectroscopic modelling we transformed the data to approximate normal distributions by taking either square roots or logarithms.

To standardize the spectra for the spectroscopic modelling, we first subtracted the reflectance of the first wavelength (with the minimum reflectance value) to correct for potential baseline shifts between the measurements. The measured reflectances were then converted to apparent absorbance as $A = \log_{10}(1/\text{Reflectance})$.

3.2.2 Denoising, compression and variable selection with wavelets

We used the discrete wavelet transform to denoise and compress the spectroscopic database, and thereby produce a more parsimonious representation of the spectra for the modelling. For this, we followed the approach described by Viscarra Rossel and Lark (2009). The decomposition was made using a Daubechies's wavelet with two vanishing moments (Daubechies, 1988). Once the wavelet decomposition was performed, we wanted to retain only those coefficients that produced the most parsimonious representation of the global spectra and would be useful in the spectroscopic modelling. Selection was based on the variance of the coefficients, regardless of the wavelet scale to which they belonged (Viscarra Rossel and Lark, 2009). The rationale is that coefficients with larger variances, which can occur at different wavelet scales, contain the systematic information in the spectra that is useful for regression, while coefficients with small variances are less likely to be useful and can be discarded. The selected coefficients were then used as the predictors in modelling and in the interpretation of the models.

3.2.3 Spectroscopic modelling for prediction of soil attributes

To train and then validate the spectroscopic models so that they could be used with confidence, the dataset for each soil attribute separately were split into a training and an independent validation set (roughly 75:25 split) by simple random sampling. To develop the spectroscopic models on the training set, we used the decision trees algorithm, Cubist. Quinlan (1992) provides detail on the algorithm, the construction of the trees and the quantification of their errors.

Briefly, Cubist partitions the response data into subsets in which their characteristics are similar with respect to their spectra and other predictors that might be used. A series of rules derived using `if`–`then` conditions defines the partitions, and these rules are arranged in a hierarchy. A condition may be simply based on one wavelength or, more often, it comprises several wavelengths. If it is true, then the next step is the prediction of the soil attribute by ordinary least-squares regression from the wavelengths in that partition. If the condition is not true, then the rule defines the next node in the tree, and the sequence of `if`–`then`–`else` is repeated. The result is that the regression equations, although general in form, are local to the partitions and their errors are smaller than they would otherwise be. It is possible that any one observation and its associated predictors satisfy more than one set of rules, in which case the average of the predictions is taken as the overall prediction. Both continuous and categorical variables are allowed in the conditions, but only numeric variables are used in the regression equations. Our implementation here is similar to that described by Viscarra Rossel and Webster (2012).

To interpret the output from the modelling, we plotted on scalograms, the wavelet coefficients that were used in more than 30 % of cases by the decision trees. We could then more clearly identify the dominant wavelengths that contributed to the models and the scales at which they occurred.

For each soil attribute we also performed the spectroscopic modelling separately on each of the six fuzzy-c-means 'crisp' classes that we described earlier. In this way we tested if a pre-classification of the spectra could further improve the spectroscopic modelling with Cubist.

3.2.4 Estimation and uncertainty

To quantify the uncertainty in the spectroscopic models that we used to estimate the soil properties, we followed the approach using the nonparametric bootstrap described in Viscarra Rossel and Hicks (2015). It involves modelling of the soil attributes with Cubist, as above, but using 100 bootstrap samples (Hastie et al., 2009; Viscarra Rossel, 2007). We assumed that the spectra in the global database would be independent and could be used with the bootstrap to measure the uncertainty in our analysis. We modelled each bootstrap realisation independently with Cubist and derived cumulative distribution functions for the predictions on the validation dataset, which we used to compute the mean estimates, their standard errors and 95% confidence limits to describe their uncertainty. When the soil data were log or square-root transformed, we computed the estimates and their 95% confidence limits on the transformed scale and then back-transformed them to assess the models in the original scale.

3.2.5 Assessment statistics

For each soil property we assessed the performance of the models by comparing the predicted values on the independent validation data set with the observed ones. In each case, the root mean squared error (RMSE) was used to quantify the inaccuracy of the estimates, the standard deviation of the error (SDE) to quantify their imprecision, and the mean error (ME) to quantify the bias. We also report the coefficient of determination (R^2), the ratio of performance to deviation (RPD) (Williams, 1987) and the concordance correlation coefficient (ρ_c) (Lin, 1989), which assesses the covariation and correspondence between our estimates and the original data. The ρ_c statistic combines measures of both precision and bias to determine how far the observed data deviate from the line of perfect concordance, which is the 1:1 line. It ranges from -1 to 1, where a value of 1 denotes perfect agreement, values >0.90 suggest excellent agreement, values between 0.80 and 0.90 substantial agreement, between 0.65 and 0.80 moderate agreement, and values <0.65 poor agreement. These categories are only indicative. We encourage readers, in their assessments of our results, to use the R^2, RPD and ρ_c values in conjunction with the measures of bias, imprecision and inaccuracy that we provide.

3.3 Harmonising the global soil database

We remodelled each soil attribute using 100 bootstraps and Cubist as described above, but this time using all of the available spectra with matching analytical data. These models were used to predict onto the entire database ($N = 23361$) for each of the nine soil properties. For each set of predictions, as above, we calculated the average estimates from the bootstraps, their standard errors and their 95% confidence limits. Thus for each of the nine soil properties in the database, we produced a complete set of soil attribute data that was harmonised by the spectroscopic method.

All of the spectroscopic and statistical analyses and modelling described above were performed using the R software (R Development Core Team, 2008).

4 Results

4.1 Characterizing soil globally using a spectral classification

The eigenvectors of the first three principal components are shown in Figure 5a. That of the first component is dominated by negative loadings around wavelengths that show characteristic absorptions for hematite and kaolinite. The eigenvector of the second component has positive loadings near wavelengths for the characteristic absorptions of 2:1 clay minerals and possibly organic matter, near 640 and 1850 nm (Viscarra Rossel and Hicks, 2015). The eigenvector of the third component has large positive loadings near 640 nm, which is attributed to organic matter and large negative loadings that are due largely to illitic and smectitic clays. In the eigenvectors of the second and third components there are also small loadings near 2340 and 2450 nm, which may be attributed to illite (Post and Noble, 1993), other minerals with metal–OH bonds and carbonates (Hunt and Salisbury, 1971).

Figure 5 PCA and fuzzy

Figure 5b shows scatter diagrams of the scores from the first three principal components, coloured by the six (crisp) classes from the fuzzy-*c*-means classification. We selected six classes because the partition coefficient, C_p, was maximized and the partition entropy, E_p, minimized at this partition (Table 7), which was then taken to represent the most satisfactory classification for the data.

Table 7 Fuzzy validity near here

The first principal component describes variations in the clay and iron oxide mineralogy of the global samples. On the left most (negative) parts of the first principal component axis (Figures 5b(i–ii)), there is class 2, the average spectrum of which (Figure 5e) is characterized by absorptions that depict weathered soil with abundant kaolinite (a 1:1 clay mineral) and hematite.

The next class along this same axis is class 3 the average spectrum of which (Figure 5g) is similar to that of class 2, but its absorptions are less intense. Class 4 is next along the first principal component axis (near the zero value on Figure 5b(i–ii)), its average spectrum (Figure 5i) is characterized by absorptions from goethite and 2:1 clay minerals. Unlike in classes 2 and 3 the presence of kaolinite in the samples of class 4, with its doublet absorption near 2160 and 2200 nm (Hunt and Salisbury, 1970), is not very apparent. Compared with the spectra in classes 2 and 3, the broad iron oxide absorption is smaller with its centre slightly shifted towards longer wavelengths near 950 nm, which are indicative of goethite (Sherman and Waite, 1985).

To end the mineralogical sequence on the first axis, there are classes 1 and 6. They appear in a similar position on the axis (Figures 5b(i–ii)) and represent soil with mainly 2:1 clay minerals but also some carbonate, that is, generally less weathered soil. The average spectrum of class 1 (Figure 5c) is characterized by absorptions that depict soil with abundant smectite (Clark et al., 1990), while the average spectrum of class 6 (Figure 5m) depicts soil with abundant illite (Post and Noble, 1993). The average spectra of both class 1 and class 6 also show small absorptions due to goethite.

The second and third principal components describe variations of the samples in terms of their mineralogy and organic matter contents. These components also differentiate between the 2:1 clay minerals. On the negative ends of the second and third axes (Figures 5b(i–iii)), there is class 5, the average spectrum of which (Figure 5k) is characterized by a small overall reflectance with a broad absorption between 400 and 1200 nm, which is characteristic of dark soils containing large amounts of soil organic matter.

The fuzzy memberships of the global spectra to each individual class are shown on the scatter diagrams of the scores of the first three principal components (Figures 5d, f, h, j, l, n (i–iii)). The points, coloured by the membership value, show the transitions and overlap between the classes. We think that they demonstrate the continuous, complex and diverse nature of the information in the spectra.

The relative associations between the first two most dominant memberships to each class are shown in Figure 6, and they support our results above.

Figure 6 Memberships dominant and associated

There are strong associations between class 1 and class 6, the classes that depict soil with 2:1 clay minerals and between them and class 4, which represents soil containing goethite as the dominant iron oxide (Figure 6). Memberships in class 2 are associated with those of class 3, both classes depicting weathered soils with abundant kaolinite and hematite. Class 3 has associations with class 4. Soil with large amounts of organic matter represented by class 5 shows weak associations with soil containing abundant iron oxides, classes 3 and 4, and smectite, class 1 (Figure 6).

4.1.1 Associations between spectra and soil type and land cover

The ordination diagrams from the correspondence analysis (CA) between the six spectral classes and soil type and land cover are shown in Figures 7a and 7b.

Figure 7 CA plots

The first two components explained 80 % and 93 % of the variance in the associations between the six classes and soil type and land cover, respectively. Figures 7a and b can be interpreted together with the respective CA contingency tables (Tables 8 and 9).

Table 8 and 9 correspondence contingencies soil and land cover near here

Vertisols are most closely associated with class 1 (Figure 7a; Table 8) which represents soil with abundant smectite and some carbonate. Rendzinas, which are most often derived from carbonate rocks plot in the upper right quadrant of Figure 7a, but nearest to class 1. Soil in class 1 is mostly associated with non-vegetated lands and pastures, but also with mixed farming and cropping (Figure 7b; Table 9).

Solonchaks and Arenosols occur in arid and semi-arid climates, while Nitosols and Ferralsols develop as a consequence of deep weathering. They are closely associated with classes 2 and 3 (Figure 7a; Table 8), which represent soil with abundant kaolinite and hematite. Soil in class 2 is associated mostly with non-vegetated lands and pastures, but also with mixed farming and forested areas (Figure 7b; Table 9). Class 3 soil, however, is most closely associated with non-vegetated lands and forests, but also pastures, cropping and forested land.

Cambisols, Fluvisols and Andosols are generally young soil types with little profile development, and are most closely associated with classes 4 and 6 (Figure 7a; Table 8), which represents soil with abundant goethite. Soil in class 4 is most closely associated with forested and cropped land, but also with pastures and land used for mixed farming (Figure 7b; Table 9).

Histosols and Phaeozems are most closely associated with class 5 (Figure 7a; Table 8), which represents soil with abundant organic matter. Soil in class 5 is fairly evenly distributed among land used for cropping, mixed farming, forests and pastures (Table 9).

Gleysols, Podzols, Fluvisols and Cambisols occur in wetter environments from either, fluvial, alluvial, colluvial or aeolian deposits, and are most closely associated with class 6 (Figure 6a; Table 8), which represents soil with abundant illite. Class 6 soil is fairly evenly distributed among land used for cropping, mixed farming and forests, but it is also associated with pastures (Table 9).

4.1.2 Associations between spectra and geography

The first two components from the CA explained 94 % of the variance in the associations between the six classes and the continents and 78 % between the six classes and the countries. The ordination diagrams from the CA between the six classes and the continents and countries are shown in Figures 7c and 7d, respectively. These diagrams can be interpreted together with the CA contingency table (Table 10).

Table 10 CA correspondence contingency geographic near here

The soil spectra that we have for Africa and South America are closely associated and are represented by classes 2 and 3 (upper right quadrant of Figure 7c), which depict weathered soil, typical of the tropics and with abundant kaolinite and hematite. Their spectra are shown in Figure 4. They have the largest proportions of samples in these classes (Table 10). The exception are the spectra from Argentina and Uruguay where Phaeozems and Vertisols occupy large areas. Their spectra are mostly represented by classes 1, 4, 5 (Table 10).

Antarctica is associated with classes 4 and 6 (bottom left quadrant of Figure 7c), which represent soil containing goethite and 2:1 clay minerals, respectively. The Antarctic soils in the database appear not to be deeply weathered and to contain small amounts of organic matter. They do not have samples in classes 2 and 3 and five, respectively (Table 10). The average spectrum from soil of the Ross Dependency in Antarctica (Figure 4) shows a much younger age of the soil with broad absorption characteristics of goethite near 1000 nm and those near 1400, 1900, and 2200 nm that might be attributed to micaceous minerals such as illite and swelling smectitic clays (Figure 4). This agrees with the mineralogical assessment of the region by Claridge (1965).

Asia and Europe are most closely associated with classes 6 and 4 (upper left quadrant of Figure 7c). These classes represent soil that is younger and contain abundant illite and goethite. Asia and Europe have the largest proportions of samples in these classes (Table 10), and their spectra show absorptions of illite near 2200 and 2340 nm (Figure 4).

The average spectra of the soil samples from North America and Oceania cover large latitudinal extents from equatorial, tropical to temperate and arctic regions (Figure 3) and their spectra represent varied mineral and organic soil composition. North American samples are represented largely by classes 1, 3, 4 and 6, representing soil with predominantly 2:1 clay minerals and iron oxides (Table 10). North America plots in the bottom right quadrant of Figure 7c, whilst Oceania plots on the bottom right quadrant of Figure 7c, towards the centre of the graph. Its samples are evenly represented by all six classes, pointing to the diversity of the Australian soil in the database.

Similar interpretations can be made for the soil from the countries in the database (Figure 7d). We note however, that our interpretations are based only on the soil spectra that are in the database and acknowledge that the composition of soil globally is likely to be more varied.

4.2 Global prediction of soil attributes

Following from the above analyses, it makes sense that vis–NIR spectra can be used to derive spectroscopic models that predict soil attributes. Table 5 lists the number of data that we had for the modelling of the soil attributes and their statistics. Correlations between the soil attributes and the spectra, described by the scores of their first three principal components, are given in Table 11.

Table 11 Correlations near here

The eigenvectors of the first principal component, which explains 55% of the variation in the spectra relates primarily to weathered soil mineralogy (Figure 4), particularly iron oxides and kaolinite. It has the strongest positive correlations to silt, inorganic C, and Fe, while the strongest negative correlations are to CEC and organic C (Table 11). The second principal component explains 16% of the spectral variation and represents smectitic mineralogy and organic matter in the soil (Figure 5). It is positively correlated to clay content, pH and CEC, and negatively correlated to organic C (Table 11). The eigenvectors of the third principal component explains 15% of the variation in the spectra and it relates to illitic and kaolinitic mineralogy. It is positively correlated to pH and CEC and negatively correlated to clay content (Table 11).

4.2.1 Validation of the global spectroscopic models

Treating the spectra with wavelets greatly reduced the dimensionality of the data by removing noise and irrelevant information for the modelling of the soil attributes. For example, to model organic C we needed only 125 wavelet coefficients, instead of 216 wavelengths, to model clay content we needed 71 coefficients and to model Fe we only needed 32 wavelet coefficients (Table 12). Therefore, wavelets improved the parsimony in the spectroscopic modelling with Cubist. Table 12 lists the overall number of data used to train the models and the number used to validate them.

Table 12 validation near here

The statistics in Table 12 are for the best predictions on the validation samples, some of which were obtained by modelling the data by spectral class. That is, for some attributes, such as organic C, pH, clay, and sand contents, modelling the data separately by spectral class (Figure 5), improved the estimates overall, while for inorganic C, extractable Fe, CEC and silt content, the pre-classification of the data before modelling with Cubist was inconsequential.

In Figure 8 we show the estimates of organic C obtained from Cubist with and without the pre-classification.

Figures 8 and 9 validation near here

The validation of the spectroscopic models to estimate organic C (Figure 8) and extractable Fe (Figure 9) were excellent with ρ_c of 0.92 and 0.91, respectively. Estimates of organic C were unbiased and their RMSE was 1.11 %. The validations of inorganic C and CEC were very good with ρ_c values of 0.87 and 0.82. Predictions of clay and silt contents were also very good with ρ_c values of 0.80 for clay and 0.79 for silt, and RMSEs of 10.26 % and 10.33 %, respectively (Table 12, Figure 9). The spectroscopic model for estimating pH was somewhat less precise and its estimates were moderately accurate producing an RMSE of 0.8 pH units and a ρ_c of 0.76. The model for sand was imprecise and the RMSE of its estimates was 18.83 % (Table 12, Figure 9).

The estimates of the soil properties were generally unbiased but, as with any regression, there was a tendency to overestimate smaller and underestimate larger values (Figures 8 and 9). The pre-classification of the spectra prior to the spectroscopic modelling reduced the smoothing of the Cubist estimates of organic C (Figures 8). The imprecision of the estimates (Table 12, Figures 8 and 9), is likely to be due to the diverse origin of the analytical data, the (unquantified) imprecision of the laboratory measurements and the absence of any replication in the analysis. The data comes from different laboratories from around the world, with measurements made using different analytical methods (Table 4).

4.2.2 Interpretation of spectroscopic models

Figure 10 shows scalograms, which display the wavelet coefficients that were used by Cubist to predict the soil properties, the scales at which they vary and their respective wavelengths.

Figure 10 Scalograms near here

In Figure 10, the abscissa on the bottom depict the particular wavelet coefficients used in the models and on the top their corresponding wavelengths and a sample A spectrum. The ordinate represents the wavelet scale. The third dimension, represented by colour intensity, indicates the amplitude (or degree of importance) of a particular coefficient at a particular scale.

For the soil attributes that we considered, except extractable Fe, the wavelet coefficients that were most important in the modelling with Cubist were those that occurred at the coarse scales (≥ 32) (Figure 10). They correspond to broad or complex absorptions in the visible and in the near infrared and contain the lower frequency systematic information in the spectra that are useful in the regressions (Viscarra Rossel and Lark, 2009). Fewer wavelet coefficients from medium and fine scales (≤ 16) were retained for the modelling of the soil attributes and when used, they were generally less important compared to those from the coarse scales (Figure 10). The coefficients from finer scales represent the high frequency, often uncorrelated random noise elements in the spectra and many of these were discarded and not used in the modelling. Hence there were generally fewer coefficients that were used at the finest scales (areas of green Figure 10 where the amplitude of the coefficients is 0). When coefficients at the finer scales (≤ 4) were used, they were at discrete locations that correspond to specific and mostly known absorptions of soil constituents. The exception is the model for sand, which appears to have fairly evenly used coefficients from all scales (Figure 10f).

At the coarsest scale, Cubist used the wavelet coefficients that correspond to absorptions throughout the visible range and in the near infrared near 1000 nm and beyond 2200 nm (Figure 10). At scale 32, the coefficients used correspond to absorptions between 400–500 nm, and near 680 nm. In the near infrared, the coefficients used correspond to absorptions near 1000 nm, near 1625 nm, and between 2200 nm and 2400 nm. Surprisingly, the model for CEC, did not use coefficients in the 2200–2500 nm range (Figure 10e), although this region contains absorptions that relate to the soil's mineral and organic composition. The model for extractable Fe did not use coefficients at the coarsest scale, but at scale 32 it used coefficients that correspond to absorptions in the visible and the short-wave near infrared up to around 1100 nm, that are likely to be due to iron oxides (Figure 10g).

Wavelet coefficients that correspond to absorptions in the visible range (Figure 10) up to the short-wave infrared around 1100 nm may be attributed to electronic transitions in atoms of iron oxides, primarily hematite and goethite (Sherman and Waite, 1985), but also organic matter (Viscarra Rossel and Hicks, 2015). Absorptions between 1000–1600 nm may be attributed to overtones of O–H vibrations in clay minerals and water, while those between 1600–1900 nm may be due to overtones of C–H and C–OH and O–H vibrations inorganic structures. The absorptions near 1400 nm and 1900 nm are due to a H–O–H vibrations of water adsorbed on mineral surfaces and in the structures of 2:1 clay minerals like smectite (Clark et al., 1990). Absorptions between 2000–2500 nm are due to soil clay minerals, carbonate and organic matter. Kaolinite absorbs near 2160 nm and 2200 nm, illite near 2200 nm, 2340 nm and 2450 nm, smectite absorbs near 2200 nm but also near 2230 nm, depending on the lattice metal configuration. Carbonates absorb near 2340 nm, and there are absorptions in this range that result from overtones and combination vibrations of organic matter compounds, including those of amines near 2100 nm, amides near 2030 nm, polysaccharides near 2140 nm, aliphatics near 2275 nm, carbohydrates near 2380 nm and methyls in the range between 2300–2500 nm (Viscarra Rossel and Behrens, 2010).

4.3 Harmonising the global soil attribute database

We used the spectroscopic models, described above, to predict onto the entire spectroscopic database. These predictions provide a harmonised set of soil attribute data because they were derived using a single method, vis–NIR spectroscopy, and come with estimates of uncertainty that are described by 95% confidence limits. Descriptive statistics for the harmonised data are given in Table 13.

Table 13 Statistical summary of harmonised dataset.

The statistics in Table 13 are summarised by continent as an indicative means to provide continental information and to show that our predictions for each attribute produced sensible values. In doing so, we acknowledge that the sampling over many of the worlds regions is sparse and strongly biased so that statistical comparisons of continental means and variances may not be entirely appropriate.

The spatial distribution of the harmonised data is sensible (Figure 11). For instance, it shows that there is more organic C in the soil of cooler and wetter environments towards the highest latitudes, except in Antarctica (Figure 11). The soil near the equator and that which occurs at mid-latitudes in either hemisphere, where most agriculture occurs also has more organic C. Soil in these regions also generally has larger CEC and is generally more acidic (Figure 11). The soil in Europe and Asia has more silt than soil that is older and deeply weathered near the equator, in Australia and Africa. Deeply weathered soil of the tropics, near the equator also have larger amounts of extractable Fe. Soil in the arid regions of southern United States, in Europe and Australia have more inorganic C.

Figure 11 Spatial distribution of sampled data near here

The harmonised soil data can also be interpreted considering the six spectral classes from the fuzzy-*c*-means classification (Figure 12).

Figure 12 Harmonised boxplots near here

Soil belonging to class 1, dominated by smectitic mineralogy, contains the most clay—on average around 45 % clay, the least amount of sand, and the largest pH and CECs (Figure 11). Although there was little evidence of carbonates in the average spectra, which commonly produce an absorption near 2335 nm (Clark et al., 1990), soil belonging to this class also contained more inorganic C than soil in other classes.

Soil with large CECs were associated with soil that has either large amounts of clay (class 1) or organic carbon (class 5). Weathered soil in classes 2 and 3, rich in kaolinite and hematite contain more sand—on average 60 % and 50 %, respectively—more extractable iron, smaller CEC and pH. On average, these soils have around 20 % clay and around 10 % and 20 % silt, respectively. Soil in classes 4 and 6, rich in goethite and illite, have the largest silt contents—on average around 25 % and 30 %, respectively, around 40 % sand and 15 % clay (Figure 12). The large amount of sand in the soil of these classes might be due to poor dispersion of stable micro-aggregates in strongly weathered soils. Class 4, like classes 2 and 3, also has larger amounts of extractable iron. Soil in class 5 contains the most organic carbon, the second largest CEC, after class 1, low pH, the least inorganic C and the least amount of iron (Figure 12).

5 Discussion

5.1 The information content of the global spectra

Stoner and Baumgardner (1981) and Price (1990) suggested that the diversity of soil reflectance spectra could be explained with four or five characteristic reflectance curves. We now know that generalised spectral curves or a spectral classification, like we did above, are useful for organising and then describing the information content of soil spectra. The spatial (and temporal) variation in soil reflectance in the vis–NIR cannot be adequately described by such general descriptions—this is confirmed by our analyses. For example, none of the six spectral classes can describe soil type variability and each soil type is associated with more than one spectral class (Figure 7; Table 8).

Soil reflectance, like other soil properties, varies continuously and the resulting spectra represent complex compositional mixtures of soil materials from diverse origins, that are also affected by their environments. The membership functions from the fuzzy-c-means classification (Figure 5) show the continuous nature of soil spectral variation. Similar to soil classification, spectral classifications can help to understand, explain, teach and communicate, but they are not useful for adequately describing the variability of soil spectra.

5.2 Relative accuracy of the global predictions

The accuracies of the global spectroscopic models predictions were comparable to those reported in the literature for prediction at different scales (Figure 13; Tables with the review data in appendix C).

Figure 13 Review of literature

The accuracies of our estimates are similar to those of other studies conducted at continental and global scales (Figure 13). In some cases they are better and in other a little worse, although direct comparisons are difficult because there are no other studies made using a global dataset that is as large or diverse as the one here.

Our results show that vis–NIR spectroscopy can be used to predict soil attributes using historical soil spectroscopic databases, developed by different people and for different applications. Filtering and standardising the global spectra with wavelets helped account for the inconsistencies in sample preparation, different measurement protocols and instruments used in the many laboratories (Table 2). Modelling of the global spectra with a data mining machine learning algorithm helped to find local relationships in the data to produce relatively accurate predictions of the soil attributes studied (Figure 13). Soil spectroscopy is a highly reproducible analytical method so the inaccuracies in the spectroscopic models are largely due to the inconsistencies of the reference soil analyses.

We have not tested the use of the global database for predictions of soil attributes at local scales (e.g. Ramirez-Lopez et al., 2013; Guerrero et al., 2016). This was not our objective here, however, we believe that more research is needed to optimally use large spectroscopic databases for local predictions of soil attributes. Large databases such as this one, should at the very least help to improve the robustness of local spectroscopic models.

5.3 The global database for soil mapping, modelling and monitoring

Both the models and the harmonised data and uncertainties, may be used in different applications and for different purposes. For instance, the harmonised soil attribute data, with estimates of uncertainty, could be used to complement and help to improve regional, continental and global soil resource assessment at those scales. The spectra and spectroscopic models could be used to make predictions of soil attributes, mineralogy and soil type, where these measurements are lacking and would be too expensive to make by conventional laboratory means. The global spectroscopic database should also help to reduce the number of soil samples that need to be measured with the reference analytical method, thereby making the assessments of soil more affordable. The spectra can also be used as a proxy for classifying soil (Viscarra Rossel and Webster, 2011; Vasques et al., 2014) and could form the basis for a unifying and objective global soil classification system to organise our understanding and to help communicate and teach.

