Use of near-infrared spectroscopy to assess phosphorus fractions of different plant availability in forest soils

J. Niederberger[1,]*, B. Todt[1,]*, A. Boča[2], R. Nitschke[1], M. Kohler[1], P. Kühn[3] and J. Bauhus[1]

1 Chair of Silviculture, Institute of Forest Science, University of Freiburg, Freiburg, Germany.

2 Wildland Resources Department, Utah State University, Logan, UT, United States.

3 Department of Geosciences, Physical Geography and Soil Science, University of Tübingen, Tübingen, Germany.

* These authors contributed equally to this work

Correspondence to: J. Niederberger (joerg.niederberger@waldbau.uni-freiburg.de)

Abstract

The analysis of soil phosphorus (P) in fractions of different plant availability is a common approach to characterize the P status of forest soils. However, quantification of organic and inorganic P fractions in different extracts is labor-intensive and therefore rarely applied for large sample numbers. Therefore, we examined whether different P fractions can be predicted using near-infrared spectroscopy (NIRS).

We used the Hedley sequential extraction method (modified by Tiessen & Moir, 2008) with increasingly strong extractants to determine P in fractions of different plant availability and measured NIR spectra for soil samples from sites of the German forest soil inventory and from a nature reserve in south-eastern China.

The R^2 of NIRS calibrations to predict P in individual Hedley fractions ranged between 0.08 and 0.85. When these fractions were combined into labile, moderately labile and stable P pools, R^2 of calibration models was between 0.38 and 0.88 (all significant). Model prediction quality was higher for organic than for inorganic P fractions and increased with the homogeneity of soil properties in soil sample sets. Usable models were obtained for samples originating from one soil type in subtropical China, whereas prediction models for sample sets from a range of soil types in Germany were only moderately useable or not useable.

Our results indicate that prediction of Hedley P fractions with NIRS can be a promising approach to replace conventional analysis, if models are developed for sets of soil samples with similar physical and chemical properties, e.g. from the same soil type or study site.

Keywords: Soil P, Hedley fractions, P pools, model calibration, model validation

1 Introduction

Phosphorus (P) is limiting plant growth and ecosystem productivity in many parts of the world (e.g. Elser et al., 2007, Vitousek et al., 2010). Unlike nitrogen, the world's P stores for the production of fertilizer are finite and therefore will decrease in the future (i.a. Cordell et al., 2009, Edixhoven et al. 2013). It is therefore of paramount importance to use P as efficiently as possible in agricultural and forestry production systems. In several parts of Europe and other parts of the world, concerns have been expressed that a substantial proportion of forest stands may suffer from P limitation (Fox et al., 2011, Khanna et al., 2007, Lorenz et al., 2003). A number of hypotheses, to explain this phenomenon of wide-spread phosphorus limitation have been put forward (e.g. nitrate uptake suppressing P uptake, soil acidification reduces rate of organic P mineralization and increases inorganic P sorption capacity) (Mohren et al., 1986, Paré and Bernier, 1989, Gillespie and Pope, 1991). In long-term development of forest ecosystems following disturbance, P (in contrast to N) can only diminish through processes such as erosion and timber harvest and therefore becomes limiting to biological activity. Consequently, increasing P deficiency is also a natural process of ecosystem retrogression (Walker et al., 1976, Wardle, 2009), which seems to be accelerated by anthropogenic activities. The increasing demand for forest biomass, such as for renewable energy, might exacerbate the P nutrition problem in many forests (Vanguelova et al., 2010). Therefore, suitable methods for monitoring the current status and the medium to long-term trends of P nutrition in forest ecosystems are urgently needed.

For forest soils, often solely total P contents are measured. But total P is not a meaningful variable as it is not very sensitive to environmental or management influences over time and is not indicative of the amount of plant-available P in forest soils, where P can occur in many organic and inorganic forms of different availability (Khanna et al., 2007). Especially for forest soils, it is important to differentiate between organic and inorganic P forms, since organic P, although not directly available for plant uptake, plays a major role in tree nutrition (Turner et al., 2005, Shen et al., 2011, Rennenberg and Herschbach, 2013). Analytical approaches, such as the Hedley fractionation (Hedley et al. 1982), have been developed to quantify organic and inorganic P in fractions of different plant availability. These fractions have been successfully applied to characterize the dynamic nature of P availability in forest ecosystems (e.g. Richter et al., 2006). Several authors discussed the usefulness of this commonly used sequential extraction method to determine P in fractions of distinctly different solubility (Cross and Schlesinger, 1995, Condron and Newman, 2011, Yang and Post, 2011). The ecological relevance of chemically derived Hedley fractions has been critically discussed in some cases (Turner et al., 2005) and it is still unclear which fractions are relevant for plant growth (Rennenberg and Herschbach, 2013). However, the application of this sequential extraction method in forest soils has proven to be useful. In contrast to agricultural soils, the slowly cycling P pool in forest soils contributes substantially to plant nutrition, whereas labile P pools stay relatively constant, indicating that higher amounts of P from more slowly cycling pools “feed” the labile P pool (Compton and Cole, 1998, Richter et al., 2006). However, the Hedley fractionation method is very time consuming, labor intensive and therefore costly, which renders it unsuitable for the analysis of large numbers of samples (Cécillon et al., 2009). Therefore, it would be useful, if Hedley P fractions could be determined by means of an indirect and less expensive technique. For this purpose, the use of near-infrared spectroscopy (NIRS) may be a promising approach.

NIRS is a rapid and non-destructive analytical method routinely used in a wide range of fields, mostly in quality control in the chemical, forage or food industry. In the field of soil sciences, NIRS has been successfully applied to predict soil chemical and soil biological parameters like carbon (in the form of total C, SOC, microbial C, and mineralizable C) or nitrogen (in the form of total N, microbial N, and mineralizable N) contents (Chang et al., 2001, Ludwig et al. 2002, Zornoza et al., 2008) and was also applied to agricultural soils to predict available P (Olsen P, Mehlich III P) (eg Maleki et al., 2006, Dhawale et al., 2013). An overview of available reflectance spectra for different soil properties is available at the global soil vis-NIR spectral library (World Agroforestry Centre – ICRAF and ISRIC-World soil Information, 2010).

The mineral fractions of soil, which commonly constitute the major portion of soils, are usually hardly visible for NIRS (Malley et al., 2004). Despite the complexity and heterogeneity of chemical and physical soil properties, the spectral information of soils is surprisingly uniform. For example, 30 soil types of the USDA soil classification system could be grouped into only 5 different spectra classes (Stoner and Baumgardner, 1981). Owing to the low dipole moment between P and oxygen, phosphates cannot be excited by infrared radiation and therefore not be detected directly with NIRS. However, P can be detected indirectly, if it is bound organically or through correlation with the above stated other soil properties such as C and N content that are detectable by NIRS.

To develop robust NIRS models, it is essential to produce a reliable quality in spectral data sets. Several publications on NIRS deal with the influence of the quality of soil samples on spectral information. Stenberg et al. (2010) and Brunet et al. (2007) discussed the influence of ground soil samples in comparison with sieved samples (<2 mm) and found increasing model quality for more homogeneous ground samples. The quality of the data with regard to the heterogeneity of samples was addressed by Abrams et al. (1987) and Shenk and Westerhaus (1993). Global models used for the prediction of forage quality in hay, where “global” means that the model is based on the largest sample population comprising a wide variety of hay samples and hay composition from many sites, produced predictions that were as good as local models derived from one single site. However, in other studies using plant material (pine needles), both, improvements and reductions in model quality for more heterogeneous data sets (addressed as global) when compared with homogeneous (local) data, were observed (Gillon et al., 1999). Reductions in model quality were observed for the prediction of C content in sample sets with high variation in chemical composition associated with high spectral variation. Since high variation in chemical composition and spectral properties is true for forest soil samples, it might be a challenging process to build robust models. The problem of detecting soil characteristics using NIRS methods is not new (Stoner and Baumgardner, 1981, Ben-Dor and Banin, 1995, Udelhofer et al., 2003, Cécillon et al., 2009, Genot et al. 2011). To our knowledge, up to now only few publications (Guerrero et al., 2010, Sankey et al., 2008) have addressed the problem of NIRS model calibration and prediction quality for forest soil samples and none of these has focused on soil P.

Currently, there are only few NIRS models available to predict soil-P parameters. These includes NIRS-prediction models for Mehlich-III extractable P (Chang et al., 2001), for total P and exchangeable P (Malley et al., 2004). Hitherto, there have been no studies focusing on the use of NIRS to predict organic and inorganic Hedley P fractions in forest soils. This study aimed to examine the general ability of NIRS methods to assess these P fractions. Specifically we addressed the following hypotheses:

1. Hedley fractions of different P availability can be sufficiently well predicted through NIRS models. The criteria used to quantify the quality of NIRS models will be introduced in the Material and Methods section.
2. The quality of predictions through NIRS models is higher for organic than for inorganic P fractions.
3. The quality of predictions of P fractions depends on the variability in Phosphorus content and soil properties of soil sample data sets.

