



**HAL**  
open science

## **Determination of potential denitrification in a range of tropical topsoils using near infrared reflectance spectroscopy (NIRS)**

Bernard G. Barthès, Didier Brunet, Alain Brauman, Nathalie Fromin, Robert Lensi, Aurélie Volant, Jean-Paul Laclau, Didier Blavet, Lydie Chapuis-Lardy

### ► To cite this version:

Bernard G. Barthès, Didier Brunet, Alain Brauman, Nathalie Fromin, Robert Lensi, et al.. Determination of potential denitrification in a range of tropical topsoils using near infrared reflectance spectroscopy (NIRS). *Applied Soil Ecology*, 2010, 46 (1), pp.81-89. <10.1016/j.apsoil.2010.06.009>. <hal-02126990>

**HAL Id: hal-02126990**

**<https://hal.science/hal-02126990v1>**

Submitted on 1 Aug 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-ND 4.0 - Attribution - No Derivative Works - International License

## **Determination of potential denitrification in a range of tropical topsoils using near infrared reflectance spectroscopy (NIRS)**

Bernard G. Barthès<sup>1\*</sup>, Didier Brunet<sup>1</sup>, Alain Brauman<sup>1</sup>, Nathalie Fromin<sup>2</sup>, Robert Lensi<sup>2</sup>, Aurélie Volant<sup>1</sup>, Jean-Paul Laclau<sup>3</sup>, Didier Blavet<sup>1</sup>, Lydie Chapuis-Lardy<sup>4</sup>

<sup>1</sup> UMR Eco&Sols (CIRAD-INRA-IRD-Montpellier SupAgro), 2 place Viala, bâtiment 12, 34060 Montpellier cedex 1, France.

<sup>2</sup> CEFÉ, CNRS UMR 5175, 1919 Route de Mende, 34293 Montpellier cedex 5, France.

<sup>3</sup> CIRAD, UPR80, TA B-80/D, 34398 Montpellier cedex 5, France.

<sup>4</sup> UMR Eco&Sols (CIRAD-INRA-IRD-Montpellier SupAgro), LEMSAT / Centre Bel Air, IRD-ISRA-UCAD, BP 1386, CP 18524 Dakar, Senegal.

\* corresponding author

### **Abstract**

Microbial denitrification plays a central role in nitrous oxide (N<sub>2</sub>O)-emitting processes, which are involved in ecosystem services such as crop production and climate regulation. Field characterization of N<sub>2</sub>O-emitting processes being time consuming due to great variability, laboratory determination of potential denitrification (upon incubation) is often used as a valuable test. Near infrared reflectance spectroscopy (NIRS) is a time- and cost-effective approach that has been reported to allow accurate determination of several soil properties. The objective of the present study was to assess the interest of NIRS for determining potential denitrification over a set of 460 topsoils sampled under crops, tree plantations, savanna or rainforest, originating from Madagascar, Congo-Brazzaville, Brazil, and French Guiana.

Prediction of potential denitrification using NIRS was satisfying over the total set, especially with LOCAL calibration, which builds a model for each sample separately using its spectral neighbours in the calibration subset ( $R^2 = 0.79$  for validation). For the other sets, either textural or geographical, global calibration only was performed, involving for each set a unique prediction model built with all calibration samples. The accuracy of NIRS determination depended on the sample set, decreasing in the following order: Malagasy clayey set > total set  $\approx$  Brazilian sandy loam set > coarse textured set (Congo-Brazil)  $\approx$  Guianese sandy clay loam set  $\approx$  Congolese sandy set > non-clayey set (Congo-Brazil-Guiana), with cross validation  $R^2$  ranging from 0.88 to 0.44 (external validation was not

carried out for small-sized sets). Thus NIRS prediction was more accurate over the clayey homogeneous set than over the non-clayey heterogeneous set.

As a result of global calibration, potential denitrification was expressed as a linear combination of absorbance at every wavelength. Wavelengths that contributed most to NIRS prediction of soil potential denitrification corresponded to wavelengths that literature has assigned to organic nitrogenous compounds, amide-containing ones especially, and to carbonaceous compounds such as cellulose or including CH<sub>3</sub> or CH<sub>2</sub> groups. This related to the importance of amides in soil organic nitrogen and microbial biomass, and to the dependence of denitrification on soil organic matter.

In short, NIRS is a time- and cost-effective approach that proved relevant for determining soil potential denitrification with acceptable accuracy, especially for clayey samples or when LOCAL calibration was performed.

### **Keywords**

Potential denitrification; near infrared reflectance spectroscopy (NIRS); global calibration; LOCAL calibration.

### **Introduction**

Over 100 years, nitrous oxide (N<sub>2</sub>O) has a global warming potential approximately 300 times that of carbon dioxide and is thus a powerful greenhouse gas (Forster et al., 2007). As a consequence from intensification of agriculture and silviculture practices, increasing nitrogen (N) inputs to the soil result in increasing emissions of N<sub>2</sub>O from the soil, which is considered of great environmental concern (Tilman et al., 2002). While N<sub>2</sub>O is produced during numerous N transformations in soils (Robertson and Tiedje, 1987), on most occasions biological denitrification is a central process (Bremner, 1997; Barnard et al., 2005; Chapuis-Lardy et al., 2007). Denitrification is a microbial process in which oxidized N compounds are used as alternative electron acceptors for energy production when oxygen is limiting. This process is carried out by a wide group of bacteria, which belong to more than 50 genera (Tiedje et al., 1984). During the process the nitrate (NO<sub>3</sub><sup>-</sup>) and nitrite (NO<sub>2</sub><sup>-</sup>) are reduced in a stepwise manner to nitric oxide (NO), nitrous oxide (N<sub>2</sub>O), and finally into nitrogen gas (N<sub>2</sub>). Each step of the process is catalyzed by a specific enzyme. Generally, in non-disturbed systems, most nitrate that enters the process is fully reduced into N<sub>2</sub>, a harmless gas; but depending on environmental conditions, a variable proportion is not fully reduced and escapes as N<sub>2</sub>O (Wallenstein et al., 2006). Denitrification has also been considered a model for

addressing regulation mechanisms of microbial functions or the effects of microbial diversity on soil functioning and ecosystem dynamics (Lensi et al., 1992). Soil denitrification thus relates to both the “fertility” functions required by farmers and agronomists, and the “environmental” functions required by ecologists and the whole society.

As in situ characterization of N<sub>2</sub>O emissions is tedious due to high variability in time and space (Mathieu et al., 2006; Chapuis-Lardy et al., 2009), laboratory measurement of potential denitrification has often been used as a valuable indicator for discriminating soils with high potential for N<sub>2</sub>O emissions. However, this laboratory determination is also time-consuming as it involves the incubation of soil samples with organic carbon (OC) and nitrate additions in oxygen-free atmosphere. Near infrared reflectance (NIR) spectroscopy (NIRS) is a physically non-destructive, rapid, reproducible and low-cost method that characterizes materials according to their reflectance in the wavelength range between 800 and 2500 nm (Roberts et al., 2004). The analysis of NIR spectra relies on calibration, which is a multivariate regression procedure that expresses a given property, determined using a conventional method, as a function of absorbance at all or selected wavelengths of the NIR region. The calibration equation can then be used to predict that property on new samples from their NIR spectra only, the acquisition of which is time- and cost-effective (< 1 min per sample, no consumables required). The application of NIRS to soil has been mentioned from the 1960s (Bowers and Hanks, 1965) and it has been used extensively to determine soil content in OC and total N (TN; Reeves et al., 1999; Chang et al., 2001; Ludwig et al., 2002; Barthès et al., 2006). It has also proven useful for characterizing soil organic fractions such as microbial biomass (Reeves et al., 1999; Chang et al., 2001; Ludwig et al., 2002), organic size fractions (Barthès et al., 2008), and even microbial activities such as OC and N mineralization (Reeves et al., 1999; Chang et al., 2001; Ludwig et al., 2002). However, attempts to use NIRS for determining potential denitrification have been rare and limited to small or homogeneous sample sets (Schimann et al., 2007; Cécillon et al., 2008).

