General comments

The paper evaluates the use of NIRS in forest soil phosphorus research. NIRS would make soil P research more cost and time efficient. Up to now, NIRS has not been used to quantify Hedley P fractions in forest soils. Hence, the paper presents a novel and potentially useful application of NIRS. The title reflects the contents of the paper. The authors have to conclude that only some of the Hedley fractions could be quantified by NIRS and that datasets used for NIRS calibration have to fulfill particular prerequisites (e.g., homogeneity of datasets). However, the description of these prerequisites of datasets is confusing and should be more precise.

The methods and assumptions are largely valid, but are not clearly outlined. For example, the selection criteria for the soil sample subsets are not comprehensible. In addition, the description of the NIRS method is too rough. Therefore, reproduction by fellow scientists would not be possible.

The results are sufficient to support the interpretations and conclusions, but the phrasing is partly misleading. The authors give proper credit to related work and clearly indicate their own contribution. However, they should add that NIRS is frequently applied in agricultural soil P research to quantify plant-available P.

The overall presentation is well structured, but could be clearer; especially the language could be more precise. Some sentences are nested and hard to understand. The number of references is high (approx. 70 references) and could be reduced. However, some few references concerning the use of NIRS in agricultural soil P research could be added.

Specific comments

<u>Abstract</u>
Page 1
L 20/21 There are different modifications of the Hedley method. Therefore, the particular fractions should be named in the abstract.
L 26 what is meant with "homogeneity of soil sample sets"? -> explain
L 27 what is meant with "useful models"? -> explain
Page 2
L 4 how similar do they have to be? In which respect similar? What are the most important properties that have to be similar? -> explain in more detail

<u>Introduction</u>
L 17 describe the hypotheses shortly
Page 2 L 28 - Page 4 L 4 this paragraph is too long and should be subdivided, e.g., 1. Role of P fractions in tree nutrition 2. Usefulness of NIRS
Page 3
L 9/10 to which part of the sentence does the phrase "particularly in forest soils" refer?
L 21 total C and N contents or which fractions?
L 26 NIRS is usually applied to dried and ground samples. Thus, the different liquid and gas status of soils should be of minor importance.
Page 4
L 2 describe the "other soil properties being detectable by NIRS"
L 4 describe what "high quality in spectral datasets" means in the context of NIRS (e.g., homogeneity of soil samples (ground vs. sieved); homogeneity of the sample sets (one soil type vs. different soil types); origin of the sample sets (regional vs. global); homogeneity of the soil sample composition (mineral soil samples with low soil organic matter content vs. mineral soil samples with various contents of soil organic matter), …)
L 4-25 rephrase this paragraph

L 16 “prediction of C content and sample sets” -> is “and” the right word here? If yes, I do not understand the meaning of the sentence.
L 17 isn’t high variation in chemical composition a cause of high spectral variation? Then “or” wouldn’t be suitable here.
L 26-28 there are several studies on NIRS models for different P forms (e.g. microbial P); in agricultural soil P research NIRS is used to quantify different P fractions

Material and methods
Page 5
L 14/15 did you select a subset of the BZE dataset?
L 25-27 Explain your selection criteria. If there were “clear correlations” between total P and P fractions, how could that help you to create subsets?

Page 6
L 22 and 25 volume to volume ratio or volume to mass ratio? Better write 2.5 ml : 1 ml or 2.5 ml : 1 g

Page 7
L 5-10 this paragraph fits better to the introduction
L 11 here, you do not write that you used replicate soil samples, but later you write something about replicate soil samples
L 12 please add type of resin (counterion)
L 14 please add energy level of ultrasonic treatment
L 23 write “PO_4-P” or “molybdate reactive P”; what kind of photometer did you use (continuous flow, microplate reader, …)? At what wavelength did you measure?

Page 8
L 3 rephrase; it is not clear from this sentence whether you summed up Pi and Po of the NaOH and S-NaOH fractions or if you summed up 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
L 7/8 Although all acids can act as oxidants, persulfate is by far a stronger oxidant than HCl as it is a source of sulfate radicals. HCl is used for hydrolytic degradation of organic matter, whereas persulfate is a “true” oxidizing agent. Yet, for the degradation of organic P compounds, both treatments might be equally efficient. Please correct your statement and check the literature if others also found no organic P in conc. HCl-extracts. If it is true that Po in 1 M HCl extracts is negligible, why did you measure TP in HCl conc. extracts?
L 9 what is meant with “satisfying”?