2 Material and methods

2.1 Soil samples

Two different sets of mineral soil samples were used for this study. Approximately half of the soil samples originated from the German nationwide forest soil inventory (BZE). This data set covers 282 archived soil samples from 145 sites belonging to the BZE inventory grid. Including 70 sites in Baden-Württemberg (70 samples 0-5 cm, 69 samples 10-30 cm), 23 sites in Hessen (23 samples from both depths), 33 sites in Lower Saxony (32 samples 0-5 cm, 28 samples 10-30 cm) and 19 sites in Saxony Anhalt (18 samples 0-5 cm, 19 samples 10-30 cm). The area covered by the BZE samples was a region of approximately 200 km width and 700 km length. The BZE plots were part of Germen Forest Soil inventory net, which is realized as a grid with a size of 8 x 8 km. From each site, if available, two samples (0-5 cm and 10-30 cm depth) were analyzed. BZE samples are composite samples from 8 replicates (BZE 2006). The samples included in this data set were selected to cover the majority of forest soil types in Germany and a large variation in total phosphorous concentration. Development of robust NIRS-models requires sample populations that cover the whole calibration range with an approximately even distribution of samples across the range of the variable to be predicted. In contrast, sample populations with normally distributed samples tend to overestimate low values and underestimate high values in model calibration (Williams, 2001). The BZE soil samples were grouped by soil type in eleven major groups: Regosols (n = 20), Gleysols (n = 6), Luvisols (n = 59), Vertisols (n = 2), Podzols (n = 10), Chernozem (n = 2), Stagnosols (n = 38), chromic Cambisols (n = 10), Anthrosols (n = 4), cambic Umbrisol (n = 78), and Cambisol (n = 53) (IUSS Working Group WRB 2006). The total P content covered by the BZE samples ranged from 58 to nearly 2800 mg kg^{-1} (Tab. 1).

The other half of the soil samples originated from the international research project "Biodiversity and Ecosystem Functioning in China (BEF China)" (Bruelheide et al., 2011). The 294 soil samples from the BEF China project were collected from 27 so-called Comparative Study Plots (= CSP, 30 x 30 m) in the Gutianshan National Nature Reserve, which has a size of 81.07 km² (mean distance between all CSPs = 3.4 km; Min = 0.04 km; Max = 8.98 km). At all 27 CSPs the soil type was a Cambisol (reference soil group after WRB, 2006) comparable to the "typical Brown Earth" after the German classification. Thus the 294 BEF samples represented a more homogeneous sample set than the BZE samples. At each CSP, nine soil cores that were evenly distributed across the plot were collected. One bulk sample was then combined from the nine cores, for each of the three depth increments (0-5 cm, 5-10 cm and 10-20 cm), i.e. 81 samples in total. At twelve CSPs, an additional 37 samples were collected from the three topmost diagnostic soil horizons, predominantly Ah and Bw horizons (at one plot, the four topmost horizons were sampled) of representative soil profiles of the CSPs (ranging in depth between 47 and 120 cm). In addition, 176 samples (0-20 cm) were collected in four replicates from 44 so-called Neighborhood Diversity Tree Clusters, which were located in the proximity of some of the permanent Comparative Study Plots. Each of the four samples of tree clusters, were also composite samples from three cores each. Each composite sample represents different conditions within the cluster; they were collected at the base of individual trees belonging to different or the same tree species and in the center of a triangle between these trees. We cannot rule out a spatial correlation between these samples. Total P content of all BEF samples ranged from 26 to 330 mg kg^{-1} (Tab. 1).

Both sample sets, BZE and BEF, contained samples that originated from different depth but the same location (soil cores). Since these samples might not have been spatially independent, the autocorrelation among samples could have influenced the quality of NIRS models. For example, the validation may not be truly independent, if calibration and validation sample sets contain samples that are systematically correlated with each other. To address this issue, we calculated the Durbin-Watson coefficient (d) as a measure of autocorrelation between measured and predicted values and compared this coefficient between sample sets from all soil depths (containing samples from the same location) and from one soil depth only (spatially independent). This test did not indicate autocorrelation within our sample sets (d values between 1.75 and 2.23). In addition, we found no differences for the Durban-Watson coefficient for all BZE samples (depth: 0-5 cm and 10-30 cm) and BZE samples sets from one depth only. Since there was no indication of autocorrelation between samples of different depths, we included all samples in our calibration and validation steps.

The data for C and N content as well as the pH values of the BZE samples were provided by the Northwest German Research Station and the Forest Research Institute Baden-Wuerttemberg. Samples were prepared and measured according to the German Forest science standard (BMVEL, 2009). Total soil C and N contents of the horizon-wise samples and the three depth increments of the BEF samples were determined after dry combustion (1150°C) with an element analyzer (Elementar Vario EL III) (DIN ISO 10964). Since the material is non calcareous, total soil C equals soil organic C. For these samples, pH values were derived in H_2O (2.5 ml : 1 g water solution to soil) with a Sentix 81 electrode and pH meter (WTW, Germany). C and N contents of the 176 neighbourhood diversity samples of the BEF dataset were determined with a TruSpec Elemental Analyzer (Leco, USA) and pH values were measured in H_2O (2.5 ml : 1g water solution to soil) with a blueline electrode and pH meter (Schott, Germany).

All soil samples were dried at 40°C for 24 h and sieved (< 2 mm). For C and N analyses, samples were finely ground. The intensity of sieving respectively grinding has an impact on the reproducibility and the distribution of extractable P over the fractionation process (Tiessen and Moir 2008) and the NIR spectra (Brunet et al., 2007). Due to the more “natural” extractability of P compared to ground samples and the ecological importance of soil aggregates, we used < 2 mm sieved soil samples for the determination of P fractions.

Before measurement of the NIR spectra, the samples were ground to powder by a mixer mill MM400 (Retsch, Germany) and dried over night at 40°C.

2.2 Phosphorus fractionation

Soil samples were analyzed by the sequential extraction method according to Hedley (Hedley et al., 1982b). We used 24 samples in one batch consisting of 23 individual samples including six replicate samples (as random quality check) and one in-house soil standard (as standard quality check). In the protocol applied to the samples, 0.5 g soil was extracted by different extractants with increasing strength (Fig. 1), starting with distilled water containing an anion exchange resin - Dowex 1x8; 20-50 mesh; Sigma-Aldrich, Taufkirchen, Germany (Resin), followed by sodium bicarbonate, 0.5 M ($NaHCO_3$), sodium hydroxide, 0.1 M (NaOH), sodium hydroxide, 0.1 M with ultrasonic treatment - ultrasonic bath RK510H, 35kHz; 23W/l; Bandelin, Berlin, Germany; (S-NaOH), hydrochloric acid, 1M (1M HCl) and a final acid digestion (Residual). Tiessen and Moir (2008) introduced an additional extraction step with hot concentrated hydrochloric acid (HCl conc.) before the final digestion. Our fractionation scheme considered all of these 7 fractionation steps, but used 25 ml solution for the extractions and 2.2 ml concentrated nitric acid for the final acid digestion. The whole fractionation process is described in detail by Tiessen and Moir (2008).

Total P content as index value for the sum of all fractions was measured separately after nitric acid digestion, which is a standard method recommended by the German forest soil survey (BMVEL, 2009).

Phosphorus was determined colorimetrically with the spectrophotometer UV mini-1240 by Shimadzu through molybdate reactive P method described by Murphy and Riley (1962) modified by John (1970), but keeping the ammonium molybdenum concentration at 10 g l^{-1}. Every extract needs to be adjusted for pH, as development and stability of the blue phosphomolybdic complex is pH dependent (Drummond and Maher, 1995, Huang and Zhang, 2008, Tiessen and Moir, 2008). The P extracted within every step is only partly detectable as free phosphate. To determine the total phosphate, the $NaHCO_3$, NaOH and the HCl conc. extracts were oxidized with ammonium persulfate in an autoclave. The difference between total P after digestion (TP) and the free phosphate in the extract (Pi) provides the organically bound P (Po) (Rowland and Haygarth, 1997, Tiessen and Moir, 2008). The amount of organically bound P in the resin and the 1 M HCl extracts were considered negligible (Tiessen and Moir, 2008) and therefore not analyzed. We summed up the values Pi of the NaOH and Pi of the S-NaOH fraction as well as Po of the NaOH and Po of the S-NaOH fraction because the ultrasonic treated NaOH step is chemically hardly different from the main NaOH fraction and a differentiation between these two fractions with NIRS did not seem meaningful. Similar considerations applied to the distinction between Pi and Po in the HCl conc. extract, since concentrated HCl is a strong oxidant and an additional oxidation with persulfate is chemically comparable. In our analysis, the precision of the Pi value of the HCl conc. extract lacked reproducibility within replicates compared to the TP value and this might be an indicator for an incomplete reaction during the boiling procedure. Also several organic phosphates are unstable under these acid conditions (Turner et al., 2005) and the results of Pi and Po of the HCl conc. fraction lacked also reproducibility within our standard. The total P content of the HCl conc. fractions showed good reproducibility and therefore was used for calibration.

In summary, we selected the following fractions to perform calibrations with NIRS: Pi Resin, Pi $NaHCO_3$, Po $NaHCO_3$, Pi NaOH, Po NaOH, Pi 1M HCl, P HCl conc., P residual. For the purpose of subsequent statistical analyses, the Hedley fractions were grouped by their solubility in labile (Pi Resin, Pi and Po $NaHCO_3$), moderately labile (1M HCl, Pi and Po NaOH) and stable P pools (P HCl conc. and P residual) as commonly found in literature (Cross and Schlesinger, 1994, Stevenson and Cole, 1999, De Schrijver et al., 2011, Yang and Post, 2011). Several tests were conducted to ensure that Hedley fractions were reproducible and accurately determined. One indicator of the reliability of the Hedley fractionation method is the recovery rate of total P in the sum of fractions, which was very good for the soils in this study (Fig. 2). Only for samples with high P contents, the sum of fractions may either systematically exceed the total P content or the amount of measured total P may be underestimated. Errors for samples with high P content could occur either due to necessary dilution steps for measuring fractions with very high P concentration, where small differences during colorimetric measurements lead to relatively high errors for the final value, or to incomplete digestion of P with nitric acid. To avoid these sources of error, we performed repeated measurements of these samples and we reduced the sample weight to 400 mg and increased the amount of acid to 3 ml in these samples. In addition, this problem occurred only in a very low proportion of samples (10 out of 576) and removing these samples from the data set did not change the calculated recovery rate.