The objective of the present study was to assess the potential of NIRS for characterizing potential denitrification over a large set of topsoil samples representing contrasting tropical environments and land uses.

## **Materials and methods**

### *Sites and samples*

The total set studied included 460 soil samples collected at 0-5, 5-10 or 0-10 cm depths, originating from four tropical sites: Bemasoandro, near Antsirabe, in the highlands of

Madagascar (108 samples); vicinity of Pointe-Noire, in south-western Congo-Brazzaville (144 samples); Itatinga, in southern Brazil (São Paulo State; 52 samples); and the Paracou forest station, in northern French Guiana (156 samples). The studied sites and sample sets are presented in Table 1. All samples were collected using 100-cm<sup>3</sup> (Bemasoandro, Itatinga) or 200-cm<sup>3</sup> (Pointe-Noire, Paracou) cylinders, and were air-dried then gently crushed to pass a 2-mm sieve. More information on the studied sites and soils has been provided by Baudoin et al. (2009) for Bemasoandro, D'Annunzio et al. (2008) and Laclau et al. (2010) for Pointe-Noire, Laclau et al. (2008, 2010) for Itatinga, and Gourlet-Fleury et al. (2004) and Hättenschwiler and Bracht Jørgensen (2010) for Paracou.

*Conventional determination of soil properties (i.e. reference determination)*

Potential denitrification was determined by measuring denitrifying enzyme activity (DEA) of soil in anoxic conditions with non-limiting OC and nitrate supply, according to Tiedje (1984) and Lensi et al. (1985). For each sample (i.e. collected at 0-5, 5-10 or 0-10 cm, depending on the site, cf. Table 1), 10 g of 2-mm sieved air-dried soil (20 g for Bemasoandro and Paracou samples) was placed in a flask and sealed with rubber stopper. In each flask, air was removed with a vacuum-pump and replaced by 90% of helium and 10% of acetylene to inhibit the N<sub>2</sub>O-reductase and allow the accumulation of denitrified N as N<sub>2</sub>O. Complete saturation of the soil (i.e. 100% water holding capacity) was achieved by adding a solution including 2 mg C g<sup>-1</sup> dry soil as 50/50 (w/w) glucose and glutamic acid, and 0.1 mg N g<sup>-1</sup> dry soil as KNO<sub>3</sub> (0.2 mg N g<sup>-1</sup> dry soil for Bemasoandro samples). These OC and N additions were considered non-limiting for denitrifying enzyme functioning. Potential denitrification was calculated as the rate of N-N<sub>2</sub>O accumulated in the flask during 4 h incubation at 28°C (25°C for Paracou samples). This short incubation time was used in order to avoid any de novo enzymatic synthesis or cellular growth. The concentration of N<sub>2</sub>O in the flask was analyzed using a gas chromatograph equipped with an electron capture detector (Varian Star 3800, Walnut Creek, CA, USA). Though slight differences, all procedures were assumed to provide equivalent results.

For Bemasoandro and Paracou samples, concentrations in total carbon and N were determined on 0.2-mm ground aliquots using a CHN elemental analyzer (Carlo Erba NA 2000, Milan, Italy). In the absence of carbonates, all carbon was considered organic.

### *NIRS analyses*

Reflectance of the soil samples was measured between 1100 and 2500 nm at 2-nm intervals with a Foss NIRSystems 5000 spectrometer (Silver Spring, MD, USA), on two 5-g subsamples per sample (2-mm sieved, oven-dried at 40°C), using a ring cup (5-cm diam.). Each spectrum, averaged from 32 scans, was recorded as absorbance ( $\log [1/\text{reflectance}]$ ) then duplicate spectra were averaged to one spectrum. In order to reduce spectral data, spectra were condensed by keeping the first out of four adjacent spectral points, yielding 173 data points per spectra (Barthès et al., 2006). Data analyses were conducted using the WinISI III-v.1.61 software (Infrasoft International, LCC, State College, PA, USA).

Several mathematical pretreatments were evaluated for spectrum pre-processing: first or second derivatives with 4-, 5- or 10-point gap and smoothing (denoted 14, 15, 110, 24, 25, and 210, respectively), which reduce baseline variation and enhance spectral features (Reeves et al., 2002); derivatives were applied either alone or in conjunction with (i) standard normal variate transform (SNV), which reduces the particle-size effect and was possibly followed by detrending (SNVD) for additionally removing linear or curvilinear trends (Barnes et al., 1989), or (ii) with multiplicative scatter correction (MSC), which removes additive and/or multiplicative signal effects (Geladi et al., 1985). A principal component analysis was carried out on spectral data for calculating the Mahalanobis distance  $H$ , and samples with  $H > 3$  were considered spectral outliers and eliminated from further investigations (Mark and Tunnell, 1985).

Several sample sets were considered, geographical sets being studied grouped or separately (i.e. Bemasoandro, Pointe-Noire, Itatinga and Paracou, respectively). Large sets were divided into a calibration subset, which included the most representative samples (spectral neighbours being represented by one sample only, according to a procedure proposed by the WinISI software), and a remaining validation subset (Shenk and Westerhaus, 1991a). For sets including more than 300 samples, the calibration subset included 200 samples; for sets including 160 to 300 samples, the calibration subset included 100 samples. For geographical sets, each including 52 to 156 samples, calibration only was carried out. Calibration models deriving reference values from absorbance spectra were built using partial least square regression (PLS): PLS reduces the spectral data to a few orthogonal combinations of absorbance, called factors, that account for most spectral information and covary with reference values; cross validation being recommended for estimating the optimal number of factors in order to avoid overfitting (Shenk and Westerhaus, 1991a, 1991b). Cross validation was performed by dividing the calibration subset into four to six groups (depending on the

subset size), all but one being used for developing the model and one for prediction, the procedure being performed several times (four to six) to use all samples for both model development and prediction. The residuals of all predictions were pooled to calculate the standard error of cross validation (SECV) and the corresponding coefficient of determination  $Q^2$  (i.e. cross-validated  $R^2$ ). Calibration outliers (i.e. with residual  $> 2.5$  times SECV) were removed and another cross validation was performed, the procedure being carried twice. The number of factors after which final SECV no longer decreased meaningfully determined the optimal number of factors of the model (Bjørsvik and Martens, 2001). The cross validation was appreciated by SECV,  $Q^2$  and  $RPD_{cal}$  (ratio of standard deviation to SECV). For large sets, the prediction accuracy of the model was evaluated on the validation subset (which had not been used for model development) using the standard error of prediction between predicted and measured values (SEP) and the corresponding  $R^2$  ( $R^2_{val}$ ) and RPD ( $RPD_{val}$ ; ratio of standard deviation to SEP). Standard errors (SE, i.e. SECV and SEP) were expressed in absolute values ( $\mu\text{g N}_2\text{O-N g}^{-1} \text{ soil h}^{-1}$ ) and as proportions of the mean reference value of the corresponding subset (in %). According to Dunn et al. (2002), NIRS predictions of soil properties were considered excellent when  $RPD > 2.0$ , and acceptable when  $1.6 < RPD < 2.0$ . The above-described calibration procedure has been called “global” because a unique model is used to predict the given property for all samples of the validation subset (Shenk et al., 1997). In addition, LOCAL calibration was also carried out on the total set (Shenk et al., 1997). In that procedure, proposed in the software WinISI, a specific calibration equation is built for each sample to be characterized, using samples selected from the calibration subset according to their spectral similarity with it. This similarity is assessed by correlation coefficient between the spectrum of the sample to be characterized and those of the calibration subset samples. The number of calibration samples was varied from 30 to 50 in steps of 10 in order to determine its optimum. The optimal number of factors used in PLS regression was not determined through cross validation, which is not appropriate for LOCAL calibration (Shenk et al., 1997). Instead, predictions were calculated as the weighed average of the predicted values generated with three to ten PLS factors, weights being computed based on the inverse of spectral data residuals (X residuals) and standard deviation of regression coefficients (Shenk et al., 1997; Fernández Pierna and Dardenne, 2008).

