Article

Estimation of Secondary Soil Properties by Fusion of Laboratory and On-Line Measured Vis–NIR Spectra

Muhammad Abdul Munnaf [ID]**, Said Nawar** [ID] **and Abdul Mounem Mouazen ***

Department of Environment, Faculty of Bioscience Engineering, Ghent University, Coupure Links 653, 9000 Ghent, Belgium; Munnaf.mabdul@ugent.be (M.A.M.); said.nawar@ugent.be (S.N.)
* Correspondence: abdul.mouazen@ugent.be

Abstract: Visible and near infrared (vis–NIR) diffuse reflectance spectroscopy has made invaluable contributions to the accurate estimation of soil properties having direct and indirect spectral responses in NIR spectroscopy with measurements made in laboratory, in situ or using on-line (while the sensor is moving) platforms. Measurement accuracies vary with measurement type, for example, accuracy is higher for laboratory than on-line modes. On-line measurement accuracy deteriorates further for secondary (having indirect spectral response) soil properties. Therefore, the aim of this study is to improve on-line measurement accuracy of secondary properties by fusion of laboratory and on-line scanned spectra. Six arable fields were scanned using an on-line sensing platform coupled with a vis–NIR spectrophotometer (CompactSpec by Tec5 Technology for spectroscopy, Germany), with a spectral range of 305–1700 nm. A total of 138 soil samples were collected and used to develop five calibration models: (i) standard, using 100 laboratory scanned samples; (ii) hybrid-1, using 75 laboratory and 25 on-line samples; (iii) hybrid-2, using 50 laboratory and 50 on-line samples; (iv) hybrid-3, using 25 laboratory and 75 on-line samples, and (v) real-time using 100 on-line samples. Partial least squares regression (PLSR) models were developed for soil pH, available potassium (K), magnesium (Mg), calcium (Ca), and sodium (Na) and quality of models were validated using an independent prediction dataset (38 samples). Validation results showed that the standard models with laboratory scanned spectra provided poor to moderate accuracy for on-line prediction, and the hybrid-3 and real-time models provided the best prediction results, although hybrid-2 model with 50% on-line spectra provided equally good results for all properties except for pH and Na. These results suggest that either the real-time model with exclusively on-line spectra or the hybrid model with fusion up to 50% (except for pH and Na) and 75% on-line scanned spectra allows significant improvement of on-line prediction accuracy for secondary soil properties using vis–NIR spectroscopy.

Keywords: vis–NIR spectroscopy; on-line soil measurement; partial least squares regression; hybrid-spectral calibration; fusion ratio; spectral discrepancy

1. Introduction

Accurate and high-resolution data on soil properties are essential for optimal soil management site-specifically with the aim of maximizing land production at minimum environmental footprints. Proximal soil sensing has made soil analysis convenient, easier, cheaper and faster [1,2] over the conventional laboratory soil analysis methods, which are laborious, costly, time consuming and destructive [3,4]. One of the best proximal soil sensors is visible and near infrared (vis–NIR) diffuse reflectance spectroscopy, which is a simple, non-destructive and rapid technique, needs no sample preparation for field applications, and can be used for off-line (the sensor is in a fixed position in the laboratory or the field) [5–8] and on-line (the sensor is being on the move during measurement) [9–12] measurement modes. The unique feature of the on-line mode is that it offers high sampling resolution

($\approx$ 2000 samples per ha) data, compared to the laboratory and in situ modes, provided with sufficient prediction accuracy for several precision agricultural applications [13].

Vis–NIR has attracted the attention of soil scientists since it was first proposed as potential means of measuring primary soil properties, which have direct spectral response (e.g., organic carbon and moisture content) in the NIR spectroscopy [14]. The direct responses originate from the combinations and overtones of fundamental vibrations of, for example, organic functional groups, and water content [15]. The vis–NIR technique also allows estimation of soil properties without direct spectral responses in NIR range, designated as secondary soil properties. This is attributed to the covariation of absorption features with one or more primary soil properties although lower prediction accuracy ranges are to be expected [16]. Reports in the literature cited good prediction accuracy for soil pH, cation exchange capacity (CEC), extractable calcium (Ca_{ex}), and extractable magnesium (Mg_{ex}) for laboratory and *in situ* conditions [8]. Earlier studies showed that the extractable sodium (Na_{ex}) and potassium (K_{ex}) are among the most difficult properties to be accurately determined using NIR spectroscopy [13,15,17–20].

Models derived with spectra collected in laboratory are used for on-line prediction of soil properties, assuming that laboratory and on-line measured spectra are mutually substitutable and ignoring spectral discrepancies between them [12,21,22]. However, many external factors such as variation of soil-to-sensor distance and angle, noise due to mechanical vibrations, presence of soil debris, and ambient light can have great influences on overall spectral quality [21] during on-line measurement, hence, introduce considerable discrepancies between one-line and laboratory collected spectrum of the same soil. Laboratory collected spectra are usually quite free from the effects of these external factors [23–25].

We hypothesize that the combination effects of all external factors during on-line measurement creates a considerable discrepancy between laboratory and on-line measured spectra for a particular soil sample. That is why a laboratory spectral-based calibration model would not always be the optimal model for on-line soil measurement. Several studies have attempted to neutralize the effects of theses external factors on spectra quality. Solutions include adopting of specific spectra pre-processing that includes maximum normalization, followed by first derivative and smoothing, a same pre-processing method followed by the same research group in several papers [25–27]. Other spectra pre-processing solutions include direct standardization (DS) and external parameter orthogonalization (EPO) adopted by other authors to remove the influence of known factors e.g., moisture content [28], while orthogonal signal correction (OSC) is suggested to reduce external influence that is orthogonal to response variable [29]. Mouazen et al. [25] also suggested a method for spectral jump correction, with a view to remove the negative effects of the variation of soil-to-sensor distance and angle that takes place during on-line measurement. These spectral transformation attempts to remove the influence of external factors during on-line measurement had led to some degree of success [25–27]. However, it is also well known that spectral pre-processing removes not only the external influences but perhaps valuable featured information [19,30]. Therefore, an alternative method to replace or minimise spectra pre-processing is recommended.

Previous studies confirmed that models developed with laboratory spectra offer better accuracy for laboratory estimations, compared to the corresponding on-line estimation [8,9,14]. The present study hypothesizes that the low prediction accuracy for on-line measurement of secondary soil properties using laboratory scanned spectra is mainly due to spectra discrepancies between laboratory and on-line scanned spectra. Furthermore, it is hypnotized that spectra pre-processing of on-line spectra cannot remove all the effects of external factors, and thus entails spectral discrepancies that introduce source of errors in the calibration. To the best of our knowledge, no previous study attempted to evaluate spectral discrepancy between laboratory and on-line scanned spectra. In addition, to date no study has reported calibration of vis–NIR spectra using fusion of laboratory and on-line measured spectra for predicting secondary soil properties. Therefore, this study aims to (i) characterize the discrepancy between laboratory and on-line measured vis–NIR spectra, and (ii) evaluate on-line estimation accuracy of secondary soil properties (pH, K, Mg, Ca, and Na) using fusion of laboratory spectra with different ratios of on-line scanned spectra.

2. Materials and Methods

2.1. Experimental Sites

Study sites included six fields (shown in Figure 1), namely, Bottelare (5 ha), Thierry (3 ha), Watermachine (6 ha), Gingelomse (11 ha), Kattestraat (5 ha), and Dal (6 ha), which belong to four different commercial farms at Melle (50.985093°N, 3.819075°E), Moeskroen (50.746902°N, 3.270746°E), Veurne (51.021793°N, 2.586106°E), and Landen (50.751915°N, 5.101041°E for Gingelomse; 51.05252°N, 3.707634°E for Kattestraat; and 50.748622°N, 5.095949°E for Dal), in Falnders, Belgium. The dominating soil texture (shown in Table 1) varied across different fields, with light to heavy loam for Gingelomse, Kattestraat and Dal, sandy to sandy-loam for Thierry, clay to clay-loamy for Watermachine's, clay to loam for Bottelare. All fields were rather flat except Gingelomse, where the elevation is higher in the middle part than the remaining parts of the field. Watermachine might have a problem associated with salt-water intrusion, since this field is located very close to the North Sea. All fields have an annual crop rotation of wheat/barley, maize, potato, and sugar beets with a short duration intermediate cover crop.

Table 1. Characteristic information of the six experimental fields located in four regions of Flanders in Belgium.

Field	Location	Period (2018)	Area (ha)	Number of Samples	Crop Type	Soil Texture	Average MC (%)	Average OC (%)
Bottelare	Melle	November	5	25	Maize	Light loam to light clay	14.64	1.60
Thierry	Moeskroen	August	3	13	Wheat	Light sandy to sandy loam	15.56	1.66
Watermachine	Veurne	August	6	19	Wheat	Heavy clay	19.86	1.35
Gingelomse	Landen	December	11	38	Barley	Light to heavy loam	22.79	1.34
Kattestraat	Landen	December	5	20	Sugar beet	Light to heavy loam	23.02	1.38
Dal	Landen	August	6	23	Barley	Light to heavy loam	8.75	1.47

MC: Moisture content; OC: Organic carbon.