2.3 Near-infrared spectroscopy

Molecules with dipole moments can be excited to stretching, bending, and rotating oscillations by absorbing infrared radiation. While an excitation of fundamental oscillations occurs in the mid infrared range of the light, the near-infrared spectrum of a substance shows overtone peaks and combination bands of molecular vibrations, mainly from bonds like O-H, C-H and N-H (Burns and Ciurczak, 2001, Malley et al., 2004, Cécillon et al., 2009). This makes NIRS especially suitable for organic compounds (Weyer, 1985, Miller, 2001). Phosphates and other inorganic P compounds are hardly detectable by NIRS due to the weak P-O dipole moment (Malley et al., 2004). The detection of nutrients such as in P compounds seems to be possible, if they are correlated with other NIRS-detectable soil properties such as soil texture, water content, organic compounds, and mineralogy, in particularly soil organic matter or organic carbon (Stenberg et al., 2010).

Each type of molecule has characteristic absorption bands for near-infrared radiation, but there is much interference and overlap between the overtone and combination oscillations. Therefore molecules cannot be identified by single peaks. The whole spectral information or at least several regions of the spectrum are needed to determine a molecule (Burns and Ciurczak, 2001, Günzler and Gremlich, 2003, Roberts et al., 2004).

The spectral region of NIR radiation ranges from 800 nm to 2500 nm wavelength, following directly the visible light region. In infrared spectroscopy, commonly the wavenumber ν in cm^{-1} is used, which is the reciprocal of the wavelength λ.

$$\nu = \lambda^{-1}$$

Since the O-H bond has a strong influence on near-infrared absorption (Schmitt, 2000, Malley et al., 2004) ground samples were dried at 40°C and measured under constant humidity conditions. For each soil sample, five spectra were taken with a Fourier Transform Mid- and Near Infrared combination instrument (Tensor 37, Brukeroptics, Ettlingen, Germany). Each spectrum is a mean of 64 individual scans over the range of 12000 to 4000 cm^{-1} wave numbers with a resolution of 16 cm^{-1}. Model development was done with the OPUS 6.5 spectroscopy software (Brukeroptics, Ettlingen, Germany). Before statistical analyses, a number of mathematical data pre-processing options were tested. The pre-processing options providing the best results were first derivative, vector normalization, or a combination of these two. These were used as data treatment before partial least square regressions were conducted and applied to a variety of frequency ranges to find the best fit with the value of the respective component; spacing was 1; several smoothing point values (5, 9, 13, 17, 21, 25) for the derivatives were tested, too. Calibration was performed with cross validation, a common approach for small data sets (Williams, 2001, Conzen, 2005). Here, a defined number of samples, in our case one sample, were stepwise excluded from the calibration process. The rest of the samples were used to predict the excluded samples. This procedure was performed until all samples were excluded once, and the best models to predict all samples were found (Conzen, 2005). For each dataset, only 80% of the samples were used for this procedure and the development of models, the remaining 20% were left out for later independent validation of the models to test their prediction quality. This distribution of samples was carried out with an automatic function in the Opus software (Conzen, 2005, Bruker Optik User Manual, 2006) selecting samples for each individual model equally over the whole range of each dataset. The min and max values were always assigned to the calibration set and the samples second to min and max were always assigned to the validation set. To avoid confusion with the final validation process of independent samples, we refer to the cross-validation process as calibration.

Our initial intent was to generate a global prediction model for P-concentration that integrated the whole variety of forest soil types. The aim was to test, if such a global model could be used to estimate P-concentration of all BZE samples and for future inventories. Since this approach proved to be difficult, we also tried to develop NIRS models for a sub-set of the total BZE sample pool with reduced variation in soil properties including P. Therefore we selected the BZE samples belonging to the German soil classification group “brown earth”. This was the only sub-set of soil types (using soil type as an indication of similar soil properties) in the BZE data set with sufficient number of samples for NIRS model development. Additionally, we used soil samples from the above mentioned BEF China project, which represented a data set from a small geographic region and similar soil properties.

To address the hypothesis that the predictive quality of NIRS models increases with the homogeneity of sample sets, we finally established calibrations for sample populations of varying heterogeneity (Fig. 3):

1) a full dataset containing all soil samples with the greatest variability (n = 576)
2) the complete BZE dataset with almost similar variability comprising a wide range of different soil types (n = 282)

3) the less variable BZE sub-set of “Brown Earths” soil samples (n = 84)
4) the BEF-China sample set comprising soils from one soil type and from one small region (n = 294).

To assess whether the prediction of P by NIRS depended on direct detection of soil P - regardless of organic or inorganic form - or rather indirectly on properties of soil organic matter, we used regression analyses to see whether the quality of NIRS models to predict P could be explained by the relationships between P and soil C and N. As mentioned above, it was presumed that P bound to C or N could be modelled with better quality.

The BZE Brown Earths dataset (Set 3) with 82 samples was too small for a robust validation with external samples that were not used in the calibration. Still, we included it in our study, because it was the only meaningful and less variable subset of samples that could be separated from the whole BZE data set (Set 2). The sizes of the sample Sets 1 and 2 were comparable with those used in other studies to develop “global” models (Abrams et al., 1987;, Shenk and Westerhaus, 1993, Gillon et al., 1999). Sample Set 4 could be regarded as an extensive but local data set. A principal component analysis (Fig. 3) (OPUS 6.5 software) was carried out to assess spectral properties of the selected subsamples. The general spectral properties of the soil samples indicated that the BEF China sample set (Set 4) was more homogenous than the BZE sample sets and clearly separated from the BZE sample set as a whole as well as from the subset BZE-“Brown-Earth”. The spectral properties of the two subsets of BZE samples were not different

To evaluate the prediction quality of NIRS models, the statistical parameters R^2 and the ratio of performance to deviation (RPD = ratio of standard deviation to standard error of prediction) were used. The usefulness of these statistical parameters to characterize the quality of NIRS models was discussed controversially (Chang et al., 2001, Williams, 2001, Malley et al., 2004, Zornoza et al., 2008). Especially for soil analysis, high values for RPD and R^2 are hardly achieved (Malley et al., 2004). To judge our RPD and R^2 values, we followed the classification suggested by Malley et al. (2004), who provided an overview and summary on this topic. We applied the following 4 levels to classify model quality:

- Level A: $R^2 > 0.95$ and RPD > 4, excellent calibrations, usable for all applications

- Level B: R^2 between 0.90 and 0.95, RPD between 3 and 4, successful calibrations usable for most applications some with caution, e.g. quality assurance
- Level C: R^2 between 0.80 and 0.90, RPD between 2.25 and 3, moderately successful calibrations, usable with caution, e.g. research
- Level D: R^2 between 0.70 and 0.80, RPD between 1.75 and 2.25, moderately useful calibrations, usable for screening of samples

Calibrations with lower values might still be useful for rough screening purposes (Chang et al., 2001).

All statistical analyses shown in this study are performed with SPSS Version 20.

3 Results

3.1 Soil phosphorus contents

Total P contents were higher and showed a wider range in the BZE samples than in the BEF-China samples for all pools of P availability and individual fractions (Tab. 1). The sub group “Brown Earths” was characterized by a large variation in total P content (130 to 2200 mg kg^{-1}, as the sum of all Hedley fractions), which had a smaller range of values compared to the pool of all BZE samples but, as indicated by the standard deviation, a similar variation in total P. The same applied to P in the labile, moderately labile, and stable pools. The mean percentage of organic P, here calculated as the sum of all Po fractions without the HCl conc. fraction, was higher in the BEF-China samples (57%) than in the BZE and BZE “Brown Earths” samples (39% and 43%). However, there was no difference in the range of organic P (10-70% of the total) for the three datasets. The organic P fraction with the highest concentrations in all three datasets was the one extracted with NaOH (mean values for BZE 155 mg kg^{-1} , equivalent to 28% of P_t, BZE “Brown Earth” 196 mg kg^{-1}, equivalent to 31% of P_t, and BEF-China 35 mg kg^{-1}, 37% of P_t). The P concentrations in the $NaHCO_3$-extractable and the 1M HCl-extractable Pi fraction in the BEF samples were so low as to fall below the detection limit of the Murphy & Riley (1962) method.

3.2. NIRS models for P fractions

Initially, NIRS models were calibrated for both, the complete sample set as well as for three different subsets (BZE, BZE Brown Earths, and BEF-China). The global model for all samples produced calibrations that were only below D level quality ($R^2 = 0.08–0.68$; RPD = 1.04–1.74). The BZE model, which was developed for samples from different soil types, showed only for the P HCl conc. fraction a moderately useful calibration, while all other fractions could not be calibrated sufficiently well (Tab. 3). For all fractions except the inorganic NaOH, better calibrations were achieved with the BZE "Brown Earths" subset than with the complete BZE dataset. However, calibration models achieved only D level quality for two of the fractions (Po NaOH, P HCl conc.). The more homogenous BEF-China dataset yielded the best calibrations. Only the inorganic $NaHCO_3$ and the inorganic 1M HCl fraction were, owing to the very low contents, not calibrated successfully. Calibrations to predict the Pi Resin, Po $NaHCO_3$, Pi NaOH and P Residual fraction produced models of level D quality, while calibrations for the organic NaOH and the P HCl conc. fractions yielded models of level C quality. Overall, best calibration models were achieved for the BEF-China dataset especially for the organic P fractions extracted with $NaHCO_3$, NaOH, and the P HCl conc. fraction (Tab. 3).

Models providing level D calibrations were subsequently tested with a validation dataset comprising samples not included in the calibration. The validation of the level D calibration model of the BZE set for the HCl conc. P fraction was not successful. Validation of this fraction in samples of the more homogenous subset BZE "Brown Earths" yielded a model quality of level C and validation models for the P HCl conc. fraction in BEF soil samples achieved a level B quality (Fig. 4).

