

Dear Editor

Please find enclosed our revised manuscript and the detailed responses to the reviewers. Unfortunately, the responses turned out to be a bit complicated, because all comments by referee #1 and #3 were based on the original version of the manuscript that was submitted to Biogeosciences on September 26th 2014. After the initial review we included the first comments of the three referees and changed our manuscript accordingly. This revised version was then published as a discussion paper in “Biogeosciences Discussion” on January the 9th 2015. Therefore many comments by the referees were already addressed at an earlier stage. Obviously, there was a communication problem so that these referees used the older version for their detailed reviews. However, we tried to document now all the previous changes that were already included and of course also the additional changes that appeared to be necessary. Inevitably, our responses to all points raised by the reviewers ended up being quite long.

Before we address the individual issues raised by the reviewers, we would like to point out that we had never intended with this manuscript to use NIRS to develop a mechanistic understanding of the Hedley-P fractions, nor did we want to create a globally valid model to predict the different Hedley-P with NIRS. Some of the comments received seem to indicate that some reviewers might have thought that this was our intention or would have liked the manuscript to achieve just that. That might have triggered some of the comments, which we find difficult to address in the context of our study.

Below we have reproduced the greatly appreciated comments of anonymous reviewer #3 and inserted our responses in italics.

Referee #3

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.

Response

This general comment by referee 3 is rather a summary of all comments & technical comments that were also listed in a table (below). Therefore we addressed all of these comments specifically below.

page/line	referee comment	our comment
P1 L20/21	There are different modifications of the Hedley method. Therefore, the particular fractions should be named in the abstract.	The modification is named in parentheses. Since we do not present results for individual fractions in the abstract, this information should be sufficient.
1 L26	what is meant with “homogeneity of soil sample sets”? -> explain	“Soil properties” was added to qualify the term homogeneity
P1 L27	what is meant with “useful models”? -> explain	Changed to “usable”
P2 L4	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	We added “e.g. same soil type, one study site.”
P2 L17	describe the hypotheses shortly	Hypotheses have been described
P2 L28 - P4 L4	this paragraph is too long and should be subdivided, e.g., 1. Role of P fractions in tree nutrition 2. Usefulness of NIRS	The paragraph was divided (P3L18)
P3 L9/10	to which part of the sentence does the phrase “particularly in forest soils” refer?	It refers to the application of the method. Therefore we rephrased this sentence to make it more clear.
P3 L21	total C and N contents or which fractions?	We refer now to the various C and N forms, that were analyzed within these studies.
P3 L26	NIRS is usually applied to dried and ground samples. Thus, the different liquid and gas status of soils should be of minor importance.	We deleted this sentence and rephrased this section.
P4 L2	describe the “other soil properties being detectable by NIRS”	Other soil properties were specified in the text.

P4 L4	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), ...)	The sentence was clarified by changing high quality to reliable quality. This is further defined in the methods section.
P4 L4-25	rephrase this paragraph	The paragraph had a misleading section in the middle, which was also pointed out by referee #1 (P4L17). We changed this section according to the comment by ref. 1.
P4 L16	“prediction of C content and sample sets” -> is “and” the right word here? If yes, I do not understand the meaning of the sentence.	“and” was changed to “in”
P4 L17	isn’t high variation in chemical composition a cause of high spectral variation? Then “or” wouldn’t be suitable here.	Sentence was rephrased “chemical composition associated with high spectral variation”
P4 L26-28	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	Further references have been included in the manuscript.
P5 L14/15	did you select a subset of the BZE dataset?	Yes. Information was added.
P5 L25-27	Explain your selection criteria. If there were “clear correlations” between total P and P fractions, how could that help you to create subsets?	This section was extensively rephrased and the specific section was deleted.
P6 L22 and 25	volume to volume ratio or volume to mass ratio? Better write 2.5ml : 1 ml or 2.5 ml : 1 g	2.5 ml : 1 g was used
P7 L5-10	this paragraph fits better to the introduction	The paragraph was reduced to one sentence and other sentences were added to the introduction as suggested.
P7 L11	here, you do not write that you used replicate soil samples, but later you write something about replicate soil samples	Information was added