2.2. On-Line Sensing Platform, Soil Scanning, and Sampling

On-line soil sensing surveys were carried out on different dates in 2018 using the on-line sensing platform developed and patented by Mouazen [31]. It consists of a subsoiler fitted to a frame, which is attached to the three-point linkage of a tractor. The subsoiler makes 15 to 25 cm deep trench in the soil, the bottom of which is smoothened by the subsoiler itself, due to the downwards vertical forces acting mainly on the chisel (see supplement). An optical probe hosted in a mild steel lens holder was appended to the backside of the subsoiler chisel to measure soil spectra in diffuse reflectance mode from the smoothened bottom of the trench. A mobile, fibre type, vis–NIR spectrophotometer (CompactSpec from Tec5 Technology for spectroscopy, Germany) with a spectral range of 305–1700 nm was used to record on-line soil spectra. A digital global positioning system (DGPS) (Trimble AG25, USA) recorded the position of spectra, which were logged together with GPS readings at a frequency of 1 Hz, using a semi-rugged laptop computer (Toughbook, Panasonic UK Ltd., Bracknell, UK) through a standard data logging and acquisition system called MultiSpec pro-II software (Tec5 Technology for spectroscopy, Germany). A 100% ceramic disc was used as the white reference, which was scanned once every 30 min. Field sensing was carried out along 12 m parallel transects at an average forward travel speed of around 3.5 km/h.
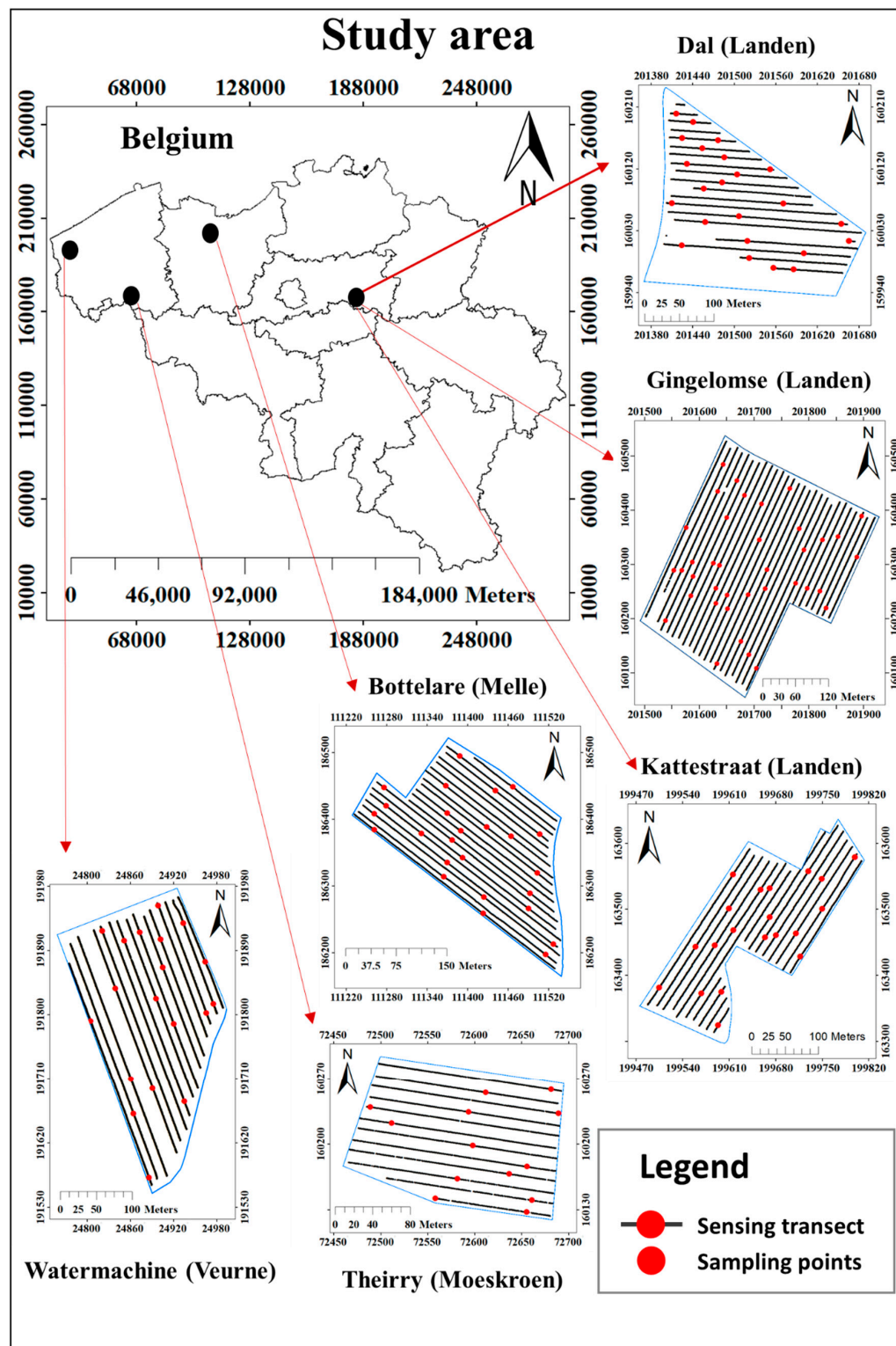
Figure 1. Location of the six experimental fields of Gingelomse, Watermachine, Theirry, Bottelare, Dal, and Kattestraat, belonging to four farms in Flanders, Belgium. The plots also show the on-line sensing transects and soil sampling points (both the calibration and validation points).

During the on-line sensing a total of 138 soil samples were collected randomly from the bottom of the trenches created by the subsoiler chisel, at an average sampling frequency of 3.83 samples per ha.

2.3. Laboratory Optical Scanning and Chemical Analyses

Each soil sample was well mixed and the sample size was reduced to around 300 g by following the standard coning and quartering method [32]. The fresh soil samples were cleaned manually by removing debris such as grass, stubble, stone/gravel, and any other foreign objects. Each sample was divided into two parts of about 150 g each, with one part used for laboratory optical measurement and the other portion for laboratory chemical analyses. The first part of each sample was placed into three Petri dishes of 2 cm in diameter and 1 cm deep. Each soil in the Petri dishes was pressed gently after levelled by a spatula, which was necessary as a smooth surface ensures maximum diffuse reflection and high signal-to-noise ratio [33]. The soil samples were scanned in diffuse reflectance mode using the same mobile, fibre type, vis–NIR spectrophotometer (CompactSpec from Tec5 Technology for spectroscopy, Germany), used for the on-line soil measurement. The same 100% white reference was used before scanning, and this reference measurement was repeated every 30 min. Ten spectra were collected per Petri dish, and these were averaged into one spectrum.

The second portion of all samples were delivered to the Soil Survey of Belgium (BDB, Heverlee, Belgium) for the chemical analysis. The soil pH was measured in the supernatant, after shaking and equilibration for 2 h in mol/l potassium chloride solution (KCl), using 1:2.5 soil:solution ratio. The available K, Mg, Ca, and Na were measured in ammonium lactate extract with inductively coupled plasma atomic emission spectroscopy (ISO 11885; CMA 2/I/B1).

2.4. Spectral Pre-Processing and Charecterization

The same pre-processing for both the on-line and laboratory measured spectra were carried out using the prospect-R package in RStudio [34]. Several combinations of spectra pre-processing were tested including smoothing, scatter corrections, first derivative, and standard normal variate (SNV) and de-trending (DT). The best performing pre-processing per individual soil property was preserved (Table 2).

Table 2. Spectra pre-processing combined steps followed for different soil properties.

Pre-Processing	Pre-Processing Order of Sequences	Soil Properties
Set-1	Moving average (w = 19) > SNV > Smoothing (SG) (w = 9; p = 2; m = 0)	pH, K
Set-2	Moving average (w = 19) > SNV de-trending > First derivative (SG) (w = 9; p = 2; m = 1) > Smoothing (SG) (w = 9; p = 2; m = 0)	Mg
Set-3	Moving average (w = 19) > Normalization (0–1)	Ca
Set-4	Moving average (w = 19) > Normalization (0–1) > GS derivative (GSD) (m = 1; w = 11; s = 5) > Smoothing (SG) (w = 11; p = 2; m = 0)	Na

GSD: Gap segment derivative; SG: Savtizky and Golay; SNV: Standard normal variate; w: Window size; m: Order of derivative; s: Gap size; p: Order of polynomial fitting; K, Mg, Ca, and Na: Extractable potassium, magnesium, calcium, and sodium.

Firstly, the raw spectra were reduced to a spectral range of 405–1660 nm. Spectral jump at the joining points of the two detectors at 1045 nm was corrected according to Mouazen et al. [25]. The moving average was used for reducing spectral noise, while maximum normalization followed conforms spectra into the same 0 to 1 scale and creates an even distribution of variances. SNV de-trending is used as a means of base-line correction [35] after comparable data scaling. Savitzky–Golay [36] and gap-segment derivatives [37] were used to reduce noise and improve the signal-to-noise ratio [9]. The moving average was used for all sets as the control pre-processing, while smoothing with Savitzky–Golay was also used for all sets except for set-3, as shown in Table 2.

After spectra cut and jump removal, principal component analysis (PCA) was conducted to investigate spectral discrepancy between laboratory and on-line scanned spectra. The analyses were done on raw and pre-processed spectra to investigate whether or not this spectra discrepancy can be minimised by spectra pre-processing. The PCA provides a set of explanatory orthogonal vectors, known as principal components (PCs) with regard to the proportion of variance explained [38]. We

considered the first two principal components (PC1 and PC2) in this study, as they captured the majority of variance in the spectral data.

2.5. Datasets Assigning, Model Building, and Quality Assessment

Five different calibration models were developed with detailed scanning mode (laboratory and on-line) and ratios of samples (Table 3): (i) Standard, (ii) hybrid-1, (iii) hybrid-2, (iv) hybrid-3, and (v) real-time.

Table 3. Dataset assignment for the five different calibration models developed in this study.

Dataset	Calibration Datasets (72%)				
Types of Calibration	**Standard**	**Hybrid-1**	**Hybrid-2**	**Hybrid-3**	**Real-Time**
Laboratory measured samples	100	75	50	25	0
On-line measured samples	0	25	50	75	100
Total samples in calibration *	100	100	100	100	100
% hybridization with on-line samples	0%	25%	50%	75%	100%
	Prediction dataset (28%)				
Samples in the prediction set	38				

* Sum of laboratory and on-line measured samples.

The Kennard–Stone (KS) algorithm [39] was used with argument metric "mahal" for selecting the calibration (72%) and prediction (28%) datasets with 100 and 38 samples, respectively. For fair comparison among the different calibration models, the same prediction dataset was used for the five models. Standard and real-time calibration involves only laboratory (100 samples) and on-line (100 samples) scanned spectra, respectively. Hybrid calibration datasets were created from fusion of laboratory and on-line scanned samples with three different ratios of 25%, 50%, and 75% on-line scanned samples, as described in Table 3. Figure 2 illustrates the different steps considered during the development of calibration models and the validation of these models for on-line predictions. The standard calibration is illustrated in Figure 2 by a solid line, whereas the proposed hybrid and real-time calibrations are illustrated by dotted lines.