The validation of models for the Po $NaHCO_3$ (R^2=0.72, RPD=1.87), Pi NaOH (R^2=0.68, RPD=1.77) and P Residual (R^2=0.75, RPD=1.98) fractions of the BEF China dataset reached level D quality, while the model for the Pi Resin fraction could be validated at level C quality (R^2=0.83, RPD=2.39). The validation of the model for the Po NaOH fraction reached level B (R^2=0.91, RPD=3.23).

3.3 NIRS models for P pools

Grouping of the Hedley fractions into labile, moderately labile and stable P fractions did result in good models for the BEF-China dataset, while only the stable fractions of the other three datasets (BZE+BEF, BZE, BZE Brown Earth) could well be predicted with NIRS models (Fig. 5). Validation of the best calibration models for the stable P pool in the three datasets BZE+BEF, BZE and BZE “Brown Earths” yielded a level D model (R^2=0.79, RPD=2.16) for the complete set (BZE+BEF) as well as the BZE dataset (R^2=0.73, RPD=1.91) and a level C model (R^2=0.88, RPD=2.83) for the BZE “Brown Earths” dataset.

Validations for the BEF China datasets achieved a level C quality for the labile and the moderately labile P pool, and level D quality for the stable P pool (Fig. 6).

To assess whether P content in Hedley fractions was determined indirectly through relationships with soil C and N, we calculated relationships between NIRS model quality and coefficient of determination for the relationship between P in Hedley fractions and soil total C and total N (Fig. 7). For organic P fractions high coefficients of determination (r^2) between P in Hedley fractions and total soil C or total N coincided with high r^2 values of NIRS models. In contrast, the P HCl conc. fraction showed the best model quality for all three data sets (r^2 0.72-0.85) but very poor r^2 for the relationship with total C and total N. For inorganic P fractions for the BZE samples, coefficients of determination (r^2) were poor for both, the relationships between soil C and N and P fractions and NIRS models (Fig. 7).

4 Discussion

4.1 NIRS models for P fractions and pools

Calibration and validation models of individual fractions achieved good to satisfactory results for some but not all individual P fractions. Generally, the accuracy and reproducibility of the reference method (here Hedley fractionation) is of crucial importance for the quality of predictions with NIRS models (Williams 2001). To increase reproducibility in our study, all samples were measured according to the same protocol in the same laboratory at the University of Freiburg. In addition, the strong agreement between the recovery of total soil P and the sum of all individual fractions indicated a high reproducibility of the method. This reproducibility, however, does not provide information about the distribution of Pi and Po within individual fractions. For the individual fractions, it is widely acknowledged that significant potential errors may occur owing to variation in alkalinity or acidity in the extracts leading to acid hydrolysis and precipitation of inorganic P. This can influence the accurate distinction between organic and inorganic forms of P and may thus lead to over- or underestimation of these forms within individual fractions (Magid et al., 1996, Condron et al., 2005, Turner et al., 2005, Tiessen and Moir, 2008, Condron and Newman, 2011). However, we achieved good reproducibility for Pi and Po in repeatedly analyzed samples for $NaHCO_3$ and NaOH fractionation steps. For example, the standard errors of 16 repeated measurements of the same sample were 5.0% for Pi NaOH, 8.3% for Po NaOH, and 7.5% for P_{total} NaOH .

Other than these methodological limitations of the Hedley fractionation method, the level of P content of each fraction is also of importance. The P content in the Pi $NaHCO_3$ and 1M HCl fractions of the soil samples from subtropical China were close to the detection limit. Therefore the reference data of these two fractions could not be regarded as reliable and hence no meaningful calibration models with NIRS could be produced.

Using all Hedley fractions, the description of soil P availability becomes easily confusing. Some fractions could not be calibrated owing to very low P concentrations or only with very low prediction quality. In case of the BEF China data set, grouping the fractions into labile, moderately labile, and stable pools showed a slight increase for the quality parameters R^2 and RPD, but the quality of models remained at the same level. These pools incorporated the inorganic $NaHCO_3$ and 1 M HCl fractions, which could not be predicted individually. As part of the labile respectively moderately labile P pool, a reasonable prediction of the total P content for the BEF China samples could still be possible. While pooling these fractions does not reduce the work involved in reference analysis of the Hedley fractionation, it is still a significant reduction in the modelling process, reducing the number of models from 8 to 3 for each dataset.

4.2 Calibration of organic and inorganic P fractions

Whether P is in organic or inorganic form seemed to be of importance for the quality of predictions of Hedley fractions with NIRS. Regardless of the datasets and sample origin, models predicting the organic P fractions performed better than models predicting the inorganic P fractions (Table 3). Similar results have been obtained for N (Ludwig et al., 2002). The superior quality of models for organic P fractions is related to the underlying mechanisms generating NIR spectra, because organic compounds can be more easily excited by the irradiation than inorganic ones. Spectral variation in relation to soil P content is often associated with organic matter or crystal water (Ben-Dor and Banin, 1995, Chang et al., 2001, Malley et al., 2005, Stenberg et al., 2010). Therefore, we also calibrated NIRS models for the sum of P in all organic fractions (data not shown). However, we could not observe any improvement for prediction of the sum of assigned organic P in comparison with single organic fractions or the pools. This might be caused by the large diversity in organic P compounds (Turner, 2008). In contrast to organic P fractions, the relationships between inorganic P fractions and other NIRS-detectable soil properties were obviously not strong enough to aid the prediction of inorganic P using NIRS.

In our study, we were not able to identify spectral regions to be specific for a P signal as was found in other studies (Malley et al., 2004). Therefore we had also assessed, if focusing on typical NIR spectral regions for C-H, N-H and O-H bonds could influence NIRS model quality. The organic residual which is connected with the phosphate molecule could be dominated by CH, NH, OH bonds or a mixture of them. For this purpose we compared NIRS models based on optimized spectral regions (automated procedure by OPUS software), on the whole spectral range and on specific spectral regions, which are known to represent C-H, N-H and O-H bonds (Conzen, 2005). We found that in all cases, the OPUS-software optimized spectral selection yielded superior models followed by models covering the whole spectral area. Models for selected bonds were in all cases of substantially lower quality, and were thus not presented in detail. The best results based on r^2 and RPD were obtained for O-H bonds for the Po-HCO_3 and P-HCL conc. fractions. This was followed by models focusing on C-H bonds and. The lowest quality models were obtained for models focusing on N-H bonds.

There was no general pattern for the relationship between NIRS model quality (r^2) for different P fractions and the coefficient of determination of relationships between soil C or N and P in these fractions. This indicated that NIRS models were based mostly on original properties of organic P compounds and the predicted P content was not simply a function of soil organic matter content. This corroborates the underlying assumptions of the Hedley fractionation that P in the individual organic fractions is of different availability owing to the different properties of the organic matter dominating the respective fraction. In turn, this indicates that total organic soil C and N can be used only to a limited extent to predict organic P in different fractions.

4.3 Homogeneity of datasets

The quality of predictions of P fractions depended on the homogeneity of individual datasets. The question of homogeneity has also been discussed in the use of global and local calibration models by Abrams et al. (1987) and Shenk & Westerhaus (1993), who found that global models, with a higher number of samples per dataset, were potentially as accurate as more local calibrations for mixtures of hay consisting of different plant species and origins. Variable results were obtained for calibrations of the total C, N and P between global and local models of heterogeneous plant material (Gillon et al., 1999). The global model, which combined all individual datasets of pine needle categories (needles on trees, fallen needles and litter), was more accurate or similar than the “local” individual datasets for each category of needles (Gillon et al., 1999). However, for global models predicting carbon, a decline in model quality was observed. Also, a comparison of NIRS models for soil organic C, inorganic C and clay content showed a decreasing accuracy from local to global calibration models (Sankey et al., 2008). In contrast Brunet et al. (2007) found for local NIRS models higher model quality predicting total C or total N for local models compared with global models. This increasing effect was stronger for ground samples (< 0.2mm) than for sieved samples (<2mm). For our soil data sets, a similar incline in model quality with increasing homogeneity of sample origin could be observed. The general observation that an increase in the range of values leads to an increase in the robustness of NIRS models (Abrams et al., 1987, Shenk and Westerhaus, 1993) did not apply in our study. Our results indicated that regardless of the organic or inorganic origin or the availability of P, the quality of NIRS models to predict P was better for more homogenous than heterogeneous sets of soil samples. For example, reducing the soil samples from the nationwide forest soil monitoring program to the subset BZE “Brown Earths”, we produced a more homogenous subset that improved model calibrations for all but the inorganic NaOH fraction. Although an upgrade in model category (into category D) was only observed for the organic NaOH fraction. The most homogenous BEF China sample set produced even better calibration models, including improvements of model category (into D as well as from D to C), than the BZE subset.

Since P is not “recognized” directly by near-infrared radiation but produces signals in the spectra that are related to P compounds, it appears that the signals for the same group of P compounds that relate to a specific Hedley fraction vary greatly among different soil types. For organic P, these groups of compounds are probably present in more than one fraction and the various extracts are unlikely to be either exhaustive or unique with respect to the target compounds (Turner et al., 2005). To our knowledge, the chemical nature of the organic phosphorus within the operationally defined fractions is still poorly understood (Magid et al., 1996, Turner et al., 2005). In contrast, the inorganic P forms represented in the distinct P fractions are more specific in their chemical nature and well known (Stevenson and Cole, 1999, Tiessen and Moir, 2008). This variation of P forms within Hedley fractions may be greatly reduced through restricting the development of NIRS models to comparable soil types.