## Results

### *Reference data*

Potential denitrification ranged from values close to 0 up to  $0.115 \mu\text{g N-N}_2\text{O g}^{-1} \text{ soil h}^{-1}$  (Table 1). In average, it was higher in the clayey Ferralsol of Bemasoandro (Madagascar) than in the sandy Arenosol of Pointe-Noire (Congo), with intermediate values in the sandy clay loam and sandy loam Ferralsols of Paracou (Guiana) and Itatinga (Brazil), respectively. As a matter of fact, soil OC content was higher at Bemasoandro than at Pointe-Noire (ca. 50 vs.  $< 10 \text{ g kg}^{-1}$  at 0-10 cm depth, respectively), with intermediate values at Paracou and Itatinga (data not shown).

### *NIRS predictions*

Using global calibration, NIRS prediction of potential denitrification was carried out on seven sample sets: the total set; the non-clayey set (total set minus Bemasoandro); the coarse-textured set (Pointe-Noire and Itatinga); the four geographical sets separately. Separating textural or geographical sets aimed at assessing the effect of set heterogeneity on prediction accuracy. Validation was carried out on the three former sets but not on the four later ones, due to their relatively small size. In addition, NIRS prediction using LOCAL calibration was carried out on the total set. In the eight cases, several mathematical pretreatments of spectra were tested but results are presented only for the pretreatment that maximized prediction accuracy (Table 2). For geographical sets, which involved calibration only, prediction accuracy was assessed according to the number of calibration outliers and SECV (as low as possible), and  $Q^2$  and  $\text{RPD}_{\text{cal}}$  (as high as possible). For predictions that involved validation, accuracy was assessed according to SEP (as low as possible), and  $R^2_{\text{val}}$  and  $\text{RPD}_{\text{val}}$  (as high as possible).

As regarded mathematical pretreatments, SNV and derivatives (first or second) generally resulted in the most accurate predictions, with or without detrending (D). The number of spectral outliers ranged from 1% to 8% of the set size and was  $< 5\%$  in general. In global calibrations, the number of calibration outliers ranged from 5 to 10% of the calibration subset, which is not negligible.

Graphs comparing NIRS predictions and conventional determinations of potential denitrification are presented in Figures 1 and 2. For geographical sets (Figure 2), prediction models have been applied on calibration samples, outliers being excluded (thus  $R^2_{\text{cal}}$  differed from  $Q^2$ , which took calibration outliers into account).

Using global calibration, fairly accurate calibration and validation were achieved over the total set ( $Q^2$  and  $R^2_{\text{val}} \approx 0.7$ ,  $SE \approx 50\%$  of the mean,  $RPD = 1.8-1.9$ ; Table 2), though the validation subset was larger than the calibration subset. Prediction accuracy was lower over the non-clayey set ( $Q^2$  and  $R^2_{\text{val}} \approx 0.4$ ,  $SE \approx 60\%$ ,  $RPD = 1.3$ ). It was also lower over the coarse-textured set than over the total set, but to a smaller extent ( $Q^2$  and  $R^2_{\text{val}} \approx 0.6$ ,  $SE \approx 50\%$ ,  $RPD = 1.6-1.7$ ). Thus increasing set homogeneity did not improve prediction accuracy. Moreover, separate calibration over geographical sets was not better at Itatinga ( $Q^2 \approx 0.7$ ,  $SECV = 26\%$ ,  $RPD_{\text{cal}} = 1.8$ ) and was clearly worse at Pointe-Noire or Paracou ( $Q^2 \approx 0.6$ ,  $SECV = 46-47\%$ ,  $RPD_{\text{cal}} = 1.6$ ) than over the total set. However, calibration was very accurate for Bemasoandro ( $Q^2 \approx 0.9$ ,  $SECV < 20\%$ ,  $RPD_{\text{cal}} = 3$ ). LOCAL calibration clearly increased prediction accuracy over the total set, when compared with global calibration ( $R^2_{\text{val}} \approx 0.8$  vs.  $0.7$ ,  $SEP = 40\%$  vs.  $53\%$ ,  $RPD_{\text{val}} = 2.2$  vs.  $1.8$ ).

#### *Most contributing wavelengths and their chemical attribution*

The ten wavelengths (out of 173) that contributed most to NIRS prediction models of potential denitrification over the total set, the non-clayey set and the coarse-textured set are listed in Table 3. For the total set, half of them were between 1588 and 1644 nm, including regions that have been attributed to the first overtone of N-H stretching from amides in proteins (around 1600 nm; Workman and Weyer, 2007) and to the first overtone of C-H stretching in  $=\text{CH}_2$  (1620 nm; Shenk et al., 2001). Other heavily contributing wavelengths could be attributed to N-containing compounds (2044 nm, possibly relating to amides, and 2060 nm to proteins), to cellulose (1820 nm), and less specifically, to organic molecules (1700 nm possibly relating to  $\text{CH}_3$  and 2420 nm to  $\text{CH}_2$ ; Shenk et al., 2001; Shui et al., 2004). Over the non-clayey set, some of the most contributing wavelengths could also be attributed to N-containing compounds (1596 and 1604 nm, possibly relating to amides from proteins), others to organic molecules (1188 and 1708 nm relating to  $\text{CH}_3$ , 1620 nm to  $\text{CH}_2$ , and 1756 nm to  $\text{CH}_2$  in aliphatic hydrocarbons; Shenk et al., 2001; Workman and Weyer, 2007). The regions around 1252-1260 nm and 1316 nm might be assigned to carboxylic acids (González-Sáiz et al., 2008). The 2076 nm wavelength could be attributed to O-H in carbohydrates (Kays and Barton, 2002), or to amine or amide N-H in proteins (2080 nm; Cozzolino and Morón, 2003).

Over the coarse-textured set, nine out of the ten most contributing wavelengths were between 1732 and 1860 nm, and six between 1772 and 1820 nm. Again, some heavy contributors

could be associated with N-containing compounds (1588 nm, possibly relating to amides, and 1732 and 1740 nm to amines), cellulose (1780, 1804, 1812 and 1820 nm) and other organic molecules (1772 nm possibly relating to CH<sub>2</sub>; Shenk et al., 2001; Workman and Weyer, 2007). The 1860 nm wavelength might be attributed to C-H in HC=CH (Morimoto et al., 2003) and the 1788 nm wavelength to water (1790 nm; Shenk et al., 2001). It is worth mentioning that the spectral region associated with water (1788 nm) had the heaviest weight by far over the coarse-textured set but was not among the 20 heaviest contributors over total and non-clayey sets (data not shown).