Although the spectroscopic estimates might be less accurate than laboratory measurements, the models provide rapid and inexpensive measures, with estimates of uncertainty, compared to the traditional wet chemistry and laboratory physics (Nanni and Demattê, 2006; Viscarra Rossel et al., 2006; Nocita et al., 2015a). If one cannot afford many conventional laboratory measurements then vis–NIR spectroscopy should provide harmonised soil data that are sufficiently accurate for mapping, modelling and for use in data-model assimilation techniques for monitoring. The accuracies of spectroscopic estimates made from using large continental spectroscopic databases, have been shown to be sufficient for large scale mapping of soil mineralogy (Viscarra Rossel et al., 2010; Viscarra Rossel, 2011), organic carbon (Viscarra Rossel et al., 2014) and other soil attributes across the whole of Australia and Africa (Viscarra Rossel et al., 2015; Vågen et al., 2016).

The largest uncertainty in the use of soil information is in predicting soil behaviour from measured soil attributes. Future efforts might therefore be best focused on relating soil behaviour and management responses directly to soil spectral, which provide a reliable and composite measure of mineral and organic composition.

5.4 Linking the global database with proximal and remote sensing

Spectroscopy, in the laboratory and in the field by proximal and remote sensing, has become an indispensable tool for soil, earth and environmental scientists who need soil information.

There is significant advantage to be gained by combining the use of laboratory spectra with proximal and remote sensing. Spectra measured in the laboratory provide a useful basis for proximal and remote sensing measurements. The advantages of proximal sensing are that one can measure the soil with minimal preparation, no interferences (*e.g.* from atmosphere, clouds or vegetation) and to depth. Remote sensing enables measurements over larger areas (and scales) and at potentially finer temporal resolutions. There are examples of laboratory, proximal, and remote vis–NIR spectroscopic sensing research, but few that combine their use (Ben-Dor and Banin, 1995; Ben-Dor et al., 2002; Gomez et al., 2008; Stevens et al., 2008). The reason might be that there are significant challenges posed by the inherent differences between the standardised laboratory measurements and those made under natural conditions, in the field.

Approaches are being developed to account for the effects of water and other environmental factors on proximally sensed spectra so that they may be used together with spectroscopic databases measured in the laboratory (e.g. Minasny et al., 2011; Ji et al., 2015). Accounting for these effects on remote sensing spectra are fundamentally more difficult because of the interferences mentioned above and because the natural roughness of the soil surface creates anisotropic patterns that cast shadows, which affect the measurements (e.g Pinty et al., 1989; Chappell et al., 2006; Croft et al., 2009). The global database might help to develop methodologies that bridge the gap between laboratory, proximal and remote sensing.

The global spectra might form the basis for new developments in hyperspectral remote sensing of soil, or at least, they may enable appropriate validation of the reflectance information extracted from the remote sensing products of current and future airborne and satellite hyperspectral spectrometers. The spectra might also be used for downscaling of the coarse resolution remote sensing images to help with regional assessments of soil condition. As we have shown here, vis–NIR spectra measure the inherent composition of the soil, so using the global spectra as baselines and proximal and remote sensors for monitoring changes in the spectra at those locations, might form a strong base for the development of an effective global soil monitoring network. This approach might be increasingly important to maintain the soil resource, human activities and food supply, as global population continues to grow.

6 Conclusions and future considerations

The global vis–NIR soil spectroscopic database we developed and analyzed is the largest and most diverse currently available. Its spectra can effectively describe global soil composition and our understanding of soil type. Information encoded in the spectra can be associated to land cover and its global geographic distribution, which may be acting as a surrogate for global climate variability.

We have shown that a global vis–NIR spectroscopic database describes soil variation and that the spectra provide an integrative measure of the soil, which can be used for both qualitative and quantitative soil analyses. We derived global spectroscopic models for prediction of individual soil attributes and their uncertainties. Using wavelets as a pretreatment before the spectroscopic modeling helped to remove unwanted background noise from the spectra. This allowed us to analyze a fairly inconsistent database with spectra and soil attributes that were measured in different laboratories using different spectrometers and methods and derive spectroscopic models that were parsimonious and robust.

We found that globally, modeling a diverse set of spectra with a data mining algorithm can, for most attributes, find the local relationships in the data to produce accurate predictions. Modeling regionally did not always help, except for some attributes (*e.g* soil organic C) where grouping the spectra into more homogeneous spectral classes (irrespective of geographical position) improved the modeling by removing bias in the predictions. Our results show that the global spectroscopic database can accurately estimate soil organic and inorganic C and extractable Fe and fairly accurately estimate CEC, clay and silt contents and pH. Using these spectroscopic models, we derived a harmonized global soil attribute dataset, which might facilitate research on soil and biogeochemical cycles at regional and global scales.

Soil vis–NIR spectroscopy is a versatile tool that can provide harmonized data with sufficient accuracy for different applications. It can help to overcome the world-wide shortage of soil data and it could also help to assess and monitor global changes in soil condition. We hope this work and the global vis–NIR spectroscopic database might reinvigorate our community's discussion of scientific practice towards larger, more coordinated collaboration and encourage other contributions. We also hope that use of the database will deepen our understanding of soil (so that we might sustainably manage it) and push the research outcomes of the soil, earth and environmental sciences towards applications that we have not yet dreamed of.

Acknowledgements

We thank the following people for helping to organise and contribute soil samples and/or spectra to the global library: H. Abbaslou for spectra from Iran, V. Barrón and colleagues for the soil samples from Andalucia, Spain, D. Brunet for spectra from Senegal, C. Castilla for spectra from Colombia, J. Eriksson for spectra from Sweden, S.T. Drummond and N.R. Kitchen for spectra from the USA, T. Kemper and S. Sommer for spectra from the Natura 2000 sites in Italy, B. Kusumo for spectra from New Zealand, A. Ringrose-Voase for spectra from the Philipinnes, S. Shibusawa and M. Kodaira for spectra from Japan, A. Sila for spectra from Kenya, and A. Thomsen for spectra from Denmark.

We thank also the two anonymous referees for their useful comments that helped to improve our manuscript.

References

Abdi, D., Tremblay, G. F., Ziadi, N., Bélanger, G., Parent, L. E., 2012. Predicting soil phosphorus-related properties using near infrared reflectance spectroscopy. Soil Science Society of America Journal 76 (6), 2318–2326.

Aïchi, H., Fouad, Y., Walter, C., Lili Chabaane, Z., Sanaa, M., 2013. Spatial quantification of total soil carbon, in Djerid arid area, by merging Vis-NIR laboratory data to ASTER image data. In: D. Arroays, N., McKenzie, N. J., Hempel, J., de Forges, A.-C. R., McBratney, A. B. (Eds.), Global Soil Map. CRC Press–Balkema, The Netherlands, pp. 461–463.

Aïchi, H., Fouad, Y., Walter, C., Viscarra Rossel, R. A., Chabaane, Z. L., Sanaa, M., 2009. Regional predictions of soil organic carbon content from spectral reflectance measurements. Biosystems Engineering 104 (3), 442–446.

Al-Abbas, A. H., Swain, P. H., Baumgardner, M. F., 1972. Relating organic matter and clay content to the multispectral radiance of soils. Soil Science 114 (6), 477–485.

Amare, T., Hergarten, C., Hurni, H., Wolfgramm, B., Yitaferu, B., Selassie, Y. G.,

2013. Prediction of Soil Organic Carbon for Ethiopian Highlands Using Soil Spectroscopy. ISRN Soil Science 2013 (Article ID 720589), 11.

Ångström, A., 1925. The albedo of various surfaces of ground. Geografiska Annaler 7, 323.

Awiti, A. O., Walsh, M. G., Shepherd, K. D., Kinyamario, J., 2008. Soil condition classification using infrared spectroscopy: A proposition for assessment of soil condition along a tropical forest-cropland chronosequence. Geoderma 143 (1–2), 73–84.

Bartholomeus, H. M., Schaepman, M. E., Kooistra, L., Stevens, A., Hoogmoed, W. B., Spaargaren, O., 2008. Spectral reflectance based indices for soil organic carbon quantification. Geoderma 145, 28–36.

Bartholomeus, H. M., Schaepman-Strub, G., Blok, D., Sofronov, R., Udaltsov, S., 2012. Spectral estimation of soil properties in siberian tundra soils and relations with plant species composition. Applied and Environmental Soil Science 2012, 241535.

Baumgardner, M. F., Silva, L. F., Biehl, L. L., Stoner, E. R., 1985. Reflectance properties of soils. Advances in Agronomy 38, 1–44.

Bayer, A. D., Bachmann, M., Müller, A., Kaufmann, H., 2012. A Comparison of Feature-Based MLR and PLS Regression Techniques for the Prediction of Three Soil Constituents in a Degraded South African Ecosystem. Applied and Environmental Soil Science 2012, Article ID 971252.

Bellinaso, H., Demattê, J. A. M., Araújo, S. R., 2010. Spectral library and its use in soil classification. Brazilian Journal of Soil Science 34, 861–870.

Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.-M., McBratney, A., 2010. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by nir spectroscopy. TrAC Trends in Analytical Chemistry 29 (9), 1073–1081.

Ben-Dor, E., Banin, A., 1994. Visible and near-infrared (0.4–1.1 μm) analysis of arid and semiarid soils. Remote Sensing of the Environment 48, 261–274.

Ben-Dor, E., Banin, A., 1995. Quantitative analysis of convolved thematic mapper spectra of soils in the visible near-infrared and shortwave-infrared spectral regions (0.4–2.5 μm). International Journal of Remote Sensing 16 (18), 3509–3528.

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

Ben-Dor, E., Patkin, K., Banin, A., Karnieli, A., 2002. Mapping of several soil properties using DAIS-7915 hyperspectral scanner data–a case study over clayey soils in Israel. International Journal of Remote Sensing 23 (6), 1043–1062.

Bezdek, J. C., Ehrlich, R., Full, W., 1984. FCM: The fuzzy c-means clustering algorithm. Computers and Geosciences 10 (2–3), 191–203.

Bilgili, A. V., van Es, H., Akbas, F., Durak, A., Hively, W., 2010. Visible–near infrared reflectance spectroscopy for assessment of soil properties in a semi-arid area of Turkey. Journal of Arid Environments 74 (2), 229–238.

Böttcher, K., Kemper, S., Machwitz, M., Mehl, W., Sommer, S., 2008. Chemometric modelling and remote sensing of arable land soil organic carbon as Mediterranean land degradation indicator–A case study in Southern Italy. Tech. Rep. EUR 23513 EN, Office for Official Publications of the European Communities, Luxembourg.

Bowers, S., Hanks, R., 1965. Reflection of radiant energy from soils. Soil Science 100 (2), 130–138.

Brodský, L., Klement, A., Penížek, V., Kodešová, R., L., B., 2011. Building soil spectral library of the Czech soils for quantitative digital soil mapping. Soil and Water Research 6, 165–172.

Brooks, F. A., 1952. Atmospheric radiation and its reflection from the ground. Journal of Meteorology 9, 41–52.

Brown, D. J., Bricklemyer, R. S., Miller, P. R., 2005. Validation requirements for diffuse reflectance soil characterization models with a case study of VNIR soil C prediction in Montana. Geoderma 129 (3–4), 251–267.

Brown, D. J., Shepherd, K. D., Walsh, M. G., Mays, M. D., Reinsch, T. G., 2006. Global soil characterization with VNIR diffuse reflectance spectroscopy. Geoderma 132 (3–4), 273–290.

Brunet, D., Bernoux, M., Barthès, B. G., 2008. Comparison between predictions of C and N contents in tropical soils using a vis–NIR spectrometer including a fibre-optic probe versus a NIR spectrometer including a sample transport module. Biosystems Engineering 100 (3), 448–452.

Butkuté, B., Šlepetiené, 2004. Near infrared reflectance spectroscopy as a fast method for simultaneous prediction of several soil quality components. Chemistry 15 (2), 12–20.

Cambule, A. H., Rossiter, D. G., Stoorvogel, J. J., Smaling, E. M. A., 2012. Building a near infrared spectral library for soil organic carbon estimation in the Limpopo National Park, Mozambique. Geoderma 183, 41–48.

Carter, W. T., 1931. Color analysis of soils with spectrophotometer. American Soil Survey Association Bulletin B12, 169–170.

Chang, C.-W., Laird, D. A., 2002. Near-infrared reflectance spectroscopic analysis of soil C and N. Soil Science 167 (2), 110–116.

Chang, C.-W., Laird, D. A., Hurburgh, G. R., 2005. Influence of soil moisture on near-infrared reflectance spectroscopic measurement of soil properties. Soil Science 170, 244–255.

Chang, C.-W., Laird, D. A., Mausbach, M. J., Hurburgh, C. R., 2001. Near-Infrared reflectance spectroscopy–principal components regression analyses of soil properties. Soil Science Society of America Journal 65 (2), 480–490.

Chappell, A., Zobeck, T., Brunner, G., 2006. Using bi-directional soil spectral reflectance to model soil surface changes induced by rainfall and wind-tunnel abrasion. Remote Sensing of the Environment 102 (3–4), 328–343.

Chodak, M., Ludwig, B., Khanna, P., Beese, F., 2002. Use of near infrared spectroscopy to determine biological and chemical characteristics of organic layers under spruce and beech stands. Journal of Plant Nutrition and Soil Science 165 (1), 27–33.

Chodak, M., Niklinska, M., Beese, F., 2007. Near-infrared spectroscopy for analysis of chemical and microbiological properties of forest soil organic horizons in a heavy-metal-polluted area. Biology and Fertility of Soils 44 (1), 171–180.

Christy, C., 2008. Real-time measurement of soil attributes using on-the-go near infrared reflectance spectroscopy. Computers and Electronics in Agriculture 61 (1), 10–19.

Chung, H., Hong, S., Song, W., Kim, Y., Hyun, B., Minasny, B., 2012. Predicting organic matter content in Korean soils using regression rules on visible—near infrared diffuse reflectance spectra. Korean Journal of Soil Science and Fertilizer 45 (4), 497–502.

Claridge, G., 1965. The clay mineralogy and chemistry of some soils from the Ross Dependency, Antarctica. New Zealand Journal of Geology and Geophysics 8 (2), 186–220.

Clark, R. N., King, T. V. V., Klejwa, M., Awayze, G. A., 1990. High spectral resolution reflectance spectroscopy of minerals. Journal of Geophysical Research 95 (B8), 12653–12680.

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

Coates, J., 2014. A review of new small-scale technologies for near infrared measurements.
URL www.americanpharmaceuticalreview.com/Featured-Articles/163573-A-Review-of-New-Small-Scale-Technologies-for-Near-Infrared-Measurements/

Cohen, M., Mylavarapu, R. S., Bogrekci, I., Lee, W., Clark, M. W., 2007. Reflectance spectroscopy for routine agronomic soil analyses. Soil Science (6), 469–485.

Cozzolino, D., Moron, A., 2003. The potential of near-infrared reflectance spectroscopy to analyse soil chemical and physical characteristics. The Journal of Agricultural Science 140 (1), 65–71.

Croft, H., Anderson, K., Kuhn, N. J., 2009. Characterizing soil surface roughness using a combined structural and spectral approach. European Journal of Soil Science 60 (3), 431–442.

Curcio, D., Ciraolo, G., D'Asaro, F., Minacapilli, M., 2013. Prediction of soil texture distributions using VNIR–SWIR reflectance spectroscopy. Procedia Environmental Sciences 19, 494–503.

Da-Costa, L. M., 1979. Surface soil color and reflectance as related to physicochemical and mineralogical soil properties. Ph.D. thesis, University of Missouri, Columbia.

Dalal, R. C., Henry, R. J., 1986. Simultaneous determination of moisture, organic carbon, and total nitrogen by near infrared reflectance spectrophotometry. Soil Science Society of America Journal 50, 120–123.

Daubechies, I., 1988. Orthonormal bases of compactly supported wavelets. Communications on Pure and Applied Mathematics 41 (7), 909–996.

Demattê, J. A. M., Campos, R. C., Alves, M. C., Fiorio, P. R., Nanni, M. R., 2004. Visible–nir reflectance: a new approach on soil evaluation. Geoderma 121 (1–2), 95–112.

Dick, W. A., Thavamani, B., Conley, S., Blaisdell, R., Sengupta, A., 2013. Prediction of β-glucosidase and β-glucosaminidase activities, soil organic C, and amino sugar N in a diverse population of soils using near infrared reflectance spectroscopy. Soil Biology and Biochemistry 56, 99–104.

Dong, Y.-W., Yand, S.-Q., Xu, C.-Y., Li, Y.-Z., Bai, W., Fan, Z.-N., Wang, Y.-N., Li, Q.-Z., 2011. Determination of soil parameters in apple-growing regions by near- and mid-infrared spectroscopy. Pedosphere 21 (5), 591–602.

Dunn, B. W., Batten, G. D., Beecher, H. G., Ciavarella, S., 2002. The potential of near infrared reflectance spectroscopy for soil analysis; a case study from the Riverine Plain of south-eastern Australia. Australian Journal of Experimental Agriculture 42 (5), 607–614.

Fernández Pierna, J. A., Dardenne, P., 2008. Soil parameter quantification by NIRS as a chemometric challenge at 'Chimiométrie 2006'. Chemometrics and Intelligent Laboratory Systems 91 (1), 94–98.

Fidêncio, P. H., Poppi, R. J., de Andrade, J. C., 2002. Determination of organic matter in soil using near-infrared spectroscopy and partial least squares regression. Communications in Soil Science and Plant Analysis 33 (9–10), 1607–1615.

Fontán, J. M., Calvache, S., López-Bellido, R. J., López-Bellido, L., 2010. Soil carbon measurement in clods and sieved samples in a mediterranean vertisol by visible and near-infrared reflectance spectroscopy. Geoderma 156 (3), 93–98.

Forouzangohar, M., Cozzolino, D., Kookana, R. S., Smernik, R. J., Forrester, S. T., Chittleborough, D. J., 2009. Direct comparison between visible near-and mid-infrared spectroscopy for describing diuron sorption in soils. Environmental Science & Technology 43 (11), 4049–4055.

Fystro, G., 2002. The prediction of C and N content and their potential mineralisation in heterogeneous soil samples using Vis–NIR spectroscopy and comparative methods. Plant and Soil 246 (2), 139–149.

Genot, V., Colinet, G., Bock, L., Vanvyve, D., Reusen, Y., Dardenne, P., 2011. Near infrared reflectance spectroscopy for estimating soil characteristics valuable in the diagnosis of soil fertility. Journal of Near Infrared Spectroscopy 19 (2), 117–138.

Gomez, C., Viscarra Rossel, R. A., McBratney, A. B., 2008. Soil organic carbon prediction by hyperspectral remote sensing and field vis–NIR spectroscopy: An Australian case study. Geoderma 146 (3–4), 403–411.

Greenacre, M., 2007. Correspondence Analysis in Practice, second edition Edition. London.

Gubler, A., 2011. Quantitative estimations of soil properties by Visible and Near Infrared Spectroscopy–applications for laboratory and field measurements. Ph.D. thesis, University of Bern.
URL `http://www.zb.unibe.ch/download/eldiss/11gubler_a.pdf`

Guerrero, C., Viscarra Rossel, R. A., Mouazen, A. M., 2010. Preface Special issue 'Diffuse reflectance spectroscopy in soil science and land resource assessment'. Geoderma 158 (1–2), 1–2.

Guerrero, C., Wetterlind, J., Stenberg, B., Mouazen, A. M., Gabarrón-Galeote, M. A., Ruiz-Sinoga, J. D., Zornoza, R., Viscarra Rossel, R. A., 2016. Do we really need large spectral libraries for local scale SOC assessment with NIR spectroscopy? Soil and Tillage Research 155, 501—509.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer Series in Statistics, Springer.

He, Y., Huang, M., Garcia, A., Hernandez, A., Song, H., 2007. Prediction of soil macronutrients content using near-infrared spectroscopy. Computers and Electronics in Agriculture 58 (2), 144–153.

Heinze, S., Vohland, M., Joergensen, R. G., Ludwig, B., 2013. Usefulness of nearinfrared spectroscopy for the prediction of chemical and biological soil properties in different longterm experiments. Journal of Plant Nutrition and Soil Science 176 (4), 520–528.

Hester, J. B., 1939. The relation of soil texture and color to the organic matter. Soil Science Society of America Proceedings 3, 112–114.

Hill, J., Schütt, B., 2000. Mapping complex patterns of erosion and stability in dry mediterranean ecosystems. Remote Sensing of Environment 74 (3), 557–569.

Hunt, G., 1977. Spectral signatures of particulate minerals, in the visible and near-infrared. Geophysics 42, 501–513.

Hunt, G. R., Salisbury, J. W., 1970. Visible and near-infrared spectra of minerals and rocks. I. Silicate minerals. Modern Geology 1 (4), 283–300.

Hunt, G. R., Salisbury, J. W., 1971. Visible and near infrared spectra of minerals and rocks. II. Carbonates. Modern Geology Geology 2, 23–30.

ICRAF, 2015. A globally distributed soil spectral library visible near infrared diffuse reflectance spectra.
URL `http://worldagroforestry.org/sites/default/files/Description_ICRAF-ISRIC%20Soil%20VNIR%20Spectral%20Library.pdf`

Islam, K., Singh, B., McBratney, A. B., 2003. Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared reflectance spectroscopy. Australian Journal of Soil Research 41 (6), 1101–1114.

IUSS Working Group WRB, 2006. World reference base for soil resources 2006. World Soil Resources Report 103, Food and Agriculture Organisation, Rome.

Ji, W., Shi, Z., Huang, J., Li, S., 2014. In situ measurement of some soil properties in paddy soil using visible and near–Infrared spectroscopy. PLoS ONE 9 (8), e105708.

Ji, W., Viscarra Rossel, R. A., Shi, Z., 2015. Accounting for the effects of water and the environment on proximally sensed vis–NIR soil spectra and their calibrations. European Journal of Soil Science 66 (3), 555–565.

Johnson, R. C., 2015. 1st MEMS spectrometer debuts.
URL `http://www.eetimes.com/document.asp?doc_id=1325630`

Kemper, T., Sommer, S., 2002. Estimate of heavy metal contamination in soils after a mining accident using reflectance spectroscopy. Environmental Science & Technology 36, 2742—2747.

Knadel, M., Deng, F., Thomsen, A., Greve, M., 2012. Development of a Danish national Vis–NIR soil spectral library for soil organic carbon determination. In: Minasny, B., Malone, B. P., McBratney, A. B. (Eds.), Digital Soil Assessments and Beyond: Proceedings of the 5th Global Workshop on Digital Soil Mapping 2012, Sydney, Australia. CRC Press, pp. 403–408.

Knadel, M., Stenberg, B., Deng, F., Thomsen, A., Greve, M., 2013. Comparing predictive abilities of three visible-near infrared spectrophotometers for soil organic carbon and clay determination. Journal of Near Infrared Spectroscopy 21 (1), 67–80.

Kooistra, L., Wanders, J., Epema, G. F., Leuven, R. S. E. W., Wehrens, R., Buydens, L., 2003. The potential of field spectroscopy for the assessment of sediment properties in river floodplains. Analytica Chimica Acta 484 (2), 189–200.

Kooistra, L., Wehrens, R., Leuven, R. S. E. W., Buydens, L. M. C., 2001. Possibilities of visible–near-infrared spectroscopy for the assessment of soil contamination in river floodplains. Analytica Chimica Acta 446 (1), 97–105.

Krishnan, P., Alexander, D. J., Butler, B., Hummel, J. W., 1980. Reflectance technique for predicting soil organic matter. Soil Science Society of America Journal 44, 1282–1285.

Kusumo, B., Hedley, C., Hedley, M., Hueni, A., Tuohy, M., Arnold, G., 2008. The use of diffuse reflectance spectroscopy for in situ carbon and nitrogen analysis of pastoral soils. Soil Research 46 (7), 623–635.

Leamer, R. W., Myers, V. I., Silva, L. F., 1973. A spectroradiometer for field use. Review of Scientific Instruments 44 (5), 611–614.

Lee, K. S., Lee, D. H., Sudduth, K. A., Chung, S. O., Kitchen, N. R., Drummond, S. T., 2009. Wavelength identification and diffuse reflectance estimation for surface and profile soil properties. Transactions of the ASABE 52 (3), 683–695.

Lee, K. S., Sudduth, K. A., Drummond, S. T., Lee, D. H., Kitchen, N. R., Chung, S. O., 2010. Calibration methods for soil property estimation using reflectance spectroscopy. Transactions of the ASABE 53 (3), 675–684.

Leone, A., Viscarra Rossel, R. A., Amenta, P., Buondonno, A., 2012. Prediction of Soil Properties with PLSR and vis-NIR Spectroscopy: Application to Mediterranean Soils from Southern Italy. Current Analytical Chemistry 8 (2), 283–299.

Lin, L. I.-K., 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics 45 (1), 255–268.

Lu, P., Wang, L., Niu, Z., Li, L., Zhang, W., 2013. Prediction of soil properties using laboratory VIS–NIR spectroscopy and Hyperion imagery. Journal of Geochemical Exploration 132, 26–33.

Luce, M. S., Ziadi, N., Zebarth, B. J., Grant, C. A., Tremblay, G. F., Gregorich, E. G., 2014. Rapid determination of soil organic matter quality indicators using visible near infrared reflectance spectroscopy. Geoderma 232, 449–458.

Lyons, D., Rayment, G., Hill, R., Daly, B., Marsh, J., Ingram, C., 2011. Aspac soil proficiency testing program report 2007–08. Tech. rep., ASPAC, Melbourne, Victoria.
URL http://www.aspac-australasia.com/index.php/documents/upload-documents/doc_download/232-annual-review-soil-07-08

Ma, Z. Y., Du, C. W., Zhou, J. M., Zhou, G. Q., Viscarra Rossel, R. A., 2012. Characterization of infrared reflectance and photoacoustic spectra of loess soil. Soils 44 (5), 862–867.

Madari, B. E., Reeves III, J. B., Machado, P. L. O. A., Guimarães, C. M., Torres, E., McCarty, G. W., 2006. Mid- and near-infrared spectroscopic assessment of soil compositional parameters and structural indices in two Ferralsols. Geoderma 136 (1–2), 245–259.

Malley, D. F., Martin, P. D., Ben-Dor, E., 2004. Application in analysis of soils. American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, pp. 729–784.

Martens, H., Næs, T., 1989. Multivariate Calibration. John Wiley & Sons.

Martin, P. D., Malley, D. F., Manning, G., Fuller, L., 2002. Determination of soil organic carbon and nitrogen at the field level using near–infrared spectroscopy. Canadian Journal of Soil Science 82 (4), 413–422.

McCarty, G. W., III, J. B. R., Reeves, V. B., Follett, R. F., Kimble, J. M., 2002. Mid-Infrared and near infrared diffuse reflectance spectroscopy for soil carbon measurement. Soil Science Society of America Journal 66, 640–646.

McCarty, G. W., Reeves III, J. B., 2006. Comparison of near infrared and mid infrared diffuse reflectance spectroscopy for field-scale measurement of soil fertility parameters. Soil Science 171 (2), 94–102.

Minasny, B., McBratney, A. B., 2008. Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. Chemometrics and Intelligent Laboratory Systems 94 (1), 72–79.

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

Morgan, C. L., Waiser, T. H., Brown, D. J., Hallmark, C. T., 2009. Simulated in situ characterization of soil organic and inorganic carbon with visible near-infrared diffuse reflectance spectroscopy. Geoderma 151 (3), 249–256.