5 Conclusions

The approach to predict soil P Hedley fractions of different availability with NIRS can be regarded successful only when certain limitations are considered. NIRS models for prediction of individual inorganic P-fractions were at best moderately useful but unsuitable for soil inventories. We found that the quality of NIRS models for prediction of P, for single fractions as well as for pools, depended on the homogeneity of sample sets. The homogeneity may not only be related to the same soil types (improved models after grouping of soil types in the BZE subset), but also to the same geology that has a large influence on soil properties (better calibration models of the BEF China sample set - with the same geology - than the BZE subset originating from different geologies). The approach may therefore be well suited for large experiments that require a great number of spatial and temporal replication of samples within well-defined soil types. For large-scale, grid based soil inventories such as those conducted in the context of environmental monitoring programs for entire regions, like nation-wide soil surveys, it appears that NIRS-models to predict P fractions need to be developed for individual soil groups or other classes of samples according to specific soil properties. While the development of NIRS models requests a considerable initial investment, these efforts are likely to pay off only for large sample numbers and in repeated inventories.

Acknowledgments

B. Todt was financed by a DFG grant (BEF China, FOR 891, BA 2821/10-1) to Jürgen Bauhus. J. Niederberger was financed through a grant provided by the Federal Ministry of Food, Agriculture and Consumer Protection, administered through the Von-Thünen Institute. We wish to thank K. Winschu for the technical support, the Forest Research Institute Baden-Wuerttemberg and the Northwest German Forest Research Station for providing the BZE samples and C. Geißler and S. Flaiz of the University of Tübingen for preparing and providing a share of the BEF China samples. We greatly appreciate the comments of Dr. Partap K. Khanna and three anonymous reviewers on an earlier version of the manuscript.

References

Abrams, S. M., Shenk, J. S., Westerhaus, M. O., and Barton F. E. II: Determination of forage quality by near infrared reflectance spectroscopy efficacy of broad-based calibration equations, J. Dairy Sci., 70, 806-813, 1987.

Attiwill, P. M. and Adams, M. A.: Nutrient cycling in forests. New Phytol., 124, 561-582, 1993.

Ben-Dor, E. and Banin A.: Near infrared analysis (NIRA) as a method to simultaneously evaluate spectral featureless constitutions in soil, Soil Sci., 159, 259-270, 1995.

BMELV: Handbuch Forstliche Analytik, HFA-Teil A, Boden und Humusuntersuchungen, 2009. http://www.bmelv.de/SharedDocs/Downloads/Landwirtschaft/Wald-Jagd/Bodenzustandserhebung/Handbuch/A-Boden-und-Humusuntersuchungen.pdf?__blob=publicationFile/ , Accessed 1. April 2014

Bruelheide, H., Böhnke, M., Both, S., Fang, T., Assmann, T., Baruffol, M., Bauhus, J., Buscot, F., Chen, X-Y., Ding, B., Durka, W., Erfmeier, A., Fischer, M., Geißler, C., Guo, D., Guo, L-D., Härdtle, W., He, J-S., Hector, A., Kröber, W., Kühn, P., Lang, A., Nadrowski, K., Pei, K., Scherer-Lorenzen, M., Shi, X., Scholten, T., Schuldt, A., Trogisch, S., von Oheimb, G., Welk, E., Wirth, C., Wu, Y-T., Yang, X., Zeng, X., Zhang, S., Zhou, H., Ma, K. and Schmid, B.: Community assembly during secondary forest succession in a Chinese subtropical forest, Ecol. Monogr., 81, 25–41, 2011.

BRUKER OPTIK GmbH: User Manual Quant, Opus Spectroscopy Software, Version 6, Ettlingen, 2006.

Brunet, D., Barthès, B. G., Chotte, J, L, and Feller, Ch.: Determination of carbon and nitrogen contents in Alfisols, Oxisols and Ultisols from Africa and Brazil using NIRS analysis: Effects of sample grinding and set heterogeneity, Geoderma, 139, 106-117, 2007.

Burns, D. A. and Ciurczak, E. W. (eds): Handbook of Near-Infrared Analysis, Second Edition, in: Practical Spectroscopy a Series, Vol 27, Marcel Dekker Inc New York, 2001

Cécillon, L., Barthès, B. G., Gomez, C., Ertlen, D., Genot, V., Hedde, M., Stevens, A. and Brun, J. J.: Assessment and monitoring of soil quality using near-infrared reflectance spectroscopy (NIRS), Eur. J. Soil Sci., 60(5), 770–784, 2009.

Chang, C.-W., Laird, D. A., Mausbach, M. J and Hurburgh, C. R.: Near-Infrared Reflectance Spectroscopy - Principal Components Regression Analyses of Soil Properties Journal Paper no. J-18766 of the Iowa Agric. and Home Econ. Exp. Stn., Ames, IA. Soil. Sci. Soc. Am. J. 65, 480-490, 2001.

Compton, J. E. and Cole, D. W.: Phosphorus cycling and soil P fractions in Douglas-fir and red alder stands, For. Ecol. Manag., 110, 101-112, 1998.

Condron, L. M., Turner, B. L. and Cade-Menum, B. J.: Chemistry and dynamics of soil organic phosphorus. In: Sims, T. and Sharpley, A. N. (eds): Phosphorus: Agriculture and the Environment, Am. Soc. Agron., Madison, Wisconsin, USA, 87-121, 2005.

Condron, L. M. and Newman, S.: Revisiting the fundamentals of phosphorous fractionation of sediments and soil, J. Soil Sediment, 11, 830-840, 2011.

Conzen, H.-P.: Multivariante Kalibration, Ein praktischer Leitfaden zur Methodenentwicklung in der analytischen Analytik, Bruker Optik GmbH, 2005.

Cordell, D., Drangert, J.-O. and, White, S.: The story of phosphorus: Global food security and food for thought, Glob. Environ. Chang., 19, 292-305, 2009.

Cross, F. and Schlesinger, W. H.: A literature review and evaluation of the Hedley fractionation: Applications to the biogeochemical cycle of phosphorous in natural ecosystems, Geoderma, 64, 197-214, 1995.

De Schrijver, A., Vesterdal, L., Hansen, K., De Frenne, P., Augusto, L., Achat, D. L., Staelens, J., Baeten, L., De Keersmaeker, L., De Neve, S. and Verheyen, K.: Four decades of post-agricultural forest development have caused major redistributions of soil phosphorus fractions, Oecologia, 167, 221-234, 2012.

Dhawale, N. M., Adamchuck, V. I., Viscarra Rossel, R. A., Prascher, S. O., Whalen, J. K. and Ismail, A. A.: Predicting extractable soil phosphorus using visible / near-infrared hyperspectral soil reflectance measurements, The Canadian Society for Bioengineering 2013 Annual Meeting, Saskatoon, Saskatchewan, 7-10 July 2013, Paper No. CSBE13-047.

Drummond, L. and Maher, W.: Determination of phosphorus in aqueous solution via formation of the phosphoantimonylmolybdenum blue complex. Re-examination of optimum conditions for the analysis of phosphate, Anal. Chim. Acta., 302(1), 69-74, 1995.

Duckworth, J.: Mathematical Data Preprocessing, in: Roberts, C. A., Workmann, J. J. and Reeves J. B. III (eds): Near-Infrared Spectroscopy in Agriculture, in: Agronomy Series, Vol 44, Am. Soc. Agron. Inc, Crop. Sci. Soc. Am. Inc, Soil Sci. Soc. Am. Inc, Madison: 729 – 784, 2004.

Edixhoven, J. D., Gupta, J., and Savenije, H. H. G.: Recent revisions of phosphate rock reserves and resources: reassuring or misleading? An in-depth literature review of global estimates of phosphate rock reserves and resources, Earth Syst. Dynam. Discuss., 4, 1005-1034, doi:10.5194/esdd-4-1005-2013, 2013.

Elser, J. J., Bracken, M. E. S., Cleland, E. E., Gruner, D. S., Harpole, W. S., Hillebrand, H., Ngai, J. T., Seabloom, E. W., Shurin, J. B. and Smith, J. E.: Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems, Ecol. Lett., 10, 1135-1142, 2007.

Fox, T. R., Miller, B. W., Rubilar, R., Stape, J. L. and Albaugh, T. J.: Phosphorus Nutrition of Forest Plantations: The Role of Inorganic and Organic Phosphorus, in: Soil Biology 26, Chapter 13, 317-338, Springer-Verlag, Berlin/Heidelberg, Germany, 2011.

Genot, V., Colinet, G., Bock, L., Vanvyve, D., Reusen, Y. and Dardenne, P.: Near infrared reflectance spectroscopy for estimating soil characteristics valuable in the diagnosis of soil fertility, J. Near Infrared Spectrosc., 119, 117-138, 2011.

Gillespie, A. R. and Pope, P. E.: Consequences of rhizosphere acidification on delivery and uptake kinetics of soil phosphorus, Tree Physiol., 8, 195-204, 1991.

Gillon, D., Houssard, C. and Joffre R.: Using near-infrared reflectance spectroscopy to predict carbon, nitrogen and phosphorus content in heterogeneous plant material, Oecologia, 118, 173-182, 1999.

Gruselle, M.-C. and Bauhus, J.: Assessment of the species composition of forest floor horizons in mixed spruce-beech stands by Near Infrared Reflectance Spectroscopy (NIRS). Soil Biol. Biochem., 42, 1347-1354, 2010.

Günzler, H. and Gremlich H.-U.: IR Spektroskopie eine Einführung, Wiley-VCH Verlag, Weinheim, 2003

Hättenschwiler, S., Aeschlimann, B., Coûteaux, M.-M., Roy, J. and, Bonal, D.: High variation in foliage and leaf litter chemistry among 45 tree species of a neotropical rainforest community, New Phytol., 179, 165-175, 2008.

Hedley, M. J., Stewart, J. W. B. and Chauhan, B. S.: Changes in Inorganic and Organic Soil Phosphorus Fractions Induced by Cultivation Practices and by Laboratory Incubations, Soil Sci. Soc. Am. J., 46, 970-976, 1982a.

Hedley, M. J., White, R. E. and Nye, P. H.: Plant-Induced changes in the rhizosphere of rape (Brassica Napus Var. Emerald) seedlings, III. Changes in L value, soil phosphate fractions and phosphatase activity, New Phytol., 91, 45-56, 1982b.