Partial least squares regression (PLSR) models were developed with leave-one-out cross validation (LOOCV) using the pls package [40] in the R software and models were also validated using the prediction set. The number of latent variables (LV) were selected based on the plot of LOOCV residual variance against the number of LV. The performance of the models was evaluated using the coefficient of determination (R^2), root mean square error of prediction (RMSEP), residual prediction deviation (RPD), and the ratio of performance to inter-quartile range (RPIQ). Regarding the RPD value, models can be ranked into six categories such as (i) excellent (RPD > 2.5), (ii) very good (RPD = 2.5–2.0), (iii) good (RPD = 2.0–1.8), (iv) fair (RPD = 1.8–1.4), (v) poor (RPD = 1.4–1.0), and (vi) very poor (RPD < 1.0) performing models [41]. The current study adopted the above criterion to compare the quality of different models in both cross-validation and prediction.
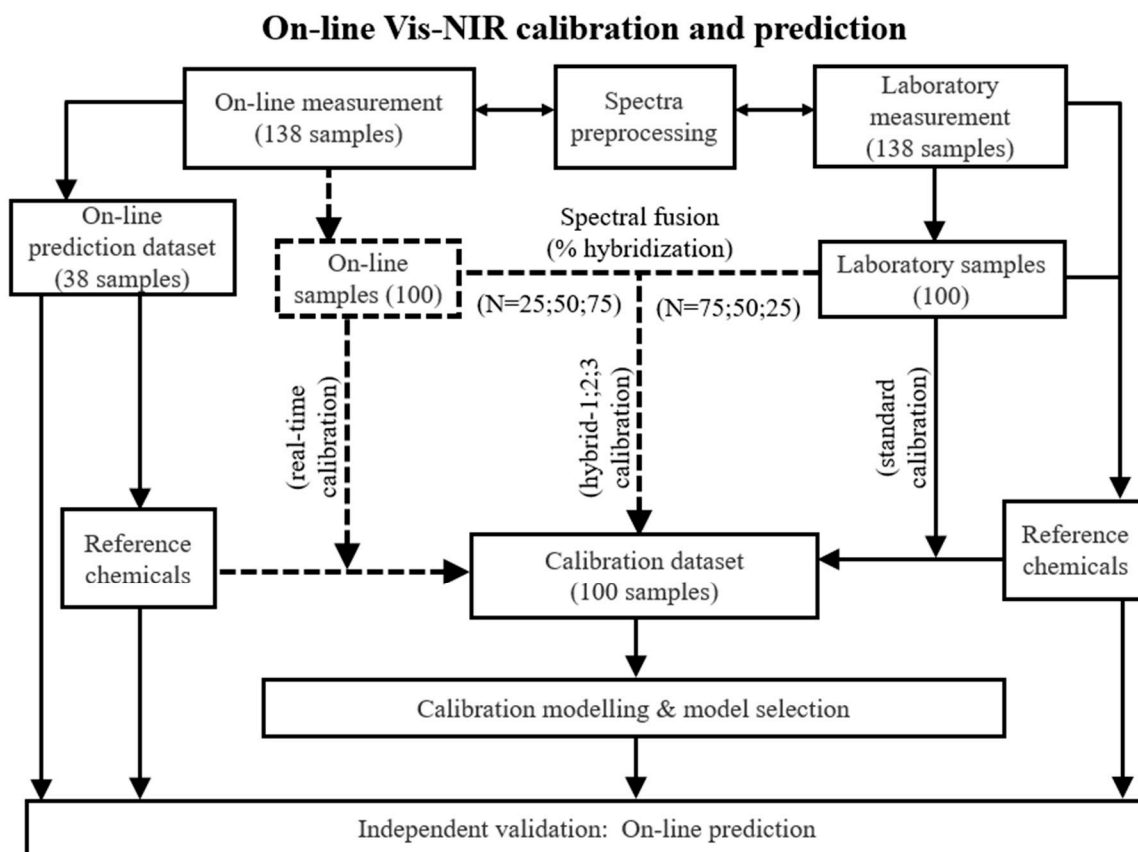
On-line Vis-NIR calibration and prediction

Figure 2. Flow diagram of steps considered for building the visible and near infrared (vis-NIR) calibration models for on-line prediction of the studied secondary soil properties. Flow paths drawn with the solid lines represent the laboratory calibration and dashed line paths correspond to the hybrid and real-time calibrations.

3. Results

3.1. Laboratory Measured Soil Data

Distribution and size of the calibration dataset critically influence the overall quality of calibration models for soil measurement, and the range of variation in the prediction set has to be approximately equal to, or lies within the range of that of the calibration set [42]. Figure 3 illustrates the descriptive statistics, Pearson correlations (*r*) with scatter plot matrices, and density distributions of laboratory measured soil pH, K, Mg, Ca, and Na both for calibration (n = 100 soil samples) and prediction (n = 38 soil samples) sets.
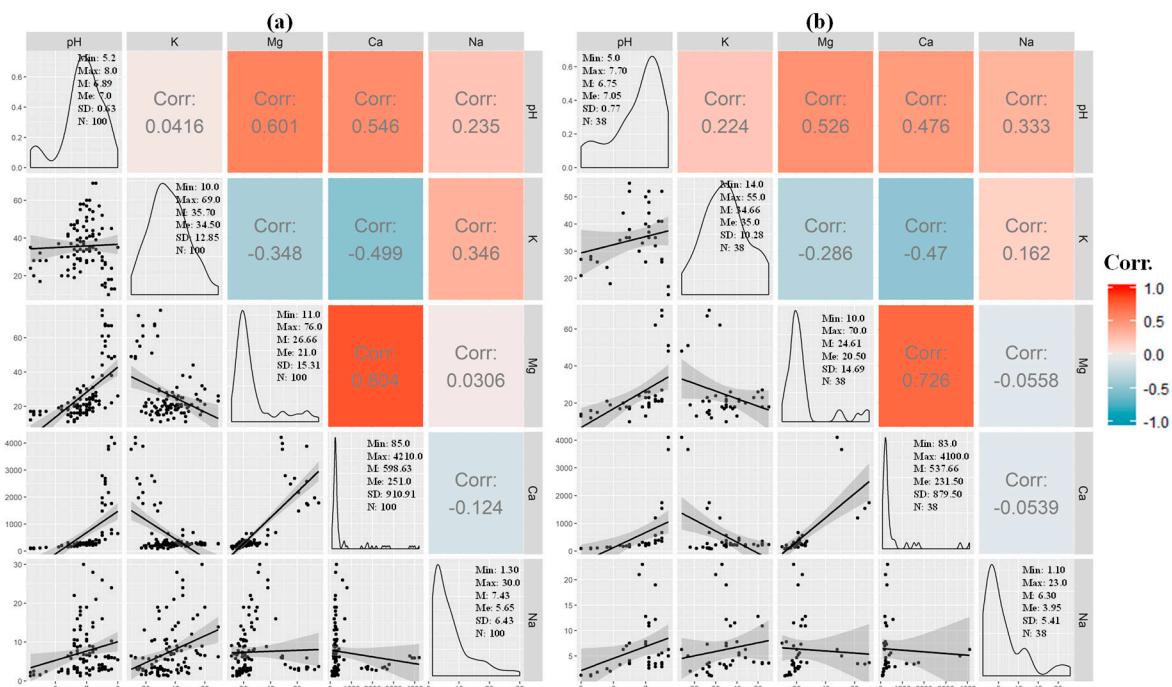
Figure 3. Descriptive statistics, correlation matrix with scatter plots, and density distribution of laboratory measured soil pH, K (potassium), Mg (magnesium), Ca (calcium), and Na (Sodium). Illustration (**a**) stands for calibration dataset (N = 100) and (**b**) for prediction dataset (N = 38), where SD = standard deviation, Min = minimum, Max = maximum, M = mean, and Me = median value.

The diagonal of Figure 3 illustrates the density plots with descriptive statistics; the upper quadrant of diagonal shows correlation matrixes with gradient colour ramps while the lower quadrant reveals scatter plots between properties. It shows that the data ranges of all individual properties in the prediction set are similar to the corresponding data range of the calibration set, though slightly smaller ranges are noticeable for the prediction set. The highest range is observed for soil Ca and successively followed by Mg, K, Na, and pH. Since the differences between the mean and median values indicates a non-normal data distribution, the differences in this study indicate slight to moderate positively skewed distributions for Mg, Ca, and Na, both in the calibration and prediction sets. The soil K data distribution shows a relatively close to normal (mean ≈ median) distribution, while data for pH is negatively skewed. Biological observations from soil data show skewed distribution [43] frequently, which is clearly visible in the density distributions charts. The correlation matrix shows a similar correlation trend both in the calibration and prediction sets, where a good correlation is found between Mg and Ca ($r \approx 0.80$) and Mg and pH ($r \approx 0.60$). In the calibration set, soil pH shows almost no correlation with K ($r = 0.042$) and weak but positive correlation with Na ($r = 0.235$). Na does not show correlation with Mg ($r = 0.031$) but weakly correlated with K ($r = 0.346$) and negatively correlated with Ca ($r = -0.124$). Moreover, K is negatively correlated with Mg ($r = -0.348$) and Ca ($r = -0.499$).

3.2. Discrepancy between Laboratory and On-Line Scanned Vis–NIR Spectra

Figure 4 presents score plots of on-line measured spectra (138 samples) against respective laboratory samples over the projected space obtained from the first two principal components (PCs) of PCA carried out before and after data pre-processing. The first two PCs (PC1 and PC2) cumulatively explain more than 99% of data variances for raw spectra, whereas smaller cumulative variances are explained for the pre-processed spectra, with ascending order for pre-processing set-4 (85.10%), set-2 (86.59%), set-1 (91.97%), and set-3 (94.02%) (see the meaning of sets in Table 2). Points scattering over the PC space indicates that the current dataset contained significant variations, including featured information. Individual group of each field confirms mutual homogeneity within that field and

heterogeneity among different fields. In this context, soils in Dal, Thierry, and Watermachine are highly heterogeneous while Bottelare, Kattestraat, and Gingelomse soils are moderately diverse. Before spectral pre-processing, it was hard to find similar grouping pattern between on-line and laboratory measured spectra, since the on-line groups are located far from the laboratory group for all fields except for Watermachine. Figure 4i reveals that generally laboratory measured samples are more homogeneously located surrounding the origin of PC plot, whereas on-line samples are more scattered and randomly distributed over the PC space. On-line samples are more scattered, possibly due to spectral alterations due to the influences of external factors such as stones and roots in the soil, and ambient light and temperature [23–25]. In the plot for the raw spectra, overlap of samples between different fields is considerable. The on-line and laboratory samples for the Watermachine field are located very closely with great deal of overlap, which may be attributed to smaller influences of the ambient conditions during on-line measurement. The highest spectral discrepancy is seen for Thierry field, followed by Dal field.