Moros, J., de Vallejuelo, S. F.-O., Gredilla, A., de Diego, A., Madariaga, J. M., Garrigues, S., de la Guardia, M., 2009. Use of Reflectance Infrared Spectroscopy for Monitoring the Metal Content of the Estuarine Sediments of the Nerbioi-Ibaizabal River (Metropolitan Bilbao, Bay of Biscay, Basque Country). Environmental Science & Technology 43 (24), 9314–9320.

Morra, M. J., Hall, M. H., Freeborn, L. L., 1991. Carbon and nitrogen analysis of soil fractions using near infrared reflectance spectroscopy. Soil Science Society of America Journal 55, 288–291.

Mouazen, A. M., De Baerdemaeker, J., Ramon, H., 2005. Towars development of on-line soil moisture sensors using fibre-type nir spectrophotometer. Soil and Tillage Research 80, 171–183.

Nanni, M. R., Demattê, 2006. Spectral reflectance methodology in comparison to traditional soil analysis. Soil Science Society of America Journal 70, 393–407.

Nduwamungu, C., Ziadi, N., Tremblay, G. F., Parent, L. E., 2009. Near-infrared reflectance spectroscopy prediction of soil properties: Effects of sample cups and preparation. Soil Science Society of America Journal 73 (6), 1896–1903.

Nocita, M., Kooistra, L., Bachmann, M., Müller, A., Powell, M., Weel, S., 2011. Predictions of soil surface and topsoil organic carbon content through the use of laboratory and field spectroscopy in the Albany Thicket Biome of Eastern Cape Province of South Africa. Geoderma 167, 295–302.

Nocita, M., Stevens, A., van Wesemael, B., Aitkenhead, M., Bachmann, M., Barthés, B., Dor, E. B., Brown, D. J., Clairotte, M., Csorba, A., Dardenne, P., Demattê, J. A., Genot, V., Guerrero, C., Knadel, M., Montanarella, L., Noon, C., Ramirez-Lopez, L., Robertson, J., Sakai, H., Soriano-Disla, J. M., Shepherd, K. D., Stenberg, B., Towett, E. K., Vargas, R., Wetterlind, J., 2015a. Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring. Vol. 132 of Advances in Agronomy. Academic Press, Ch. Four, pp. 139–159.

Nocita, M., Stevens, A., van Wesemael, B., Brown, D. J., Shepherd, K. D., Towett, E., Vargas, R., Montanarella, L., 2015b. Soil spectroscopy: an opportunity to be seized. Global Change Biology 21 (1), 10–11.

Obukhov, A. I., Orlov, D. S., 1964. Spectral reflectivity of the major soil groups and possibility of using diffuse reflection in soil investigations. Soviet Soil Science 2, 174–184.

O'Neal, A. M., 1927. The effect of moisture on the color of certain Iowa soils. American Soil Survey Association Bulletin B8, 158–174.

Ouerghemmi, W., Gomez, C., Naceur, S., Lagacherie, P., 2011. Applying blind source separation on hyperspectral data for clay content estimation over partially vegetated surfaces. Geoderma 163 (3–4), 227–237.

Patzold, S., Mertens, F., Bornemann, L., Koleczek, B., Franke, J., Feilhauer, H., Welp, G., 2008. Soil heterogeneity at the field scale: a challenge for precision crop protection. Precision Agriculture 9 (6), 367–390.

Pietrzykowski, M., Chodak, M., 2014. Near infrared spectroscopy–a tool for chemical properties and organic matter assessment of afforested mine soils. Ecological Engineering 62, 115–122.

Pimstein, A., Notesco, G., Ben-Dor, E., 2011. Performance of three identical spectrometers in retrieving soil reflectance under laboratory conditions. Soil Science Society of America Journal 75 (2), 746–759.

Pinty, B., Verstraete, M. M., Dickinson, R. E., 1989. A physical model for predicting bidirectional reflectances over bare soil. Remote Sensing of Environment 27, 273–288.

Pirie, A., Singh, B., Islam, K., 2005. Ultra-violet, visible, near-infrared, and mid-infrared diffuse reflectance spectroscopic techniques to predict several soil properties. Australian Journal Soil Research 43 (6), 713–721.

Post, J. L., Noble, P. N., 1993. The near ifnrared combination band frequencies of dioctahedral smectites, micas and illites. Clays and Clay Minerals 41 (6), 639–644.

Price, J., 1990. On the information content of soil reflectance spectra. Remote Sensing of Environment 33, 113–121.

Quinlan, J., 1992. Learning with continuous classes. In: Adams, A., Sterling, L. (Eds.), Proceedings AI'92, 5th Australian Conference on Artificial Intelligence. World Scientific, Singapore, pp. 343–348.

R Development Core Team, 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. ISBN 3-900051-07-0. URL `http://www.R-project.org`

Rabenarivo, M., Chapuis-Lardy, L., Brunet, D., Chotte, J. L., Rabeharisoa, L., Barthès, B. G., 2013. Comparing near and mid-infrared reflectance spectroscopy for determining properties of Malagasy soils, using global or LOCAL calibration. Journal of Near Infrared Spectroscopy 21 (6), 495–509.

Ramirez-Lopez, L., Behrens, T., Schmidt, K., Stevens, A., DemattÃ, J. A. M., Scholten, T., 2013. The spectrum-based learner: A new local approach for modeling soil vis–NIR spectra of complex datasets. Geoderma 195–196, 268–279.

Rawlins, B., Kemp, S., Milodowski, A., 2011. Relationships between particle size distribution and VNIR reflectance spectra are weaker for soils formed from bedrock compared to transported parent materials. Geoderma 166 (1), 84–91.

Reeves III, J. B., 2010. Near- versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: Where are we and what needs to be done? Geoderma 158 (1–2), 3–14.

Reeves III, J. B., Follett, R. F., McCarty, G. W., 2006. Can near or mid-infrared diffuse reflectance spectroscopy be used to determine soil carbon pools? Soil Science and Plant Analysis 37 (15–20), 2307–2325.

Reeves III, J. B., McCarty, G. W., 2001. Quantitative analysis of agricultural soils using near infrared reflectance spectroscopy and fibre-optic probe. Journal of Near Infrared Spectroscopy 9, 25–34.

Reeves III, J. B., McCarty, G. W., Mimmo, T., 2002. The potential of diffuse reflectance spectroscopy for the determination of carbon inventories in soils. Environmental Pollution 116, 227–284.

Reeves III, J. B., Smith, D. B., 2009. 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. Applied Geochemistry 24 (8), 1472–1481.

Richter, N., Jarmer, T., Chabrillat, S., Oyonarte, C., Hostert, P., Kaufmann, H., 2009. Free iron oxide determination in mediterranean soils using diffuse reflectance spectroscopy. Soil Science Society of America Journal 73 (1), 72–81.

Roudier, P., Palmer, A., Aislabie, J. M., McLeod, M., Hedley, C., Snape, I., June 2013. The potential of spectroscopy to map microbial habitats in the soils of Antarctica. In: Proceedings of NIR 2013, the 16th International Conference on Near Infrared Spectroscopy. La GrandeMotte, France, pp. 248–255.

Sanchez, P. A., Ahamed, S., Carré, F., Hartemink, A. E., Hempel, J., Huising, J., Lagacherie, P., McBratney, A. B., McKenzie, N. J., Mendonça-Santos, M. d. L., Minasny, B., Montanarella, L., Okoth, P., Palm, C. A., Sachs, J. D., Shepherd, K. D., Vågen, T.-G., Vanlauwe, B., Walsh, M. G., Winowiecki, L. A., Zhang, G.-L., 2009. Digital soil map of the world. Science 325 (5941), 680–681.

Sankey, J. B., Brown, D. J., Bernard, M. L., Lawrence, R. L., 2008. Comparing local vs. global visible and near-infrared (VisNIR) diffuse reflectance spectroscopy (DRS) calibrations for the prediction of soil clay, organic C and inorganic C. Geoderma 148 (2), 149–158.

Sarkhot, D. V., Grunwald, S., Ge, Y., Morgan, C. L. S., 2011. Comparison and detection of total and available soil carbon fractions using visible/near infrared diffuse reflectance spectroscopy. Geoderma 164 (1), 22–32.

Shepherd, K. D., Walsh, M. G., 2002. Development of reflectance spectral libraries for characterization of soil properties. Soil Science Society of America Journal 66 (3), 988–998.

Sherman, D. M., Waite, T. D., 1985. Electronic spectra of Fe^{3+} oxides and oxyhydroxides in the near infrared to ultraviolet. American Mineralogist 70, 1262–1269.

Shi, T., Chen, Y., Liu, H., Wang, J., Wu, G., 2014a. Soil organic carbon content estimation with laboratory-based visible-near-infrared reflectance spectroscopy: Feature selection. Applied spectroscopy 68 (8), 831–837.

Shi, Z., Wang, Q., Peng, J., Ji, W., Liu, H., Li, X., Viscarra Rossel, R. A., 2014b. Development of a national VNIR soil-spectral library for soil classification and prediction of organic matter concentrations. Science China Earth Sciences 57 (7), 1671–1680.

Shibusawa, S., Anom, S. W. I., Sato, S., Sasao, A., Hirako, S., 2001. Soil mapping using the real-time soil spectrophotometer. In: Grenier, G., Blackmore, S. (Eds.), ECPA 2001, Third European Conference on Precision Agriculture. Agro Montpellier, Montpellier, France, pp. 497–508.

Shonk, J., Gaultney, L., Schulze, D. G., Scoyoc, G. E. V., 1991. Spectroscopic sensing of soil organic matter content. Transactions of the ASAE 34 (5), 1978–1984.

Sörensen, L. K., Dalsgaard, S., 2005. Determination of clay and other soil properties by near infrared spectroscopy. Soil Science Society of America Journal 69 (1), 159–167.

Soriano-Disla, J. M., Janik, L. J., Viscarra Rossel, R. A., Macdonald, L. M., McLaughlin, M. J., 2014. The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. Applied Spectroscopy Reviews 49 (2), 139–186.

Stenberg, B., 2010. Effects of soil sample pretreatments and standardised rewetting as interacted with sand classes on vis–NIR predictions of clay and soil organic carbon. Geoderma 158, 15–22.

Stenberg, B., Nordkvist, E., Salomonsson, L., 1995. Use of near infrared reflectance spectra of soils for objective selection of samples. Soil Science 159 (2), 109–114.

Stenberg, B., Viscarra Rossel, R. A., Mouazen, A. M., Wetterlind, J., 2010. Visible and near infrared spectroscopy in soil science. Advances in Agronomy 107, 163–215.

Stevens, A., Nocita, M., Tóth, G., Montanarella, L., van Wesemael, B., 2013. Prediction of Soil Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy. PLoS ONE 8 (6), e66409.

Stevens, A., van Wesemael, B., Bartholomeus, H., Rosillon, D., Tychon, B., Ben-Dor, E., 2008. Laboratory, field and airborne spectroscopy for monitoring organic carbon content in agricultural soils. Geoderma 144 (1–2), 395–404.

Stoner, E. R., Baumgardner, M. F., 1981. Characteristic variations in reflectance of surface soils. Soil Science Society of America Journal 45, 1161–1165.

Sudduth, K., Hummel, J., 1993. Portable near infrared spectrophotometer for rapid soil analysis. Transactions of the ASAE 36 (1), 187–195.

Sudduth, K., Kitchen, N., Sadler, E., Drummond, S., Myers, D., 2010. VNIR spectroscopy estimates of within-field variability in soil properties. Springer Science+Business Media, New York, pp. 153–163.

Sudduth, K. A., Hummel, J. W., 1991. Evaluation of reflectance methods for soil organic matter sensing. Transactions of the ASAE 34 (4), 1900–1909.

Sudduth, K. A., Hummel, J. W., Funk, R. C., 1989. NIR soil organic matter sensor. American Society of Agricultural Engineers 89-1035, 23.

Summers, D., Lewis, M., Ostendorf, B., Chittleborough, D., 2011. Visible near-infrared reflectance spectroscopy as a predictive indicator of soil properties. Ecological Indicators 11 (1), 123–131.

Tekin, Y., Tumsavas, Z., Mouazen, A. M., 2012. Effect of moisture content on prediction of organic carbon and pH using visible and near-infrared spectroscopy. Soil Science Society of America Journal 76 (1), 188–198.

Terhoeven-Urselmans, T., Schmidt, H., Joergensen, R. G., Ludwig, B., 2008. Usefulness of near-infrared spectroscopy to determine biological and chemical soil properties: Importance of sample pre-treatment. Soil Biology and Biochemistry 40 (5), 1178–1188.

Udelhoven, T., Emmerling, C., Jarmer, T., 2003. Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial least-square regression: A feasibility study. Plant and Soil 251 (2), 319–329.

UN Sustainable Development Knowledge Platform, 2015. Open working group proposal for sustainable development goals.
URL https://sustainabledevelopment.un.org/content/documents/1579SDGs%20Proposal.pdf

van Groenigen, J. W., Mutters, C. S., Horwath, W. R., van Kessel, C., 2003. NIR and DRIFT-MIR spectrometry of soils for predicting soil and crop parameters in a flooded field. Plant and Soil 250 (1), 155–165.

Van Vuuren, J. A. J., Meyer, J. H., Claassens, A. S., 2006. Potential use of near infrared reflectance monitoring in precision agriculture. Communications in Soil Science and Plant Analysis 37 (15–20), 2171–2184.

Van Waes, C., Mestdagh, I., Lootens, P., Carlier, L., 2005. Possibilities of near infrared reflectance spectroscopy for the prediction of organic carbon concentrations in grassland soils. The Journal of Agricultural Science 143 (6), 487–492.

Vasques, G. M., Demattê, J. A. M., Viscarra Rossel, R. A., Ramírez-López, L., Terra, F. S., 2014. Soil classification using visible–near infrared diffuse reflectance spectra from multiple depths. Geoderma 223–225, 73–78.

Vasques, G. M., Grunwald, S., Harris, W. G., 2010. Spectroscopic models of soil organic carbon in Florida, USA. Journal of Environmental Quality 39 (3), 923–934.

Vendrame, P. R. S., Marchão, R. L., Brunet, D., Becquer, T., 2012. The potential of NIR spectroscopy to predict soil texture and mineralogy in Cerrado Latosols. European Journal of Soil Science 63 (5), 743–753.

Viscarra Rossel, R. A., 2007. Robust modelling of soil diffuse reflectance spectra by bagging-partial least squares regression. Journal of Near Infrared Spectroscopy 15 (1), 39–47.

Viscarra Rossel, R. A., 2009. The Soil Spectroscopy Group and the development of a global soil spectral library. NIR News 20 (4), 14–15.

Viscarra Rossel, R. A., 2011. Fine-resolution multiscale mapping of clay minerals in Australian soils measured with near infrared spectra. Journal of Geophysical Research: Earth Surface 116, F04023.

Viscarra Rossel, R. A., Behrens, T., 2010. Using data mining to model and interpret soil diffuse reflectance spectra. Geoderma 158 (1–2), 46–54.

Viscarra Rossel, R. A., Bui, E. N., de Caritat, P., McKenzie, N. J., 2010. Mapping iron oxides and the color of Australian soil using visible–near infrared reflectance spectra. Journal of Geophysical Research: Earth Surface 115.

Viscarra Rossel, R. A., Cattle, S., Ortega, A., Fouad, Y., 2009. In situ measurements of soil colour, mineral composition and clay content by vis–NIR spectroscopy. Geoderma 150 (3-4), 253–266.

Viscarra Rossel, R. A., Chen, C., Grundy, M. J., Searle, R., Clifford, D., Campbell, P. H., 2015. The Australian three-dimensional soil grid: Australia's contribution to the *GlobalSoilMap* project. Soil Research 53 (8), 845–864.

Viscarra Rossel, R. A., Hicks, W. S., 2015. Soil organic carbon and its fractions estimated by visible–near infrared transfer functions. European Journal of Soil Science 66 (3), 438–450.

Viscarra Rossel, R. A., Lark, R. M., 2009. Improved analysis and modelling of soil diffuse reflectance spectra using wavelets. European Journal of Soil Science 60 (3), 453–464.

Viscarra Rossel, R. A., McBratney, A. B., 1998. Laboratory evaluation of a proximal sensing technique for simultaneous measurement of soil clay and water content. Geoderma 85 (1), 19–39.

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

Viscarra Rossel, R. A., Webster, R., 2011. Discrimination of Australian soil horizons and classes from their visible–near infrared spectra. European Journal of Soil Science 62 (4), 637–647.

Viscarra Rossel, R. A., Webster, R., 2012. Predicting soil properties from the Australian soil visible–near infrared spectroscopic database. European Journal of Soil Science 63 (6), 848–860.

Viscarra Rossel, R. A., Webster, R., Bui, E. N., Baldock, J. A., 2014. Baseline map of organic carbon in Australian soil to support national carbon accounting and monitoring under climate change. Global Change Biology 20 (9), 2953–2970.

Vohland, M., Emmerling, C., 2011. Determination of total soil organic C and hot water-extractable C from VIS–NIR soil reflectance with partial least squares regression and spectral feature selection techniques. European Journal of Soil Science 62 (4), 598–606.

Vohland, M., Ludwig, M., Thiele-Bruhn, S., Ludwig, B., 2014. Determination of soil properties with visible to near-and mid-infrared spectroscopy: Effects of spectral variable selection. Geoderma 223, 88–96.

Vågen, T.-G., Shepherd, K. D., Walsh, M. G., 2006. Sensing landscape level change in soil fertility following deforestation and conversion in the highlands of Madagascar using vis–NIR spectroscopy. Geoderma 133 (3–4), 281–294.

Vågen, T.-G., Winowiecki, L. A., Tondoh, J. E., Desta, L. T., Gumbricht, T., 2016. Mapping of soil properties and land degradation risk in Africa using MODIS reflectance. Geoderma 263, 216–225.

Waiser, T. H., Morgan, C. L. S., Brown, D. J., Hallmark, C. T., 2007. In situ characterization of soil clay content with visible–near Infrared diffuse reflectance spectroscopy. Soil Science Society of America Journal 71 (2), 389–396.

Waruru, B. K., Shepherd, K. D., Ndegwa, G. M., Kamoni, P. T., Sila, A. M., 2014. Rapid estimation of soil engineering properties using diffuse reflectance near infrared spectroscopy. Biosystems Engineering 121, 177–185.

Wetterlind, J., Stenberg, B., 2010. Near-infrared spectroscopy for within-field soil characterization: small local calibrations compared with national libraries spiked with local samples. European Journal of Soil Science 61 (6), 823–843.

Wetterlind, J., Stenberg, B., Söderström, M., 2010. Increased sample point density in farm soil mapping by local calibration of visible and near infrared prediction models. Geoderma 153 (3–4), 152–160.

Wetterlind, J., Stenberg, B., Viscarra Rossel, R. A., 2013. Soil analysis using visible and near infrared spectroscopy. In: Maathuis, F. J. (Ed.), Plant Mineral Nutrients. Vol. 953 of Methods in Molecular Biology. Humana Press, pp. 95–107.

Williams, P. C., 1987. Variables affecting near infrared reflectance spectroscopic analysis. In: Williams, P., Norris, K. (Eds.), Near infrared technology in the agricultural and food industries. American Association of Cereal Chemists Inc., Saint Paul, MN, pp. 143–167.

Winowiecki, L., 2008. Soil biogeochemical patterns in the Talamanca foothills, Costa Rica: Local soil knowledge and implications for agroecosystems. Ph.D. thesis, University of Idaho, USA and CATIE, Costa Rica.

Xie, H. T., Yang, X. M., Drury, C. F., Yang, J. Y., Zhang, X. D., 2011. Predicting soil organic carbon and total nitrogen using mid and near infrared spectra for Brookston clay loam soil in Southwestern Ontario, Canada. Canadian Journal of Soil Science 91 (1), 53–63.

Xie, X., Pan, X., Sun, B., 2012. Visible and near infrared diffuse reflectance spectroscopy for prediction of soil properties near a copper smelter. Pedosphere 22 (3), 351–366.

Yang, X. M., Xie, H. T., Drury, C. F., Reynolds, W. D., Yang, J. Y., Zhang, X. D., 2012. Determination of organic carbon and nitrogen in particulate organic matter and particle size fractions of Brookston clay loam soil using infrared spectroscopy. European Journal of Soil Science 63 (2), 177–188.

Zornoza, R., Guerrero, C., Mataix-Solera, J., Scow, K., Arcenegui, V., Mataix-Beneyto, J., 2008. Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in mediterranean soils. Soil Biology and Biochemistry 40 (7), 1923–1930.

Figure captions

Figure 1. Soil vis–NIR spectra can be measured at points or by imaging, from different platforms; by proximal sensing in the field, in the laboratory using sampled material, or from remote sensing systems with multi- and hyperspectral capabilities. The graph shows typical spectra for soil noting absorptions to minerals and organic matter in the visible (vis) and near infrared, separating the regions where overtones (OT) and combination vibrations occur.

Figure 2 The soil vis–NIR spectroscopy timeline showing important developments, early publications and a small but important sample of the published research to date. The black disc in 2008 represents the conception of the global soil spectroscopy project..

Figure 3 Locations of the 12 509 unique sites with reflectance spectra that are in the global database.

Figure 4 Average reflectance spectra by continent (see Table 1 for abbreviations) and their standard deviations. Upper curves are the continuum removed (CR) spectra..

Figure 5 Principal component analysis (PCA) and fuzzy-c-means classification: (a) first three loadings of the PCA analysis, (b, i–iii) PCA scores coloured by the six 'crisp' fuzzy-c-means classes, (c, e, g, i, k, m) average continuum removed (CR) spectra (solid curves) and standard deviations (broken curves) for the six fuzzy-c-means classes (colours correspond to the classes shown in (b)), (d, f, h, j, l, n, i–iii) are the membership functions for each class, which show that the information content of the soil spectra vary continuously.

Figure 6 Associations between the six different spectral classes derived by classifying the scores from the principal component analysis of the global spectra with the fuzzy-c-means algorithm.

Figure 7 Ordination diagrams from the correspondence analysis (CA) between the six spectral classes and (a) soil type, (b) land cover, (c) continent and (d) country

(see Table 2 for country name abbreviations).

Figure 8 Independent test set validation of the soil organic C predictions from CUBIST showing the observed values against the predicted ones and their uncertainties coloured by continent, using (a) a pre-classfication of the spectra into the six spectral classes, and (b) without the pre-classication. The statistics shown are the concordance correlation coefficient (ρ_c) and the root mean square error (RMSE).

Figure 9 Independent test set validations of the best soil attribute predictions from CUBIST showing the observed values against the predicted ones and their uncertainties coloured by continent. The statistics shown are the concordance correlation coefficient (ρ_c) and the root mean square error (RMSE).

Figure 10 Scalograms showing the wavelet coefficient that were used by CUBIST to predict the soil properties, the scales at which they vary and their respective wavelengths. The abscissa on the bottom depict the particular wavelet coefficient used in the models and on the top their corresponding wavelengths and a sample $A = \log(1/\text{Reflectance})$ spectrum. The ordinate represents the wavelet scale. The third dimension, represented by colour intensity, indicates the amplitude (or degree of importance) of a particular coefficient at a particular scale..

Figure 11 Spatial distribution of the predicted soil attribute data harmonised with the spectroscopic method. The attributes shown are (a) soil organic C (SOC) in the log scale, (b) pH_w, (c) cation exchange capacity (CEC), (d) extractable Fe in the log scale, (e) soil inorganic C (SIC) in the log scale, (f) clay content, (g) sand content and (h) silt content. For (a), (d) and (e) we transformed the predictions to the log scale to help visualise their global variability.

Figure 12 Boxplots of the predicted harmonised soil properties by spectral class. Class 1 represents soil with smectitic mineralogy and with some carbonates, classes 2 and 3 represent weathered soil rich in kaolinite, hematite and sand, classes 4 and 6 represent soil with goethite and illite and soil in class 5 represents soil with organic material.

Figure 13 Review of literature showing quantile boxplots of the root mean square error (RMSE) of predictions made with visiblenear infrared spectroscopic models, grouped by scale. RMSE$_v$ are from the independent test set validations and RMSE$_c$ are from the cross validations. The local scale comprises studies on single or several fields, or small areas with similar soil types, the regional scale comprises studies over larger geographical areas than local, or including several soil types, the country scale comprises studies over entire countries or from many regions across a country, or many soil types, and the global and continental scale comprises studies across several countries and across diverse soil types. The black diamonds represent the RMSEs obtained in the modelling of the global data. The data used to derive the boxplots are given in Appendix C.

Table 1: Contributors to the global spectral database, their affiliations and citations for the spectra that are included in the database

.

Contributor	Continent, Country	Abbreviation	Reference
	Africa	**AF**	
K. Shepherd, A. Sila	Kenya	KE	ICRAF
H. Aïchi	Tunisia	TN	Aïchi et al. (2013)
B.G Barthès, M. Bernoux	Madagascar	MG	IRD
M. Bernoux, D. Brunet	Senegal	SN	IRD
A. Bayer	South Africa	ZA	Bayer et al. (2012)
M. Nocita			Nocita et al. (2011)
A. Demattê	Angola	AO	São Paulo University
	Antarctica	**AN**	
C. Hedley, P. Roudier	Ross Dependency	RD	Roudier et al. (2013)
	Asia	**AS**	
E. Ben Dor	Israel	IL	Ben-Dor and Banin (1994)
Z. Shi,	China	CN	Shi et al. (2014b)
D. Changwen	China	CN	Ma et al. (2012)
H. Abbaslou	Iran	IR	Uni. Shiraz
R. Viscarra Rossel	Brunei	BN	CSIRO
A. Ringrose-Voase	Philippines	PH	CSIRO
S. Y. Hong and E. Choi	South Korea	KR	Chung et al. (2012)
S. Shibusawa, M. Kodaira	Japan	JP	TUAT
	Europe	**EU**	
B. Stenberg, J. Eriksson	Sweden	SE	Stenberg (2010)
M. Knadel, A. Thomsen,	Denmark	DK	Knadel et al. (2012)
H. Bartholomeus	Netherlands	NL	WUR
H. Bartholomeus	Russia	RU	Bartholomeus et al. (2012)
A. Stevens, V. Genot	Belgium	BE	Genot et al. (2011)
Y. Fouad, C. Walter,	France	FR	Aïchi et al. (2009)
C. Gomez	France	FR	Ouerghemmi et al. (2011)
C. Guerrero, V Barrón	Spain	ES	Uni. M. H. de Elche
T. Behrens	Germany	DE	Uni. of Tüebingen
K. Böttcher,	Italy	IT	Böttcher et al. (2008)
T. Kemper, S. Sommer			
M. Sellito			
B. Rawlins, A. Chappell	United Kingdom	UK	Rawlins et al. (2011)
A. Gubler	Switzerland	CH	Gubler (2011)
L. Brodsky	Czech Republic	CZ	Brodský et al. (2011)
	North America	**NA**	
D. Brown	United States (+ other)	US	Brown et al. (2006)
K. Sudduth, N.R. Kitchen,	United States	US	Lee et al. (2010)
S.T. Drummond, S. Grunwald			
P. Sanborn,	Canada	CA	Uni. Northern British Columbia
V. Adamchuk			Uni. McGill
B.G Barthès, M. Bernoux	Martinique	MQ	IRD
L. Winowiecki	Costa Rica	CR	Winowiecki (2008)
	Oceania	**OC**	
R. Viscarra Rossel	Australia	AU	Viscarra Rossel and Webster (2012)
C. Hedley, B. Kusumo	New Zealand	NZ	Kusumo et al. (2008)
	South America	**SA**	
A. Demattê	Brazil	BR	Bellinaso et al. (2010)
L. Ramirez Lopez	Colombia	CO	CORPOICA
C. Castilla			
H. J.M. Morrás	Argentina	AR	CIRN-INTA
E. Rufasto Campos	Perú	PE	UNPRG
	Other		
ISRIC World Soil Information	Other countries, see Table 2		ICRAF (2015)

Table 2: Number of spectra by continent and country

.