Huang, X. L. and Zhang, J. Z.: Rate of phosphoantimonylmolybdenum blue complex formation in acidic persulfate digested sample matrix for total dissolved phosphorus determination: Importance of post-digestion pH adjustment, Talanta, 77(1), 340-345, 2008.

John, M. K.: Colorimetric Determination of Phosphorus in Soil and Plant Materials with Ascorbic Acid, Soil Sci., 109, 214-220, 1970.

Khanna, P. K., Bauhus, J., Meiwes, K. J., Kohler, M., Rumpf, S. and Schönfelder, E.: Assessment of changes in the Phosphorus status of forest ecosystems in Germany – Literature review and analysis of existing data. Report to the German Federal Ministry of Food, Agriculture and Consumer Protection, 2007.

Lei, P. and Bauhus, J.: Use of near-infrared reflectance spectroscopy to predict species composition in tree fine-root mixtures, Plant Soil, 333, 93-103, 2010.

Lorenz, M., Mues, V., Becher, G., Müller-Edzards, C., Luyssaert, S., Raitio, H., Fürst, A. and Langouche, D.: Forest Condition in Europe - Results of the 2002 Large-scale Survey. UNECE and EC, Geneva and Brussels, 2003.

Ludwig, B., Khanna, P. K., Bauhus, J. and Hopmans, P.: Near infrared spectroscopy of forest soils to determine chemical and biological properties related to soil sustainability. For. Ecol. Manag., 171, 121-132, 2002.

Magid, J., Tiessen, H. and Condron, L. M.: Dynamics of organic phosphorus in soils under natural and agricultural ecosystems, in: Piccolo, A. (ed): Humic Substances in Terrestrial Ecosystems, Elsevier Sci., Oxford, 429-466, 1996.

Maleki, M. R., Van Holm, L., Ramon, H., Merckx, R., De Baerdemaeker, J. and Mouazen, A. M.: Phosphorus Sensing for Fresh Soils using Visible and Near Infrared Spectroscopy, Biosystems Engineering, 95, 425-436, 2006.

Malley, D. F., Martin, P. D. and Ben-Dor, E.: Application in Analysis of Soils, in: Roberts, C. A., Workmann, J. J. and Reeves, J. B. III (eds): Near-Infrared Spectroscopy in Agriculture, in: Agronomy Series, Vol 44, Am. Soc. Agron. Inc, Crop Sci. Soc. Am. Inc, Soil Sci. Soc. Am. Inc, Madison, 729 – 784, 2004.

Miller, C. E.: Chemical principles of near infrared technology, in Williams, P. and Norris, K.: Near-infrared technology in the agricultural and food industries, Am Association Cereal Chemist. Inc, 19-37, 2001.

Mohren, G. M. J., Van den Burg, J. and Burger, F. W.: Phosphorus deficiency induced by nitrogen input in Douglas fir in the Netherlands, Plant Soil, 95, 191-200, 1986.

Murphy, J. and Riley, J. P.: A modified single solution method for the determination of phosphate in natural waters, Anal. Chim. Acta., 27, 31-36, 1962.

Paré, D. and Bernier, B.: Origin of the phosphorus deficiency observed in declining sugar maple stands in the Quebec Appalachians, Can. J. For. Res., 19, 24-34, 1989.

Rennenberg, H. and Herschbach, C.: Phosphorus nutrition of woody plants: many questions – few answers, Plant Biol., 15, 785–788, 2013.

Richter, D., Allen, H. L., Li, J., Markewitz, D. and Raikes, J.: Bioavailability of slowly cycling soil phosphorus: major restructuring of soil P fractions over four decades in an aggrading forest, Oecologia, 150, 259-271, 2006.

Roberts, C. A., Workmann, J. J. and Reeves, J. B. III (eds): Near-Infrared Spectroscopy in Agriculture, in: Agronomy Series, Vol 44, Am. Soc. Agron. Inc, Crop Sci. Soc. Am. Inc, Soil Sci. Soc. Am. Inc, Madison, 729–784, 2004.

Rowland, A. P. and Haygarth, P. M.: Determination of total dissolved phosphorus in soil solutions, J. Environ. Qual., 26(2), 410-415, 1997.

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

Schmidt, W.: Optische Spektroskopie eine Einführung, Wiley-VCH Verlag, Weinheim, 2000.

Shen, J., Yuan, L., Zhang, J., Li, H., Bai, Z., Chen, X., Zhang, W. and Zhang, F.: Phosphorus Dynamics: From Soil to Plant, Plant Physiol., 156, 997-1005, 2011.

Shenk, J. S. and Westerhaus, M. O.: Near infrared reflectance analysis with single- and multiproduct calibrations, Crop Sci., 33, 582-584, 1993.

Stenberg, B., Viscarra-Rossel, R. A., Mouazen, A. M. and Wetterlind, J.: Visible and Near Infrared Spectorscopy in Soil Science, in: Sparks, D. L. (ed): Adv. Agron., 107, Burlington: Academica Press, 163-215, 2010.

Stevenson, F. and Cole, M.: Phosphorus, in: Cycles of soil: Carbon, nitrogen, phosphorus, sulfur, and micronutrients, John Wiley & Sons, Inc., New York, 2nd edition, 279-329, 1999.

Stoner, E. R. and Baumgardner, M. F.: Characteristic variations in reflectance of surface soils, Soil Sci. Soc. Am. J., 45, 1161-1165, 1981.

Tiessen, H. and Moir, J. O.: Characterization of available P by sequential extraction, in: Carter, M. R. and Gregorich, E. G. (eds): Soil sampling and methods of analysis, chapter 25, 293-306, Can. Soc. Soil Sci., Boca Raton Fl, USA, 2nd edition, 2008.

Turner, B. L., Mahieu, N. and Condron, L. M.: The phosphorus composition of temperate grassland soils determined by NAOH-EDTA extraction and solution 31P NMR spectroscopy, Org. Geochem., 34, 1199-1210, 2003.

Turner, B. L., Cade-Menun, B. J., Condron, L. M. and Newman, S.: Extraction of soil organic phosphorus, Talanta, 66, 294-306, 2005.

Turner, B. L., Newman, S. and Reddy, K. R.: Overestimation of organic phosphorus in wetland soils by alkaline extraction and molybdate colorimetry, Environ. Sci. Technol., 40, 3349-3354, 2006.

Turner, B. L. Resource partitioning for soil phosphorus: a hypothesis, J. Ecol., 96, 698-702, 2008.

Udelhofer, T., Emmerling, Ch. and Jarmer, T.: Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial least-square regression: feasibility study, Plant Soil, 251, 319-329, 2003.

Vanguelova, E., Pitman, R., Luiro, J. and Helmisaari, H.-S.: Long term effects of whole tree harvesting on soil carbon and nutrient sustainability in the UK, Biogeochem., 101, 43-59, 2010.

Vitousek, P. M., Porder, S., Houlton, B. Z. and Chadwick, O. A.: Terrestrial phosphorus limitation: mechanisms, implications, and nitrogen-phosphorus interactions, Ecol. Appl., 20, 5-15, 2010.

Walker, T. W. and Syers, J. K.: The fate of phosphorus during pedogenesis, Geoderma, 15: 1-19, 1976.

Wardle, D.: Aboveground and belowground consequences of long-term forest retrogression in the timeframe of millennia and beyond, in: Wirth, C., Gleixner, G. and Heimann, M. (eds): Old-growth forests: Function, fate and value, Ecol. Stud., 207, 193-209, Springer Verlag Berlin Heidelberg, 2009.

Weyer, L. G.: Near-Infrared Spectroscopy of Organic Substances. Appl. Spectrosc. Rev., 21(1), 1-43, 1985.

Williams, P. C.: Implementation of near-infrared technology. in: Williams, P. C. and Norris, K. H. (eds): Near-Infrared Technology in the Agricultural and Food Industries, Am. Assoc. Cereal Chemist, St. Paul, Minnesota, USA, 145–169, 2001.

Workman, J. J.: NIR Spectroscopy Calibration Basis, in: Burns, D. A. and Ciurczak, E. W. (eds): Handbook of Near-Infrared Analysis, Second Edition, in: Practical Spectroscopy a Series, Vol 27, Marcel Dekker Inc New York, 2001

World Agroforestry Centre (ICRAF) and ISRIC-World Soil Information: ICRAF–ISRIC Soil vis–NIR spectral Library, World Agroforestry Centre (ICRAF), Nairobi, Kenya, 2010.

WRB: IUSS Working Group World Reference Base for Soil Resources 2006, World Soil Resources Report No. 103, FAO, Rome, 2006.

Yang, X. and Post, W. M.: Phosphorus transformations as a function of pedogenesis: A synthesis of soil phosphorus data using Hedley fractionation method, Biogeosciences, 8, 2907-2916, 2011.

Zornoza, R., Guerrero, C., Mataix-Solera, J., Scow, K. M., Arcenegui, V. and Mataix-Beneyto, J.: Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils, Soil Biol. Biochem., 40, 1923-1930, 2008.

Table 1. Descriptive statistical parameters of soil P content (mg kg^{-1}) as total sum of Hedley P fractions and in pools of different plant availability, grouped by data sets used for NIRS modelling.