P7 L12	please add type of resin (counterion)	Added: Dowex 1x8 20-50 mesh (Sigma-Aldrich, Taufkirchen, Germany)
P7 L14	please add energy level of ultrasonic treatment	Added: ultrasonic bath RK510H, 35kHz; 23W/l (Bandelin, Berlin, Germany)
P7 L23	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?	Molybdate reactive P. We measured at 882nm wavelength with the spectrophotometer - Shimadzu UV mini-1240 (information was added)
P8 L3	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	We clarified this to: we 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.
P8 L7/8	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?	We did not state that we or others did not find organic P in the concentrated HCl fractionation step. What we stated was that the reproduction of inorganic P is very poor, since in the fractionation step with concentrated acid, before adding the persulfate solution, could already degrade organic matter and therefore the distinction between Pi and Po could be not reliable. Therefore we decided to use only TP for the concentrated HCl fraction when developing NIRS models.
P8 L9	what is meant with "satisfying"?	Satisfying was changed to reproducibility within replicates.
P9 L3-20	In part, this has already been mentioned in the introduction, some general remarks may be shortened.	This is the main Material & Methods section on NIRS, which is already quite short. We believe that we cannot shorten it further without losing important content. Since none of the other referees suggested to shorten this section and we already shortened other sections of the M&M section substantially .
P9 L6/7	O-H, C-H and N-H are bounds and not functional groups	Was changed to bonds
P9 L11	NIRS detectable soil properties -> describe them	Soil properties were specified in the text
P9 L27	Why did you not test the second or third derivative? According to Barnes et al. (1989) spectra should be detrended to	Regarding mathematical data pre-processing we specified: "Before statistical analyses, a number of

	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	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.”
P9 L28	rewrite this sentence; you did not do these treatments for the PLS	We clarified that the data treatment was done before the PLS
P9 L29-31	please give more details (size of gaps, amount of smoothing)	Information on spacing (1) and range of smoothing points (5, 9,13,17, 21, 25) was added.
P9 L31	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?	Because with small datasets and complex sample matrix, too few calibration samples could lead to models which are not robust in validation and application. The advantage of cross-validation is the increase of information since more samples could be integrated for model building. Therefore the cross-validation is an appropriate method to reduce the number of samples which are necessary for model development. References to explain just this were added at the specific position in the manuscript.
P10 L4-10	The criteria for this automated selection do not get clear from this.	This automated selection is part of the software package. Since this is a standard procedure in NIRS software, it is not further explained in the software documentation.
P10 L11-17	move this section to 2.1 Soil samples; did you consider to group samples by parent material?	This section was extensively restructured and the focus changed so it fits into this part of the Material & Methods section. The number of samples originating from soils with the same parent material was too small to develop meaningful NIRS models.
P10 L20	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	As stated above, this section was extensively restructured and the misleading sentence deleted.
P10 L26-30	move this section to 2.1 Soil samples and give the number of samples in each sample set	Number of samples in each sample set was added. The section was not moved to the section 2.1 Soil samples. Even though the sample sets consists of soil samples, it is clearly a description of the dataset used for NIRS modelling and

		therefore should remain in this section.
P11 L 9	please correct: RDP=ratio of SD to standard error of prediction	Corrected
P12 L14ff	You should always write cross-validation instead of calibration	The cross-validation is in fact a calibration since a sample set was used to create a prediction model for unknown samples. All samples of the cross validation are part of the process. The actual validation is performed with independent samples which were at no stage part of the calibration/cross-validation process. To avoid confusion between the actual calibration and the cross-validation we decided to keep the term calibration which was explained in the m&m section. We have used this terminology in a number of previous publications, and this has always been accepted.
P12 L16	do you mean worse than level D when you write “produced no useful calibrations”?	Yes, we clarified this by adding: “(lower than quality level D)”
P13 L11-19	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).)	Paragraph was rephrased accordingly
P13 L19	useful -> best?	Changed to best
P13 L26	do you mean the levels defined on p 11 with “goodness of fit of calibration models”; in Fig. 7 you use the R2 of the calibration model	Please see re-written section on the interpretation of the new Fig. 7.
P13 L28	"...were best for the Po fractions... “ I couldn't find any good relationship in Fig. 7	We defined r^2 values > 0.7 for NIRS models as indicative of useable models. Some of these high values coincided with high values for the relationship between soil C or N and P in Hedley fractions. We agree that we cannot speak about relationships here, but perhaps of indications. The text was modified