Page 9
L 3-20 In part, this has already been mentioned in the introduction, some general remarks may be shortened.
L 6/7 O-H, C-H and N-H are bounds and not functional groups
L 11 NIRS detectable soil properties -> describe them
L 27 why did you not test the second or third derivative? According to Barnes et al. (1989) spectra should be detrended to remove scatter effects. Please consider this.
Barnes et al. (1989) Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Applied Spectroscopy 43
L 28 rewrite this sentence; you did not do these treatments for the PLS
L 29-31 please give more details (size of gaps, amount of smoothing)
L 31 cross validation is used to avoid overfitting and to obtain the optimal number of terms in the calibration; why is it a common approach to replace the calibration step by cross validation for small data sets? References?

Page 10

L 4-10 The criteria for this automated selection do not get clear from this.
L 11-17 move this section to 2.1 Soil samples; did you consider to group samples by parent material?
L 20 I didn't understand the sentence before I saw the results; you do not mean the relationship between P and soil C and N but the quality of the relationship
L 26-30 move this section to 2.1 Soil samples and give the number of samples in each sample set

Page 11
L 9 please correct: RDP=ratio of SD to standard error of prediction

Results
Page 12
L 14ff You should always write cross-validation instead of calibration
L 16 do you mean worse than level D when you write "produced no useful calibrations"?

Page 13
L 11-19 rephrase this paragraph, it is really hard to understand (e.g., 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).)
L 19 useful -> best?
L 26 do you mean the levels defined on p 11 with "goodness of fit of calibration models"; in Fig. 7 you use the R^2 of the calibration model
L 28 „…were best for the Po fractions…" I couldn't find any good relationship in Fig. 7
L 26-7 (page 14) It makes no sense to correlate a R^2 and a Spearman rank correlation coefficient (r_s). r_s is a non-parametric measure which may not simply be related to a parametric measure like R^2. If your only rationale behind this approach is to test whether NIRS models for P fractions are a result of C-P or N-P relationships, why don't you simply test if your NIRS models for P fractions have a similar predictive power for C and N as for P fractions?

Page 14
L 7 the given correlation coefficients are not for the dataset presented in Fig. 7, but for a dataset with some fractions removed, right?

Discussion
Page 14
L 12 you might use the data of your "random quality check" to calculate coefficients of variation for individual fractions
L 13 reference method = Hedley fractionation?
L 25/L 26 "repeatedly analyzed" and "random quality check" -> describe in the material and methods section how many replications you did; did you repeat the analysis or the fractionation?
L 30/31 the reason for the bad NIRS models might also include other factors
L 31 what is meant with "valid"?

Page 15
L 13 what is "a reasonable prediction"?
L 29 "total organic P" is an inadequate term for the sum of Hedley organic P fractions, since not all organic P is extracted during the Hedley procedure

Page 16
L 8/9 rephrase
L 13-16 Explain why global models are potentially as accurate as more local calibrations.
L 12-23 This paragraph is a bit confusing, since you compare studies dealing with organic material with studies dealing with soil. Due to numerous reasons (which you partly mentioned) soils are more

complex than organic material and to create "global models" for soils is potentially less successful. Please rather refer to studies dealing with soils. For instance, Brunet et al. (2007) also found better predictions for total C when using subsets of soils compared to a "global model".

Page 17
L 8-10 Evidence on these questions is limited, but there are for instance combined Hedley fractions/31P NMR studies dealing with these questions. See Negassa and Leinweber (2009) JPNSS 172:305-325
L 11/12 even in soils of comparable soil type the variation in P forms within Hedley fractions may be high due to other reasons like tree species -> differing litter quality, climate -> soil humidity -> soil microorganisms
L 12 the development of NIRS models for specific subgroups of soils is probably more promising, but why only create subgroups according to soil types and not parent material?
L 12-14 rewrite the sentence "The possible … individual dataset."

Tables and figures
Table 1 and 2 Dataset "all" is missing
Figure 1 Check the presentation of the modelled NaOH fractions? What did you combine?