Continent/Country	N	Continent/Country	N	Continent/Country	N	Continent/Country	N
Africa (AF)	**1621**	**Antarctica (AN)**	**144**	**Europe (EU)**	**3518**	**North America (NA)**	**5198**
Angola (AO)	109	Ross Dependency (AQ)	144	Albania (AL)	29	Belize (BZ)	6
Benin (BJ)	26			Belgium (BE)	262	Canada (CA)	144
Botswana (BW)	15	**Asia (AS)**	**3097**	Bulgaria (BU)	24	Costa Rica (CR)	104
Burkina Faso (BF)	5	Brunei (BN)	147	Czechoslovakia (CZ)	42	Cuba (CU)	124
Cameroon (CM)	8	China (CN)	1810	Denmark (DK)	210	Jamaica (JM)	29
Congo (CG)	14	India (IN)	67	Estonia (EE)	6	Martinique (MQ)	67
Congo, the Democratic Republic of (CD)	11	Indonesia (ID)	248	Finland (FI)	36	Mexico (MX)	22
Cote d'Ivoire (CI)	41	Iran (IR)	142	France (FR)	257	Nicaragua (NI)	77
Egypt (EG)	3	Israel (IL)	220	Germany (DE)	235	United States (US)	4625
Gabon (GA)	28	Japan (JP)	25	Greece (GR)	29		
Ghana (GH)	11	Korea (KR)	95	Hungary (HU)	134		
Kenya (KE)	365	Malaysia (MY)	98	Iceland (IS)	7	**Oceania (OC)**	**8646**
Madagascar (MG)	18	Mongolia (MN)	5	Ireland (IE)	39	Australia (AU)	8274
Malawi (MW)	17	Nepal (NP)	5	Italy (IT)	209	New Zealand (NZ)	346
Mali (ML)	48	Oman (OM)	11	Latvia (LV)	4	Samoa (WS)	26
Morocco (MA)	9	Pakistan (PK)	50	Lithuania (LT)	6		
Mozambique (MZ)	43	Philippines (PH)	47	Netherlands (NL)	210		
Namibia (NA)	51	Russia (RU)	20	Norway (NO)	16	**South America (SA)**	**1407**
Niger (NE)	31	Sri Lanka (LK)	29	Poland (PL)	60	Argentina (AR)	77
Nigeria (NG)	202	Taiwan (TW)	28	Romania (RO)	27	Brazil (BR)	722
Rwanda (RW)	6	Thailand (TH)	50	Slovakia (SK)	39	Colombia (CO)	283
Senegal (SN)	72			Spain (ES)	606	Ecuador (EC)	107
Somalia (SO)	5			Sweden (SE)	423	Peru (PE)	168
South Africa (ZA)	193			Switzerland (CH)	160	Uruguay (UY)	47
Togo (TG)	20			Turkey (TR)	56	Venezuela (VE)	3
Tunisia (TN)	89			United Kingdom (GB)	392		
Uganda (UG)	11						
Zambia (ZM)	79						
Zimbabwe (ZW)	91					**Total**	**23 631**

Table 3: Number of samples in the global soil spectroscopic database by World Reference Base (WRB) major soil groups and land cover type

.

WRB Soil type	Count	% Total	Land use type	Count	% Total
Acrisols	1690	7	Bare	353	1
Albeluvisols	86	< 1	Cropland	4743	20
Andosols	657	3	Forest	4199	18
Arenosols	1508	6	Grassland-Shrubland	6709	28
Cambisols	2306	10	Mixed farming	2616	11
Chernozems	218	1	Native vegetation	549	2
Ferralsols	915	4	Other	224	1
Fluvisols	703	3	Paddy	60	< 1
Gleysols	1422	6	Not recorded	4178	18
Gypsisols	940	4			
Histosols	121	< 1			
Kastanozems	880	4			
Leptosols	671	3			
Luvisols	3665	16			
Nitosols	488	2			
Phaeozems	1102	5			
Planosols	1290	5			
Podzols	1014	4			
Regosols	368	2			
Solonchaks	175	< 1			
Solonetz	686	3			
Vertisols	1981	8			
Not recorded	745	3			

Table 4: Metadata of analytical methods. N is the number of data, M is the total number of data with a record of the analytical method and m is the number of records with the specific method used

Soil attribute	N	M	Method	m
Organic C /%	17 931	9757	Walkley-Black	7509
			Oxidation with H_2O_2	978
			Loss on ignition	671
			CHN Pyrolysis	269
			Tyurin method	134
			Springer-Klee	110
			Dry combustion	86
Inorganic C /%	2690	1388	HCl treatment and manometer	1363
			Volumetric calcimeter	25
pH	20 515	20 515	1:5 Water	14 820
			1:5 0.01M Calcium chloride	5695
CEC /cmol(+)kg^{-1}	9588	5014	Ammonium acetate pH 7	4262
			Ammonium chloride pH 7	584
			Silver thiourea	130
			Compulsive exchange	31
			Ammonium chloride pH 8.5	7
Fe /%	4151	3311	Citrate-Dithionite	3239
			DTPA	67
			Oxalate	5
Clay /%	17 463	10 064	Pipette	5389
Sand /%	12 058	3395	Hydrometer	3395
Silt /%	9542	1280	Laser granulometer	572
			Plummet balance	358
			Bouyoucos	298
			Spectroscopic	52

Table 5: Statistical summary of the soil data. N is the number of units for which measurements exist.

Continent	N	Mean	St. dev.	Min.	Med.	Max.	Skew.	Coeff. var.
Organic C /%								
Africa	1606	1.193	1.714	0.01	0.62	24.91	4.648	143.7
Antarctica	144	0.09	0.107	0.01	0.05	0.59	2.359	118.8
Asia	2516	2.11	3.996	0.02	1.18	55.7	6.938	189.4
Europe	3187	2.633	3.552	0.02	1.8	50.6	6.494	134.9
North America	4821	1.595	4.198	0.01	0.54	55.27	8.084	263.2
Oceania	4315	2.069	3.163	0.01	1.17	46	6.702	152.8
South America	1339	4.871	5.97	0.01	1.96	28.16	1.553	122.6
All	17928	2.163	3.917	0.01	1.00	55.7	6.133	181.2
Inorganic C /%								
Africa	132	3.341	6.135	0.024	1.4	30.7	3.506	183.6
Asia	594	7.962	8.338	0.024	7.15	69.1	2.984	104.7
Europe	354	10.794	13.769	0.012	3.3	69.8	1.617	127.6
North America	1286	2.555	5.218	0.012	1.338	75.7	6.634	204.2
Oceania	11	4.536	2.008	2.3	4.3	9.4	1.087	44.3
South America	156	3.739	4.144	0.024	2.2	20.4	1.777	110.8
All	2533	5.097	8.341	0.012	2.1	75.7	3.478	188.5
Fe /%								
Africa	453	3.831	4.908	0.04	1.28	20.2	1.524	128.1
Asia	331	1.347	1.386	0.02	1.00	8.7	2.612	102.9
Europe	288	0.711	0.952	0.01	0.45	11.5	6.239	134.0
North America	2543	1.67	1.776	0.1	1.2	15.8	3.588	106.3
Oceania	452	0.757	1.151	0.006	0.33	13.7	4.366	151.9
South America	32	0.98	0.717	0.09	1.00	2.7	0.786	73.2
All	4099	1.709	2.38	0.006	1.00	20.2	3.77	139.9
CEC /cmol$_c$ Kg^{-1}								
Africa	918	11.384	13.047	0.2	6.9	84.1	2.693	114.6
Asia	1777	17.365	14.301	0.1	13.3	104.2	2.118	82.4
Europe	807	17.438	12.381	0.2	15	75.1	1.03	71.0
North America	4048	19.582	13.642	0.2	16.1	98.3	1.199	69.7
Oceania	1329	18.078	16.198	0.1	12.3	84	1.047	89.6
South America	619	12.527	13.226	0.2	8.3	77.6	1.875	105.6
All	9498	17.522	14.219	0.1	13.7	104.2	1.439	82.3
pH$_{water}$								
Africa	905	5.863	1.117	3.6	5.6	9.1	0.821	19.0
Antarctica	144	7.861	0.514	6.71	7.89	9.22	-0.143	6.5
Asia	1767	6.097	1.335	3.5	5.81	10	0.544	21.9
Europe	1028	6.575	1.397	3.4	6.6	10	-0.156	21.2
North America	4042	6.614	1.36	3.4	6.6	9.9	-0.087	20.6
Oceania	6055	6.818	1.251	3.5	6.6	10.03	0.263	18.40
South America	621	5.881	1.291	3.6	5.5	9.6	0.77	21.9
All	14562	6.568	1.338	3.4	6.4	10.03	0.15	20.62
Clay /%								
Africa	1266	31.964	19.781	0.2	29.7	90.8	0.51	61.9
Asia	1692	32.071	20.187	0.2	29.05	95.6	0.857	62.9
Europe	3165	21.92	16.042	0.2	17.8	96.8	1.324	73.2
North America	5040	26.029	18.056	0.1	23.9	90	0.78	69.4
Oceania	4896	28.011	19.766	0.3	23.4	85	0.495	70.6
South America	1223	35.719	24.345	0.2	29	92.7	0.524	68.2
All	17282	27.55	19.441	0.1	23.4	96.8	0.793	71.2
Sand /%								
Africa	1005	50.088	26.003	1.4	50.1	99.4	-0.109	51.9
Asia	1077	29.918	23.875	0.1	25.2	98.9	0.725	79.8
Europe	1762	36.062	26.762	0.2	29.67	99.2	0.618	74.2
North America	4088	35.71	26.351	0.1	31.35	99	0.513	73.8
Oceania	3095	57.995	25.539	1.07	61	99	-0.314	44.0
South America	609	42.498	27.169	1.3	39.4	97.2	0.272	63.9
All	11636	42.752	27.952	0.1	40	99.4	0.237	65.5
Silt /%								
Africa	896	16.096	14.334	0.2	11.8	84.3	1.721	89.0
Asia	974	32.704	19.324	0.4	28.2	88.3	0.434	59.1
Europe	1938	36	21.892	0.9	32.4	90.5	0.363	60.8
North America	403	34.794	16.941	0.4	32.9	81	0.313	48.7
Oceania	4636	13.68	10.003	0.4	12.0	80.3	1.367	73.1
South America	603	25.661	15.685	1.00	23	79.8	0.599	61.1
All	9450	22.112	18.172	0.2	17	90.5	1.211	82.6

Table 6: Principal component analysis of the continuum removed spectra

Principal component	Eigenvalue	% Total	cumulative %
1	0.33	55	55
2	0.10	16	71
3	0.09	15	86
4	0.02	4	90
5	0.02	3	93

Table 7: Fuzzy validity indices, the partition coefficient, C_{p}, and the partition entropy, E_{p}

Class	C_{p},	E_{p},
2	0.695	0.642
3	0.699	0.578
4	0.705	0.561
5	0.709	0.558
6	0.710	0.551
7	0.664	0.771
8	0.590	0.856
9	0.570	0.919
10	0.579	0.917

.

Table 8: Correspondence contingency table for soil type

Soil type	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
Acrisols	4	17	22	24	5	28
Andosols	4	10	15	33	18	19
Arenosols	4	30	20	21	13	13
Cambisols	3	9	14	32	17	26
Chernozems	35	0	3	15	23	24
Ferralsols	1	27	37	18	6	10
Fluvisols	6	5	11	35	5	38
Gleysols	7	2	4	20	13	54
Greyzems	0	0	0	0	0	100
Histosols	0	6	3	18	45	27
Kastanozems	33	13	13	10	12	18
Lithosols	8	18	17	26	7	24
Luvisols	15	10	19	27	8	21
Nitosols	6	32	27	15	8	12
Phaeozems	21	2	11	19	36	11
Planosols	15	6	15	23	14	27
Podzols	5	5	7	16	26	40
Podzoluvisols	0	2	56	19	2	21
Rankers	20	0	80	0	0	0
Regosols	24	21	24	12	5	14
Rendzinas	50	11	21	7	0	11
Solonchaks	12	33	21	12	8	15
Solonetz	21	13	13	13	16	24
Vertisols	57	6	9	9	9	10
Xerosols	13	24	34	18	4	6
Yermosols	23	30	21	15	1	10

Table 9: Correspondence contingency table for land cover
.

Land cover	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
Cropping	13	6	14	24	13	29
Forested	7	10	20	25	11	27
Mixed farming	18	10	13	18	13	27
Non-vegetated	30	24	30	9	0	7
Pastures/Grasses/Shrublands	26	17	14	19	9	15

Table 10: Correspondence contingency table for continent and country

.

Continent/Country	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Continent/Country	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
Africa (AF)	**7**	**39**	**32**	**16**	**5**	**2**	**Europe (EU)**	**4**	**3**	**9**	**31**	**13**	**41**
Angola (AO)	0	39	26	30	4	0	Albania (AL)	21	10	10	31	0	28
Benin (BJ)	25	50	0	25	0	0	Belgium (BE)	2	0	5	48	3	42
Botswana (BW)	0	38	46	0	15	0	Bulgaria (BU)	26	13	9	39	13	0
Burkina Faso (BF)	0	0	100	0	0	0	Czechoslovakia (CZ)	5	0	0	7	12	76
Cameroon (CM)	0	0	50	50	0	0	Denmark (DK)	1	0	1	16	30	51
Congo (CG)	0	14	71	0	14	0	Estonia (EE)	2	0	6	41	10	41
Cote d'Ivoire (CI)	0	67	33	0	0	0	Finland (FI)	0	0	33	17	0	50
Egypt (EG)	33	0	0	67	0	0	France (FR)	0	7	20	27	0	47
Gabon (GA)	0	17	50	33	0	0	Germany (DE)	0	12	23	36	0	28
Ghana (GH)	0	33	33	0	33	0	Greece (GR)	25	13	38	25	0	0
Kenya (KE)	4	62	22	0	12	0	Hungary (HU)	0	0	14	29	29	29
Madagascar (MG)	0	0	100	0	0	0	Iceland (IS)	0	2	15	27	6	51
Malawi (MW)	0	67	33	0	0	0	Ireland (IE)	0	0	25	50	25	0
Mali (ML)	33	33	33	0	0	0	Italy (IT)	0	42	32	16	5	5
Morocco (MA)	20	20	40	0	20	0	Latvia (LV)	0	0	25	50	0	25
Mozambique (MZ)	0	75	25	0	0	0	Lithuania (LT)	0	17	17	33	0	33
Namibia (NA)	10	60	30	0	0	0	Netherlands (NL)	0	2	0	31	24	44
Niger (NE)	0	100	0	0	0	0	Norway (NO)	0	0	0	33	67	0
Nigeria (NG)	0	60	20	15	5	0	Poland (PL)	0	0	0	67	22	11
Rwanda (RW)	0	100	0	0	0	0	Romania (RO)	0	0	0	100	0	0
Senegal (SN)	0	0	0	0	0	100	Slovakia (SK)	20	0	0	60	0	20
Somalia (SO)	40	20	20	20	0	0	Spain (ES)	17	12	35	27	8	2
South Africa (ZA)	0	15	43	38	3	1	Sweden (SE)	1	0	1	20	11	67
Togo (TG)	0	100	0	0	0	0	Switzerland (CH)	0	0	0	0	0	100
Tunisia (TN)	19	30	36	14	0	1	Turkey (TR)	31	15	38	15	0	0
Uganda (UG)	0	78	11	0	0	11	United Kingdom (GB)	1	0	5	37	19	38
Zambia (ZM)	0	50	40	0	10	0							
Zimbabwe (ZW)	4	44	16	20	4	12	**North America (NA)**	**23**	**9**	**18**	**21**	**9**	**20**
							Belize (BZ)	23	43	15	10	3	6
Antarctica (AN)	**25**	**0**	**0**	**44**	**0**	**31**	Canada (CA)	15	19	0	0	35	31
Ross Dependency (AQ)	25	0	0	44	0	31	Costa Rica (CR)	0	0	47	33	7	13
							Cuba (CU)	50	20	20	5	5	0
Asia (AS)	**6**	**6**	**8**	**28**	**8**	**44**	Jamaica (JM)	17	50	17	0	0	17
Brunei (BN)	0	0	2	34	36	27	Martinique (MQ)	100	0	0	0	0	0
China (CN)	4	3	4	29	9	51	Mexico (MX)	33	39	11	11	0	6
India (IN)	32	23	36	5	0	5	Nicaragua (NI)	0	44	33	0	22	0
Indonesia (ID)	9	30	30	17	9	6	United States (US)	23	9	18	21	9	21
Iran (IR)	0	0	0	100	0	0							
Israel (IL)	37	10	24	27	0	1	**Oceania (OC)**	**21**	**15**	**17**	**17**	**13**	**17**
Japan (JP)	0	0	50	50	0	0	Australia (AU)	22	16	18	16	12	15
Korea (KR)	0	0	100	0	0	0	New Zealand (NZ)	0	0	4	26	25	43
Malaysia (MY)	0	50	29	7	7	7	Samoa (WS)	17	17	33	33	0	0
Mongolia (MN)	0	0	0	40	0	60							
Nepal (NP)	0	0	67	0	0	33	**South America (SA)**	**8**	**21**	**35**	**17**	**13**	**7**
Oman (OM)	50	0	0	50	0	0	Argentina (AR)	14	0	0	50	29	7
Pakistan (PK)	5	5	11	26	0	53	Brazil (BR)	4	58	19	12	4	4
Philippines (PH)	11	21	53	16	0	0	Colombia (CO)	3	19	53	15	5	4
Russia (RU)	0	0	7	33	7	53	Ecuador (EC)	10	10	25	15	30	10
Sri Lanka (LK)	0	50	33	17	0	0	Peru (PE)	14	14	27	18	5	23
Taiwan (TW)	11	18	21	7	0	43	Uruguay (UY)	20	10	30	0	40	0
Thailand (TH)	13	38	19	19	0	13	Venezuela (VE)	50	0	0	0	50	0

Table 11: Correlations between the soil attributes and the principal component scores of the spectra
.

	Organic C	Inorganic C	Fe	CEC	pH$_W$	Clay	Sand	Silt	PC1 (55%)	PC2 (16%)	PC3 (15%)
Organic C	1.00								-0.26	-0.36	-0.01
Inorganic C	0.10	1.00							0.18	-0.03	0.11
Fe	0.04	-0.29	1.00						0.17	0.06	-0.05
CEC	0.37	0.00	0.34	1.00					-0.34	0.15	0.32
pH$_W$	-0.18	0.28	-0.09	0.27	1.00				-0.13	0.17	0.34
Clay	0.10	0.11	0.24	0.46	0.08	1.00			0.04	0.36	-0.18
Sand	-0.16	-0.20	-0.42	-0.58	-0.07	-0.71	1.00		-0.16	-0.13	-0.10
Silt	0.14	0.17	0.31	0.33	0.02	0.07	-0.68	1.00	0.27	-0.11	0.25

Table 12: Assessment statistics for model validation on the independent test data set. $\mathcal{W}$ are the number of wavelet coefficients used in the models, $\mathcal{C}$ refers to the modelling with (Y) and without (N) the pre-classification using the six spectral classes, $\mathcal{T}$ is the number of training data, $\mathcal{V}$ is the number of independent validation data, Mean$_\mathcal{V}$ is the mean of the validation data, SD$_\mathcal{V}$ is the standard deviation of the validation data. The statistics reported are the coefficient of determination (R^2), the concordance correlation coefficient (ρ_c), the root mean squared error (RMSE), mean error (ME), the standard deviation of the error (SDE) and the ratio of performance to deviation (RPD).

Soil attribute	$\mathcal{W}$	$\mathcal{C}$	$\mathcal{T}$	$\mathcal{V}$	Mean$_\mathcal{V}$	SD$_\mathcal{V}$	R^2	ρ_c	RMSE	ME	SDE	RPD
Organic C /%	125	Y	14 343	3585	2.07	3.66	0.89	0.92	1.11	0.18	1.10	3.3
Inorganic C /%	138	N	1794	739	5.02	8.24	0.77	0.87	3.96	0.025	3.96	2.1
extractable Fe /%	32	N	2868	1248	1.67	2.30	0.86	0.91	0.89	0.007	0.89	2.6
CEC /cmol(+)kg^{-1}	107	N	6672	2826	17.60	14.06	0.73	0.82	6.86	-0.22	6.85	2.0
pH$_{\text{Water}}$	79	Y	10 192	4389	6.58	1.33	0.62	0.76	0.82	0.01	0.82	1.6
Clay /%	71	Y	12 125	5177	27.53	19.03	0.71	0.80	10.26	-0.01	10.26	1.9
Sand /%	88	Y	8130	3525	43.51	28.28	0.57	0.68	18.83	-0.11	18.82	1.5
Silt /%	84	N	6647	2823	22.37	18.22	0.68	0.79	10.33	-0.13	10.33	1.8

Table 13: Statistical summary of the predictions—the harmonised global dataset. $N=$ 23631.