Data Set	P-Pool	n	Min	Mean	Max	SD	Skewness	Kurtosis
BZE all	total sum	282	57.8	564.0	2796.1	413.4	1.9	6.2
	labile		6.0	123.6	850.0	110.7	2.4	8.9
	moderately labile		8.1	296.7	2307.7	266.2	3.2	18.5
	stable		6.7	141.1	805.2	127.4	1.	4.1
BZE “Brown Earths”	total sum	84	128.1	657.7	2178.0	440.59	1.2	1.7
	labile		11.0	166.3	850.0	146.89	2.1	5.6
	moderately labile		46.9	357.3	1227.0	256.9	1.2	1.4
	stable		6.7	134.3	534.3	118.4	1.4	1.9
BEF China	total sum	294	25.7	94.8	328.6	47.9	1.7	4.5
	labile		1.2	17.0	78.9	10.6	1.8	6.5
	moderately labile		10.8	43.7	161.2	21.8	1.7	5.6
	stable		10.8	34.1	174.8	23.1	3.0	11.8

Table 2. Descriptive statistical parameters of carbon (g kg^{-1}), nitrogen (g kg^{-1}) and pH (H_2O) in soil, grouped by data sets used for NIRS modelling.

		n	Min	Mean	Max	SD	Skewness	Kurtosis
BZE	**carbon**	282	0.2	42.9	314.7	39.7	2.0	7.6
	nitrogen		0.1	2.7	14.2	2.3	1.6	3.3
	pH		3.2	4.5	8.2	0.8	1.9	4.9
BZE „Brown Earths“	**carbon**	84	3.6	56.2	178.0	45.2	0.8	-0.3
	nitrogen		0.2	3.4	10.0	2.5	0.8	-0.2
	pH		3.4	4.4	7.8	0.7	2.7	11.3
BEF-China	**carbon**	294	3.4	30.9	118.1	17.2	1.4	3.0
	nitrogen		0.3	1.6	5.9	0.8	1.5	4.5
	pH		3.5	4.5	5.1	0.2	-0.7	1.7

Table 3. Quality parameters for NIRS model calibration (R^2, RPD) for all Hedley fractions in the three data sets: BZE, BZE “Brown Earths”, and BEF-China.

Hedley Fraction	**BZE**		**BZE „Brown Earths“**		**BEF-China**	
	R^2	RPD	R^2	RPD	R^2	RPD
Pi Resin	0.26	1.16	0.51	1.44	0.77	2.08^d
Pi $NaHCO_3$	0.08	1.04	0.19	1.10	0.48	1.39
Po $NaHCO_3$	0.61	1.60	0.63	1.64	0.78	2.13^d
Pi NaOH	0.23	1.14	0.15	1.09	0.72	1.88^d
Po NaOH	0.61	1.60	0.75	1.98^d	0.82	2.35^c
Pi 1M HCl	0.30	1.19	0.38	1.26	0.34	1.23
P HCl conc.	0.72	1.88^d	0.76	2.05^d	0.85	2.57^c
P Residual	0.56	1.50	0.57	1.52	0.69	1.79^d

^{d}) model quality level D; ^{c}) model quality level C;

Figure 1. Sequential extraction scheme of Hedley fractionation modified by Tiessen and Moir, grouped by pools of availability; $NaHCO_3$, NaOH and HCl conc. steps provide both organic and inorganic compounds for this fraction step.

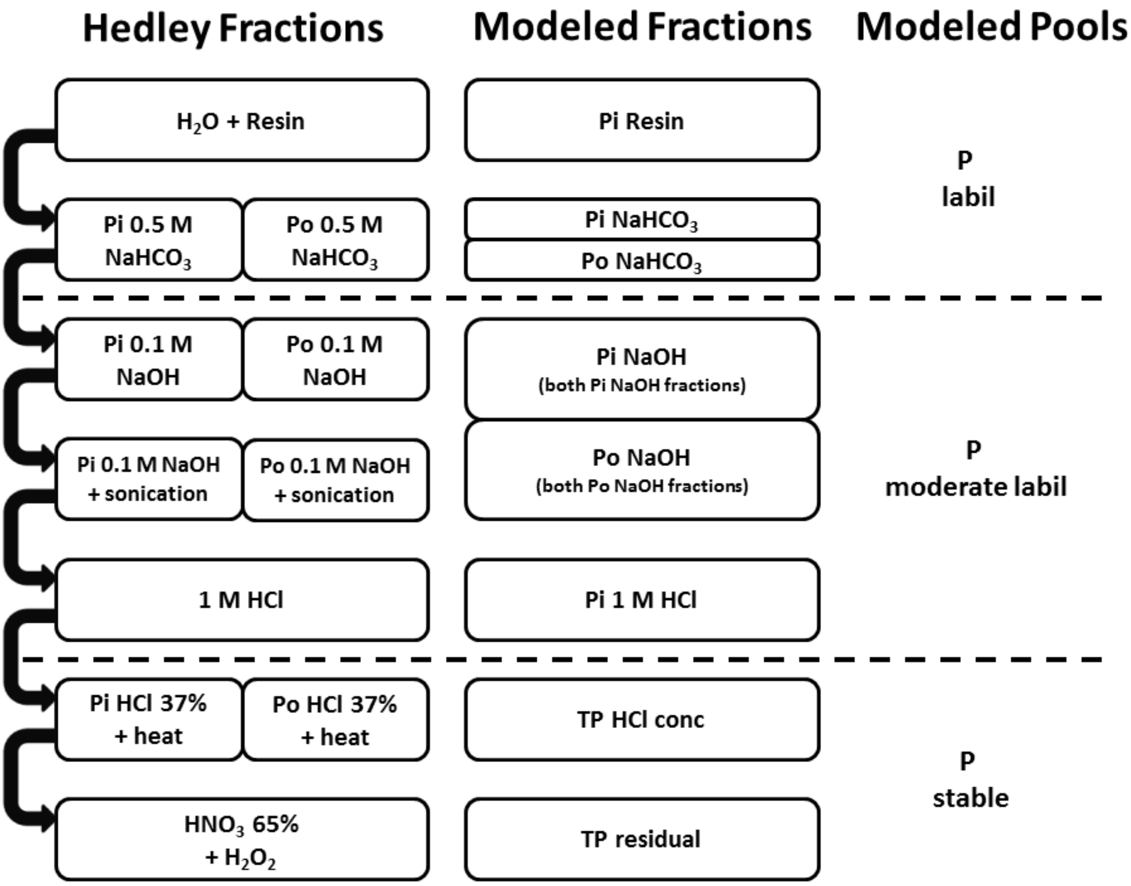

Figure 2. Relationship between the sum of recovered Hedley P fractions and independently determined total P (mg kg^{-1}) in soil samples.

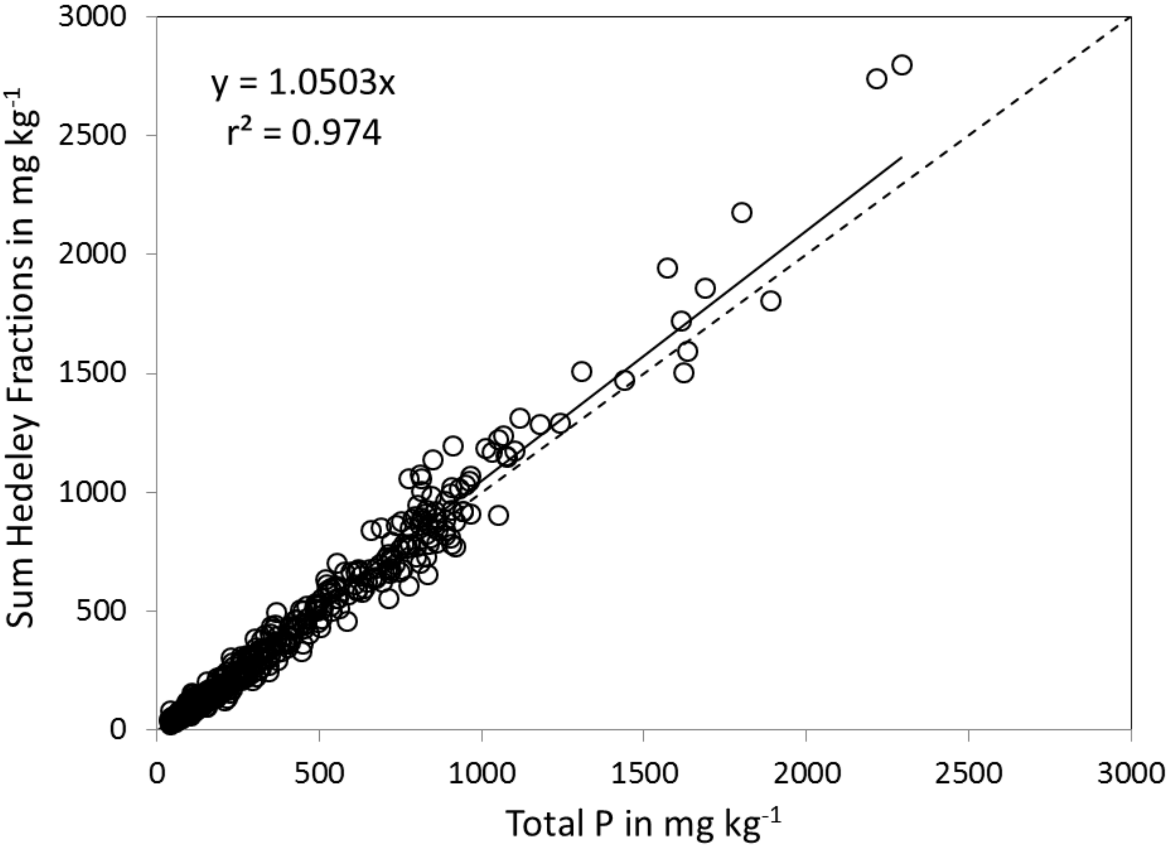

Figure 3. Score plot of principal component analysis (total of five components derived; principal component 1 vs. principal component 2) for the three datasets BZE (white, set 2), BZE "Brown Earths" (grey, set 3) and BEF China (black, set 4).