Some heavily contributing regions were similar over several sets: the spectral region around 1588-1596 nm, which could be attributed to amides, was a heavy contributor over the three sets; 1620 nm, associated with CH<sub>2</sub>, and 1700-1708 nm, with CH<sub>3</sub>, over total and non-clayey sets; 1820 nm, associated with cellulose, over total and coarse-textured sets. On the whole, NIRS prediction of potential denitrification thus depended heavily on amides (sometimes as proteins), CH<sub>2</sub>, and CH<sub>3</sub> and/or cellulose.

## **Discussion**

### *NIRS predictions using global calibration*

The accuracy of predictions achieved using global calibration decreased from the homogeneous clayey set (Bemasoandro:  $Q^2 = 0.9$ ,  $SECV < 20\%$ ,  $RPD = 3$ ) to the heterogeneous non-clayey set (all except Bemasoandro:  $Q^2 = 0.4$ ,  $SECV = 60\%$ ,  $RPD = 1.3$ ), with intermediate results for the other studied sets ( $Q^2 = 0.6-0.7$ ,  $SECV = 25-50\%$ ,  $RPD = 1.6-1.9$ ).

Results are consistent with those reported by Schimann et al. (2007) regarding NIRS calibration of potential denitrification on two neighbouring chronosequences representing gradients of reforestation in French Guiana:  $RPD_{cal}$  was 2.02, 2.05 and 2.15 and  $SECV$  was 59%, 55% and 48% over the total set, the clayey set and the sandy set, respectively. Relatively higher  $SECV$  (in proportion of the mean) and  $RPD$  (i.e. standard deviation  $SD / SECV$ ) in that study might be attributed to the higher variability of potential denitrification, due to very contrasting treatments (from unplanted area of open-cast mine to natural rainforest): coefficient of variation was around 100-130% (vs. 60-80% in the geographical sets of the present study), which resulted, for a given  $SECV$ , in higher  $SD / SECV$  (i.e.  $RPD$ ) and  $SECV / \text{mean}$  ratios than in the present study. It should be noticed that results from the sites studied by Schimann et al. (2007) and from Paracou, both located in

French Guiana but separated by more than 100 km, were similar as regarded mean potential denitrification (0.012 vs. 0.014  $\mu\text{g g}^{-1} \text{h}^{-1}$ ) and SECV (0.007 vs. 0.006  $\mu\text{g g}^{-1} \text{h}^{-1}$ ).

Cécillon et al. (2008) studied potential denitrification on topsoil samples and earthworm casts originating from 25 neighbouring forest plots variably affected by wildfire, on loamy sand Cambisol in southern France. Depending on the proportion of wavelengths selected for calibration (from 100% to 16%),  $Q^2$  ranged from 0.38 to 0.91, SECV from 67% to 25%, and RPD from 1.3 to 3.4, respectively. Results without wavelength selection were similar to the worst of the present study, regarding an heterogeneous set (non-clayey set). By contrast, results achieved after wavelength selection (according to their explanatory importance) were similar to the best of the present study, regarding a clayey set under rather homogeneous land use (Bemasoandro).

Considering cross-validation results more precisely, the sample sets studied here could be ranged as follows: Bemasoandro > total set  $\approx$  Itatinga > coarse-textured set  $\approx$  Paracou  $\approx$  Pointe Noire > non-clayey set. Several authors have reported more accurate NIRS prediction of soil properties, OC content especially, (i) for homogeneous than for heterogeneous sample sets, in terms of localization, texture or land use, and (ii) for clayey than for sandy samples (Van Waes et al., 2005; Brunet et al., 2007). Schimann et al. (2007) also observed that the accuracy of NIRS prediction of total C, N and substrate-induced respiration decreased from the clayey set to the total set then to the sandy set; however their results were not clear for potential denitrification. Barthès et al. (2008) suggested that less accurate NIRS prediction of soil organic and biological properties on coarse- than on fine-textured samples could be caused by their lower content in organic matter and its more heterogeneous repartition, which might lead to discrepancies between subsamples used for conventional and spectral analyses. As regarded geographical sets, set heterogeneity and soil texture could explain better prediction for Bemasoandro than for Itatinga: sampling over a small area under rather homogeneous crop cover and clayey soil texture vs. sampling over a larger area under tree plantation and savanna, with sandy loam texture, respectively (ratio of SD to mean potential denitrification was 57% vs. 70%, respectively, cf. Table 1). Similarly, set heterogeneity and texture could explain better prediction for Itatinga than for Pointe Noire, where an even larger area was sampled, also under tree plantation and savanna but with sandy soil texture (SD / mean = 83%). Worse results than expected for Paracou (sampling over a small area, with sandy clay loam texture) could be attributed to huge rainforest plant diversity (SD / mean = 79%). Better prediction for the coarse-textured set than for the non-clayey set was also consistent considering set heterogeneity, as the former included two sites and the latter three

sites, but not considering texture, as the former was more sandy. Moreover, soil texture could, but set homogeneity could not account for better prediction on the total set than on coarse-textured and non-clayey sets. The total set was characterized by much varying potential denitrification than both others (cf. Figure 1), thus by higher mean and SD; this tended to decrease SECV as a proportion of the mean and to increase  $Q^2$  and  $RPD_{cal}$ . Indeed, when looking at Figure 1, higher prediction quality for the total set than for non-clayey and coarse-textured sets was not obvious, and seemed to result firstly from wider range of potential denitrification. This underlined the difficulty of comparing prediction accuracy between sets that differed greatly in the variability of the property studied. In the same way, comparing prediction accuracy between geographical sets and total or textural sets might be somewhat questionable.

In addition, the question arises whether NIRS prediction of potential denitrification was direct or whether it was indirect, resulting from NIRS prediction of soil OC (or TN) content and correlation between potential denitrification and OC (or TN). Soil OC and TN were determined for Bemasoandro and Paracou (data not shown). For Bemasoandro, determination coefficient ( $R^2$ ) between potential denitrification and OC was 0.60 for the complete set, 0.73 when removing three outliers (according to the graph that matches potential denitrification vs. OC), and 0.81 when removing eight outliers;  $R^2$  between potential denitrification and TN was 0.61 for the complete set, and 0.73 and 0.82 when removing three and eight outliers, respectively. By contrast,  $Q^2$  was 0.88 for NIRS prediction of potential denitrification, with six outliers removed. For Paracou,  $R^2$  between potential denitrification and OC or TN was  $< 0.3$  whatever the number of outliers removed, whereas  $Q^2$  was 0.61 for NIRS prediction of potential denitrification. Thus at both sites,  $R^2$  between potential denitrification and soil OC or TN was not sufficient for explaining indirect NIRS prediction of potential denitrification as a result of NIRS prediction of OC or TN.