Continent	Mean	St. Dev.	Min.	Med.	Max.
Organic C /%					
Africa	1.24 (0.90, 1.72)	1.42 (1.03, 1.97)	0.05 (0.04, 0.07)	0.81 (0.60, 1.10)	24.82 (18.38, 33.52)
Antartica	0.1 (0.08, 0.13)	0.08 (0.06, 0.10)	0.02 (0.01, 0.03)	0.07 (0.05, 0.09)	0.35 (0.27, 0.46)
Asia	1.75 (1.30, 2.36)	2.47 (1.71, 3.63)	0.06 (0.04, 0.08)	1.10 (0.82, 1.48)	48.22 (30.65, 75.87)
Europe	2.59 (1.93, 3.49)	2.68 (1.90, 3.83)	0.11 (0.09, 0.15)	2.00 (1.49, 2.66)	40.62 (25.82, 63.91)
North America	1.46 (1.07, 2.00)	3.48 (2.34, 5.24)	0.03 (0.02, 0.04)	0.64 (0.48, 0.86)	74.62 (47.43, 117.40)
Oceania	1.71 (1.26, 2.32)	1.49 (1.07, 2.10)	0.09 (0.07, 0.11)	1.29 (0.96, 1.72)	34.23 (21.76, 53.85)
South America	5.6 (3.72, 8.47)	7.12 (4.60, 11.09)	0.06 (0.04, 0.07)	2.13 (1.55, 2.87)	38.73 (24.62, 60.93)
All	1.98 (1.44, 2.74)	3.08 (2.07, 4.64)	0.02 (0.01, 0.03)	1.18 (0.89, 1.58)	74.62 (47.43, 117.40)
Inorganic C /%					
Africa	1.83 (0.89, 4.06)	2.25 (1.29, 4.47)	0.04 (0.02, 0.07)	1.25 (0.57, 2.89)	25.47 (17.32, 41.24)
Antartica	4.61 (2.11, 10.41)	4.09 (1.95, 8.99)	0.37 (0.17, 0.83)	3.06 (1.38, 6.72)	23.21 (11.48, 48.18)
Asia	4.52 (2.67, 8.25)	6.67 (4.23, 11.59)	0.04 (0.02, 0.08)	1.79 (0.88, 3.81)	69.47 (52.80, 114.51)
Europe	5.07 (2.86, 9.58)	10.47 (6.47, 17.70)	0.05 (0.01, 0.13)	1.62 (0.76, 3.51)	83.23 (54.21, 141.57)
North America	1.41 (0.83, 2.62)	3.75 (2.46, 6.11)	0.03 (0.00, 0.07)	0.48 (0.22, 1.04)	72.90 (52.22, 108.92)
Oceania	1.93 (0.84, 4.78)	3.09 (1.62, 6.58)	0.09 (0.03, 0.23)	1.08 (0.44, 2.71)	68.31 (36.20, 132.34)
South America	1.42 (0.69, 3.19)	1.90 (1.10, 3.56)	0.07 (0.03, 0.14)	0.97 (0.41, 2.32)	26.04 (15.99, 43.50)
All	2.61 (1.39, 5.37)	5.70 (3.55, 9.92)	0.03 (0.00, 0.07)	1.14 (0.50, 2.64)	83.23 (54.21, 141.57)
Fe /%					
Africa	1.49 (1.03, 2.20)	2.43 (1.72, 3.60)	0.08 (0.04, 0.14)	0.69 (0.48, 1.05)	17.84 (13.36, 24.02)
Antartica	1.09 (0.68, 1.79)	0.19 (0.12, 0.40)	0.47 (0.28, 0.80)	1.09 (0.68, 1.74)	1.65 (1.04, 3.09)
Asia	1.01 (0.69, 1.53)	0.69 (0.49, 1.02)	0.08 (0.03, 0.19)	0.93 (0.62, 1.40)	11.99 (8.70, 16.69)
Europe	0.77 (0.51, 1.20)	0.52 (0.36, 0.83)	0.06 (0.02, 0.14)	0.65 (0.43, 1.01)	10.28 (6.64, 16.07)
North America	1.28 (0.95, 1.75)	1.15 (0.84, 1.68)	0.12 (0.07, 0.16)	1.00 (0.75, 1.32)	14.64 (10.22, 22.12)
Oceania	0.73 (0.47, 1.17)	0.49 (0.32, 0.78)	0.03 (0.01, 0.07)	0.62 (0.39, 1.00)	6.97 (4.24, 11.56)
South America	1.36 (0.85, 2.22)	1.08 (0.65, 1.86)	0.08 (0.04, 0.14)	1.10 (0.69, 1.74)	7.47 (4.88, 14.73)
All	0.99 (0.67, 1.49)	1.02 (0.72, 1.52)	0.03 (0.01, 0.07)	0.76 (0.50, 1.14)	17.84 (13.36, 24.02)
CEC /cmol(+)kg^{-1}					
Africa	14.95 (11.34, 19.18)	11.92 (9.71, 14.56)	1.86 (0.67, 2.96)	10.97 (7.88, 14.21)	62.11 (52.97, 79.33)
Antartica	30.89 (23.47, 39.50)	11.60 (9.79, 13.81)	12.62 (6.99, 17.21)	29.22 (22.48, 38.36)	59.66 (49.11, 71.33)
Asia	15.68 (12.33, 19.53)	11.19 (9.47, 13.23)	1.38 (0.42, 2.93)	11.71 (8.97, 14.91)	67.86 (54.80, 84.33)
Europe	19.02 (14.48, 24.35)	9.98 (8.26, 12.25)	2.19 (0.97, 3.56)	16.75 (12.43, 21.60)	69.51 (60.59, 79.81)
North America	20.38 (16.50, 24.79)	10.68 (9.24, 12.52)	1.41 (0.45, 2.59)	18.51 (14.96, 22.36)	75.81 (65.43, 87.07)
Oceania	21.28 (16.70, 26.56)	16.87 (14.39, 19.69)	1.14 (0.26, 2.32)	15.11 (11.22, 19.63)	77.22 (66.29, 90.01)
South America	15.39 (11.82, 19.56)	12.32 (10.58, 14.29)	0.80 (0.22, 1.47)	10.97 (7.98, 14.51)	65.56 (56.11, 75.86)
All	19.28 (15.13, 24.07)	13.66 (11.60, 16.09)	0.80 (0.22, 1.47)	15.31 (11.65, 19.44)	77.22 (66.29, 90.01)
pH$_{\mathrm{Water}}$					
Africa	6.12 (5.78, 6.46)	0.93 (0.87, 1.01)	4.52 (4.17, 4.80)	5.81 (5.51, 6.13)	9.40 (9.03, 9.82)
Antartica	7.75 (7.44, 8.06)	0.36 (0.38, 0.35)	6.31 (5.88, 6.74)	7.75 (7.45, 8.05)	8.72 (8.31, 9.25)
Asia	6.57 (6.24, 6.89)	1.08 (1.07, 1.10)	4.13 (3.85, 4.40)	6.48 (6.14, 6.81)	8.87 (8.46, 9.38)
Europe	6.59 (6.19, 6.99)	0.91 (0.91, 0.94)	4.05 (3.51, 4.33)	6.45 (6.06, 6.86)	9.41 (9.02, 9.84)
North America	6.59 (6.26, 6.92)	1.01 (1.03, 1.00)	3.96 (3.37, 4.29)	6.43 (6.09, 6.78)	9.59 (9.27, 9.91)
Oceania	6.75 (6.41, 7.09)	0.80 (0.79, 0.83)	4.54 (4.06, 4.92)	6.63 (6.30, 6.96)	9.03 (8.59, 9.67)
South America	6.10 (5.77, 6.43)	0.82 (0.81, 0.85)	3.90 (3.55, 4.25)	6.00 (5.67, 6.32)	8.92 (8.55, 9.45)
All	6.59 (6.25, 6.93)	0.94 (0.94, 0.96)	3.90 (3.37, 4.25)	6.48 (6.13, 6.82)	9.59 (9.27, 9.91)
Clay /%					
Africa	31.90 (26.71, 37.79)	14.69 (13.86, 15.83)	2.87 (1.25, 4.69)	29.93 (24.18, 36.41)	77.39 (71.91, 87.39)
Antartica	16.59 (11.07, 23.38)	10.92 (8.35, 13.93)	2.70 (0.86, 5.25)	13.88 (8.73, 20.07)	46.32 (33.96, 60.71)
Asia	26.16 (22.00, 30.86)	13.38 (12.03, 15.10)	2.05 (0.95, 3.61)	22.55 (19.06, 26.41)	83.62 (74.88, 103.87)
Europe	22.76 (19.03, 26.99)	12.35 (11.13, 13.88)	1.47 (0.23, 2.47)	20.27 (17.03, 23.99)	81.67 (69.35, 98.03)
North America	26.59 (22.81, 30.82)	14.42 (13.23, 15.81)	0.72 (0.08, 1.57)	24.94 (21.45, 28.73)	82.79 (76.22, 90.42)
Oceania	26.22 (21.17, 32.00)	11.63 (10.52, 13.16)	2.55 (0.87, 4.93)	23.30 (18.61, 28.80)	77.99 (66.40, 93.37)
South America	34.62 (29.64, 40.18)	19.01 (17.69, 20.51)	1.42 (0.49, 2.54)	29.85 (25.10, 35.49)	91.58 (80.72, 103.47)
All	26.62 (22.15, 31.69)	13.68 (12.49, 15.20)	0.72 (0.08, 1.57)	23.52 (19.33, 28.26)	91.58 (80.72, 103.47)
Sand /%					
Africa	48.66 (40.56, 56.76)	14.98 (14.90, 15.41)	8.12 (3.46, 12.78)	48.68 (40.66, 56.54)	86.20 (75.81, 99.77)
Antartica	44.46 (30.53, 58.38)	9.57 (7.92, 11.65)	21.09 (12.19, 29.99)	45.38 (31.72, 59.56)	64.50 (50.35, 80.30)
Asia	38.40 (29.77, 47.04)	13.72 (13.43, 14.37)	6.39 (1.15, 11.06)	36.69 (27.85, 45.58)	82.91 (73.54, 93.39)
Europe	40.36 (31.96, 48.76)	15.29 (14.53, 16.33)	5.15 (0.56, 9.64)	40.20 (31.58, 48.75)	92.14 (85.73, 98.55)
North America	35.77 (27.96, 43.58)	15.33 (14.56, 16.39)	2.88 (0.14, 6.91)	34.53 (26.79, 42.34)	91.04 (83.26, 103.63)
Oceania	53.52 (45.29, 61.75)	13.68 (14.26, 13.47)	6.71 (1.01, 12.40)	55.50 (47.14, 63.70)	86.89 (80.88, 100.60)
South America	44.92 (36.43, 53.40)	14.51 (14.48, 14.85)	4.61 (0.44, 9.67)	44.61 (35.81, 53.04)	92.21 (83.64, 100.79)
All	44.77 (36.51, 53.02)	16.22 (16.09, 16.68)	2.88 (0.14, 6.91)	44.96 (36.24, 53.47)	92.21 (85.73, 103.63)
Silt /%					
Africa	17.31 (13.05, 22.36)	7.75 (13.05, 22.36)	3.89 (1.09, 5.47)	15.82 (11.88, 20.22)	58.33 (43.70, 77.55)
Antartica	25.82 (16.36, 37.69)	5.81 (5.39, 6.22)	10.66 (5.03, 18.44)	26.50 (16.91, 38,19)	37.10 (29.94, 50.32)
Asia	35.62 (28.46, 43.79)	15.34 (13.47, 17.61)	3.45 (0.81, 7.49)	32.64 (25.61, 40.79)	71.49 (64.24, 80.93)
Europe	30.68 (24.52, 37.81)	13.46 (12.60, 14.70)	3.33 (1.11, 6.12)	28.88 (22.62, 36.01)	74.61 (67.58, 92.52)
North America	28.64 (22.26, 35.98)	8.09 (7.42, 8.95)	6.05 (3.10, 9.39)	27.92 (21.51, 35.46)	61.56 (54.18, 69.53)
Oceania	15.18 (11.66, 19.29)	6.74 (5.68, 8.14)	3.54 (0.80, 6.16)	13.85 (10.60, 17.60)	59.42 (51.47, 68.05)
South America	19.00 (14.73, 23.95)	9.61 (8.41, 11.13)	4.15 (1.98, 5.83)	17.08 (12.97, 21.65)	61.34 (52.43, 71.07)
All	23.58 (18.42, 29.55)	12.68 (10.96, 14.82)	3.33 (0.80, 5.47)	20.56 (15.59, 26.49)	74.61 (67.58, 92.52)

Appendix A

Requirements to contribute to the global spectroscopic database.

To include new spectra in the global database they need to be recorded in the range 350 to 2500 nm at intervals of one, two, five or 10 nm, from air or oven dry soil crushed or sieved to a size fraction of ≤ 2mm. The minimum dataset requested for each spectrum is:

1. Name(s) and affiliation(s) of contributor(s).
2. Country(ies) from which the spectra originate.
3. Coordinates in latitude and longitude using the World Geodetic System (WGS-84).
4. Organic carbon (and reference laboratory method used).
5. Clay, sand and silt contents (and reference laboratory method used).

Other data, which is also desirable includes:

6. Inorganic carbon (and reference laboratory method used).
7. Cation exchange capacity and exchangeable cations (and reference laboratory method used).
8. Extractable iron content (and reference laboratory method used).
9. pH measured in water and/or calcium chloride (and reference laboratory method used).
10. Soil classification (in the FAO-WRB system).
11. Land use classified as cropping, pasture, forest, natural vegetation, other.

We also request details on how the spectroscopic measurements were made:

i. The white reference used e.g. Spectralon®.

ii. The internal standard if one is used (*e.g.* (Ben-Dor et al., 2015))

iii. The frequency of re-calibration with the white reference.

iv. The number of averaged readings per saved spectrum for both reference and samples.

v. The number of replicates per sample.

vi. The instrument brand and type.

vii. The measurement configuration and setup, *e.g.* contact probe or if an external light source, the distance and angle to sample of both the light and the detector as well as the colour temperature of the light.

viii. If spectra were collected through a sample holder, what material?

Appendix B

Measurement protocol.

The measurement protocol described below is general and will enable the recording of consistent and good quality soil vis–NIR spectra using benchtop and portable spectrometers. The instrument should have a spectral resolution of 10 nm or less across the visible and near infrared range (between 400 and 2500 nm), and spectra should be recorded in this range at 1 nm intervals. The soil samples should be air or oven dry, crushed or sieved to a size fraction of $\leq$ 2mm. If the samples are sieved, it is important to avoid preferential sieving, that is, all of the sample should pass through the mesh.

Most instruments include the necessary accessories to perform the spectroscopic measurements. Depending on the instrument, they can be specific to suit the particular instrument or they can be more general. We suggest that the analyst follow the instructions of the instrument manufacturer and if necessary adapt these for measuring soil, following the guidelines below.

This protocol was tested and is easy to follow. It was developed to be practical and straight-forward because we do not need to overcomplicate the measurement of soil vis–NIR spectra. vis–NIR spectroscopy is an attractive soil analytical technique because of, amongst other, the robustness and simplicity of its measurements.

We note that since the conception of this project a measurement protocol was published by Wetterlind et al. (2013) and recommendations on the use of internal standards by Pimstein et al. (2011) and Ben-Dor et al. (2015). For completeness, the analyst might wish to also consults the mentioned publications before measuring.

Instrument setup.

a. Turn instrument on for a minimum of one hour before measurements.

b. If separate to a. above, turn the light source on for around 30 minutes before measurements.

c. Set the instrument control and data logging software to record (and average) 30 readings per soil sample measurement, and 50 readings per calibration with the white and dark reference measurement.

d. Set the instrument control and data logging software to record in wavelength intervals of 1 nm.

Instrument calibration.

e. Use a Halon white reference (Spectralon® is a commonly used commercial product) to optimise and calibrate the sensor.

f. The spectrometer should be calibrated every 10 minutes or around once every 10 measurements if you are sequentially measuring many samples in blocks of time.

g. If the measurement configuration uses an external calibration, then ensure that this configuration is the same as that used for the soil sample measurements.

h. Check that the spectrum of the white reference represents 100 % reflection at all wavelengths across the 400–2500 nm range with no more than around approximately 0.03 reflectance units (or 3 %) of noise. Often noise is be present towards the edges of the sensors response and towards the extremes near 400 nm and near 2500 nm. If the reference spectrum shows noise that exceeds this threshold, check the setup and repeat the calibration. If it persists, check with the instrument manufacturer.

i. To track spectral sensitivity, uniformity and wavelength accuracy over time, we recommend the use of a standard material (e.g. Ben-Dor et al., 2015) or alternatively, a uniform, non-specular material such as a sample of pure kaolinite, which should represent a smooth, clean spectrum across the 350–2500 nm range. Using kaolinite, sensitivity, uniformity and wavelength accuracy may be monitored over time by tracking the absorptions at 967 nm and the doublets at 1404 nm and 2200 nm. Measurements of the standard material should be made at the start of each day after the calibration with the Halon white reference and using the same measurement configuration as that used for the soil measurements.

Soil sample preparation and measurements.

j. Different instruments will have their own sample presentation setup and compatible sample containers. We recommend to follow the manufacturer instructions for the specific instrument. Generally however, we recommend that the soil sample container be at least 6 cm in diameter and 1cm deep. If measurements are made using a bare fibre optic, the diameter of the soil sample container will depend on the distance between the fibre and the sample.

k. Ensure that the sample is thoroughly mixed in the sample container and that the instrument measures a representative sample. If the measurement area is small, we recommend to take two to four replicate measurements.

l. Ensure that the soil sample surface is smooth and if needed, use a spatula to carefully flatten the surface.

m. Fill the sample containers in the same way for all samples. Avoid packing and compressing the samples.

n. If the same container is to be used for several samples it is important to clean it between samples. You can use water, however, ensure that the sample container is completely dry before filling it with a soil sample.

o. If the measurements are made through a window that comes in direct contact with the soil, ensure that the window is thoroughly cleaned between measurements. You can use water but ensure that the window is completely dry before measuring.

p. Measure the soil samples and record the (diffuse) reflectance. Do not record the spectra in Log1/R or first derivative.

q. Check that the spectrum of the soil sample does not exceed approximately 0.03 reflectance units (or 3 %) of noise. Often noise is be present towards the edges of the sensors response and towards the extremes near 400 nm and near 2500 nm. If the spectrum shows noise that exceeds this threshold, check the setup, repeat the calibration (see above) and the measurement. If it persists, check with the instrument manufacturer.

r. Check that the smallest reflectance value of the spectrum does not exceed a reflectance value of around 0.2. If it does, check the setup, repeat the calibration (see above) and the measurement. If it persists, check with the instrument manufacturer.

s. Check that the spectrum does not have discontinuities. Some instruments that use a two or three spectrometers to cover the vis–NIR range will produce spectra with discontinuities at the interface between the different sensors. Often these can be 'spliced' using the instrument's control and data logging software to produce a continuous spectrum. Check the instrument's manual if this applies.

Measurement configuration and setup.

Different instruments will describe their particular measurement configuration and set up. They can be specific to suit the particular instrument or they can be quite general. We suggest that the analyst follow the instructions from the instrument manufacturer and if necessary adapt these for measuring soil following the guidelines below.

t. Measuring using a 'bare' fibre optic and external illumination:

- Install the bare fibre and lamps on stable surfaces and ensure that their relative position (distances and angles) to each other and to the surface of the sample are constant.

- Ensure that the fibre's field of view, the fibre to sample distance and the sample surface area are compatible and that no shading occurs.

- The light source should have a colour temperature of approximately 3000 K and be sufficiently strong with relation to the distance from the fibre optic to soil sample surface to prevent excessive noise.

- The power supply for the lamps should be from a stable voltage. Note that it must be DC power to prevent any lamp-induced modulation of the spectra, which will occur if you use AC–powered lamps

- Eliminate all other interfering light sources during measurements, *e.g.* fluorescent lights and ambient light coming through windows.

- Calibrations and measurements should be performed as described above. It is important that you use the same geometry for the calibrations and for the soil measurements.

For example, a setup might be: The fibre optic placed on a stable stand 7 cm above the soil sample surface. Two lamps with halogen bulbs (12 volt, 50 w, 24 degree illumination angle) and integral parabolic (aluminium) reflectors. Place one on each side of the fibre optic approximately 60 cm from and at 45° to the sample. Ensure that the light spot of each lamp falls on the sample. Adjust the position of the lamps (change angle and distance) so that there is no shading on the measurements

u. Measuring using different accessories:

- As before, we recommend that the analyst follow the instructions from the instrument and accessory manufacturer and if necessary, adapt these using the guidelines given in this Appendix.

- Ensure that the sensor and light source are in a stable position and that this position is constant for the calibrations and all the soil measurements.

- If measurements are made through the accessory's (*e.g.* sapphire) window, ensure that there is full contact between this and the soil sample.

- If the measurements need to be made through the sample container, load them into Duran glass or optical glass Petri dishes to 1 cm depth. To calibrate, place the white reference face down in one dedicated dish of the same type and sample batch. Wipe the bottom of reference dish to clean off any dust between samples. Maintain consistency in the type of sample container used (e.g. manufacturer, specifications).

- Calibration and measurements should be performed as described above. It is important that the setup of the accessory is constant for the calibrations and for the soil measurements.

Appendix C

Review of the literature.

The literature review in Tables 14–20, includes only studies with spectra recorded in the laboratory and in the range between 350–2500 nm. We grouped the review into four scales:

- Local: comprises studies on single or several fields, or small areas with similar soil types,
- Regional: comprises studies over larger geographical areas than local, or including several soil types,
- Country: comprises studies over entire countries or from many regions across a country, or many soil types,
- Global or Continental: comprises studies over several or many countries.

The data reported in the tables (below) are N, the number of samples used for training (T) and validating (V) the spectroscopic models (the number of samples in T and V are separated by /) , the coefficient of determination R^2, the root mean squared error (RMSE) and the ratio of performance to deviation (RPD). Each statistic is reported for the for T and V separately. Missing values in the tables indicate that those statistics were not reported.

Table 14: Literature review of organic C content (%) predictions.

Scale	Country	N	R^2_T	$RMSE_T$(%)	RPD_T	R^2_V	$RMSE_V$(%)	RPD_V	Reference
Local	USA	180	0.94						Reeves III and McCarty (2001)
Local	USA	161/83	0.88	0.4	2.7				Chang et al. (2005)
Local	Spain	91CV	0.79						Hill and Schütt (2000)
Local	Netherland	70/35	0.69	1.1					Kooistra et al. (2003)
Local	Canada	143/144				0.78	0.3	2.2	Martin et al. (2002)
Local	USA	179CV	0.97	0.1					Reeves III et al. (2002)
Local	USA	64CV	0.93	0.1					Reeves III et al. (2002)
Local	USA	136CV	0.78	0.2					Reeves III et al. (2002)
Local	USA	136CV	0.84	0.1					Reeves III et al. (2002)
Local	Belgium	117CV		0.1	2.0				Stevens et al. (2008)
Local	Australia	228CV	0.57	0.4	1.8				Summers et al. (2011)
Local	Madagascar	101	0.94	0.6		0.92	0.8		Vågen et al. (2006)
Local	Australia	118CV	0.60	0.2					Viscarra Rossel et al. (2006)
Local	Sweden	25/58				0.70	0.1	1.9	Wetterlind and Stenberg (2010)
Local	Sweden	25/112				0.85	0.2	2.6	Wetterlind and Stenberg (2010)
Local	Sweden	24/65				0.71	0.5	1.5	Wetterlind and Stenberg (2010)
Local	Sweden	25/81				0.57	0.3	1.9	Wetterlind and Stenberg (2010)
Local	Sweden	25/72				0.89	0.2	3.0	Wetterlind and Stenberg (2010)

Local	USA	181/363	0.90	0.2		0.88	0.2		McCarty and Reeves III (2006)
Local	USA	107/118				0.82	2.9	2.4	Sankey et al. (2008)
Local	USA	52				0.86	0.7	2.6	Sankey et al. (2008)
Local	USA	54				0.31	1.1	1.1	Sankey et al. (2008)
Local	USA	1548/118				0.89	2.6	2.7	Sankey et al. (2008)
Local	USA	1548/52				0.96	0.4	4.9	Sankey et al. (2008)
Local	USA	1548/54				0.60	0.8	1.6	Sankey et al. (2008)
Local	USA	1548/118				0.80	3.3	2.1	Sankey et al. (2008)
Local	USA	1548/52				0.93	0.7	2.5	Sankey et al. (2008)
Local	USA	1548/54				0.60	0.8	1.6	Sankey et al. (2008)
Local	Canada	165	0.92	0.1	3.6	0.87	0.1	2.8	Yang et al. (2012)
Local	Canada	221	0.86	0.1	3.0	0.84	0.1	2.5	Yang et al. (2012)
Local	Canada	221	0.87	0.1	3.0	0.88	0.1	2.9	Yang et al. (2012)
Local	Canada	221	0.92	0.1	3.5	0.91	0.1	3.3	Yang et al. (2012)
Local	USA	181	0.92	0.7		0.91	0.7		Dick et al. (2013)
Local	USA	181/45	0.96	0.6		0.83	0.7		Dick et al. (2013)
Local	Germany	422	0.93		3.5				Heinze et al. (2013)
Local	Germany	142	0.41		1.2				Heinze et al. (2013)
Local	China	49	0.83	0.1	2.3				Lu et al. (2013)
Local	China	66/32	0.85	0.2		0.82	0.2	2.2	Shi et al. (2014a)
Local	China	62/31	0.92	0.2		0.84	0.2	2.4	Shi et al. (2014a)
Local	Canada	150	0.91	0.8		0.86			Nduwamungu et al. (2009)
Local	Belgium	117		0.7	2.0				Stevens et al. (2008)
Local	Australia	112	0.65	0.9	1.7				Forouzangohar et al. (2009)
Local	Poland	74		0.4		0.81	0.5	1.6	Chodak et al. (2007)
Local	Spain	205	0.41		1.3	0.34		1.3	Fontán et al. (2010)
Local	Spain	205	0.37		1.3	0.48		1.6	Fontán et al. (2010)
Local	USA	360/154	0.83	0.4		0.86	0.3	2.7	Sarkhot et al. (2011)
Local	USA	360/154	0.95	0.2		0.61	0.5	1.5	Sarkhot et al. (2011)
Local	USA	360/154	0.95	0.2		0.65	0.5	1.7	Sarkhot et al. (2011)
Local	Germany	109/40	0.88	0.2	2.9	0.89	0.3	2.7	Vohland and Emmerling (2011)
Local	Germany	109/40	0.86	0.2	2.6	0.89	0.3	2.8	Vohland and Emmerling (2011)
Local	Germany	109/40	0.93	0.2	3.5	0.89	0.3	2.8	Vohland and Emmerling (2011)
Regional	USA	76/32	0.96	0.6	4.7	0.89	0.6	4.2	Chang and Laird (2002)
Regional	Brazil	140/60				0.96	0.3		Fidêncio et al. (2002)
Regional	Brazil	140/60				0.88	0.4		Fidêncio et al. (2002)
Regional	USA	237	0.8	5.3		0.82	5.5		Reeves III (2010)
Regional	USA	237	0.8	5.5		0.80	5.8		Reeves III (2010)
Regional	Spain	393CV	0.98	0.6	5.8				Zornoza et al. (2008)
Regional	Australia	146	0.71	0.5	1.9				Gomez et al. (2008)
Regional	USA	177/60	0.90	0.6		0.82	0.6		McCarty et al. (2002)
Regional	USA	177/60	0.85	0.5		0.80	0.6		McCarty et al. (2002)
Regional	France	43/21	0.91	0.4	3.4	0.83	0.5	2.4	Aïchi et al. (2009)
Regional	Norway	75/48	0.95	0.7		0.80	0.7	2.2	Fystro (2002)
Regional	USA	376/164				0.73	0.5	1.7	Morgan et al. (2009)
Regional	Germany	30CV	0.85	0.1	2.6				Patzold et al. (2008)
Regional	Germany	30CV	0.93	0.1	3.8				Patzold et al. (2008)
Regional	USA	150/35	0.88	0.4		0.78	0.8		Reeves III et al. (2006)
Regional	Sweden	346/50	0.71	0.9		0.71	0.9		Stenberg (2010)
Regional	USA	30CV	0.89	0.2					Sudduth and Hummel (1993)
Regional	USA	4761/2359	0.94	1.3		0.79	2.5	2.1	Vasques et al. (2010)
Regional	USA	4676/2306	0.97	0.2		0.97	0.7	1.8	Vasques et al. (2010)
Regional	USA	85/50	0.89	5.3		0.35	10.2	1.2	Vasques et al. (2010)
Regional	USA	4639/2294	0.96	0.2		0.67	0.7	1.7	Vasques et al. (2010)
Regional	Germany	48CV	0.83	0.3	2.4				Vohland and Emmerling (2011)
Regional	Germany	21	0.89	0.2	3.1				Vohland and Emmerling (2011)
Regional	Germany	23	0.92	0.2	3.6				Vohland and Emmerling (2011)
Regional	Brazil	120CV	0.99	0.1					Madari et al. (2006)
Regional	Australia	270/90	0.62	0.3		0.66	0.3	1.7	Dunn et al. (2002)
Regional	Australia	121/40	0.61	0.4		0.76	0.4	1.7	Islam et al. (2003)
Regional	Australia	121/40				0.81	0.4		Islam et al. (2003)
Regional	Australia	121/40				0.68	0.5		Islam et al. (2003)
Regional	Australia	195				0.76	0.5	2.0	Pirie et al. (2005)
Regional	Brazil, Martinique	67/25	0.88	0.3	2.9	0.89	0.2		Brunet et al. (2008)
Regional	Brazil, Martinique	64/27	0.96	0.2	5.1	0.70	0.3		Brunet et al. (2008)
Regional	Senegal	44/20	0.85	0.1	2.4	0.85	0.1		Brunet et al. (2008)
Regional	Senegal	46/21	0.89	0.1	2.9	0.91	0.1		Brunet et al. (2008)
Regional	Poland	74CV	0.81	5.1	1.6				Chodak et al. (2007)
Regional	Australia	72/48	0.86	0.2		0.86	0.2		Dalal and Henry (1986)
Regional	Germany	102CV	0.97	0.2	4.7				Terhoeven-Urselmans et al. (2008)
Regional	Germany	110CV	0.98	0.2	3.8				Terhoeven-Urselmans et al. (2008)
Regional	Lithuania	127	0.93	0.05		0.91	0.1		Butkuté and Šlepetiené (2004)
Regional	USA	283		0.11		0.77	0.1	2.1	Brown et al. (2005)
Regional	USA	283		0.12		0.86	0.1	2.6	Brown et al. (2005)
Regional	Germany	60	0.74	0.33	2.0				Vohland et al. (2014)
Regional	Canada	145/49	0.95	0.3	3.7	0.88	0.4	2.8	Luce et al. (2014)
Regional	Mozambique	129	0.84	0.32	1.9				Cambule et al. (2012)
Regional	Poland	36	0.98	1.18					Pietrzykowski and Chodak (2014)
Regional	Poland	36	0.98	0.23					Pietrzykowski and Chodak (2014)
Regional	USA	150/206				0.97			Rabenarivo et al. (2013)
Regional	USA	150/206				0.99			Rabenarivo et al. (2013)
Regional	Ethiopia	64/64	0.97	0.2		0.91	0.3		Amare et al. (2013)
Regional	China	138/45				0.81	0.3	2.2	Ji et al. (2014)
Regional	China	82/42				0.93	0.5	3.4	Ji et al. (2015)
Regional	Belgium	1300				0.70	1.1	1.8	Genot et al. (2011)
Regional	Turke, UK	270	0.80			0.88	1.3	2.8	Tekin et al. (2012)
Regional	USA, Canada	720	0.53	1.74	1.5				Reeves III and Smith (2009)
Regional	USA, Canada	360/360	0.58	1.8		0.34	1.9	1.2	Reeves III and Smith (2009)
Regional	South Africa	76/37	0.81	0.4		0.93	0.3	3.7	Nocita et al. (2011)
Regional	South Africa	75/36	0.88			0.87	0.3	3.0	Nocita et al. (2011)
Regional	Poland	77/77	0.96	0.31		0.94	0.3	3.4	Chodak et al. (2002)
Regional	Italy	374	0.82	0.6	2.4	0.91	1.0	3.0	Leone et al. (2012)
Regional	Italy	186	0.84	0.6	2.5	0.88	0.9	2.5	Leone et al. (2012)
Regional	Italy	67	0.78	0.2	2.1	0.84	0.3	2.5	Leone et al. (2012)
Regional	Italy	121	0.78	0.8	2.1	0.93	0.2	2.4	Leone et al. (2012)
Regional	Belgium	1038/500	0.89	0.39		0.88	0.4	2.9	Van Waes et al. (2005)

Regional	Canada	217/78	0.97	0.2	6.1	0.95	0.2	4.0	Xie et al. (2011)
Regional	Canada	165	0.92	0.1	3.6	0.87	0.1	2.8	Yang et al. (2012)
Regional	Canada	221	0.92	0.1	3.5	0.91	0.1	3.3	Yang et al. (2012)
Regional	USA	697	0.87	0.27	2.7				Lee et al. (2009)
Regional	USA	165	0.80	0.30	2.3				Lee et al. (2009)
Country	Australia	1104				0.62	1.5		Viscarra Rossel and Behrens (2010)
Country	Australia	1104				0.89	0.8		Viscarra Rossel and Behrens (2010)
Country	Australia	1122				0.86	0.9		Viscarra Rossel and Lark (2009)
Country	Australia	1122				0.74	1.3		Viscarra Rossel and Lark (2009)
Global	World	3793	0.82	0.9					Brown et al. (2006)
Global	World	3793	0.87	0.8					Brown et al. (2006)
Global	Africa	674/337	0.91	0.2		0.80	0.3		Shepherd and Walsh (2002)
Global	World	2743/900		0.8		0.68	0.8		Ramirez-Lopez et al. (2013)
Global	World	20/20	0.87	3.2	3.4		4.1	2.7	Bartholomeus et al. (2008)
Global	World	20/20	0.76	5.2	1.9		4.8	2.1	Bartholomeus et al. (2008)
Global	EU	20,000/2828				0.79	0.4	2.2	Stevens et al. (2013)
Global	EU	20,000/1383				0.87	0.6	2.7	Stevens et al. (2013)
Global	EU	20,00/1564				0.89	1.0	2.9	Stevens et al. (2013)
Global	EU	20,000/6053				0.86	0.8	2.6	Stevens et al. (2013)
Global	EU	20,000/36				0.76	5.1	2.0	Stevens et al. (2013)

Table 15: Literature review of pH_w predictions.