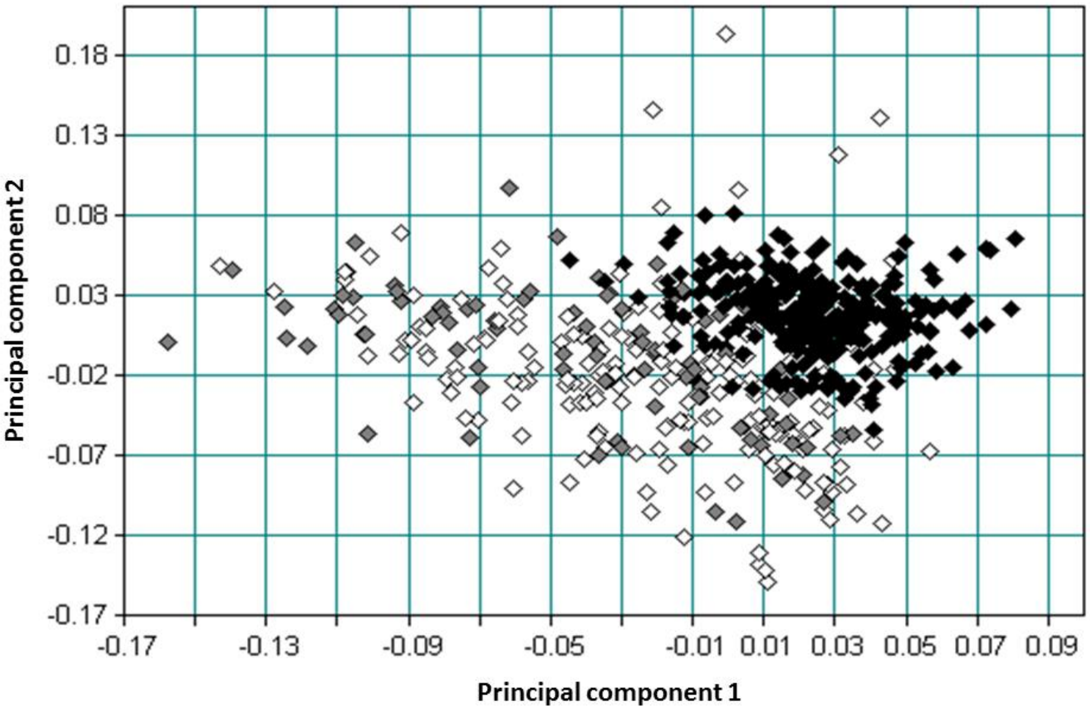

Figure 4. Validation for P HCl conc. fraction of the three datasets BZE, BZE “Brown Earths” and BEF China (mg kg^{-1}), x-axis measured, y-axis predicted values, please note the difference in scale among graphs.

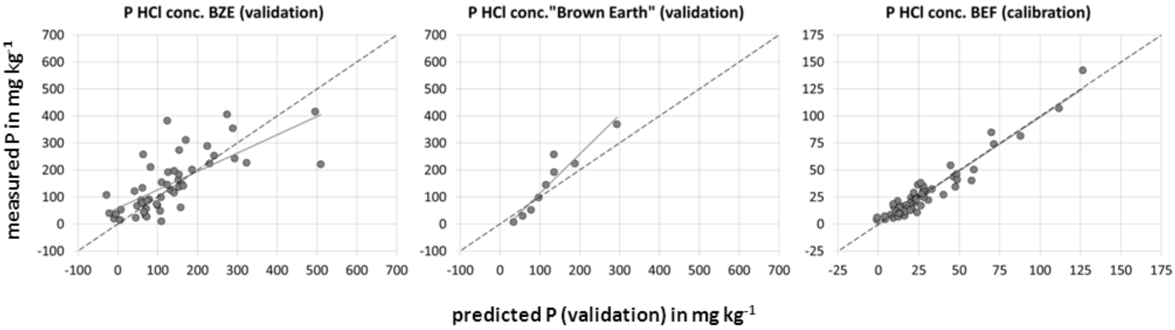

Figure 5. Calibration models for the four datasets BEF China+BZE, BZE, BZE "Brown Earths" and BEF China, pooled for labile, moderately labile (mod. labile) and stable phosphorus (mg kg^{-1}), x-axis measured, y-axis predicted values, please note the difference in scales.

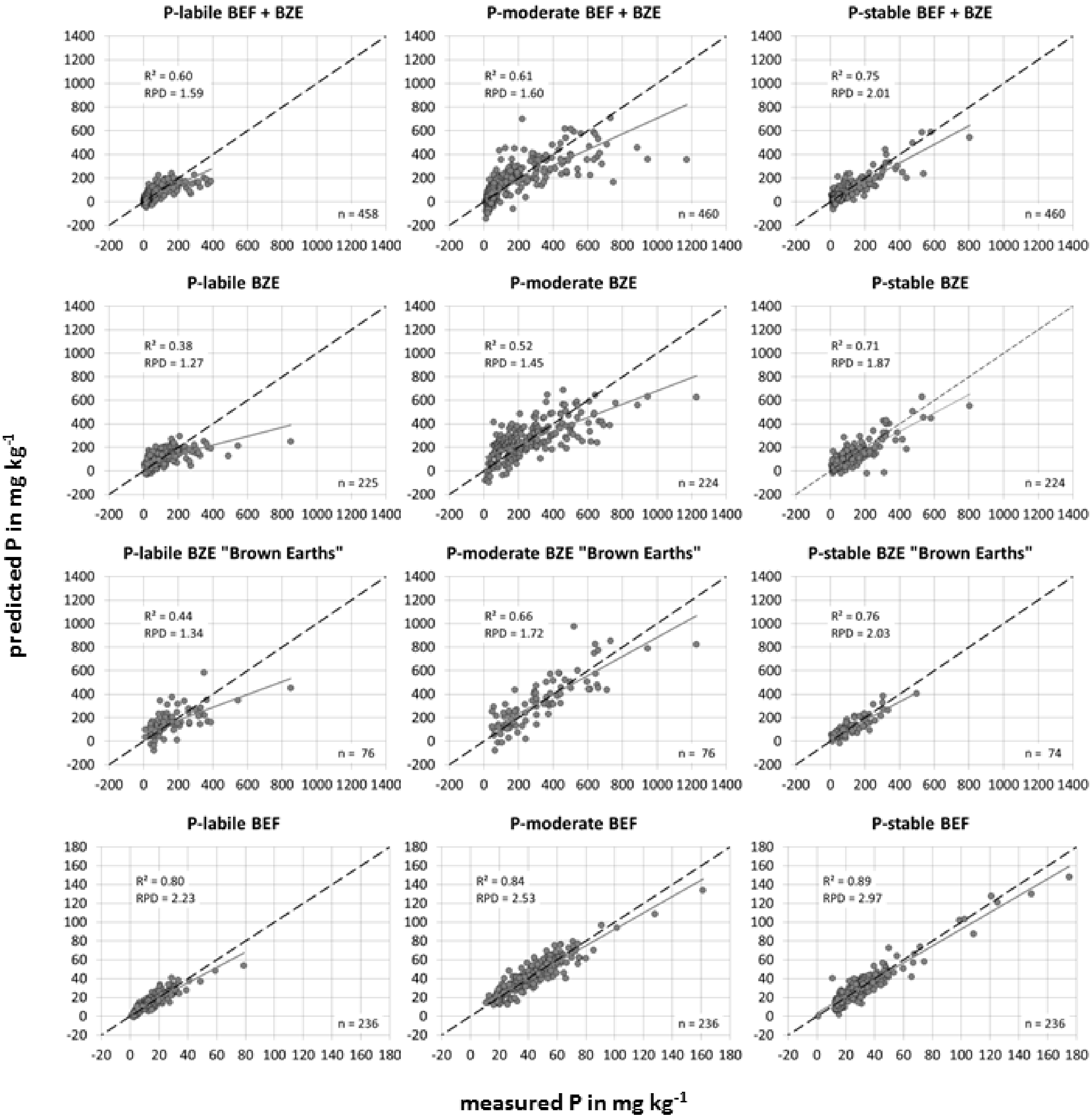

Figure 6. Validation for BEF China calibration models with external datasets pooled for labile moderately labile (mod. labile) and stable phosphorus (mg kg^{-1}), x-axis measured, y-axis predicted values.

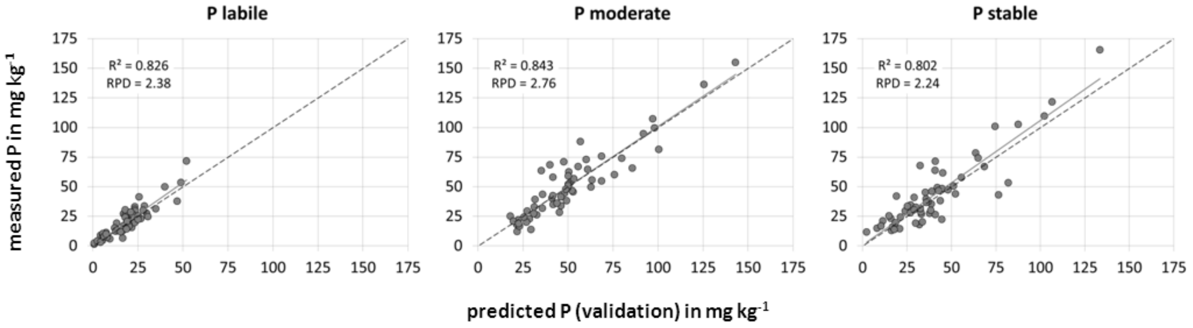

Figure 7. Goodness of fit of calibration models (r^2) for all individual P fractions and coefficient of determination (r^2) of regression analysis of total organic carbon **(a)**, total nitrogen **(b)** and the P fractions (circles=organic fractions; squares=inorganic fractions; triangles= HCl_{conc} fraction) in soils for all three datasets (black=BEF China; grey=BZE "Brown Earths"; white=BZE), dashed line indicates the minimum model quality of r^2 = 0.7.