After spectral pre-processing, overlapping of spectra from different fields is smaller compared to that for the raw spectra. It seems logical that each field conveys distinguished features originating from self pedo-genesis; e.g., soil mineralogy and soil matrix characteristics. The separation is particularly clear for Watermachine samples from those of the other fields, which may be attributed to the heavy clay texture of Watermachine containing very high percentage of Ca, since it is located very close to the North Sea. Although Dal, Kattestraat, and Gingelomse are expected to be similar in soil characteristics, since they are from the same farm, the latter two fields are of more similar spectral characteristics compared to those of Dal, whose samples are perfectly separated from those of the other two fields. This perfect separation might be due to the very dry soil conditions during the on-line measurement that took place in summer 2018 (average moisture content = 8.75%). On-line spectra show more dispersion than the corresponding laboratory spectra. For example, laboratory spectra from Gingelomse field are located at one or two quadrants, while on-line spectra are spread out over the four quadrants. The highest spectral dispersion can be observed in the case of Gingelomse and Kattestraat fields, whereas the lowest discrepancy is observed for Watermachine. Since the degree of discrepancy is highly influenced by MC and that is the reason why the highest spectral differences (between the on-line and laboratory spectra) is observed for Gingelomse (average MC = 22.79%) and Katestraat (average MC = 23.02%), since these two fields were measured at very wet soil conditions (Figure 4). However, during sample preparation for laboratory scanning, MC is lost explaining the considerable difference between laboratory and on-line scanning for these two fields. Since Watermachine was measured on-line at dry soil conditions (MC = 8.75%), and the field is of a heavy clay soil texture, this has resulted potentially in the lowest reduction in MC during laboratory scanning, explaining the smallest differences between the laboratory and on-line scanned spectra. Comparing all the pre-processing sets, it is suggested that pre-processing can successfully reduce external influences by some degree, but it is quite unable to neutralize the entire impact of external factors during the on-line spectral measurements, causing differences with the laboratory measurements.

Different degree of discrepancy between laboratory and on-line scanned spectra is also noticeable for both the raw and pre-processed spectra (Figure 5). Raw spectra revealed higher variability of on-line spectra than laboratory measurement, evidenced by the higher mean, SD, and median values of the former, compared with the latter spectra. All the plots for set-1 to 4 in Figure 5 also complement the conclusion of PCA score plot that spectra pre-processing can reduce external influences only partially, which is supported by the different mean, SD, and median values at some specific wavebands.

Distinguishable absorption peaks at 420, 575, 600, 650, 930, 1125, 1400, and 1500 nm are more prominent for the laboratory scanning mode (Figure 5 (set-4 (ii))). Specifically, the differences in the spectra are clearly visible at 420 and 575 nm, which are associated with the absorption of the blue band, strongly linked with OC; and at 1400 nm associated with O–H absorption at 1450 nm [41]. This difference between the laboratory and on-line spectra may result in errors in estimation of not only soil

MC and OC, but also those properties having possible covariation with O-H absorption and OC, such as pH and P [21].
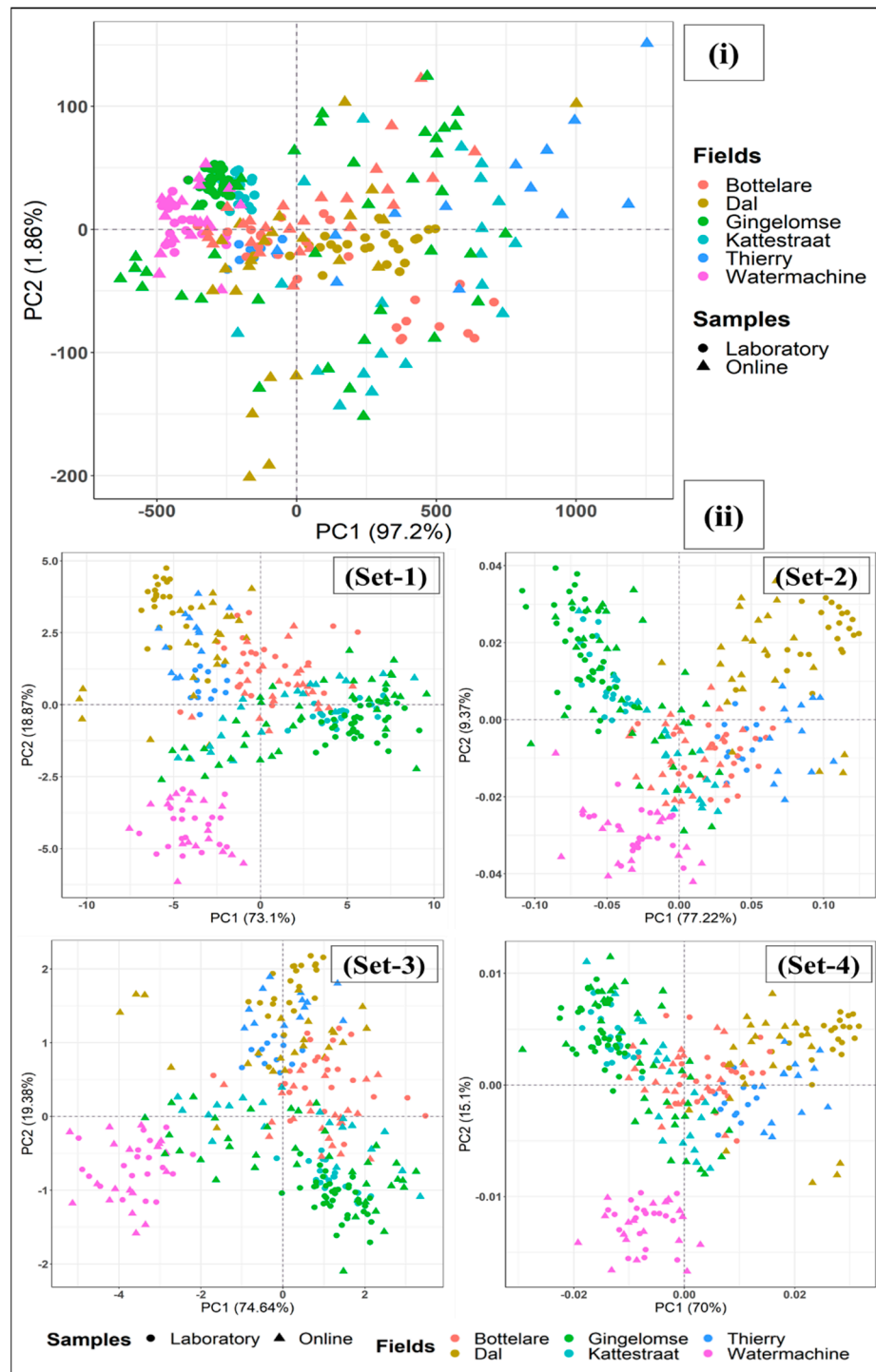

Figure 4. Characterization of the spectral discrepancy between laboratory and on-line measured samples that resulted from principal component analysis (PCA). PCA similarity maps are shown between principal component 1 (PC1) and 2 (PC2) for (i) before spectral pre-processing and (ii) after spectral pre-processing of different sets (set-1, set-2, set-3, and set-4), those described in Table 2.
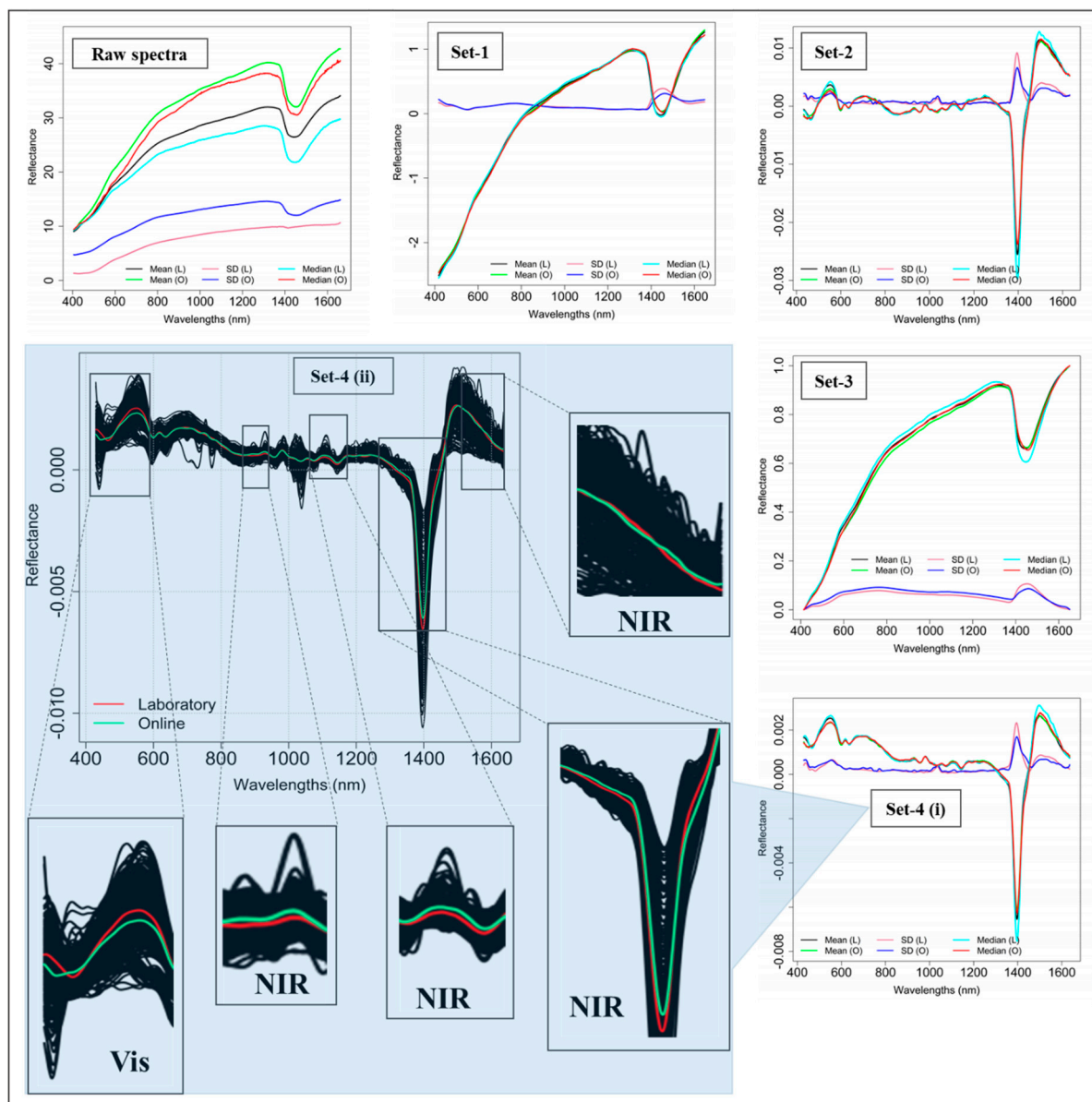
Figure 5. Spectral discrepancy, between laboratory and on-line scanning modes before (raw spectra) and after spectra pre-processing (for sets-1 to 4), shown with respect to the mean, standard deviation (SD), and median of spectra. A detail illustration is shown for pre-processing set-4 (ii), as an example, to highlight particular wavebands where a prominent discrepancy occurs. Red and green lines in the plot of set-4 (ii), respectively, stand for mean laboratory and on-line spectra while black lines represent the entire dataset. L: Laboratory; O: On-line.