Scale	Country	N	R^2_T	$RMSE_T$(%)	RPD_T	R^2_V	$RMSE_V$(%)	RPD_V	Reference
Local	Japan	25	0.54						Shibusawa et al. (2001)
Local	China	165	0.87	0.06		0.87	0.07		He et al. (2007)
Local	Turkey	359/153	0.35	0.11		0.27	0.13	1.2	Bilgili et al. (2010)
Local	Turkey	359/153	0.36	0.11		0.26	0.12	1.3	Bilgili et al. (2010)
Local	Turkey	153/359	0.5	0.11		0.21	0.13	1.0	Bilgili et al. (2010)
Local	Turkey	153/359	0.5	0.11		0.18	0.14	1.0	Bilgili et al. (2010)
Local	USA	181/363	0.73	0.24		0.53	0.31		McCarty and Reeves III (2006)
Local	Australia	118CV	0.57	0.17					Viscarra Rossel et al. (2006)
Local	Sweden	25/94				0.49	0.10	1.3	Wetterlind and Stenberg (2010)
Local	Sweden	25/112				0.33	0.19	1.1	Wetterlind and Stenberg (2010)
Local	Sweden	24/103				0.5	0.22	1.4	Wetterlind and Stenberg (2010)
Local	Sweden	25/81				0.48	0.31	1.4	Wetterlind and Stenberg (2010)
Local	Kenya	130/64	0.83	0.36		0.72	0.57		Awiti et al. (2008)
Local	Germany	422	0.89	0.25	2.6				Heinze et al. (2013)
Local	Germany	142	0.87	0.14	2.4				Heinze et al. (2013)
Local	China	49	0.63	0.21	1.6				Lu et al. (2013)
Local	Canada	150	0.6	0.2		0.89	0.18		Nduwamungu et al. (2009)
Local	Canada	151/38	0.94	0.1		0.91	0.13	3.2	Abdi et al. (2012)
Regional	USA	180/93	0.97			0.79	0.36	2.4	Cohen et al. (2007)
Regional	Spain	39/109	0.48	0.16			0.2	0.9	Moros et al. (2009)
Regional	USA	743				0.55	0.57	1.4	Chang et al. (2001)
Regional	Sweden	92/31				0.65	0.1	1.6	Wetterlind et al. (2010)
Regional	Sweden	94/31				0.85	0.15	2.8	Wetterlind et al. (2010)
Regional	Spain	393CV	0.72	0.14	1.9				Zornoza et al. (2008)
Regional	USA	1300/600	0.68	0.35		0.65	0.36	1.7	Cohen et al. (2007)
Regional	USA	1300/600	0.71	0.34		0.46	0.45	1.4	Cohen et al. (2007)
Regional	Australia	121/40	0.73	0.62		0.71	0.61	1.8	Islam et al. (2003)
Regional	Australia	121/40				0.63	0.68		Islam et al. (2003)
Regional	Australia	121/40				0.7	0.62		Islam et al. (2003)
Regional	Australia	173				0.65	0.73	1.7	Pirie et al. (2005)
Regional	China	67/33		0.28		0.79	0.21	1.8	Dong et al. (2011)
Regional	Brazil	86/44	0.27	0.4	1.2	0.25	0.6	1.1	Vendrame et al. (2012)
Regional	China	138/45				0.82	0.51	2.4	Ji et al. (2014)
Regional	Turkey, UK	270	0.59			0.65	0.70	1.7	Tekin et al. (2012)
Regional	USA	697	0.84	0.5	2.5				Lee et al. (2009)
Regional	USA	165	0.68	0.48	1.8				Lee et al. (2009)
Country	Australia	18501		0.61	2.3		0.63	2.3	Viscarra Rossel and Webster (2012)
Country	Australia	1104				0.81	0.53		Viscarra Rossel and Behrens (2010)
Country	Australia	1104				0.62	0.77		Viscarra Rossel and Behrens (2010)
Country	China	2955/225				0.69	0.64	2.6	Ji et al. (2015)
Global	Africa	758/378	0.83	0.34		0.70	0.43		Shepherd and Walsh (2002)

Table 16: Literature review of cation exchange capacity (CEC, $cmmol_c/kg$) predictions.

Scale	Country	N	R^2_T	$RMSE_T$(%)	RPD_T	R^2_V	$RMSE_V$(%)	RPD_V	Reference
Local	Turkey	359/153	0.77	1.5		0.79	1.42	2.3	Bilgili et al. (2010)
Local	Turkey	359/153	0.78	1.48		0.79	1.44	2.3	Bilgili et al. (2010)
Local	Turkey	153/359	0.79	1.46		0.68	1.88	1.7	Bilgili et al. (2010)
Local	Turkey	153/359	0.83	1.35		0.7	1.74	1.8	Bilgili et al. (2010)
Local	USA	179/82	0.86	4.85	1.7				Chang et al. (2005)
Local	Madagascar	67/34	0.8	2.55		0.68	3.12		Vågen et al. (2006)
Local	Australia	49	0.13	1.04					Viscarra Rossel et al. (2006)
Local	USA	50/50				0.83	1.36	2.4	van Groenigen et al. (2003)
Local	USA	299/74	0.73	1.4	2.0	0.87	1.22	2.3	Sudduth et al. (2010)
Local	China	49	0.47	1.24	1.4				Lu et al. (2013)
Local	Canada	150	0.93	1.4		0.89	1.8		Nduwamungu et al. (2009)
Local	USA	299/74	0.73	1.4	2.0	0.87	1.22	2.3	Sudduth et al. (2010)
Regional	Belgium	396/113				0.75	3.33		Fernández Pierna and Dardenne (2008)
Regional	Belgium	396/113				0.66	3.45		Minasny and McBratney (2008)
Regional	USA	802	0.81	3.82	2.3				Chang et al. (2001)
Regional	Australia	121/40	0.75	3.8		0.64	4.33	1.6	Islam et al. (2003)
Regional	Australia	121/40				0.68	3.92		Islam et al. (2003)
Regional	Australia	121/40				0.67	4.07		Islam et al. (2003)

Regional	Australia	193				0.52	5.83	1.4	Pirie et al. (2005)
Regional	Australia	422/139	0.88	2.19		0.9	1.88	3.3	Dunn et al. (2002)
Regional	Australia	237/79	0.71	3.27		0.8	2.74	2.3	Dunn et al. (2002)
Regional	Brazil	89/44	0.70	1.2	1.8	0.81	1	2.0	Vendrame et al. (2012)
Regional	Kenya	136/120	0.80	5.9	2.4	0.7	9.6	1.7	Waruru et al. (2014)
Regional	Poland	36	0.83	6.63	2.0				Pietrzykowski and Chodak (2014)
Regional	Belgium	1300				0.43	5.1	1.3	Genot et al. (2011)
Regional	Italy	374	0.69	5.82	1.8	0.70	6.26	1.9	Leone et al. (2012)
Regional	Italy	186	0.45	4.56	1.4	0.593	5.13	1.6	Leone et al. (2012)
Regional	Italy	67	0.78	5.2	2.1	0.74	5.69	1.9	Leone et al. (2012)
Regional	Italy	121	0.78	6.2	2.2	0.85	5.55	2.6	Leone et al. (2012)
Regional	USA	697	0.81	3.86	2.3				Lee et al. (2009)
Regional	USA	165	0.83	3.43	2.5				Lee et al. (2009)
Country	Australia	3706		6.28	2.3		7.08	2.1	Viscarra Rossel and Webster (2012)
Country	Israel	35/56	0.82	6.72		0.64	8.46		Ben-Dor and Banin (1994)
Global	world	4183	0.74	6.7					Brown et al. (2006)
Global	world	4183	0.83	5.5					Brown et al. (2006)
Global	Africa	740	0.95	2.6		0.88	3.8		Shepherd and Walsh (2002)

Table 17: Literature review of extractable Fe (%) predictions.

Scale	Country	N	R^2_T	$RMSE_T$(%)	RPD_T	R^2_V	$RMSE_V$(%)	RPD_V	Reference
Local	Spain	45CV	0.76	0.37					Richter et al. (2009)
Local	Australia	229CV	0.61	0.23	1.7				Summers et al. (2011)
Local	Germany	52CV	0.84	0.24					Udelhoven et al. (2003)
Local	China	254	0.81	0.27	2.3	0.75	0.3	2.0	Xie et al. (2012)
Local	China	254	0.79	0.13	2.2	0.83	0.13	2.3	Xie et al. (2012)
Local	China	254	0.74	0.28	1.9	0.68	0.33	1.5	Xie et al. (2012)
Local	Germany	195/211	0.96	0.08		0.94	0.08		Chodak et al. (2002)
Local	Canada	141/38	0.81			0.77		2.1	Abdi et al. (2012)
Local	Italy	119/118				0.71		2.0	Kemper and Sommer (2002)
Local	Italy	119/118				0.72		1.9	Kemper and Sommer (2002)
Regional	USA	784	0.64	0.006	1.7				Chang et al. (2001)
Regional	Australia	161	0.78	0.31		0.52	0.46	1.3	Islam et al. (2003)
Regional	Australia	161				0.48	0.49		Islam et al. (2003)
Regional	Australia	161				0.49	0.48		Islam et al. (2003)
Regional	Israel	91	0.57	1.15		0.51	1.25		Ben-Dor and Banin (1994)
Regional	USA	1300/600	0.53	1.84		0.38	2.04	1.3	Cohen et al. (2007)
Regional	USA	1300/600	0.34	2.19		0.26	2.66	1.3	Cohen et al. (2007)
Regional	Brazil	93/44	0.76	1.48	2.0	0.80	1.55	2.1	Vendrame et al. (2012)
Regional	USA & Canada	720	0.59	0.86	1.6				Reeves III and Smith (2009)
Regional	USA & Canada	360/360	0.59	0.99		0.38	0.87	1.3	Reeves III and Smith (2009)
Country	Australia	1448		1.79	1.9		0.26	1.8	Viscarra Rossel and Webster (2012)
Country	Uruguay	311	0.92	0.002			0.003		Cozzolino and Moron (2003)
Global	World	2909CV	0.73	0.96					Brown et al. (2006)
Global	World	2909CV	0.77	0.89					Brown et al. (2006)

Table 18: Literature review of clay content (%) predictions.

Scale	Country	N	R^2_T	$RMSE_T$(%)	RPD_T	R^2_V	$RMSE_V$(%)	RPD_V	Reference
Local	Turkey	359/153	0.82	3.83		0.87	4.05	2.6	Bilgili et al. (2010)
Local	Turkey	359/153	0.89	3.19		0.9	3.39	3.1	Bilgili et al. (2010)
Local	Turkey	153/359	0.88	3.54		0.83	4.03	2.3	Bilgili et al. (2010)
Local	Turkey	153/359	0.91	3.17		0.85	3.66	2.5	Bilgili et al. (2010)
Local	USA	529	0.78	1.54		0.69	1.8		McCarty and Reeves III (2006)
Local	Australia	237CV	0.66	3.13	2.0				Summers et al. (2011)
Local	Madagascar		0.93	3.31		0.72	6.1		Vågen et al. (2006)
Local	Australia	116CV	0.6	1.91					Viscarra Rossel et al. (2006)
Local	Sweden	25/61				0.61	3.5	1.3	Wetterlind and Stenberg (2010)
Local	Sweden	25/112				0.82	3.7	2.3	Wetterlind and Stenberg (2010)
Local	Sweden	24/65				0.5	3.6	1.2	Wetterlind and Stenberg (2010)
Local	Sweden	25/81				0.81	4.3	2.4	Wetterlind and Stenberg (2010)
Local	Kenya	130/64	0.87	0.35		0.77	0.404		Awiti et al. (2008)
Local	USA	42/13	0.49	3.59	1.4	0.15	2.68	1.9	Sudduth et al. (2010)
Local	USA	210/234				0.52	4.92	1.4	Sankey et al. (2008)
Local	USA	52				0.02	10.85	1.0	Sankey et al. (2008)
Local	USA	54				0.09	13.76	1.0	Sankey et al. (2008)
Local	USA	4184/234				0.38	5.63	1.2	Sankey et al. (2008)
Local	USA	4184/52				0.21	9.54	1.1	Sankey et al. (2008)
Local	USA	4184/54				0.49	10.25	1.4	Sankey et al. (2008)
Local	USA	4184/234				0.24	6.51	1.1	Sankey et al. (2008)
Local	USA	4184/52				0.19	9.62	1.1	Sankey et al. (2008)
Local	USA	4184/54				0.51	12.31	1.1	Sankey et al. (2008)
Local	Canada	150	0.98			0.97			Nduwamungu et al. (2009)
Local	South Africa	575				0.92			Van Vuuren et al. (2006)
Local	Neitherlands	69		2.39			2.65		Kooistra et al. (2001)
Regional	Sweden	92/31				0.75	3.6	2.3	Wetterlind et al. (2010)
Regional	Sweden	94/31				0.95	2.7	3.7	Wetterlind et al. (2010)
Regional	USA	743	0.67	4.06	1.7				Chang et al. (2001)
Regional	Brazil	120CV	0.94	3.24					Madari et al. (2006)
Regional	Denmark	784		0.2	2.9				Sörensen and Dalsgaard (2005)
Regional	Sweden	346/50	0.9	5.55		0.89	5.38		Stenberg (2010)
Regional	Australia	1287	0.77	8.3					Viscarra Rossel et al. (2009)
Regional	USA	188/82				0.84	6.2	2.3	Waiser et al. (2007)
Regional	Australia	121/40	0.82	7.8		0.72	8.9	1.9	Islam et al. (2003)
Regional	Australia	121/40				0.73	8.7		Islam et al. (2003)
Regional	Australia	121/40				0.75	8.7		Islam et al. (2003)

Regional	Australia	208				0.61	12.3	1.6	Pirie et al. (2005)
Regional	Brazil	93/42	0.83	6.48	2.4	0.74	6.89	2.0	Vendrame et al. (2012)
Regional	Kenya	136/120	0.5	11	1.4	0.5	16	1.1	Waruru et al. (2014)
Regional	Italy	70/30	0.87	6.6		0.87	5.8		Curcio et al. (2013)
Regional	Belgium	1300				0.41	6.74	1.3	Genot et al. (2011)
Regional	Italy	374	0.81	6.91	2.3	0.829	6.65	2.4	Leone et al. (2012)
Regional	Italy	186	0.80	5.92	2.2	0.81	5.45	2.3	Leone et al. (2012)
Regional	Italy	67	0.82	5.66	2.4	0.83	4.88	2.5	Leone et al. (2012)
Regional	Italy	121	0.77	8.79	2.1	0.88	6.05	3.0	Leone et al. (2012)
Regional	USA	697	0.8	4.69	2.2				Lee et al. (2009)
Regional	USA	165	0.76	3.74	2.1				Lee et al. (2009)
Country	Australia	1134				0.81	8.36		Viscarra Rossel and Lark (2009)
Country	Australia	1104	0.88	6.42					Viscarra Rossel et al. (2009)
Country	Australia	15205		8.54	2.4		8.49	2.4	Viscarra Rossel and Webster (2012)
Country	Australia	1104				0.75	9.44		Viscarra Rossel and Behrens (2010)
Country	Australia	1104				0.88	6.42		Viscarra Rossel and Behrens (2010)
Country	Israel	35/56	0.76	8.6		0.56	10.3		Ben-Dor and Banin (1994)
Country	Uruguay	321	0.9	3.6			3.8		Cozzolino and Moron (2003)
Global	World	4184	0.73	9.5					Brown et al. (2006)
Global	World	4184	0.91	5.4					Brown et al. (2006)
Global	World	3150/1050		7.97		0.77	12.01		Ramirez-Lopez et al. (2013)
Global	Africa	457/225	0.88	5.4		0.8			Shepherd and Walsh (2002)

Table 19: Literature review of sand content (%) predictions.

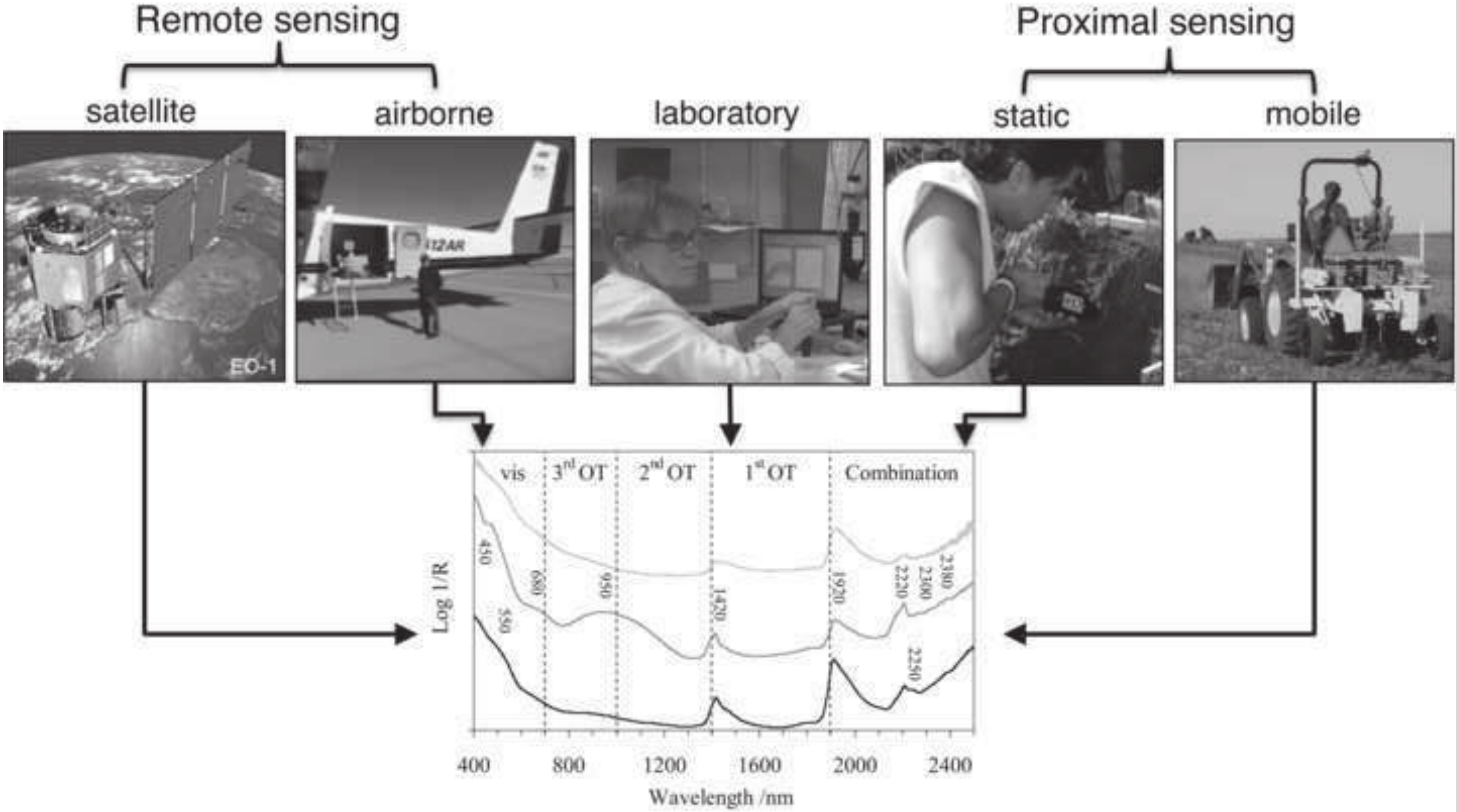
Scale	Country	N	R^2_T	RMSE$_T$(%)	RPD$_T$	R^2_V	RMSE$_V$(%)	RPD$_V$	Reference
Local	Turkey	359/153	0.81	4.33		0.84	4.45	2.5	Bilgili et al. (2010)
Local	Turkey	359/153	0.84	3.98		0.82	4.76	2.3	Bilgili et al. (2010)
Local	Turkey	153/359	0.84	4.48		0.7	5.67	1.8	Bilgili et al. (2010)
Local	Turkey	153/359	0.86	4.1		0.72	5.39	1.9	Bilgili et al. (2010)
Local	USA	187/87	0.57	12.2	0.9				Chang et al. (2005)
Local	USA	176/353	0.75	3.94		0.42	6.1		McCarty and Reeves III (2006)
Local	Australia	116CV	0.59	3.3					Viscarra Rossel et al. (2006)
Local	Sweden	25/61				0.3	2.6	0.8	Wetterlind and Stenberg (2010)
Local	Sweden	25/112				0.89	0.3	3.0	Wetterlind and Stenberg (2010)
Local	Sweden	24/65				0.53	0.5	1.5	Wetterlind and Stenberg (2010)
Local	Sweden	25/81				0.73	6.2	2.0	Wetterlind and Stenberg (2010)
Local	Kenya	130/64	0.83	0.62		0.75	0.57		Awiti et al. (2008)
Local	USA	42/13	0.04	2.53	1.0	0.76	1.91	1.4	Sudduth et al. (2010)
local	Canada	150	0.91	1.93		0.95	4.16		Nduwamungu et al. (2009)
Regional	Sweden	92/31				0.93	2.5	3.4	Wetterlind et al. (2010)
Regional	Sweden	94/31				0.91	3.8	3.3	Wetterlind et al. (2010)
Regional	USA	743	0.82	11.93	2.3				Chang et al. (2001)
Regional	Brazil	120CV	0.99	1.71					Madari et al. (2006)
Regional	Australia	199				0.28	90.9	0.3	Pirie et al. (2005)
Regional	Australia	121/40	0.72	12.2		0.53	14.5	1.5	Islam et al. (2003)
Regional	Brazil	92/42	0.67	7	1.7	0.56	6.25	1.5	Vendrame et al. (2012)
Regional	Brazil	94/44	0.66	6.23	1.7	0.72	5.47	1.9	Vendrame et al. (2012)
Regional	Italy	70/30	0.89	8		0.8	7.7		Curcio et al. (2013)
Regional	Italy	374	0.59	1.59	1.6	0.58	11.84	1.5	Leone et al. (2012)
Regional	Italy	186	0.49	9.2	1.4	0.52	8.66	1.5	Leone et al. (2012)
Regional	Italy	67	0.72	8.85	1.9	0.71	9.83	1.8	Leone et al. (2012)
Regional	Italy	121	0.71	12.26	1.9	0.81	10.02	2.1	Leone et al. (2012)
Regional	USA	697	0.78	10.89	2.1				Lee et al. (2009)
Regional	USA	165	0.79	7.74	2.2				Lee et al. (2009)
Country	Australia	11783		13.3	1.6		13.56	1.6	Viscarra Rossel and Webster (2012)
Country	Australia	12426		10.21	1.6		9.77	1.6	Viscarra Rossel and Webster (2012)
Country	Australia	11829		10.59	2.4		12	2.1	Viscarra Rossel and Webster (2012)
Country	Uruguay	319	0.8	6.8			7.2		Cozzolino and Moron (2003)
Global	Africa	682	0.91	6.1		0.76	10.8		Shepherd and Walsh (2002)

Table 20: Literature review of silt content (%) predictions.