		accordingly.
P13 L26- P14 L 7	It makes no sense to correlate a R2 and a Spearman rank correlation coefficient (rs). rs is a non-parametric measure which may not simply be related to a parametric measure like R2. 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?	It is not possible to predict C or N with models designed to predict P. Instead we picked up the suggestion of Referee #1 to perform a regression analysis. Please see our comment to Referee #1 (P13 L26 – P14 L7).
P14 L7	the given correlation coefficients are not for the dataset presented in Fig.7, but for a dataset with some fractions removed, right?	We removed figure 7 since Referee #1 and #3 had concerns about the usage of correlation coefficients and r^2 of the NIRS models. According to Referee #3 we replaced them with r^2 values from regression analysis. Compare response on previous comment (Referee #3 on P13 L26-P14 L7) See supplementary material, Figure S1 and Table S3
P14 L12	you might use the data of your “random quality check” to calculate coefficients of variation for individual fractions	We calculated instead standard errors of the repeated measured samples and displayed them in the text. Compare response on comment Referee #1 on P14 L24-26
P14 L13	reference method = Hedley fractionation?	Yes, this was further clarified in the text
P14 L25/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?	We included a description of the procedure in the Material and Methods section.
P14 L30/31	the reason for the bad NIRS models might also include other factors	Yes, that's true but values around or below the detection limit might be in this case the most important one.
P14 L31	what is meant with “valid”?	Valid replaced by “meaningful”
P15 L13	what is “a reasonable prediction”?	We rephrased the sentence for more clarity.
P15 L29	“total organic P” is an inadequate term for the sum of Hedley organic P fractions,	Changed to sum of assigned organic P fractions.

	since not all organic P is extracted during the Hedley procedure	
P16 L8/9	Rephrase	Rephrased; Compare response to comment Referee #1 on P16 L1-6ff
P16 L13-16	Explain why global models are potentially as accurate as more local calibrations.	If sample size in local calibrations are <150, an option would be to increase model quality by combining datasets to a global model. Information was added.
P16 L12-23	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".	We changed this paragraph, references and examples of studies dealing with soils were added.
P17 L8-10	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	We rephrased this section to address a comment of Rev#1. Nevertheless it was not the aim of our study to identify organic P forms incorporated in a specific organic Hedley P fraction.
P17 L11/12	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	This might be true but in our study our focus was on soil properties.
P17 L12	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?	The number of samples with the same parent material was too small to achieve meaningful NIRS models
P17 L12-14	rewrite the sentence "The possible ... individual dataset	Sentence was rephrased
Tab.1 and 2	Dataset "all" is missing	Dataset all did not lead to any useful results and therefore was mainly presented in the text. Furthermore it is a combination of the sample sets BEF and BZE and information can be extracted

		from the two tables
Fig.1	Check the presentation of the modelled NaOH fractions? What did you combine?	Figure was redesigned to clarify which NaOH fractions were combined for modelling.

Technical comments

page/line	referee comment	our comment
P1 L1	Near-Infrared -> near-infrared	Changed
P1 L1	Phosphorus -> phosphorus	Changed
P1 L15	P -> phosphorus (P)	Changed
P1 L20	Hedley method -> Hedley sequential extraction method	Changed
P2 L10	Phosphorus -> Phosphorus (P)"	Changed
P2 L16	phosphorus-limitations -> phosphorus limitation	Changed
P2 L24	P-nutrition -> P nutrition	Changed
P2 L25	monitoring of the -> monitoring the	Changed
P2 L28	rephrase "solely total P contents are often measured"	Changed
P3 L4	cite the papers of Hedley	Cited: Hedley et al. 1982
P3 L8	have been -> has been	Changed
P3 L10/11	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	Changed
P3 L14	Hedley-fractionation -> Hedley fractionation	Changed
P3 L16	Hedley-P fractions -> Hedley P fractions	Changed
P3 L18	start new paragraph after "may be a promising approach."	Changed
P3 L21	C or N -> carbon (C) or nitrogen (N)	Changed
P3 L24	bracket in bracket...	Changed
P3 L26	gas -> gases	Changed
P3 L29	of to the USDA -> of the USDA	Changed