#### *LOCAL vs. global calibration*

LOCAL calibration, which builds prediction model for each sample separately using spectral neighbours selected in the calibration subset, resulted in more accurate prediction than global calibration, which uses the whole calibration subset to build a unique prediction model. Over the total set,  $R^2_{val}$  and RPD were noticeably increased by LOCAL calibration (from 0.70 to 0.79 and from 1.8 to 2.2 respectively), while SEP decreased (from 53% to 40%). For NIRS prediction of properties of plant and animal materials over large sample sets, LOCAL PLS calibration has been reported to generally outperform global PLS calibration (Shenk et al.,

1997; Berzaghi et al., 2000) and even sophisticated non-linear regression procedures such as artificial neural networks (Dardenne et al., 2000). Nevertheless, to date, it has not been widely used with soils though promising results: more accurate predictions have been reported with LOCAL than with global PLS calibration for OC, TN, clay, silt and sand over a large and diverse set from Italy (Cabassi et al., 2005); similar results have been reported for OC, clay and cation exchange capacity (CEC) over a large and diverse set from Belgium (Génot et al., 2007); LOCAL PLS even outperformed least squares support vector machine regression for predicting OC and CEC, but not TN, over the Belgium set (Fernández Pierna and Dardenne, 2008). LOCAL PLS calibration is more successful in general than global PLS calibration but it requires the optimization of three additional parameters: the number of “calibration neighbours” and the minimum and maximum number of PLS factors. Moreover, usual information on model predictive capacity, such as  $Q^2$ , SECV and  $RPD_{cal}$ , cannot be provided for LOCAL calibration (which predictive capacity can only be assessed on a validation set). Nevertheless, LOCAL calibration using very large calibration databases can be considered a procedure for the future, especially for predicting soil properties.

#### *Most contributing wavelengths and their chemical attribution*

Over total, non-clayey and coarse-textured sets, spectral regions that contributed most to NIRS prediction of potential denitrification could be attributed to organic nitrogenous compounds, amide-containing ones especially, and carbonaceous compounds such as cellulose or including  $CH_3$  or  $CH_2$ . This might be explained by the dependence of denitrification potential upon soil OC content (Burford and Bremner, 1975; Enwall et al., 2005) and by the fact that amides represent the predominant form of soil organic N (Schulten and Schnitzer, 1998) and an important component of microbial biomass (Lehmann et al., 2008).

Some wavelengths selected by Cécillon et al. (2008) for predicting soil potential denitrification were similar to important contributing regions revealed by the present study: 1690 nm (vs. 1700-1708 nm in the present study) and 2060 nm (vs. 2060-2076 nm). Other important wavelengths selected in that study differed from the important contributors of the present study but could be attributed to similar groups: 1730 nm, which could be associated with  $CH_2$ , and 2270 and 2350 nm, associated with cellulose. The wavelength 2310 nm, also selected in that study, did not correspond to any important chemical contributor of the present study. Moreover, wavelengths selected by Cécillon et al. (2008) for optimizing NIRS prediction of potential denitrification in their loamy sand samples were predominantly in the

region 1750-2500 nm. In the present study, this spectral region included the majority of most contributing wavelengths over the coarse-textured set (Table 3), which is thus consistent with the results of Cécillon et al. (2008).

## **Conclusion**

The study was carried out over a set of topsoil samples representing a range of tropical sites and land uses, and shows the interest of NIRS analysis and PLS regression for determining soil potential denitrification. Using global calibration (i.e. unique prediction model built with all calibration samples), prediction accuracy depended on the sample set considered, decreasing in the following order: Malagasy clayey set > total set  $\approx$  Brazilian sandy loam set > coarse textured set (Congo-Brazil)  $\approx$  Guianese sandy clay loam set  $\approx$  Congolese sandy set > non-clayey set (Congo-Brazil-Guiana), with  $Q^2$  ranging from 0.88 to 0.44, SECV from 18 to 60% (of set mean), and  $RPD_{cal}$  from 3.0 to 1.3 (cross-validation results). On the whole, it was thus more relevant to develop the prediction model over a clayey homogeneous set than over a non-clayey heterogeneous set. LOCAL calibration (i.e. prediction model built for each sample separately using spectral neighbours selected in the calibration subset) was performed on the total set, and clearly outperformed global calibration ( $R^2_{val}$  and  $RPD_{val}$  increased from 0.70 to 0.79 and from 1.8 to 2.2, respectively), thus representing a very useful procedure for large and diverse sample sets.

Wavelengths that contributed most to NIRS prediction of soil potential denitrification have been assigned to organic nitrogenous compounds, amide-containing ones especially, and to carbonaceous compounds such as cellulose or including  $CH_3$  or  $CH_2$ . This confirmed that NIRS prediction models are physically based, as amides represent a predominant form of soil organic N and microbial N, and potential denitrification depends on soil OC.

## **Acknowledgments**

The authors thank the NGO Tafa and Bodovololona Rabary (FOFIFA) for access to experimental fields in Madagascar, Corinne Rouland (IRD) for managing the experiments carried out on Congolese samples, Rildo Moreira e Moreira and the entire staff at the Itatinga experimental station (Brazil) for their contribution to this work, Jean-Christophe Roggy and Anne-Marie Domenach (EcoFoG, CIRAD-CNRS-AgroParisTech-ENGREF-INRA-Université Antilles-Guyane, in Kourou, French Guiana) for access and sampling to the Paracou forest station, and Damien Landais, Benjamin Porte and Bruno Buatois (CNRS, Montpellier, France) for their technical contribution to the conventional characterization of

Brazilian and Guianese samples. Data from Congo were obtained during the Ecosylv project, which was funded by the program Agriculture et développement durable of the French Agence nationale de la recherche (ANR-ADD-010). The authors also thank four anonymous reviewers.

## References

- Barnard, R., Leadley, P., Hungate, B.A., 2005. Global change, nitrification, and denitrification: a review. *Global Biogeochem. Cy.* 19, GB1007, doi: 10.1029/2004GB002282.
- Barnes, R.J., Dhanoa, M.S., Lister, S.J., 1989. Standard normal variate transformation and detrending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc.* 43, 772-777.
- Barthès, B.G., Brunet, D., Ferrer, H., Chotte, J.L., Feller, C., 2006. Determination of total carbon and nitrogen content in a range of tropical soils using near infrared spectroscopy: influence of replication and sample grinding and drying. *J. Near Infrared Spec.* 14, 341-348.
- Barthès, B.G., Brunet, D., Hien, E., Enjalric, F., Conche, S., Freschet, G., D'Annunzio, R., Toucet-Louri, J., 2008. Determining the distributions of soil carbon and nitrogen in particle size fractions using near infrared reflectance spectrum of bulk soil samples. *Soil Biol. Biochem.* 40, 1533-1537.
- Baudoin, E., Philippot, L., Chèneby, D., Chapuis-Lardy, L., Fromin, N., Bru, D., Rabary, B., Brauman, A., 2009. Direct seeding mulch-based cropping increases both the activity and the abundance of denitrifier communities in a tropical soil. *Soil Biol. Biochem.* 41, 1703-1709.
- Berzaghi, P., Shenk, J.S., Westerhaus, M.O., 2000. LOCAL prediction with near infrared multi-product databases. *J. Near Infrared Spec.* 8, 1-9.
- Bowers, S.A., Hanks, R.J., 1965. Reflection of radiant energy from soils. *Soil Sci.* 100, 130-138.
- Bjørsvik, H.R., Martens, H., 2001. Data analysis: calibration of NIR instruments by PLS regression. In: Burns, D.A., Ciurczak, E.W. (Eds.), *Handbook of Near-Infrared Analysis*, 2<sup>nd</sup> edition. Taylor and Francis, Boca Raton, FL, USA, pp. 185-207.
- Bremner, J.M., 1997. Sources of nitrous oxide in soils. *Nutr. Cycl. Agroecosys.* 49, 7-16.
- Brunet, D., Barthès, B.G., Chotte, J.L., Feller, C., 2007. 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.