Scale	Country	N	R^2_T	RMSE$_T$(%)	RPD$_T$	R^2_V	RMSE$_V$(%)	RPD$_V$	Reference
Local	Turkey	359/153	0.41	4.63		0.4	4.43	1.4	Bilgili et al. (2010)
Local	Turkey	359/153	0.51	4.22		0.35	4.72	1.3	Bilgili et al. (2010)
Local	Turkey	153/359	0.51	3.99		0.32	5.06	1.2	Bilgili et al. (2010)
Local	Turkey	153/359	0.55	3.87		0.29	5.32	1.1	Bilgili et al. (2010)
Local	USA	187/87	0.27	7.14	0.7				Chang et al. (2005)
Local	USA	176/353	0.67	3.25		0.22	5.1		McCarty and Reeves III (2006)
Local	Madagascar	101	0.84	2.5		0.4	6.45		Vågen et al. (2006)
Local	Australia	116CV	0.41	2.35					Viscarra Rossel et al. (2006)
Local	Sweden	25/61				0.62	3.2	1.4	Wetterlind and Stenberg (2010)
Local	Sweden	25/112				0.43	5	1.2	Wetterlind and Stenberg (2010)
Local	Sweden	24/65				0.3	4.2	0.9	Wetterlind and Stenberg (2010)
Local	Sweden	25/81				0.12	4.3	1	Wetterlind and Stenberg (2010)
Local	Kenya	130/64	0.83	0.52		0.77	0.61		Awiti et al. (2008)
Local	USA	42/13	0.68	3.12	1.8	0.63	1.79	3.1	Sudduth et al. (2010)
local	Canada	150	0.91	1.93		0.97	3.26		Nduwamungu et al. (2009)
Regional	Sweden	92/31				0.73	3.4	1.8	Wetterlind et al. (2010)
Regional	Sweden	94/31				0.63	2.8	1.5	Wetterlind et al. (2010)
Regional	Brazil	120CV	0.64	3.35					Madari et al. (2006)
Regional	USA	743	0.84	9.51	2.5				Chang et al. (2001)
Regional	Australia	121/40	0.34	7.1		0.05	9.8	0.9	Islam et al. (2003)
Regional	Australia	207				0.14	130.7	0.3	Pirie et al. (2005)
Regional	Brazil	93/44	0.71	2.38	1.9	0.46	3.97	1.3	Vendrame et al. (2012)
regional	Italy	70/30	0.82	5.4		0.6	7.2		Curcio et al. (2013)

regional	Italy	374	0.57	8.58	1.5	0.51	8.84	1.4	Leone et al. (2012)
regional	Italy	186	0.24	7.36	1.1	0.16	6.78	1.1	Leone et al. (2012)
regional	Italy	67	0.53	6.07	1.4	0.44	7.24	1.2	Leone et al. (2012)
regional	Italy	121	0.28	7.21	1.2	0.48	5.7	1.4	Leone et al. (2012)
regional	USA	697	0.72	8.94	1.9				Lee et al. (2009)
regional	USA	165	0.79	6.47	2.2				Lee et al. (2009)
Country	Australia	14831		5.57	1.6		5.5	1.6	Viscarra Rossel and Webster (2012)
Country	Uruguay	317	0.84	6			6.2		Cozzolino and Moron (2003)
Global	Africa	682	0.79	3.9		0.67	4.9		Shepherd and Walsh (2002)

Figure1


Remote sensing
Proximal sensing
satellite
airborne
laboratory
static
mobile
EO-1
vis
3rd OT
2nd OT
1st OT
Combination
450
550
680
950
1420
1920
2220
2300
2380
2250
Log 1/R
400
800
1200
1600
2000
2400
Wavelength /nm

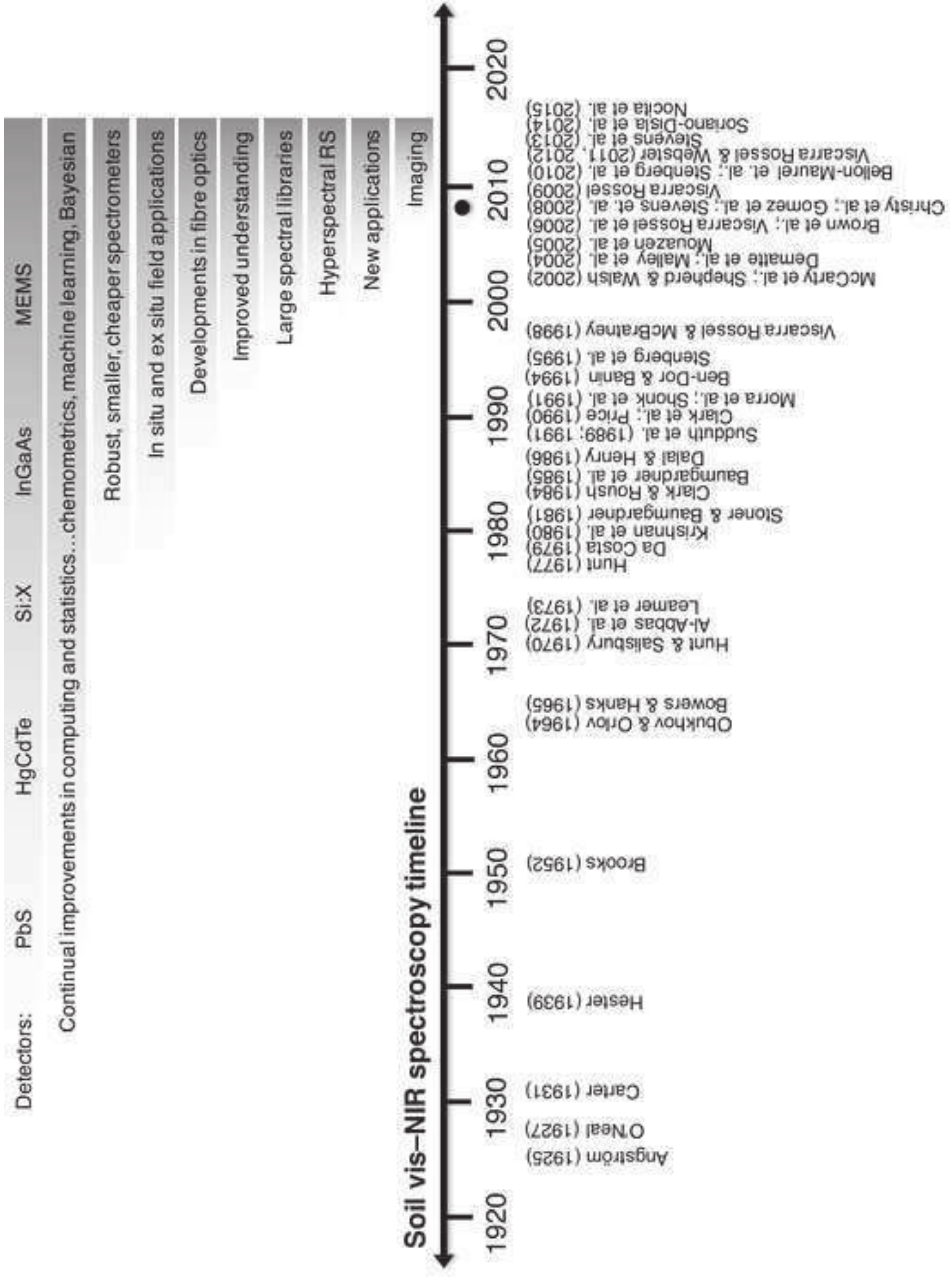

Soil vis–NIR spectroscopy timeline
1920
1930
1940
1950
1960
1970
1980
1990
2000
2010
2020
Detectors:
PbS
HgCdTe
Si:X
InGaAs
MEMS
Continual improvements in computing and statistics…chemometrics, machine learning, Bayesian
Robust, smaller, cheaper spectrometers
In situ and ex situ field applications
Developments in fibre optics
Improved understanding
Large spectral libraries
Hyperspectral RS
New applications
Imaging
Angstrom (1925)
O'Neal (1927)
Carter (1931)
Hester (1939)
Brooks (1952)
Obukhov & Orlov (1964)
Bowers & Hanks (1965)
Hunt & Salisbury (1970)
Al-Abbas et al. (1972)
Leamer et al. (1973)
Hunt (1977)
Da Costa (1979)
Krishnan et al. (1980)
Stoner & Baumgardner (1981)
Clark & Roush (1984)
Baumgardner et al. (1985)
Dalal & Henry (1986)
Sudduth et al. (1989; 1991)
Clark et al.; Price (1990)
Morra et al.; Shonk et al. (1991)
Ben-Dor & Banin (1994)
Stenberg et al. (1995)
Viscarra Rossel & McBratney (1998)
McCarty et al.; Shepherd & Walsh (2002)
Dematte et al.; Malley et al. (2004)
Mouazen et al. (2005)
Brown et al.; Viscarra Rossel et al. (2006)
Christy et al.; Gomez et al.; Stevens et. al. (2008)
Viscarra Rossel (2009)
Bellon-Maurel et. al.; Stenberg et al. (2010)
Viscarra Rossel & Webster (2011, 2012)
Stevens et al. (2013)
Soriano-Disla et al. (2014)
Nocita et al. (2015)

Figure2

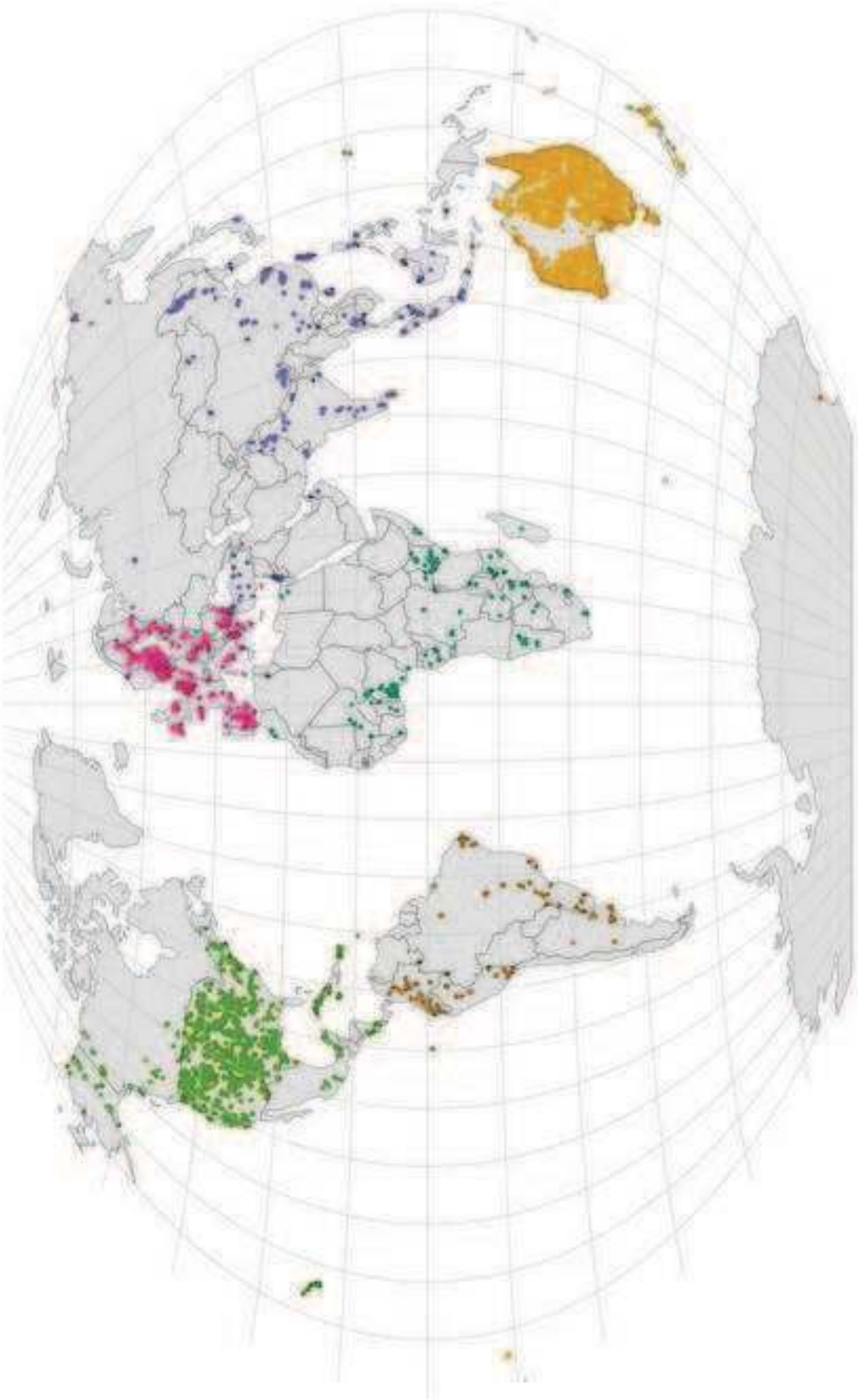


Figure3

Figure4

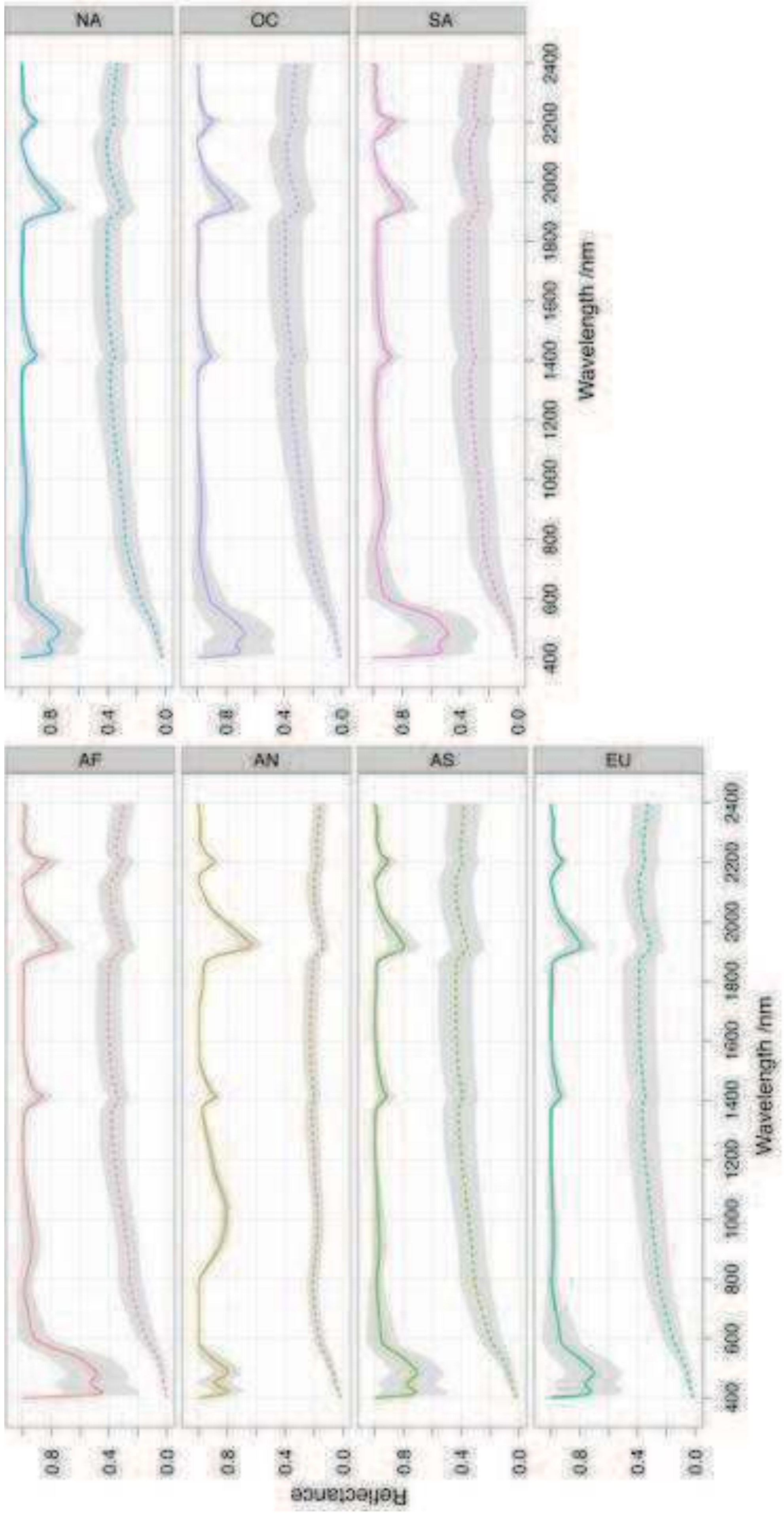

NA
OC
SA
AF
AN
AS
EU
Reflectance
0.0
0.4
0.8
Wavelength /nm
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400

Figure5

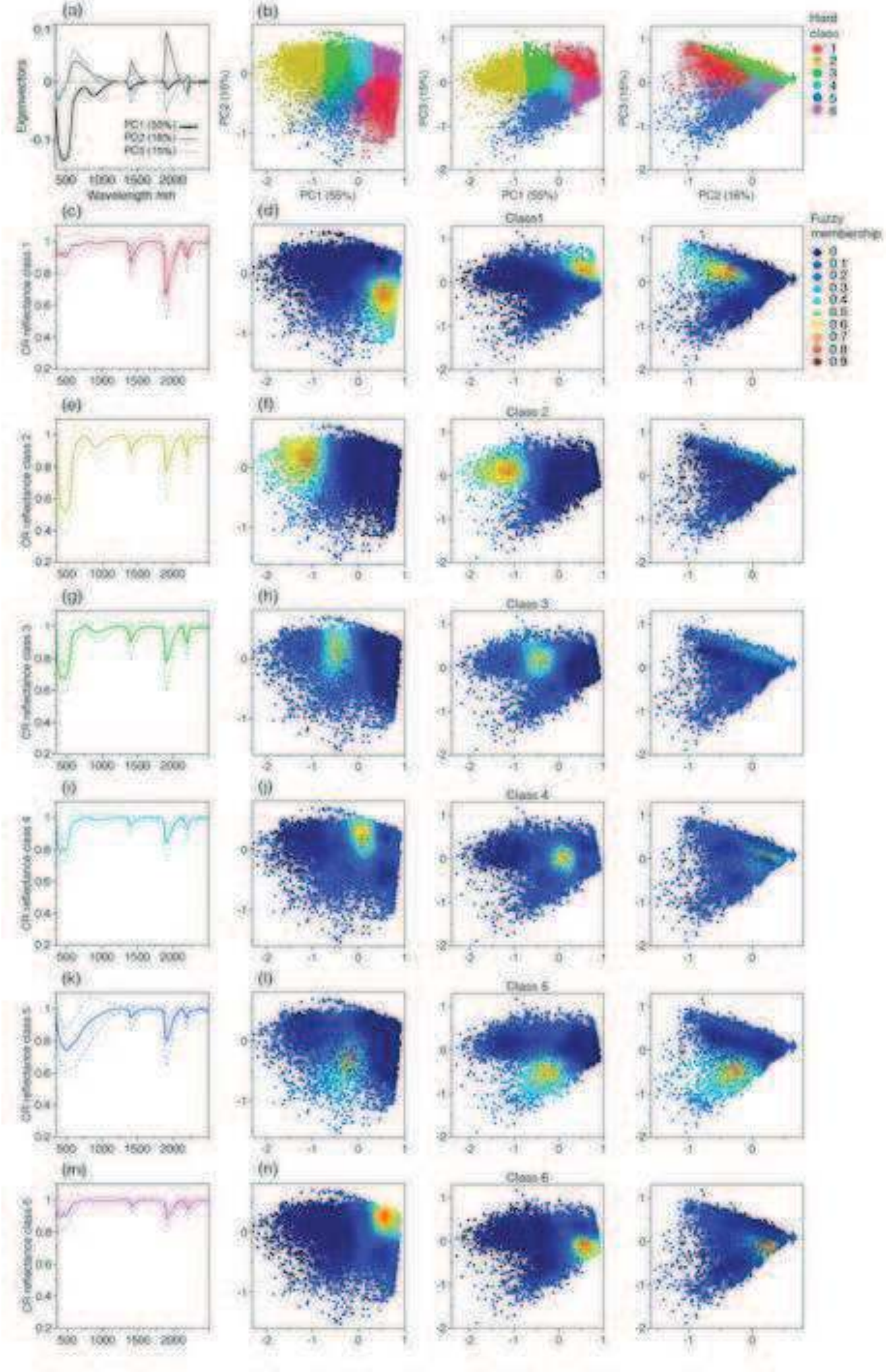

(a)
Eigenvectors
0.1
0
-0.1
500 1000 1500 2000
Wavelength /nm
PC1 (55%)
PC2 (18%)
PC3 (15%)
(b)
PC2 (18%)
PC3 (15%)
PC1 (55%)
PC2 (18%)
Hard class
1
2
3
4
5
6
(c)
(d)
Class1
CR reflectance class 1
Fuzzy membership
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
(e)
(f)
Class 2
CR reflectance class 2
(g)
(h)
Class 3
CR reflectance class 3
(i)
(j)
Class 4
CR reflectance class 4
(k)
(l)
Class 5
CR reflectance class 5
(m)
(n)
Class 6
CR reflectance class 6
0.2
0.4
0.6
0.8
1
-2
-1
0
1

Figure6

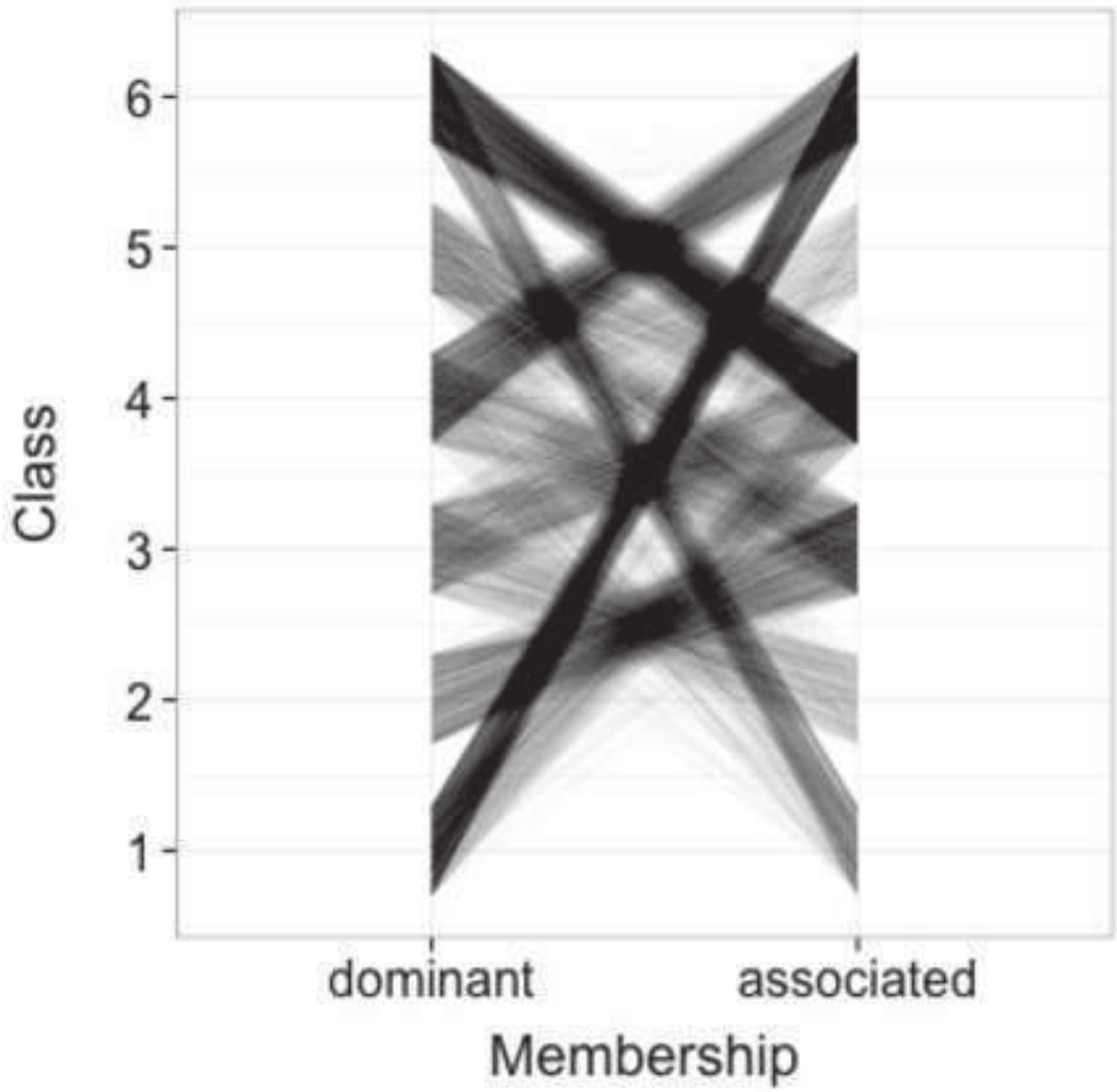


Figure7

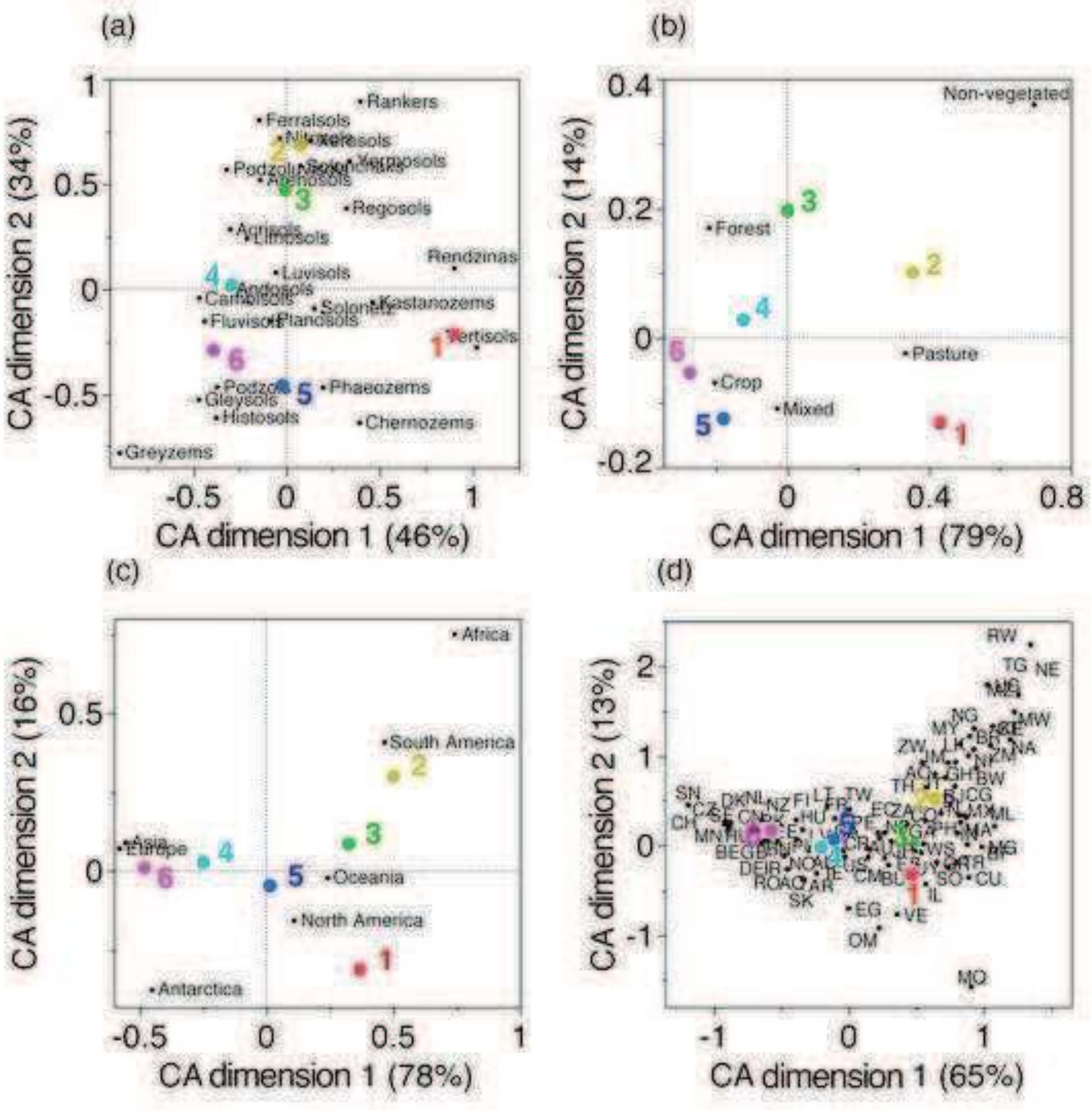

(a)
CA dimension 2 (34%)
CA dimension 1 (46%)
Rankers
Ferralsols
Nitosols
Xerosols
Podzoluvisols
Solonchaks
Yermosols
Arenosols
Regosols
Acrisols
Lithosols
Rendzinas
Luvisols
Andosols
Cambisols
Kastanozems
Solonetz
Fluvisols
Planosols
Vertisols
Podzols
Phaeozems
Gleysols
Histosols
Chernozems
Greyzems
1
2
3
4
5
6
(b)
CA dimension 2 (14%)
CA dimension 1 (79%)
Non-vegetated
Forest
Pasture
Crop
Mixed
1
2
3
4
5
6
(c)
CA dimension 2 (16%)
CA dimension 1 (78%)
Africa
South America
Asia
Europe
Oceania
North America
Antarctica
1
2
3
4
5
6
(d)
CA dimension 2 (13%)
CA dimension 1 (65%)
RW
TG
NE
NG
MW
ZW
NA
BW
SN
CZ
CH
DK
NL
NZ
FI
LT
TW
FR
HU
MN
BE
GB
DE
IR
RO
AR
SK
EG
VE
OM
IL
SO
CU
ML
MG
CG
MX
TH
MO
1
2
3
4
5
6