Technical corrections

Page 1
L 1 Near-Infrared -> near-infrared
L 1 Phosphorus -> phosphorus
L 15 P -> phosphorus (P)
L 20 Hedley method -> Hedley sequential extraction method

Page 2
L 10 Phosphorus -> Phosphorus (P)
L 16 phosphorus-limitations -> phosphorus limitation
L 24 P-nutrition -> P nutrition
L 25 monitoring of the -> monitoring the
L 28 rephrase "*solely* total P contents are *often* measured"

Page 3
L 4 cite the papers of Hedley
L 8 have been -> has been
L 10/11 Here, in contrast to agricultural soils, the slowly cycling P pool contributes -> In contrast to agricultural soils, the slowly cycling P pool in forest soils contributes
L 14 Hedley-fractionation -> Hedley fractionation
L 16 Hedley-P fractions -> Hedley P fractions
L 18 start new paragraph after "may be a promising approach."
L 21 C or N -> carbon (C) or nitrogen (N)
L 24 bracket in bracket…
L 26 gas -> gases
L 29 of to the USDA -> of the USDA
L 31 "P or" can be deleted
L 32 find a more suitable word than "subsequently" (e.g., hence)

Page 4
L 23/24 couldn't "depending on the homogeneity respectively heterogeneity of" be replaced by "for"; would make the sentence shorter and easier to understand
L 30 "to do so" -> could the sentences be rephrased so that "to do so" can be replaced?

Page 5
L 9-13 change the order of the two sentences “From each site…” and “Including 70 sites…”
L 22 delete “aimed to” and change “select” to “selected”
L 24 add a reference

Page 6
L 1 Research -> research
L 8 5-10cm -> 5-10 cm
L 14 delete “and”
L 15 pH-Values -> pH values
L 16 North Western German Forest Research Institute -> Northwest German Forest Research Station
L 17 rephrase “data was measured according to the Handbuch Forstliche Analytik”
L 18 carbon and nitrogen -> C and N
L 19 1150 °C ?
L 21 2x carbon -> C
L 22/25 rephrase “water solution”
L 25 derivedin -> derived in

Page 7
L 6 analysis –> analyses
L 20 with the -> after
L 23 Phosphorous -> Phosphorus
L 27 dot is missing
L 29 remove the different “-“
L 29 organically bound -> total
L 30 autoclave and the -> autoclave. The
L 31 P(Po) -> P (Po)

Page 8
L 19 delete the dot
L 22 Hedley Fractionation Method -> Hedley fractionation method

Page 9
L 3 exited -> excited
L 9 Phosphates and other P compounds -> Phosphates and other inorganic P compounds
L 14 either replace the comma by a dot or fill in “but” or “instead”

Page 10
L 31 set3 -> set 3

Page 11
L 6 software) -> software)
L 28 Phosphorus concentrations -> phosphorus contents
L 29 P concentrations -> P contents
L 31 P concentration -> P content

Page 12
L 8 concentrations -> contents
L 11 within -> below?
L 11 Hedley method -> Murphy & Riley (1962) method?
L 13 3.2.1 -> 3.2
L 13 NIRS models by P fractions -> NIRS models for P fractions?

L 17 soils type -> soil type
L 19 in -> with
L 20 in -> with
L 20/21 rephrase: only D level quality or only two fractions?
L 23 concentrations -> contents
L 30 replace “Whereas” by a more suitable word

Page 13
L 10 3.2.2 -> 3.3
L 28 Carbon -> C
L 28 Nitrogen -> N

Page 14
L 10 NIRS models for Hedley fractions and pools -> NIRS models for P fractions and pools
L 27 minimum -> level of the
L 30 factions -> fractions

Page 15
L 4 “In addition” is not appropriate here
L 18 Fractions -> fractions
L 19 rephrase (e.g., Whether P is in organic or inorganic form seemed to be of importance for…”)
L 21 models predicting the organic P fractions performed better than for inorganic P fractions throughout -> models predicting the organic P fractions performed better than models predicting the inorganic P fractions
L 22-25 change the order of the two sentences “The superior quality…” and “Similar results…”
L 23 in which -> because
L 27 Why “Therefore”?
L 30 to -> by

Page 16
L 3 and not simply -> and are not simply
L 9 even poorer or non-existent -> even poorer than for organic fractions or non-existent

Page 17
L 8 To our knowledge -> To our knowledge,
L 11 P-forms -> P forms
L 17 soil P in Hedley fractions of different availability -> soil P Hedley fractions of different availability with NIRS
L 30 represents -> requests

Page 18
L 6 North Western German Forest Research Institute -> the Northwest German Forest Research Station

Figure 1 provides -> provide; compound -> compounds
Figure 3 set4 -> set 4
Figure 7 add “triangles = P HCl conc. fractions”