- Burford, J., Bremner, J.M., 1975. Relationships between the denitrification capacities of soils and total, water-soluble and readily decomposable soil organic matter. *Soil Biol. Biochem.* 7, 389-394.
- Cabassi, G., Marino Gallina, P., Piombino, M., Orfeo, D., Maffioli, G., 2005. Estimation of the properties of heterogeneous soils on a multiregional scale in Italy using near infrared spectroscopy and LOCAL calibration procedures. In: Burling-Claridge, G.R., Holroyd, S.E., Sumner, R.M.W. (Eds.), *NIR 2005 – NIR in action. Making a difference*, Proceedings of the 12<sup>th</sup> International Conference on Near Infrared Spectroscopy, Auckland, New Zealand, 9-15 April 2005. New Zealand Near Infrared Spectroscopy Society Inc., Hamilton, New Zealand, pp. 341-350.
- Cécillon, L., Cassagne, N., Czarnes, S., Gros, R., Brun, J.J., 2008. Variable selection in near infrared spectra for the biological characterization of soil and earthworm casts. *Soil Biol. Biochem.* 40, 1975-1979.
- Chang, C.W., Laird, D.A., Mausbach, M.J., Hurburgh, C.R.J., 2001. Near-infrared reflectance spectroscopy - Principal components regression analyses of soil properties. *Soil Sci. Soc. Am. J.* 65, 480-490.
- Chapuis-Lardy, L., Wrage, N., Metay, A., Chotte, J.L., Bernoux, M. 2007. Soils, a sink for N<sub>2</sub>O? A review. *Glob. Change Biol.* 13, 1-17.
- Chapuis-Lardy, L., Metay, A., Martinet, M., Rabenarivo, M., Toucet, J., Douzet, J.M., Razafimbelo, T., Rabeharisoa, L., Rakotoarisoa, J. 2009. Nitrous oxide fluxes from Malagasy agricultural soils. *Geoderma* 148, 421-427.
- Cozzolino, D., Morón, A., 2003. The potential of near-infrared reflectance spectroscopy to analyse soil chemical and physical characteristics. *J. Agric. Sci.* 140, 65-71.
- D'Annunzio, R., Conche, S., Landais, D., Saint-André, L., Joffre, R., Barthès, B.G., 2008. Pairwise comparison of soil organic particle-size distributions under native savannahs and Eucalyptus plantations in Congo. *Forest Ecol. Manag.* 255, 1050-1056.
- Dardenne, P., Sinnaeve, G., Baeten, V., 2000. Multivariate calibration and chemometrics for near infrared spectroscopy: which method? *J. Near Infrared Spec.* 8, 229-237.
- Dunn, B.W., Beecher, H.G., Batten, G.D., Ciavarella, S., 2002. The potential of near-infrared reflectance spectroscopy for soil analysis – a case study from the Riverine Plain of south-eastern Australia. *Aust. J. Exp. Agr.* 42, 607-614.
- Enwall, K., Philippot, L., Hallin, S., 2005. Activity and composition of the denitrifying bacterial community respond differently to long-term fertilization. *Appl. Environ. Microbiol.* 71, 8335-8343.

- FAO-ISRIC-ISSS (Food and Agriculture Organization of the United Nations- International Soil Reference and Information Centre - International Society for Soil Science), 1998. World Reference Base for Soil Resources, World Soil Resources Report, #84. FAO, Rome, Italy.
- Fernández Pierna, J.A., Dardenne, P., 2008. Soil parameter quantification by NIRS as a Chemometric challenge at 'Chimiométrie 2006'. *Chemometr. Intell. Lab.* 91, 94-98.
- Forster, P., Ramaswamy, V., Artaxo, P., Berntsen, T., Betts, R., Fahey, D.W., Haywood, J., Lean, J., Lowe, D.C., Myhre, G., Nganga, J., Prinn, R., Raga, G., Schulz, M., Van Dorland, R., 2007. Changes in atmospheric constituents and in radiative forcing. In: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., Miller, H.L. (Eds.), *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, United Kingdom, and New York, NY, USA, pp. 129-234.
- Geladi, P., McDougall, D., Martens, H., 1985. Linearization and scatter correction for near infrared reflectance spectra of meat. *Appl. Spectrosc.* 39, 491-500.
- Génot, V., Colinet, G., Dardenne, P., Bah, B., Legrain, X., Bock, L., 2007. Etude de la performance de la spectroscopie proche infrarouge pour la mesure de caractéristiques du sol nécessaires au conseil de fumure. In : Rossignol J.P. (Ed.), *Actes des 9<sup>e</sup> Journées Nationales de l'Etude des Sols*, Angers (France), 3-5 April 2007. AFES-INH, Angers, France, pp. 379-380.
- González-Sáiz, J.M., Esteban-Díez, I., Sánchez-Gallardo, C., Pizarro, C., 2008. Monitoring of substrate and product concentrations in acetic fermentation processes for onion vinegar production by NIR spectroscopy: value addition to worthless onions. *Anal. Bioanal. Chem.* 391, 2937-2947.
- Gourlet-Fleury, S., Guehl, J.M., Laroussinie, O. (Eds.), 2004. *Ecology and Management of a Neotropical Rainforest: Lessons Drawn from Paracou, a Long-Term Experimental Research Site in French Guiana*. Elsevier, Paris, 311 p.
- Hättenschwiler, S., Bracht Jørgensen, H., 2010. Carbon quality rather than stoichiometry controls litter decomposition in a tropical rain forest. *J. Ecol.*, in press (DOI: 10.1111/j.1365-2745.2010.01671.x).
- Kays, S.E., Barton, F.E., II, 2002. Near-infrared analysis of soluble and insoluble dietary fiber fractions of cereal food products. *J. Agric. Food Chem.* 50, 3024-3029.

- Laclau, J.P., Bouillet, J.P., Gonçalves, J.L.M., Silva, E.V., Jourdan, C., Cunha, M.C.S., Moreira, M.R., Saint-André, L., Maquère, V., Nouvellon, Y., Ranger, J., 2008. Mixed-species plantations of *Acacia mangium* and *Eucalyptus grandis* in Brazil. 1. Biomass allocation and net primary production. *Forest Ecol. Manag.* 255, 3905-3917.
- Laclau, J.P., Ranger, J., de Moraes Gonçalves, J.L., Maquère, V., Krusche, A.V., Thongo M'Bou, A., Nouvellon, Y., Saint-André, L., Bouillet, J.P., de Cassia Piccolo, M., Deleporte, P., 2010. Biogeochemical cycles of nutrients in tropical Eucalyptus plantations. Main features shown by intensive monitoring in Congo and Brazil. *Forest Ecol. Manag.* 259, 1771–1785.
- Lehmann, J., Solomon, D., Kinyangi, J., Dathe, L., Wirick, S., Jacobsen, C., 2008. Spatial complexity of soil organic matter forms at nanometre scales. *Nat. Geosci.* 1, 238-242.
- Lensi, R., Gourbière, F., Jossierand, A., 1985. Measurement of small amounts of nitrate in an acid soil by N<sub>2</sub>O production. *Soil Biol. Biochem.* 17, 733-734.
- Lensi, R., Domenach, A.M., Abbadie, L., 1992. Field study of nitrification and denitrification in a wet savanna of West Africa. *Plant Soil* 147, 107-113.
- Ludwig, B., Khanna, P.K., Bauhus, J., Hopmans, P., 2002. Near infrared spectroscopy of forest soils to determine chemical and biological properties related to soil sustainability. *Forest Ecol. Manag.* 171, 121-132.
- Mark, H.L., Tunnell, D., 1985. Qualitative near-infrared reflectance analysis. *Anal. Chem.* 58: 379-384.
- Mathieu, O., Lévêque, J., Hénault, C., Milloux, M.J., Bizouard, F., Andreux, F., 2006. Emissions and spatial variability of N<sub>2</sub>O, N<sub>2</sub> and nitrous oxide mole fraction at the field scale, revealed with <sup>15</sup>N isotopic techniques. *Soil Biol. Biochem.* 38, 941-951.
- Morimoto, S., McClure, W.F., Crowell, B., Stanfield, D.L., 2003. Near infrared technology for precision environmental measurements: Part 2. Determination of carbon in green grass tissue. *J. Near Infrared Spec.* 11, 257-267.
- Reeves, J.B., III, McCarty, G.W., Meisenger, J.J., 1999. Near infrared reflectance spectroscopy for the analysis of agricultural soils. *J. Near Infrared Spec.* 7, 179-193.
- Reeves, J.B., III, McCarty, G.W., Mimmo T., 2002. The potential of diffuse reflectance spectroscopy for the determination of carbon inventories in soils. *Environ. Pollut.* 116, S277-S284.
- Roberts, C.A., Workman, J., Jr, Reeves, J.B., III, 2004. *New-Infrared Spectroscopy in Agriculture*. ASA – CSSA - SSSA, Madison, WI, USA.