Figure8

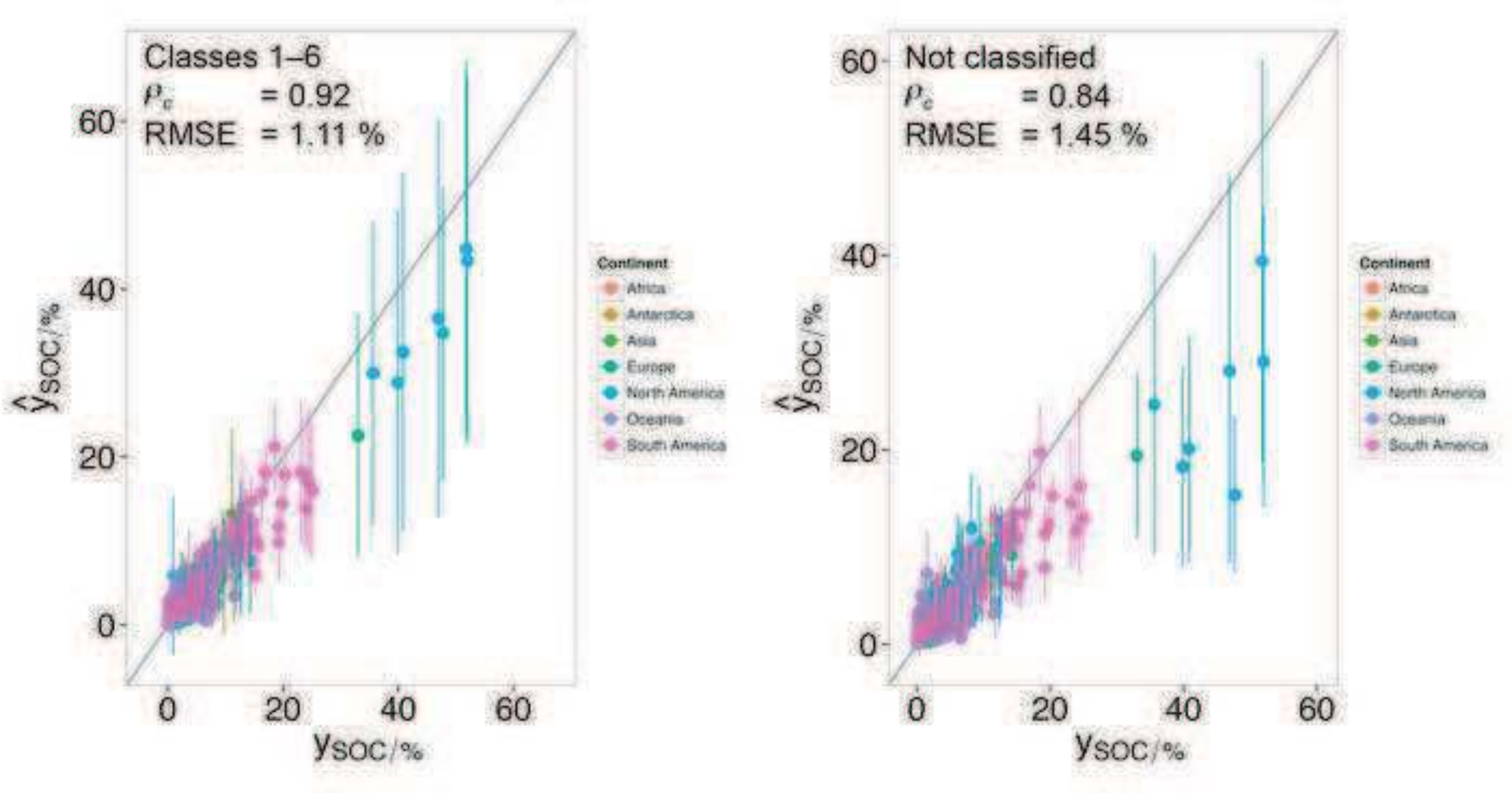

Classes 1–6
ρ_c = 0.92
RMSE = 1.11 %
$\hat{y}_{SOC}$/%
y_{SOC}/%
0
20
40
60
Continent
● Africa
● Antarctica
● Asia
● Europe
● North America
● Oceania
● South America
Not classified
ρ_c = 0.84
RMSE = 1.45 %
$\hat{y}_{SOC}$/%
y_{SOC}/%
0
20
40
60
Continent
● Africa
● Antarctica
● Asia
● Europe
● North America
● Oceania
● South America

Figure9

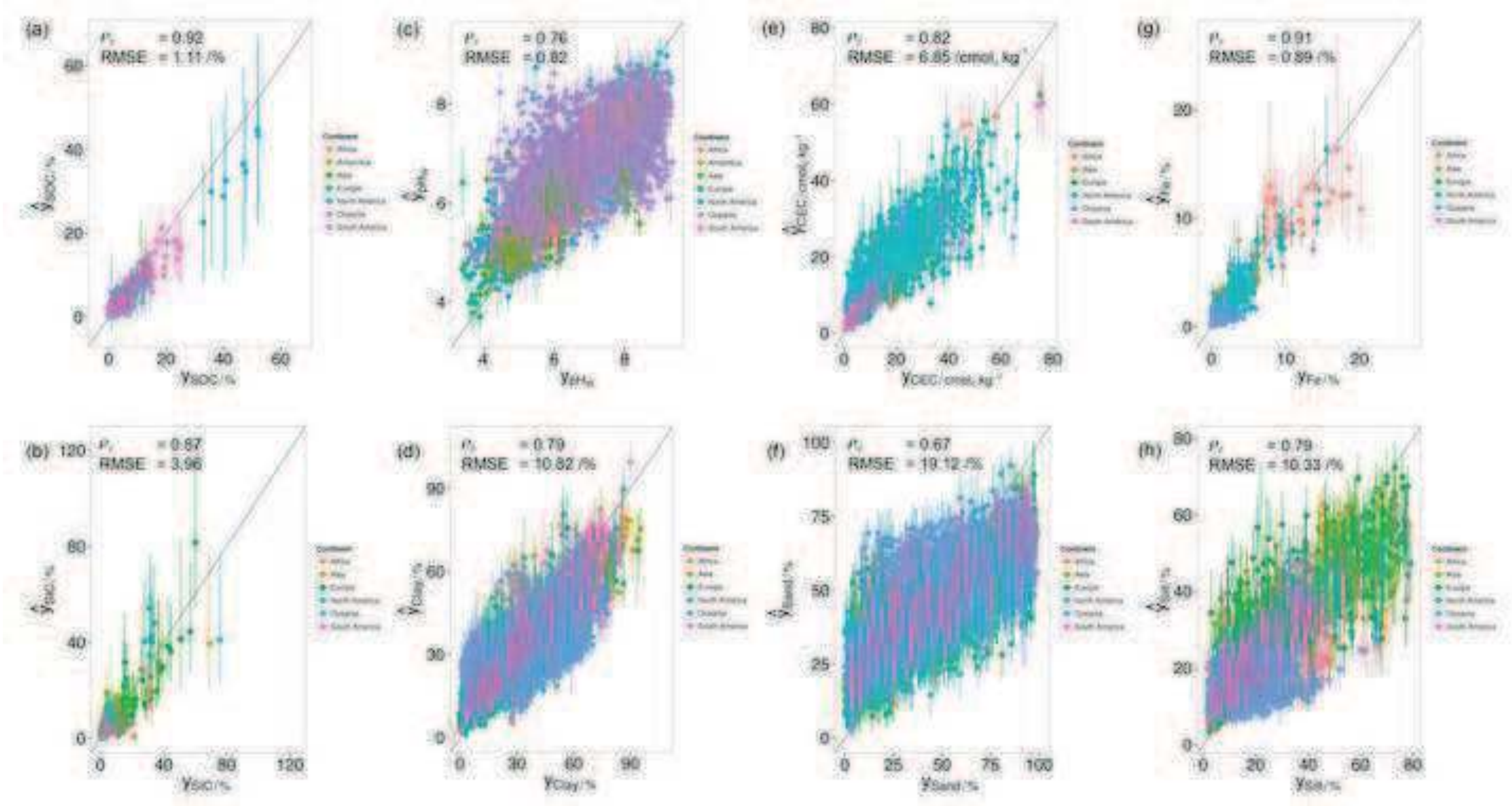

(a)
ρ_c = 0.92
RMSE = 1.11 /%
$\hat{y}_{SOC}$ /%
y_{SOC} /%
0
20
40
60
Continent
Africa
Antarctica
Asia
Europe
North America
Oceania
South America
(c)
ρ_c = 0.76
RMSE = 0.82
$\hat{y}_{pH_w}$
y_{pH_w}
4
6
8
(e)
ρ_c = 0.82
RMSE = 6.85 /cmol, kg^{-1}
$\hat{y}_{CEC}$ /cmol, kg^{-1}
y_{CEC} /cmol, kg^{-1}
0
20
40
60
80
(g)
ρ_c = 0.91
RMSE = 0.89 /%
$\hat{y}_{Fe}$ /%
y_{Fe} /%
0
10
20
(b)
ρ_c = 0.87
RMSE = 3.96
$\hat{y}_{SIC}$ /%
y_{SIC} /%
0
40
80
120
(d)
ρ_c = 0.79
RMSE = 10.82 /%
$\hat{y}_{Clay}$ /%
y_{Clay} /%
0
30
60
90
(f)
ρ_c = 0.67
RMSE = 19.12 /%
$\hat{y}_{Sand}$ /%
y_{Sand} /%
0
25
50
75
100
(h)
ρ_c = 0.79
RMSE = 10.33 /%
$\hat{y}_{Silt}$ /%
y_{Silt} /%
0
20
40
60
80

Figure10

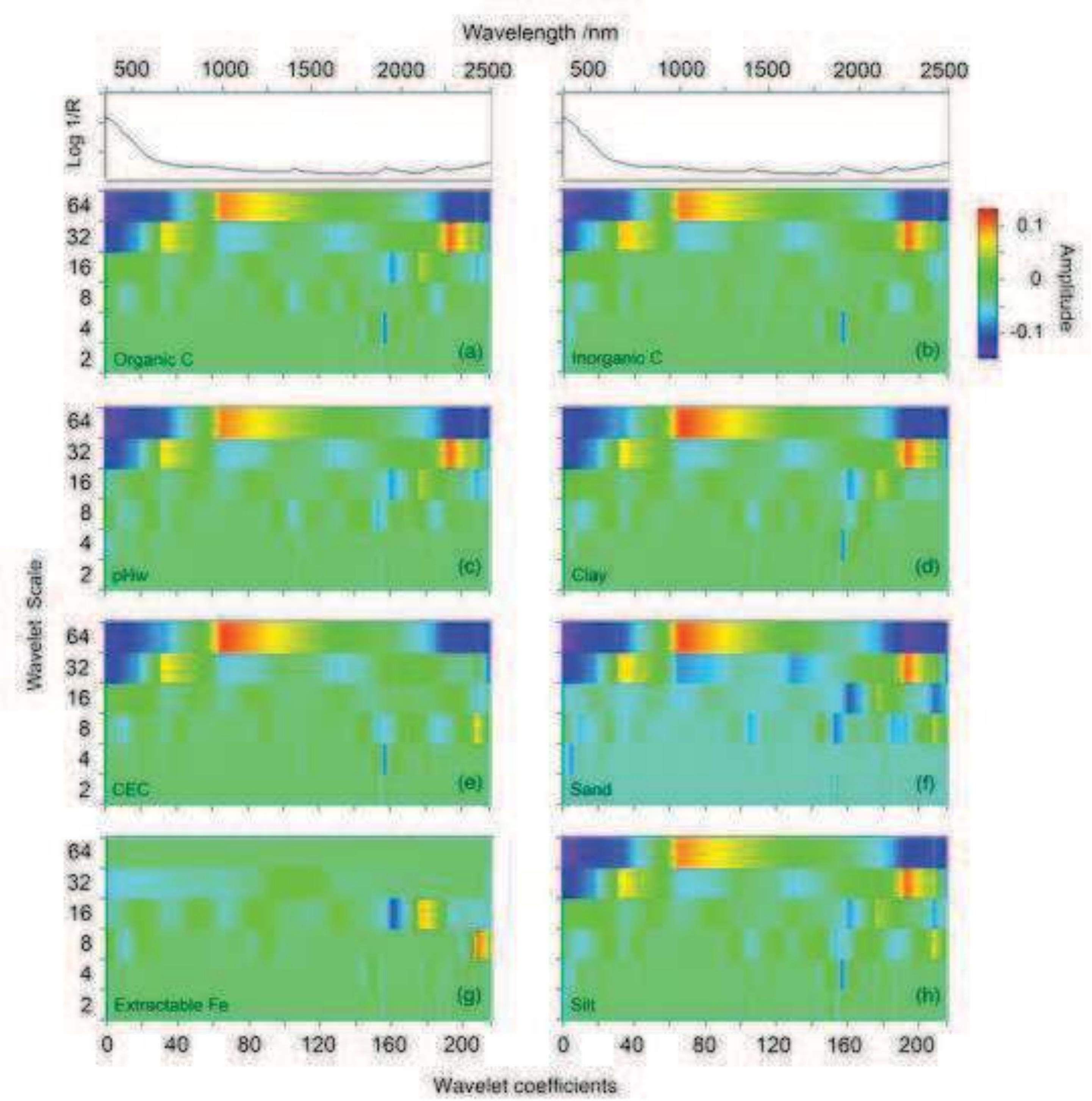

Wavelength /nm
500
1000
1500
2000
2500
Log 1/R
Wavelet Scale
64
32
16
8
4
2
Organic C
(a)
Inorganic C
(b)
pHw
(c)
Clay
(d)
CEC
(e)
Sand
(f)
Extractable Fe
(g)
Silt
(h)
0.1
0
-0.1
Amplitude
0
40
80
120
160
200
Wavelet coefficients

Figure11

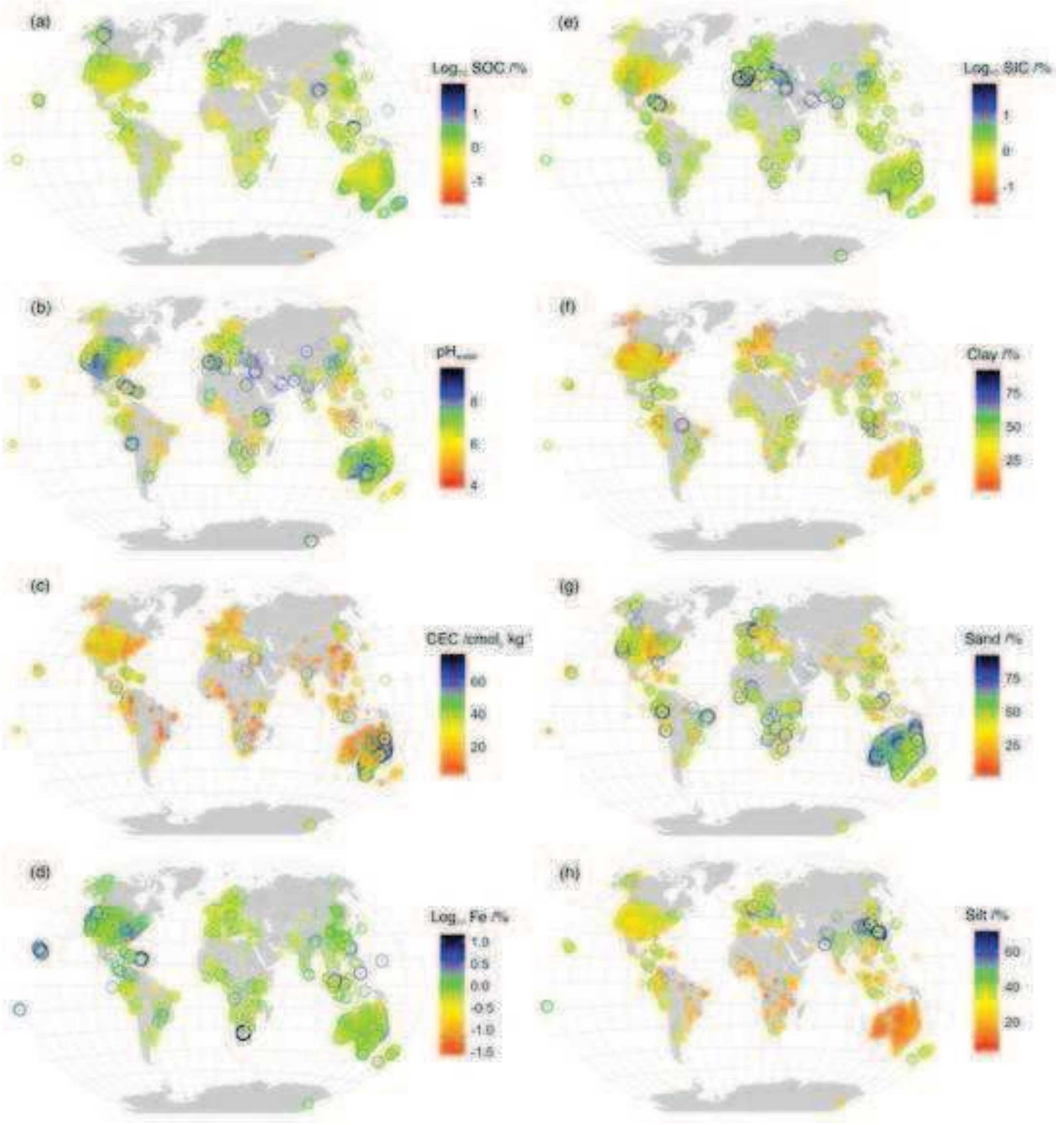
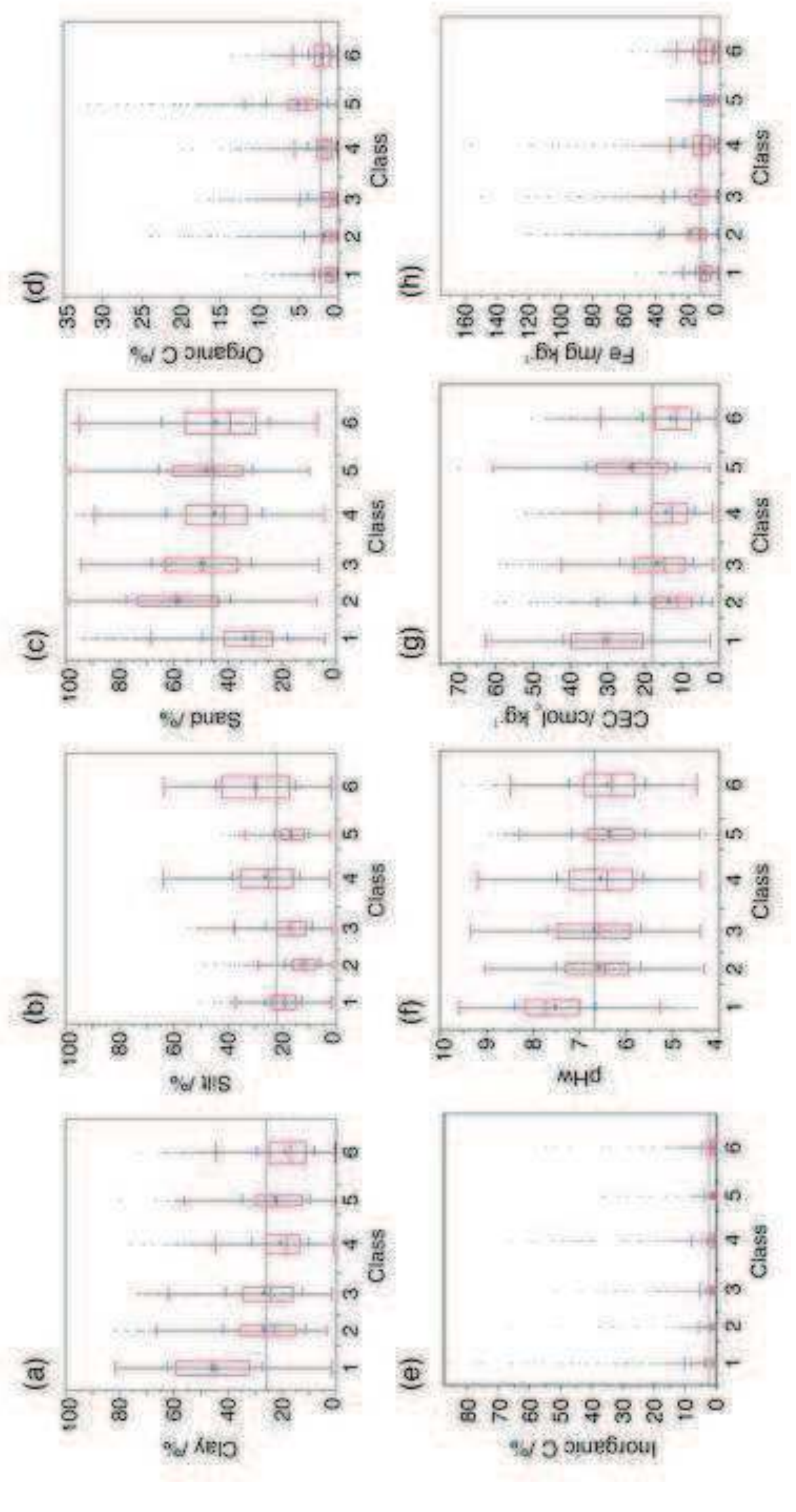

(a)
Log$_{10}$ SOC /%
1
0
-1
(b)
pH$_{water}$
8
6
4
(c)
CEC /cmol$_c$ kg^{-1}
60
40
20
(d)
Log$_{10}$ Fe /%
1.0
0.5
0.0
-0.5
-1.0
-1.5
(e)
Log$_{10}$ SIC /%
1
0
-1
(f)
Clay /%
75
50
25
(g)
Sand /%
75
50
25
(h)
Silt /%
60
40
20

Figure12


(a)
Clay /%
100
80
60
40
20
0
(b)
Silt /%
(c)
Sand /%
(d)
Organic C /%
35
30
25
20
15
10
5
0
(e)
Inorganic C /%
80
70
60
50
40
30
20
10
0
(f)
pHw
10
9
8
7
6
5
4
(g)
CEC /cmol$_c$ kg^{-1}
(h)
Fe /mg kg^{-1}
160
140
120
100
80
60
40
20
0
Class
1
2
3
4
5
6

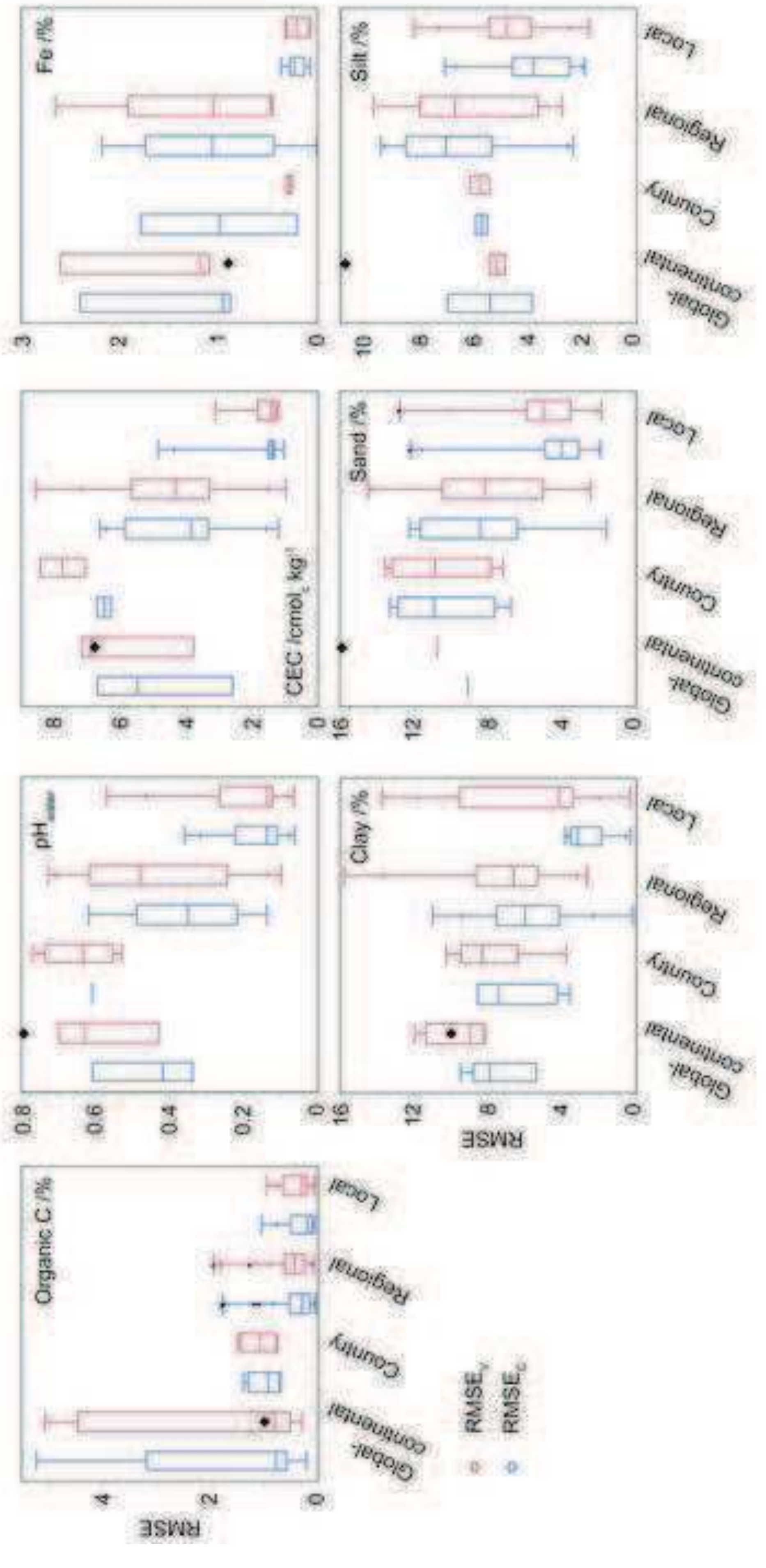

Organic C /%
RMSE
4
2
0
Global-continental
Country
Regional
Local
RMSE$_v$
RMSE$_c$
pH$_{water}$
0.8
0.6
0.4
0.2
0
Clay /%
RMSE
16
12
8
4
0
CEC /cmol$_c$ kg^{-1}
8
6
4
2
0
Sand /%
16
12
8
4
0
Fe /%
3
2
1
0
Silt /%
10
8
6
4
2
0

Figure13