3.3. PLSR Coefficients

Figure 6 illustrates PLSR coefficients obtained from the standard, hybrid, and real-time calibrations. Important absorption peaks are evenly distributed across both the visible (400–780 nm) and NIR (780–1700 nm) regions. Only few key absorption peaks are observed for pH, K, and Ca though several smaller peaks are observed for Mg and Na. Important wavelengths of 455, 772, 1361, and 1424 nm are observed for pH. Mouazen et al. [21] reported several correlation features for pH in the visible and NIR ranges, which were probably associated with amine N–H (751, 1000, and 1500 nm), hydroxyl O–H (950, 1450, and 1950 nm), and aromatic C–H (825, 1100, and 1650 nm) bonds [14]. Similarly, the 772 nm wavelength in the present study can be attributed to N–H absorption at 751 nm. The wavelengths of 1361 and 1424 nm can be associated with the second overtone of O–H absorption, whereas the 455 nm

wavelength can be associated with the blue colour absorption band that used to be at 450 nm, which can be attributed to OC and water. This explains that pH is partially being successfully measured through covariation with water and OC. For K, a moderate absorption band (456 nm), two wide peaks at around 645 and 1158 nm and a stronger absorption peak at 1425 nm are recorded.
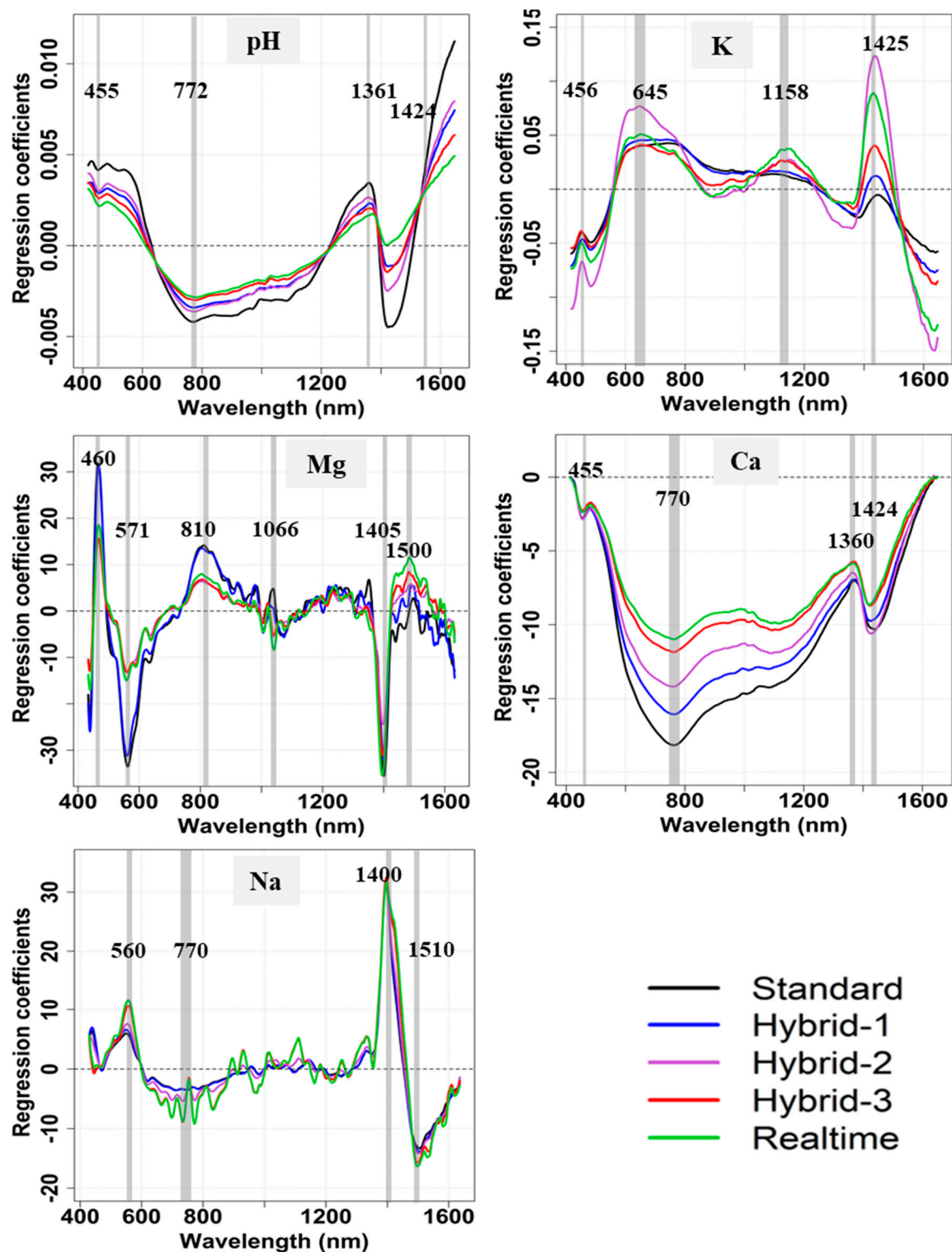
Figure 6. Regression coefficients' plots resulted from partial least squares regression (PLSR) analysis for the development of standard, hybrid-(1,2,3), and real-time calibration models for on-line prediction of soil pH, available K (potassium), Mg (magnesium), Ca (calcium), and Na (sodium).

Similar to pH, the 456 and 1425 nm wavelengths are attributed to the blue colour (associated with overall changes is soil albedo related to water and organic matter) and O–H second overtone absorptions, respectively. The absorption at 645 and 1158 can be attributed, respectively, to red colour (at 680 nm, associated with soil mineralogy and iron oxides in particular) and aromatic C–H, used to be around 1100 nm, according to Viscarra Rossel and Behrens [44], respectively. Among the wavelengths of 460, 571, 810, 1056, 1405, and 1500 nm, contributing to Mg successful prediction, significant sharp absorption features are found at 460 and 1405 nm, which can be attributed to the blue band (attributed to OC and water) and the second overtone of O–H absorptions (similar to pH and K), respectively. Fewer absorption features for Na are observable at 560, 770, 1400, and 1510 nm. The significant wavelengths at 1500 nm for Mg and 1510 nm for Na, can be linked with the absorption of amine N–H bonds. Four absorption features are seen for Ca, which are almost identical to those for pH (455, 790, 1360, and 1424 nm), and these may well be attributed to calcium carbonate. Our findings of key absorption wavelengths that have significant features at around 1400–1450 nm reveal possible covariations of pH, K, Mg, Na, and Ca with soil MC, which is attributed to the prominent O–H absorption in the second overtone region. Both Mg and K show possible co-variations with aromatic C–H (1056 nm for Mg, and 1158 for K) and amine N–H bonds for Mg at 810 and 1500 nm, while Na shows covariations with the N–H bond only, attributed to peak coefficient at 1500 nm.

Nevertheless, the general trend of PLSR coefficients indicates that the key absorption peaks discussed above happen at exactly the same wavelengths for the standard, hybrid and real-time calibrations. This indicates overall consistency of significant spectral features associated with respective properties across the models developed.

3.4. Quality of Prediction Results

Table 4 shows the prediction results for the studied soil properties using PLSR models developed for the five datasets explained above. It can be observed that the prediction accuracy is dependent on the soil property. For example, the best on-line prediction is observed for Mg with R^2 = 0.48, RMSEP = 10.42 mg/100 g, RPD = 1.41, RPIQ = 0.55, which is successively in descent order of Na, pH, K, and Ca. Based on laboratory scanned soil samples. Compared to the standard calibration using laboratory spectra only, hybrid calibrations performed better for the prediction of all the investigated properties. The best on-line prediction result (R^2 = 0.81, RMSEP = 6.25 mg/100 g, RPD = 2.35, and RPIQ = 0.92) is obtained for Mg using the hybrid-3 model (Table 4 and Figure 7). The real-time calibration shows the same prediction quality to those of hybrid-2 and hybrid-3 in particular. The best on-line prediction result for the real-time model was found for Mg (R^2 = 0.81, RMSEP = 6.38 mg/100 g, RPD = 2.30, RPIQ = 0.90), followed by Ca (R^2 = 0.75, RMSEP = 436.46 mg/100 g, RPD = 2.02, RPIQ = 0.43). The remaining models can be sorted in descending order as pH (R^2 = 0.74, RMSEP = 0.39, RPD = 1.97, RPIQ = 2.25), Na (R^2 = 0.65, RMSEP = 3.14 mg/100 g, RPD = 1.72, RPIQ = 1.43), and K (R^2 = 0.54, RMSEP = 6.85 mg/100 g, RPD = 1.50, RPIQ = 2.00).

3.5. Influences of Fusion Ratio on On-Line Prediction Quality

Figure 7 illustrates the influence of fusion ratio of on-line versus laboratory collected spectra on on-line prediction quality of soil pH, K, Mg, Ca, and Na in comparison with the standard and real-time calibrations. Results show that by increasing the percentage of on-line collected spectra in the calibration set, proportional improvement in on-line prediction can be observed. Comparing among the three hybrid models, hybrid-1 (25%) is the least performing model followed by hybrid-2 (50%) and hybrid-3 (75%), successively. Both hybrid-2 and hybrid-3 provided comparable results, except for pH and Na, where hybrid-3 outperformed hybrid-2 clearly. Hybrid-1 has resulted in slight improvements in the prediction of pH, K, and Mg, whereas significant improvement can be already observed for Ca (R^2 = 0.69, RMSEP = 483.81 mg/100 g, RPD = 1.82, RPIQ = 0.39), compared with the standard calibration (R^2 = 0.13, RMSEP = 809.13 mg/100 g, RPD = 1.09, RPIQ = 0.23). Hybrid-2, hybrid-3, and real-time models all provide the best prediction results for Mg (R^2 = 0.81, RMSEP = 6.25–6.38 mg/100 g,

RPD = 2.30–2.35, and RPIQ = 0.90–0.92), whereas the second best accurate prediction is found with hybrid-3 for Ca (R^2 = 0.77, RMSEP = 412.23 mg/100 g, RPD = 2.13, and RPIQ = 0.45). The hybrid-3 model provides the best prediction results for Na (Table 4). K is best predicted by the hybrid-2 model (R^2 = 0.58, RMSEP = 6.60 mg/100 g, RPD = 1.56, and RPIQ = 2.08), whereas pH (R^2 = 0.74, RMSEP = 0.39 mg/100 g, RPD = 1.97, RPIQ = 2.25) is best predicted by the real-time model. Results indicate that the proposed hybrid-2 (50% on-line spectra) and hybrid-3 (75% on-line spectra), both perform equally well as the real-time model, except the underperformance of hybrid-2 for pH and Na (Table 4). The on-line prediction is classified as good performing for pH (hybrid-3 and real-time models; RPD = 1.96–1.97), fair for K (hybrid-2, hybrid-3, and real-time models; RPD = 1.48–1.56) very good for Mg and Ca (hybrid-2, hybrid-3, and real-time models; RPD = 2.02–2.35), and good for Na (hybrid-3 and real-time models; RPD = 1.72–1.82), which are improved way beyond the reported results in the literature. Therefore, they can be used successfully for the prediction of the named soil properties, except for hybrid-2 for pH and Na (Table 4).