- Robertson, G.P., Tiedje, J.M., 1987. Nitrous oxide sources in aerobic soils: nitrification, denitrification and other biological processes. *Soil Biol. Biochem.* 19, 187-193.
- Schimann, H., Joffre, R., Roggy, J.C., Lensi, R., Domenach, A.M., 2007. Evaluation of the recovery of microbial functions during soil restoration using near-infrared spectroscopy. *Appl. Soil Ecol.* 37, 223-232.
- Schulten, H.R., Schnitzer, M., 1998. The chemistry of soil organic nitrogen: a review. *Biol. Fertil. Soils.* 26, 1-15.
- Shenk, J.S., Westerhaus, M.O., 1991a. Population definition, sample selection and calibration procedures for near infrared reflectance spectroscopy. *Crop Sci.* 31, 469-474.
- Shenk, J.S., Westerhaus, M.O., 1991b. Population structuring of near infrared spectra and modified partial least square regression. *Crop Sci.* 31, 1548-1555.
- Shenk, J.S., Berghazi, P., Westerhaus, M.O., 1997. Investigation of a LOCAL calibration procedure for near infrared instruments. *J. Near Infrared Spec.* 5, 223-232.
- Shenk, J.S., Workman Jr., J.J., Westerhaus, M.O., 2001. Application of NIR spectroscopy to agricultural products. In: Burns, D.A., Ciurczak, E.W. (Eds.), *Handbook of Near-Infrared Analysis*, 2<sup>nd</sup> edition. CRC Press, New York, USA, pp. 419-474.
- Shui, M., Yue, L., Xu, Z., 2004. The mechanical and NIR studies on ultrafine calcium carbonate treated by four surface modifiers. *Spectrochim. Acta A* 60, 441-447.
- Tiedje, J.M., 1984. Denitrification. In: Page, L.A., Miller, R.H., Keeney, D.R. (Eds.), *Methods of Soil Analysis - Part 2: Chemical and Microbiological Properties*, 2<sup>nd</sup> edition. ASA - SSSA, Madison, WI, USA, pp. 1011-1026.
- Tiedje, J.M., Sexstone, A.J., Parkin, T.B., Revsbech, N.P., Shelton, D.R., 1984. Anaerobic processes in soil. *Plant Soil* 76, 197-212.
- Tilman, D., Cassman, K.G., Matson, P.A., Naylor R., Polasky, S., 2002. Agricultural sustainability and intensive production practices. *Nature* 418, 671-677.
- Van Waes, C., Mestdagh, I., Lootens, P., Carlier, L., 2005. Possibilities of near infrared reflectance spectroscopy for the prediction of organic carbon concentrations in grassland soils. *J. Agric. Sci.* 143, 487-492.
- Wallenstein, M.D., Myrold, D.D., Firestone, M., Voytek, M., 2006. Environmental controls on denitrifying communities and denitrification rates: Insights from molecular methods. *Ecol. Appl.* 16, 2143-2152.
- Workman, J., Jr, Weyer, L., 2007. *Practical Guide to Interpretative Near-Infrared Spectroscopy*. CRC Press, Boca Raton, FL, USA.

**Table 1.** Presentation of the studied sites and soil samples.

Location	Mean annual rainfall and temperature	Soil type <sup>a</sup> and texture	Land use	Number of samples	Sample depth (cm)	Area sampled	Potential denitrification ( $\mu\text{g N-N}_2\text{O g}^{-1} \text{ soil h}^{-1}$ )			
							Min	Max	Mean	SD <sup>b</sup>
Bemasoandro Madagascar 19°47'S, 47°06'E	1750 mm yr <sup>-1</sup> 16°C	Ferralsol, clayey	Rotations rice/soybean	108	0-5 and 5-10	Several hectares	0.005	0.115	0.037	0.021
Pointe-Noire Congo-Brazzaville 04°S, 12°E	1200 mm yr <sup>-1</sup> 25°C	Ferralsol, Arenosol, sandy	Tree plantations (eucalyptus and/ or legumes) and savanna	144	0-5 and 5-10	Several hundreds of hectares	< 0.001	0.024	0.006	0.005
Itatinga Brazil (São Paulo) 23°02'S, 48°38'W	1350 mm yr <sup>-1</sup> 19°C	Ferralsol, sandy loam	Tree plantations (eucalyptus and/ or legumes) and savanna	52	0-5	Several tens of hectares	0.004	0.093	0.020	0.014
Paracou French Guiana 5°18'N, 52°53'W	3000 mm yr <sup>-1</sup> 25°C	Ferralsol, sandy clay loam	Rainforest	156	0-10	About one hectare	< 0.001	0.071	0.014	0.011

<sup>a</sup> FAO-ISRIC-ISSS (1998).<sup>b</sup> Standard deviation.

**Table 2.** Calibration and validation statistics for each sample set, using the mathematical pretreatment that provided best results (the non-clayey set included Pointe-Noire, Itatinga and Paracou; the coarse-textured set included Pointe-Noire and Itatinga).

Sample set	Pre-treatment	Spectral outliers	Calibration							Validation				
			Calibr. outliers	n <sub>1</sub>	n <sub>2</sub>	SEC <sub>V</sub> ( $\mu\text{g N-N}_2\text{O g}^{-1} \text{h}^{-1}$ )	Q <sup>2</sup>	RPD <sub>cal</sub>	n <sub>3</sub>	SEP ( $\mu\text{g N-N}_2\text{O g}^{-1} \text{h}^{-1}$ )	R <sup>2</sup> <sub>val</sub>	RPD <sub>val</sub>		
Global calibrations														
Total	SNVD14	13 (3%)	13 (7%)	187	187	0.009	49	0.71	1.9	247	0.008	53	0.70	1.8
Non-clayey	SNVD24	10 (3%)	15 (8%)	185	185	0.007	60	0.44	1.3	142	0.006	59	0.41	1.3
Coarse-textured	SNV15	9 (5%)	9 (9%)	91	91	0.004	51	0.63	1.7	87	0.004	52	0.59	1.6
Bemasoandro	SNVD14	1 (1%)	5 (5%)	102	102	0.006	18	0.88	3.0	0	nd	nd	nd	nd
Pointe-Noire	SNVD110	2 (1%)	15 (10%)	127	127	0.003	46	0.59	1.6	0	nd	nd	nd	nd
Itatinga	MSC110	4 (8%)	4 (8%)	44	44	0.005	26	0.71	1.8	0	nd	nd	nd	nd
Paracou	SNVD25	6 (4%)	12 (8%)	138	138	0.006	47	0.61	1.6	0	nd	nd	nd	nd
LOCAL calibrations														
Total	SNV24	12 (3%)	0	200	30	nd	nd	nd	nd	248	0.007	40	0.79	2.2
Total	SNV24	12 (3%)	0	200	40	nd	nd	nd	nd	248	0.007	40	0.79	2.2
Total	SNV24	12 (3%)	0	200	50	nd	nd	nd	nd	248	0.007	40	0.78	2.2

nd: not determined.

n<sub>1</sub> is the size of the calibration subset after elimination of calibration outliers.

n<sub>2</sub> is the number of samples used for calibration (equal to n<sub>1</sub> in case of global calibration).

n<sub>3</sub> is the size of the validation subset.

