A novel non-integer order Savitzky–Golay derivative function of visible and near-infrared spectra for improving prediction accuracy of phosphorus in pig manure

Jian Zhang, Abdul M. Mouazen*

Department of Environment, Ghent University, Coupure Links 653, 9000 Gent, Belgium
*Corresponding author. Email: Abdul.Mouazen@UGent.be

Abstract

Prediction of phosphorus (P) content in manure using visible and near infrared (vis-NIR) spectroscopy is essential for better manure management. However, accurate prediction requires better spectra pre-processing, out of which spectra derivative is frequently used. This study presents the development of a non-integer order Savitzky-Golay derivative function (NISGDF) of vis-NIR spectra for the analysis of P composition in manure. A formula for a NISGDF is developed and presented in the form of a matrix. Next, both quantitative and qualitative analyses were performed using developed software packages, all of which can be easily called by a command line. Simulated absorption bands using the Gaussian model were operated by NISGDF to assess its abilities in reduction of noise and elimination of baseline and offset. A total of 110 pig manure samples were collected and scanned by a fibre type vis–NIR spectrophotometer in the range of 305 - 1700 nm. The collected manure spectra were subjected to the NISGDF, before they were used to establish the partial least squares regression (PLSR) model for P. Results indicate that the NISGDF offers significant flexibility in interpolating between the integer order SG derivatives, and reduces baseline offsets and tilts compared to conventional integer order Savitzky-Golay derivative function (ISGDF). The PLSR prediction accuracy with NISGDF was improved by 5-8%, compared with the corresponding PLSR model using the traditional ISGDF. This work shows that the NISGDF offers the advantage of wider applicability and better performance for the prediction of manure P, hence, it is recommended as a more general and better algorithm than the traditional ISGDF for the prediction of P.

Keywords:

Manure management, vis-NIR spectroscopy, spectra-pre-processing, phosphorous.

1.Introduction

Although it is an excellent source of major plant nutrients as nitrogen (N), phosphorus (P), potassium (K), and organic matter (OM), the improper management of manure can result in environmental risks including N leaching to groundwater, P run off and ammonia emission to the air. It is estimated that there are1.6 billon pigs, cattle, and chicken in Europe, which generate 1,400 million tons of manure per year [1]. The challenge of reducing the environmental and economic effects caused by such an enormous volume of manure will be highly important and decisive in manure management. Manure management from manure generation to final use generally includes the phases of housing (feeding), storage, treatment, transport, and spreading [1]. In all of these phases, the determination of manure compositions is crucial to optimize its management [2]. However, traditional wet chemical techniques e.g., the Kjeldahl procedure [3], are time consuming, difficult and costly, and generate hazardous wastes [4]. Moreover, the reliable estimation of the nutrient concentrations in liquid manure could be problematic due to the difficulty in obtaining representative samples in storage facilities, even with traditional chemical tests [5]. Although collecting multiple subsamples from one storage device and combining them into a single composite sample for laboratory analysis is an alternative solution [6], manure nutrient variability needs also be observed during storage-unloading, and pumping up, because the solid and liquid fractions in manure tend to separate during storage [7]. Thus, high-density sampling would be helpful to offer more reliable assessment of nutrient composition, resulting in better management.

Considering the 11 million tons of N in mineral fertilizers currently applied to crops in Europe [1], substituting manure in part could substantially decrease environmental risks. Essentially, all chemical fertilizer and feed P are derived from phosphate-rich rocks which are located in a few places on Earth and are finite. Europe is highly dependent on imports of P and K fertilizers as Europe has no significant P mines [8]. Above all, the recovery of these nutrients from manures to soil would be desirable. Both N and P macronutrients are essential for crop growth and development, and they can be provided by organic fertilisers including manure. There are two major forms of N in manure, namely, inorganic (ammonium) N and organic N. Ammonium N, found in manure in the form of urea is not different from the urea synthetic fertilizer, and is available for crop uptake as soon as it is applied to the soil [9], whereas the organic N is not available to plant. Likewise, as the second most important macronutrient that limits plant growth, development and productivity [10], P can be used by the plants when applied by manure. During the calculation of manure application rate, either N or P content in manure is determined by means of a composite sample collected from the manure storage. However, this composite sample does not represent the actual variability exist in the storage place, or even during manure application in the field. All these issues imply the need for rapid and cost-effective methods for the measurement of manure nutrient composition.