Table 4. Quality of on-line prediction of soil pH, potassium (K), magnesium (Mg), calcium (Ca), and sodium (Na), obtained from partial least squares regression (PLSR) models developed for (i) standard, (ii) hybrid-1, (iii) hybrid-2, (iv) hybrid-3, and (v) real-time calibrations.

Soil Property	Calibration Type	Prediction			
		R^2	RMSEP	RPD	RPIQ
pH	Standard	0.45	0.56	1.37	1.56
	Hybrid-1	0.50	0.54	1.43	1.63
	Hybrid-2	0.57	0.50	1.54	1.76
	Hybrid-3	0.73	0.39	1.96	2.24
	Real-time	**0.74**	**0.39**	**1.97**	**2.25**
K	Standard	0.25	8.75	1.17	1.57
	Hybrid-1	0.33	8.30	1.23	1.66
	Hybrid-2	**0.58**	**6.60**	**1.56**	**2.08**
	Hybrid-3	0.53	6.93	1.48	1.98
	Real-time	0.54	6.85	1.50	2.00
Mg	Standard	0.48	10.42	1.41	0.55
	Hybrid-1	0.68	8.15	1.80	0.71
	Hybrid-2	0.81	6.29	2.33	0.91
	Hybrid-3	**0.81**	**6.25**	**2.35**	**0.92**
	Real-time	0.81	6.38	2.30	0.90
Ca	Standard	0.13	809.13	1.09	0.23
	Hybrid-1	0.69	483.81	1.82	0.39
	Hybrid-2	0.76	428.91	2.05	0.44
	Hybrid-3	**0.77**	**412.23**	**2.13**	**0.45**
	Real-time	0.75	436.46	2.02	0.43
Na	Standard	0.37	4.23	1.28	1.06
	Hybrid-1	0.26	4.58	1.15	0.98
	Hybrid-2	0.54	3.62	1.49	1.24
	Hybrid-3	**0.69**	**2.96**	**1.83**	**1.51**
	Real-time	0.65	3.14	1.72	1.43

RMSEP: Root mean square error of prediction (mg/100 g); R^2: Coefficient of determination; RPD: Residual prediction deviation; and RPIQ: Performance to inter-quartile range. The best prediction results are highlighted in bold.
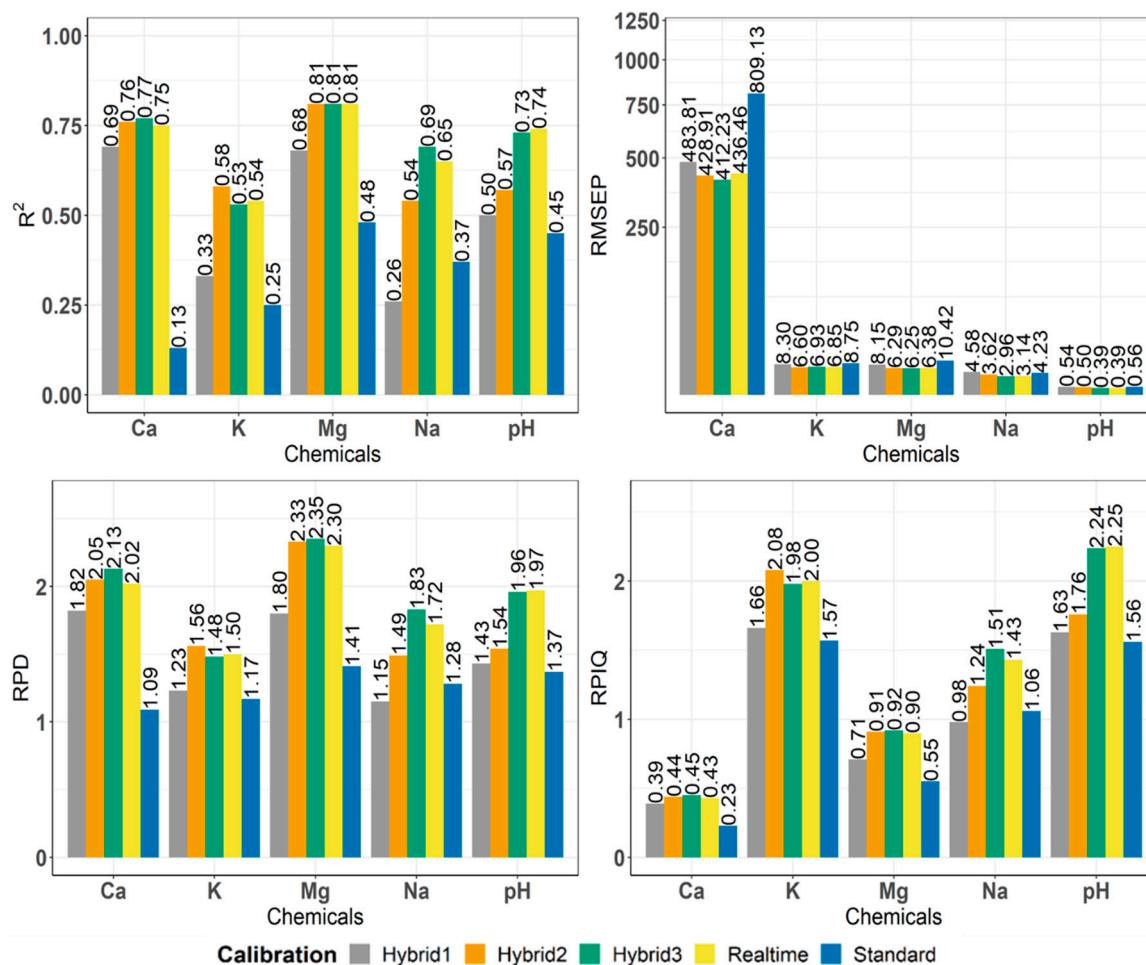
Figure 7. Variations of model performance for on-line prediction of soil pH, available K (potassium), Mg (magnesium), Ca (calcium), and Na (sodium), shown in terms of root mean square error of prediction (RMSEP), coefficient of determination (R^2), residual prediction deviation (RPD), and performance to inter-quartile range (RPIQ), obtained from the standard, hybrid-1, hybrid-2, hybrid-3, and real-time calibrations.

4. Discussions

The current study hypothesises that the on-line measured vis-NIR spectra deviate from laboratory spectra collected for the same soil samples, due to the influences of external factors on the former scanning method, such as ambient conditions, mechanical vibrations, and sensor-to-soil distance variations. As a consequence, it is assumed that the on-line measured spectra, can if included in the calibration set, improve the prediction accuracy of on-line measured soil properties (i.e., pH, K, Mg, Ca, and Na), having indirect spectral responses in the near infrared spectroscopy. However, the influence of the percentage of on-line spectra to be added to the calibration set is unknown, and requires the investigation carried out in this work.

From the spectra analysis discussed above, one can conclude that indeed spectral differences between laboratory and on-line scanning methods exist at both individual spectral level for the same sample, and at groups of spectra. The PCA similarity maps obtained from PC1 and PC2, showed clear differences between laboratory and on-line measured spectra for both the raw and pre-processed spectra. The differences become smaller after implementing the different pre-processing steps considered in the present work, indicating that spectra pre-processing can at least partially remove these differences. Overlap between the laboratory and on-line samples was observed in the PCA similarity maps due to spectral pre-processing, particularly for fields with relatively low MC. However, due to the significant

difference in the raw spectra, the same data pre-processing might not work for both the laboratory and on-line collected spectra, hence, different pre-processing is proposed to resolve this issue.

Current findings indicate that soil MC is one of the dominant properties responsible for spectra differences [23,44]. During laboratory preparation and scanning of samples collected during on-line measurement, there is a possibility for soil samples to lose MC, introducing spectral differences between laboratory and on-line scanned spectra. Discrepancy is more prominent in some particular spectral bands, shown in Figure 6. These differences may well be removed with a proper spectra pre-processing, before calibration. Similarly, the influence of noise due to vibration can be removed by a gentle smoothing, while variation of sensor-to-soil distance can also be removed by an algorithm suggested by Mouazen et al. [25]. However, aggressive spectra pre-processing may also lead to losing important feature information necessary for the successful prediction of the studied soil properties. Above all, the on-line prediction results of the standard models were of rather poor quality, compared to the hybrid and real-time models, suggesting that the spectra pre-processing is not sufficient to remove all sources of discrepancy, and that different ratios of on-line spectra should be included in the calibration set. This has resulted in improved prediction accuracy, and the degree of improvement was proportional with the ratio of on-line spectra added.

As can be revealed from the results in Table 4 that the real-time calibration performs almost as equally good as the hybrid-2 (except for pH and Na) and hybrid-3 models, while it outperforms the standard calibration and hybrid-1 models for all the studied soil properties. When comparing the hybrid against the standard calibrations, all hybrid models outperformed the standard model for on-line measurement of the studied soil properties (Table 4 and Figure 7). The hybrid calibration improved the on-line prediction quality for Mg and Ca from being poor to very good, for pH and Na from being poor to good and for K from being poor to fair, according to RPD classes proposed by Viscarra Rossel et al. [41]. This means that it is necessary to include on-line scanned spectra in the calibration set, which have special features that do not exist in the laboratory scanned spectra. These features still exist even after the spectra pre-processing methods detailed in Table 2. These results suggest that either the real-time model with on-line spectra or the hybrid model with fusion up to 50% (except for pH and Na) and 75% on-line scanned spectra allows significant improvement of the prediction accuracy for soil properties having indirect spectral response in the NIR spectroscopy for on-line vis–NIR spectral scanning mode.

Calibration models derived from laboratory spectra can predict secondary soil properties in on-line mode with relatively low accuracy. In this study, we examined whether or not there is a need to include on-line scanned spectra in the calibration data set for on-line prediction of the studied secondary soil properties. While laboratory scanning is essential to build a spectral library that is needed for future laboratory scanning-based calibrations, one should bear in mind that according to the results achieved in this work, there is a need to establish two different spectral libraries, one for laboratory scanning conditions and one for on-line scanning conditions.