SEC<sub>V</sub> and SEP are standard error of cross validation and prediction, respectively, in  $\mu\text{g N-N}_2\text{O g}^{-1} \text{h}^{-1}$  or in % of subset mean.

Q<sup>2</sup> and R<sup>2</sup><sub>val</sub> are determination coefficient for cross validation and (external) validation, respectively.

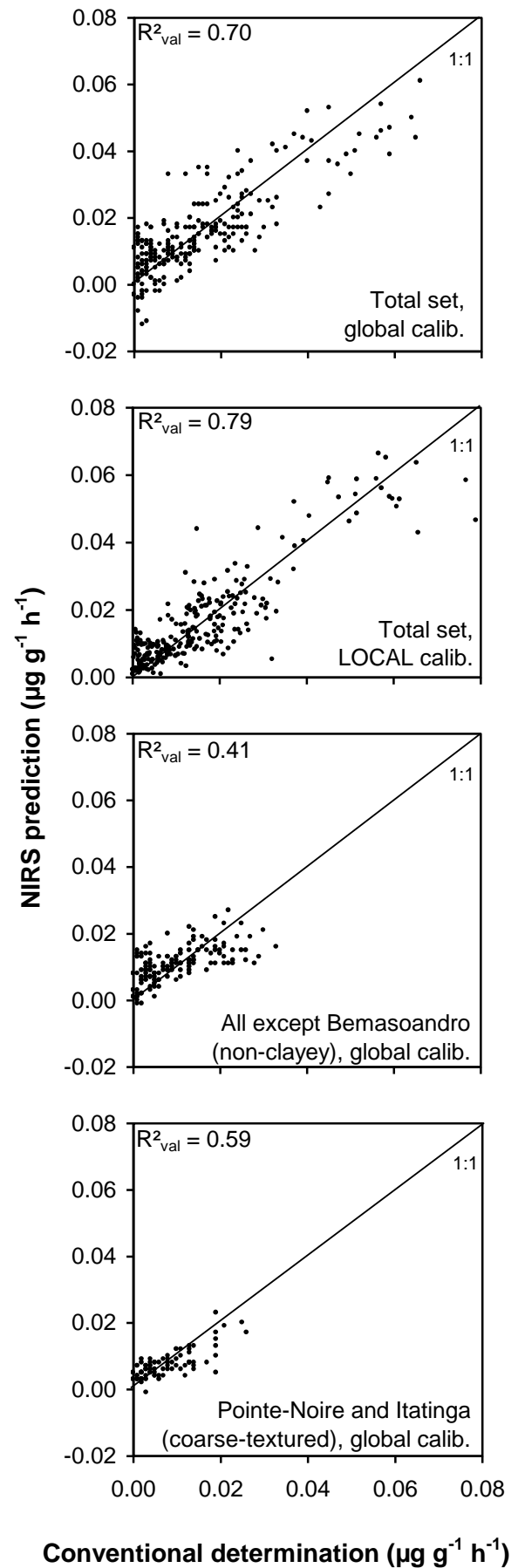
RPD<sub>cal</sub> and RPD<sub>val</sub> are ratio of standard deviation (of calibration and validation subset, respectively) to SEC<sub>V</sub> and SEP, respectively.

SNV, D, and MSC are standard normal variate transform, detrend, and multiplicative scatter correction, respectively; 14, 24, 15, 25 and 110 indicate that first or second derivatives with 4-, 5- or 10-point gap and smoothing were used.

**Table 3.** The ten wavelengths contributing most to NIRS prediction of potential denitrification over the total set, the non-clayey set and the coarse-textured set, using most relevant pretreatments, and chemical bond vibrations absorbing in similar spectral regions (according to Shenk et al., 2001, except when otherwise mentioned).

Most contributing wavelengths (nm)			Bond vibrations absorbing in similar spectral regions (nm)
Total set	Non-clayey set	Coarse-textured set	
	1188		1195 C-H 2 <sup>nd</sup> overtone in CH <sub>3</sub>
	1252		~1260 Combination of O-H stretch and O-H bend in carboxylic acid (González-Sáiz et al., 2008)
	1260		
	1316		~1320 Combination of C-H stretch and C-H deformation in carboxylic acid (González-Sáiz et al., 2008)
1588	1596	1588	~1600 N-H stretch 1 <sup>st</sup> overtone from amides in proteins (Workman and Weyer, 2007)
1612	1604		
1620	1620		1620 C-H stretch 1 <sup>st</sup> overtone in =CH <sub>2</sub>
1628			
1644			
1700	1708		1705 C-H stretch 1 <sup>st</sup> overtone in CH <sub>3</sub>
		1732	1735 C-H stretch 1 <sup>st</sup> overtone in CH <sub>3</sub> in amines (Workman and Weyer, 2007)
		1740	
	1756		1762 Symmetrical H-C-H stretch 1 <sup>st</sup> overtone in aliphatic hydrocarbons (Workman and Weyer, 2007)
		1772	1765 C-H stretch 1 <sup>st</sup> overtone in -CH <sub>2</sub>
		1780	1780 C-H stretch 1 <sup>st</sup> overtone in cellulose
		1788	1790 O-H combination in water
		1804	
		1812	
1820		1820	1820 O-H stretch / C-O stretch 2 <sup>nd</sup> overtone combination in cellulose
		1860	1860 C-H stretch 1 <sup>st</sup> overtone in HC=CH (Morimoto et al., 2003)
2044			2050 N-H bend / C-N stretch combination in CONH
			2055 Symmetrical N-H stretch / C=O stretch combination in proteins
2060			2060 N-H bend 2 <sup>nd</sup> overtone or N-H bend / N-H stretch combination in proteins
	2076		2076 O-H in carbohydrates (Kays and Barton, 2002)
			2080 Amine or amide N-H in proteins (Cozzolino and Morón, 2003)
2420			2420 Symmetrical C-H bend in CH <sub>2</sub> (Shui et al., 2004)

**Figure 1.** Comparisons between NIRS predictions and conventional determinations of potential denitrification over the validation subsets of the total set (using global or LOCAL calibration), the non-clayey set (total minus Bemasoandro; global calibration), and the coarse-textured set (Pointe-Noire and Itatinga; global calibration).



**Figure 2.** Comparisons between NIRS predictions and conventional determinations of potential denitrification over the four geographical sets separately (Bemasoandro, Pointe-Noire, Itatinga, and Paracou) using global calibration, outliers being excluded.