P3 L31	“P or” can be deleted L 32 find a more suitable word than “subsequently” (e.g., hence)	Deleted, “subsequently” changed to “therefore”
P4 L23/24	couldn't “depending on the homogeneity respectively heterogeneity of” be replaced by “for”; would make the sentence shorter and easier to understand	Replaced it with “for”
P4 L30	“to do so” -> could the sentences be rephrased so that “to do so” can be replaced?	To do so -> to assess these P fractions
P5 L9-13	change the order of the two sentences “From each site...” and “Including 70 sites...”	Changed
P5 L22	delete “aimed to” and change “select” to “selected”	Deleted and changed
P5 L24	add a reference	The section was changed, the references were added in a later part of the Material & Methods section. P10 Chapter 2.3 Near infrared spectroscopy. Compare response to comment referee #3 P9 L31
P6 L1	Research -> research	Changed
P6 L8	5-10cm -> 5-10 cm	Changed
P6 L14	delete “and”	Deleted
P6 L15	pH-Values -> pH values	Changed
P6 L16	North Western German Forest Research Institute -> Northwest German Forest Research Station	Changed
P6 L17	rephrase “data was measured according to the Handbuch Forstliche Analytik”	Rephrased: Samples were prepared and measured according to the German Forest science standard
P6 L18	carbon and nitrogen -> C and N	Changed
P6 L19	1150 C ?	1150°C
P6 L21	2x carbon -> C	Changed
P6 L22/25	rephrase “water solution”	
P6 L25	derivedin -> derived in	Changed
P7 L6	analysis -> analyses	Changed
P7 L20	with the -> after	Changed

P7 L23	Phosphorous -> Phosphorus	Changed
P7 L27	dot is missing	Included dot
P7 L29	remove the different “-“	Removed
P7 L29	organically bound -> total	Changed
P7 L30	autoclave and the -> autoclave. The	Changed
P7 L31	P(Po) -> P (Po)	Changed
P8 L19	delete the dot	Dot deleted
P8 L22	Hedley Fractionation Method -> Hedley fractionation method	Changed
P9 L3	exited -> excited	Changed
P9 L9	Phosphates and other P compounds -> Phosphates and other inorganic P compounds	Changed
P9 L14	either replace the comma by a dot or fill in “but” or “instead”	Replaced comma by a dot
P10 L31	set3 -> set 3	Changed
P11 L6	software) -> software)	Changed
P11 L28	Phosphorus concentrations -> phosphorus contents	Changed
P11 L29	P concentrations -> P contents	Changed
P11 L31	P concentration -> P content	Changed
P12 L8	concentrations -> contents	Changed
P12 L11	within -> below?	Changed to below
P12 L11	Hedley method -> Murphy & Riley (1962) method?	Changed to Murphy and Riley method
P12 L13	3.2.1 -> 3.2	Changed
P12 L13	NIRS models by P fractions -> NIRS models for P fractions?	Changed
P12 L17	soils type -> soil type	Changed
P12 L19	in -> with	Changed
P12 L20	in -> with	Changed
P12 L20/21	rephrase: only D level quality or only two fractions?	Rephrased: only D level quality for two of the fractions
P12 L23	concentrations -> contents	Changed

P12 L30	replace "Whereas" by a more suitable word	Deleted whereas. Rephrased sentence structure
P13 L10	3.2.2 -> 3.3	Changed
P13 L28	Carbon -> C	Changed
P13 L28	Nitrogen -> N	Changed
P14 L10	NIRS models for Hedley fractions and pools -> NIRS models for P fractions and pools	Changed
P14 L27	minimum -> level of the	Changed
P14 L30	factions -> fractions	Changed
P15 L4	"In addition" is not appropriate here	Deleted In addition
P15 L18	Fractions -> fractions	Changed
P15 L19	rephrase (e.g., Whether P is in organic or inorganic form seemed to be of importance for...")	Rephrased as recommended
P15 L21	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	Changed
P15 L22-25	change the order of the two sentences "The superior quality..." and "Similar results..."	Changed sentence order
P15 L23	in which -> because	Changed
P15 L27	Why "Therefore"?	This relates to the prior sentence
P15 L30	to -> by	Changed
P16 L3	and not simply -> and are not simply	Changed
P16 L9	even poorer or non-existent -> even poorer than for organic fractions or non-existent	Changed
P17 L8	To our knowledge -> To our knowledge,	Changed
P17 L11	P-forms -> P forms	Changed
P17 L17	soil P in Hedley fractions of different availability -> soil P Hedley fractions of different availability with NIRS	Changed
P17 L30	represents -> requests	Changed
P18 L6	North Western German Forest Research	Changed

	Institute -> the Northwest German Forest Research Station	
Figure 1	provides -> provide; compound -> compounds	Changed
Figure 3	set4 -> set 4	Changed
Figure 7	add "triangles = P HCl conc. fractions"	Added