Although the proposed hybrid approach can improve the overall accuracy, it has to be admitted that the vis–NIR spectroscopy is limited [15] in providing excellent prediction results for secondary soil properties unlike for properties with direct spectral responses, e.g., OC [45,46] and MC [8,47]. The on-line prediction was good for pH (hybrid-3 and real-time models; RPD = 1.96–1.97), fair for K (hybrid-2, hybrid-3, and real-time models; RPD = 1.48–1.56) very good for Mg and Ca (hybrid-2, hybrid-3, and real-time models; RPD = 2.02–2.35), and good for Na (hybrid-3 and real-time models; RPD = 1.72–1.82), which are improved results than those reported in the literature.

Based on laboratory calibrations, Chang et al. [15] reported that the vis–NIR is a limited technique for the estimation of soil K, Mg, and Ca with the least prediction accuracy obtained for Ca ($R^2 < 0.5$ and RPD < 1.4). A study by Dunn et al. [17] reported that NIR spectroscopy showed a high level of prediction accuracy with $R^2 = 0.67, 0.91, 0.87$, and 0.69, respectively for K, Mg, Ca, and Na for top soil (0 to 10 cm) using spectral data collected in the laboratory. In addition, using a full range spectrophotometer (350 to 2500 nm) in the laboratory, Mouazen et al. [19] found good measurement

accuracy for Ca (R^2 = 0.77 and RPD = 2.10), whereas poor results were reported for Mg (R^2 = 0.59 and RPD = 1.56), Na (R^2 = 0.40 and RPD = 1.29), and K (R^2 = 0.33 and RPD = 1.21). Qiao and Zhang [48] developed NIR calibration models using laboratory spectra, achieving a better accuracy for K with R^2 = 0.69 and RMSE = 0.69%, compared to that of Mouazen et al. [19]. Therefore, all the above studies showed that laboratory-based vis–NIR spectroscopy demonstrate fluctuating results for the measurement of secondary soil properties and the prediction performance will depend on the sample set available for each study. The conclusion also applies to on-line prediction of secondary soil properties, although only three studies can be found in the literature [16,49,50].

Since secondary soil properties have indirect spectral responses in NIR spectroscopy, it is not possible to determine key wavebands that contribute directly to the parameter estimations. That is the reason why previous studies [15,16,48] did not attempt to identify important wavebands for secondary soil properties. However, significant bands were identified in the present study that associate with the successful prediction of pH (455, 772, 1361, 1424 nm), K (456, 645, 1158, 1425 nm), Mg (571, 810, 1066, 1405, 1500 nm), Ca (455, 770, 1360, 1424 nm), and Na (560, 770, 1400, 1510 nm). However, these featured wavelengths may vary for different datasets, depending on the parent material of soil, weathering conditions, soil texture, colour, and mineralogy. It was possible to explain the association of these bands or group of bands with soil properties having direct spectral responses in the NIR spectroscopy. These include among others, bands associated with OC, MC, and blue and red colour absorptions. Spectroscopy bands associated with molecular bonds such as O–H, aromatic C–H, and amine N–H were assigned for each soil property studied, where considerable deal of overlap was observed for few properties. For example, the same four absorption features were observed for both pH and Ca (455, 790, 1360, and 1424 nm), which was attributed to calcium carbonate. Among these four features, three similar features at 460, 810, and 1405 nm were observed for Mg. This indicates similarity of the significant spectral features for these three properties. The laboratory chemical analysis results show a similar trend of data for these three properties. For example, they are all positively skewed (Figure 3) with a strong correlation between Mg and Ca (r ≈ 0.80) and good correlations between Mg with pH (r ≈ 0.60) and Ca and pH (r ≈ 0.55), indicating that the spectral correlations obtained from the PLS regression coefficients' plots (Figure 6) contain real information about the chemical background of the data set. This also shows that links between spectral and chemical data can be made to understand why the secondary soil properties can be measured with vis–NIR spectroscopy, successfully. However, further study is needed for in-depth evaluation of the relationship between the spectral and chemical data to quantify the individual contribution of significant bands associated with the primary soil properties including soil color in the visible range to the prediction of the different secondary properties studied.

It is interesting to discuss the advantages of the modelling approach used in this study over the conventional methods (i.e., EPO, DS, and OSC) used for removing external effects from soil spectra. Several studies have used the EPO to remove the influences of known external factors, e.g., MC [6,51], soil roughness, aggregation, and ambient temperature [52]. However, both EPO and SD have not been reported to remove the influences of unknown factors, e.g., those frequently encountered during the on-line soil sensing such as noise, and presence of stones and plant roots and residues. Both DS and EPO require a transfer sample set, consisting of identical samples but measured under different measurement conditions to account for the known influences [28,52], e.g., dry versus wet condition when the external factor under consideration is moisture content. OSC is more advantageous over EPO and DS methods, since it does not require a transfer sample set [28], hence, it can theoretically tackle unknown external influences. However, it has mathematically been proven that OSC, when coupled with PLSR (OSC–PLSR) could not improve prediction quality but rather improved model interpretability [29]. A recent study [28] reported inconsistent performance of OSC–PLSR models, with slight improvement in the prediction quality of on-line measured pH, Ca, CEC, and lime requirement in a clay soil, whereas no improvement was reported for soil organic matter, Mg, potential acidity, sum of basis, percent base saturation, and MC. This contradictory performance was difficult to explain.

Our approach based on fusion of laboratory and on-line scanned spectra can handle the influences of unknown external factors present in on-line spectra, without the need for a transfer data set to be created. This is supported by the fact that results achieved in the present work are much better than those recently reported by Franceschini et al., [28] for on-line measurement of soil properties, obtained after spectra transformation with EPO, DS, and OSC techniques. It would be interesting in a future work to combine the approach of the present study with EPO, DS, and OSC to enable removal of both the known and unknown external influences, hence, maximising the prediction accuracy for the on-line measurement of secondary soil properties. This approach can be tested further for primary soil properties having direct spectral responses in the NIR spectroscopy.

5. Conclusions

The current study introduced a novel calibration approach to model the vis–NIR spectra for on-line prediction of secondary soil properties, namely, pH, available K (potassium), Mg (magnesium), Ca (calcium), and Na (sodium). It compares the performance of partial least squares regression (PLSR) models developed using 100% laboratory scanned spectra (standard model), 100% on-line measured spectra (real-time model) and hybrid-1, hybrid-2, and hybrid-3 models, having 25%, 50%, and 75% on-line measured spectra fused with laboratory spectra of 75%, 50%, and 25%, respectively. Results obtained suggest the following conclusions:

a.   For a particular soil sample, laboratory and on-line spectra are rarely identical and spectra pre-treatments can reduce the discrepancies to some extent but cannot remove them completely. Therefore, the laboratory scanned spectra-based calibration models predict on-line soil properties with low accuracy.

b.   Inclusion of on-line collected spectra in the spectra set is necessary, which has resulted in improved prediction accuracy. The degree of improvement was proportional with the ratio of on-line spectra added. The real-time calibration performed almost equally good as the hybrid-2 model (except for pH and K) and hybrid-3 model (for all the soil properties investigated). Furthermore, the three hybrid models outperformed the standard calibration. Thus, either the real-time, the hybrid-2 (excluding pH and Na) or the hybrid-3 models should be used for successful on-line prediction of the secondary soil properties considered in this study.

c.   The current study identified key absorption wavelengths significantly contributing to the predictions of soil pH, K, Mg, Ca, and Na. These wavelengths are associated with the absorption band of the blue colour, second overtone of O–H absorption, aromatic C–H, and amine (N–H) absorptions, depending on the soil property.

To sum up, the proposed modelling approach can be successfully used for on-line measurement of secondary soil properties (e.g., pH, K, Mg, Ca, and Na). This approach is of practical use for different end users, e.g., precision farming practitioners and soil scientists, who are interested in high resolution data that is acquired rapidly, accurately, and cost-effectively. However, future study is needed to further prove that the current modelling approach is applicable on different spectral data sets having a wider range of variability in the soil attributes, compared to the current dataset collected from four different farms in Belgium.

Author Contributions: All the authors substantially contributed to this article. M.A.M. and A.M.M. conceptualized the study and developed the methodology. M.A.M. accomplished data analysis and wrote a draft of the manuscript. S.N. contributed to reviewing and editing the manuscript. In addition, corresponding author A.M.M. wrote part of the manuscript and reviewed and edited the different versions of the paper. He supervised the research, as he is the Odysseus project coordinator, who secured the project fund.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Adamchuk, V.I.; Viscarra Rossel, R.A. Development of On-the-Go Proximal Soil Sensor Systems. In *Proximal Soil Sensing*; Springer: Dordrecht, The Netherlands, 2010; pp. 15–28. ISBN 9789048188598.
2. Wang, D.; Chakraborty, S.; Weindorf, D.C.; Li, B.; Sharma, A.; Paul, S.; Ali, M.N. Synthesized use of VisNIR DRS and PXRF for soil characterization: Total carbon and total nitrogen. *Geoderma* **2015**, *243–244*, 157–167. [CrossRef]
3. McDowell, M.L.; Bruland, G.L.; Deenik, J.L.; Grunwald, S.; Knox, N.M. Soil total carbon analysis in Hawaiian soils with visible, near-infrared and mid-infrared diffuse reflectance spectroscopy. *Geoderma* **2012**, *189–190*, 312–320. [CrossRef]
4. Viscarra Rossel, R.A.; McBratney, A.B. Soil chemical analytical accuracy and costs: Implications from precision agriculture. *Aust. J. Exp. Agric.* **1998**, *38*, 765–775. [CrossRef]
5. Ackerson, J.P.; Morgan, C.L.S.; Ge, Y. Penetrometer-mounted VisNIR spectroscopy: Application of EPO-PLS to in situ VisNIR spectra. *Geoderma* **2017**, *286*, 131–138. [CrossRef]
6. Ackerson, J.P.; Demattê, J.A.M.; Morgan, C.L.S. Predicting clay content on field-moist intact tropical soils using a dried, ground VisNIR library with external parameter orthogonalization. *Geoderma* **2015**, *259–260*, 196–204. [CrossRef]
7. Chen, C.; Viscarra Rossel, R.A. Digitally mapping the information content of visible–near infrared spectra of surficial Australian soils. *Remote Sens. Environ.* **2011**, *115*, 1443–1455. [CrossRef]
8. Kuang, B.; Mahmood, H.S.; Quraishi, M.Z.; Hoogmoed, W.B.; Mouazen, A.M.; van Henten, E.J. Sensing soil properties in the laboratory, in situ, and on-line. A review. *Adv. Agron.* **2012**, *114*, 155–223. [CrossRef]
9. Nawar, S.; Mouazen, A.M. On-line vis-NIR spectroscopy prediction of soil organic carbon using machine learning. *Soil Tillage Res.* **2019**, *190*, 120–127. [CrossRef]
10. Maleki, M.R.; Mouazen, A.M.; De Ketelaere, B.; Ramon, H.; De Baerdemaeker, J. On-the-go variable-rate phosphorus fertilisation based on a visible and near-infrared soil sensor. *Biosyst. Eng.* **2008**, *99*, 35–46. [CrossRef]
11. Tekin, Y.; Tumsavaş, Z.; Ulusoy, Y.; Mouazen, A.M. On-line Vis-Nir sensor determination of soil variations of sodium, potassium and magnesium. In *IOP Conference Series: Earth and Environmental Science*; IOP Publishing: Bristol, UK, 2016; Volume 41, p. 012011.
12. Nawar, S.; Mouazen, A.M. Optimal sample selection for measurement of soil organic carbon using on-line vis-NIR spectroscopy. *Comput. Electron. Agric.* **2018**, *151*, 469–477. [CrossRef]
13. Shepherd, K.D.; Walsh, M.G. Development of Reflectance Spectral Libraries for Characterization of Soil Properties. *Soil Sci. Soc. Am. J.* **2002**, *66*, 988. [CrossRef]
14. Stenberg, B.; Viscarra Rossel, R.A.; Mouazen, A.M.; Wetterlind, J. Visible and Near Infrared Spectroscopy in Soil Science. In *Advances in Agronomy*; Sparks, D.L., Ed.; Academic Press: Cambridge, MA, USA, 2010; Volume 107, pp. 163–215.
15. Chang, C.-W.; Laird, D.A.; Mausbach, M.J.; Hurburgh, C.R. Near-Infrared Reflectance Spectroscopy–Principal Components Regression Analyses of Soil Properties. *Soil Sci. Soc. Am. J.* **2001**, *65*, 480. [CrossRef]
16. Marín-González, O.; Kuang, B.; Quraishi, M.Z.; Munóz-García, M.Á.; Mouazen, A.M. On-line measurement of soil properties without direct spectral response in near infrared spectral range. *Soil Tillage Res.* **2013**, *132*, 21–29. [CrossRef]
17. Dunn, B.W.; Batten, G.D.; Beecher, H.G.; Ciavarella, S. The potential of near-infrared reflectance spectroscopy for soil analysis—A case study from the Riverine Plain of south-eastern Australia. *Aust. J. Exp. Agric.* **2002**, *42*, 607. [CrossRef]
18. Islam, K.I.; Khan, A.; Islam, T. Correlation between Atmospheric Temperature and Soil Temperature: A Case Study for Dhaka, Bangladesh. *Atmos. Clim. Sci.* **2015**, *5*, 200–208. [CrossRef]
19. Mouazen, A.M.; De Baerdemaeker, J.; Ramon, H. Effect of Wavelength Range on the Measurement Accuracy of Some Selected Soil Constituents Using Visual-Near Infrared Spectroscopy. *J. Near Infrared Spectrosc.* **2006**, *14*, 189–199. [CrossRef]
20. Zornoza, R.; Guerrero, C.; Mataix-Solera, J.; Scow, K.M.; Arcenegui, V.; Mataix-Beneyto, J. Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. *Soil Biol. Biochem.* **2008**, *40*, 1923–1930. [CrossRef]

21. Mouazen, A.M.; Maleki, M.R.; De Baerdemaeker, J.; Ramon, H. On-line measurement of some selected soil properties using a VIS-NIR sensor. *Soil Tillage Res.* **2007**, *93*, 13–27. [CrossRef]
22. Nawar, S.; Mouazen, A.M. Predictive performance of mobile vis-near infrared spectroscopy for key soil properties at different geographical scales by using spiking and data mining techniques. *Catena* **2017**, *151*, 118–129. [CrossRef]
23. Chang, C.-W.; Laird, D.A.; Hurburgh, C.R., Jr. Influence of soil moisture on near-infrared reflectance spectroscopic measurement of soil properties. *Soil Sci.* **2005**, *170*, 244–255. [CrossRef]
24. Mouazen, A.M.; Karoui, R.; De Baerdemaeker, J.; Ramon, H. Characterization of soil water content using measured visible and near infrared spectra. *Soil Sci. Soc. Am. J.* **2006**, *70*, 1295–1302. [CrossRef]
25. Mouazen, A.M.; Maleki, M.R.; Cockx, L.; Van Meirvenne, M.; Van Holm, L.H.J.; Merckx, R.; De Baerdemaeker, J.; Ramon, H. Optimum three-point linkage set up for improving the quality of soil spectra and the accuracy of soil phosphorus measured using an on-line visible and near infrared sensor. *Soil Tillage Res.* **2009**, *103*, 144–152. [CrossRef]
26. Tekin, Y.; Kuang, B.; Mouazen, A.M. Potential of On-Line Visible and Near Infrared Spectroscopy for Measurement of pH for Deriving Variable Rate Lime Recommendations. *Sensors* **2013**, *13*, 10177–10190. [CrossRef] [PubMed]
27. Shamal, S.A.M.; Alhwaimel, S.A.; Mouazen, A.M. Application of an on-line sensor to map soil packing density for site specific cultivation. *Soil Tillage Res.* **2016**, *162*, 78–86. [CrossRef]
28. Franceschini, M.H.D.; Demattê, J.A.M.; Kooistra, L.; Bartholomeus, H.; Rizzo, R.; Fongaro, C.T.; Molin, J.P. Effects of external factors on soil reflectance measured on-the-go and assessment of potential spectral correction through orthogonalisation and standardisation procedures. *Soil Tillage Res.* **2018**, *177*, 19–36. [CrossRef]
29. Indahl, U.G. The O-PLS methodology for orthogonal signal correction-is it correcting or confusing? *J. Chemom.* **2017**, 1–14. [CrossRef]
30. Rinnan, Å.; van den Berg, F.; Engelsen, S.B. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends Anal. Chem.* **2009**, *28*, 1201–1222. [CrossRef]
31. Mouazen, A.M. Soil Survey Device 2006. In *International Publication, Published under the Patent Cooperation Treaty (PCT)*; World Intellectual Property Organization, International Bureau: Brussels, Belgium, 2006.
32. Mukhopadhyay, S.; Maiti, S.K. Techniques for Quantative Evaluation of Mine Site Reclamation Success. In *Bio-Geotechnologies for Mine Site Rehabilitation*; Elsevier: Amsterdam, The Netherlands, 2018; pp. 415–438. ISBN 9780128129876.
33. Mouazen, A.M.; Saeys, W.; Xing, J.; De Baerdemaeker, J.; Ramon, H. Near infrared spectroscopy for agricultural materials: An instrument comparison. *J. Near Infrared Spectrosc.* **2005**, *13*, 87–97. [CrossRef]
34. Stevens, A.; Ramirez Lopez, L. An Introduction to the Prospectr Package. Available online: https://cran.r-project.org/web/packages/prospectr/vignettes/prospectr-intro.pdf (accessed on 1 January 2019).
35. Barnes, R.J.; Dhanoa, M.S.; Lister, S.J. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc.* **1989**, *43*, 772–777. [CrossRef]
36. Savitzky, A.; Golay, M.J.E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. [CrossRef]
37. Norris, K.H. Understanding and Correcting the Factors Which Affect Diffuse Transmittance Spectra. *NIR News* **2001**, *12*, 6–9. [CrossRef]
38. Maxwell, A.E.; Harman, H.H. Modern Factor Analysis. *J. R. Stat. Soc. Ser. A* **2006**, *131*, 615. [CrossRef]
39. Kennard, R.W.; Stone, L.A. Computer Aided Design of Experiments. *Technometrics* **1969**, *11*, 137–148. [CrossRef]
40. Mevik, B.H.; Wehrens, R. The pls Package: Principal Component and Partial Least Squares Regression in R. *J. Stat. Softw.* **2007**, *18*, 1–24. [CrossRef]
41. Viscarra Rossel, R.A.; Walvoort, D.J.J.; McBratney, A.B.B.; Janik, L.J.; Skjemstad, J.O. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* **2006**, *131*, 59–75. [CrossRef]
42. Kuang, B.; Mouazen, A.M. Calibration of visible and near infrared spectroscopy for soil analysis at the field scale on three European farms. *Eur. J. Soil Sci.* **2011**, *62*, 629–636. [CrossRef]

43. Bellon-Maurel, V.; Fernandez-Ahumada, E.; Palagos, B.; Roger, J.-M.; McBratney, A. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *TrAC Trends Anal. Chem.* **2010**, *29*, 1073–1081. [CrossRef]

44. Viscarra Rossel, R.A.; Behrens, T. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma* **2010**, *158*, 46–54. [CrossRef]

45. Mouazen, A.M.; Al-Asadi, R.A. Influence of soil moisture content on assessment of bulk density with combined frequency domain reflectometry and visible and near infrared spectroscopy under semi field conditions. *Soil Tillage Res.* **2018**, *176*, 95–103. [CrossRef]

46. Mouazen, A.M.; Kuang, B.; De Baerdemaeker, J.; Ramon, H. Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy. *Geoderma* **2010**, *158*, 23–31. [CrossRef]

47. Slaughter, D.C.; Pelletier, M.G.; Upadhyaya, S.K. Sensing soil moisture using NIR spectroscopy. *Appl. Eng. Agric.* **2013**, *17*, 241. [CrossRef]

48. Qiao, Y.; Zhang, S. Near-infrared spectroscopy technology for soil nutrients detection based on LS-SVM. In *IFIP Advances in Information and Communication Technology*; Springer: Berlin/Heidelberg, Germany, 2012; Volume 368 AICT, pp. 325–335.

49. Kodaira, M.; Shibusawa, S. Using a mobile real-time soil visible-near infrared sensor for high resolution soil property mapping. *Geoderma* **2013**. [CrossRef]

50. Kweon, G.; Lund, E.; Maxton, C. Soil organic matter and cation-exchange capacity sensing with on-the-go electrical conductivity and optical sensors. *Geoderma* **2013**, *199*, 80–89. [CrossRef]

51. Liu, Y.; Pan, X.; Wang, C.; Li, Y.; Shi, R. Predicting Soil Salinity with Vis–NIR Spectra after Removing the Effects of Soil Moisture Using External Parameter Orthogonalization. *PLoS ONE* **2015**, *10*, e0140688. [CrossRef] [PubMed]

52. Ji, W.; Viscarra Rossel, R.A.; Shi, Z. Accounting for the effects of water and the environment on proximally sensed vis-NIR soil spectra and their calibrations. *Eur. J. Soil Sci.* **2015**, *66*, 555–565. [CrossRef]