Among available spectroscopy methods, the visible and near infrared spectroscopy (vis-NIRS) is widely used in the agricultural and food industries, including among others applications in soil [11], manure [12], plants [13], and fruits [14]. In manure, vis-NIRS was reported as a successful tool to predict key parameters [15]-[16]. The justification of successful results reported for some constitutes e.g., total N, dry matter (DM), and organic matter (OM) is that they have direct spectral responses in the NIR spectral range associated with molecule vibrations (e.g., N-H, C-H, and O-H) [4]. The constituents that are not spectrally active but display covariation through other properties that have direct spectral responses in the NIR are e.g., Phosphorus (P) and Magnesium (Mg). This is the reason why the latter group of constituents can be predicted with NIRS but with less accuracy, compared to the former group. Previous research illustrates that only few reports on the determination of P in swine manure by vis–NIRS are available [4], [17]-[18]. In these studies, the Savitsky-Golay (SG) first or second order derivatives, SG smoothing, and multiplicative scatter correction were applied for spectral pre-treatment combined with partial least square regression (PLSR) to establish predictive models of P. Therefore, there is a need to use advanced spectra pre-processing and modelling methods to maximise the accuracy of predicting the manure constitutes having indirect spectral response e.g., P.

The SG differential filter operating based on a localized linear regression of several neighbouring points with a moving window is a popular algorithm for spectra derivative. It is used to determine a best fit polynomial, whereupon this polynomial can be mathematically differentiated and evaluated at values coincident with wavelength collection points [19]. Traditionally, SG differential filters are of integer orders, e.g., first and second orders. When the order is set as zero, the SG derivative is degraded to a SG smoothing filter. SG smoothing and differentiation filters can be used to reduce high frequency noise in spectra due to their smoothing feature and reduce low frequency signals using differentiation, respectively, thereby improving the contrast of relevant spectral features. Fractional order derivatives (FOD) are generalizations, in which integer-order derivatives are only a subset of FOD that allows significant flexibility. Since the Grünwald-Letnikov (G-L) formula is the simplest FOD calculus, few studies have evaluated its potential in the pre-processing of NIR spectra by directly processing the entire spectrum without moving window [20]-[22]. Although a decimal order SG derivative based on the Riemann–Liouville (RL) calculus was also presented for noisy-signal processing [23]-[24], it would cause counter-intuitive result that the derivative of the constant term in interpolation polynomials is non-zero when derivative order is larger than 1. Furthermore, to the best of our knowledge, there is currently no work in the literature that explores the potential use of NISGDF in pre-processing of vis-NIR manure spectra to predict spectrally inactive constitutes like P in manure. The aim of this work is to evaluate the performance of NISGDF pre-treated vis-NIR

spectra-based PLSR model in the prediction of P in manure, in comparison with the conventional integer-order SG filters based PLSR model.

2. Materials and methods

2.1 Non-integer order SG filter

In the present work, the analytical formulas in the form of matrix are firstly deduced to calculate any non-integer order SG filter based on a Riemann–Liouville (R-L) fractional derivative calculus, which is a popular fractional derivative calculus. The R-L fractional derivative is the most important extension of ordinary calculus. Almost all other definitions of the fractional derivative represent a special case of the Riemann-Liouville fractional derivative [25]. A α^{th}-order R-L calculus is defined as follows:

$$D_x^{\alpha} f(x) \triangleq D^h\left(D_x^{\alpha-h} f(x)\right) = \frac{1}{\Gamma(h-\alpha)} \frac{d^h}{dx^h} \int_{\beta}^{x} \frac{f(t)}{(x-t)^{\alpha-h+1}} dt \quad (1)$$

where $h = \lceil \alpha \rceil$, $\lceil \cdot \rceil$ denotes rounding up to the next higher integer, and $\Gamma(z) \triangleq \int_0^{\infty} x^{z-1} e^{-x} dx$ is a gamma function. Let $f(x) = x^v$, and submitting it into (1) yields:

$$D_x^{\alpha}(x^v) = \frac{\Gamma(v+1)}{\Gamma(v+1-\alpha)} x^{v-\alpha} \quad (2)$$

In this work, the vis-NIRS operates in the spectral range of 305 -1700 nm with a periodic interval $T = 1$ between adjacent pixels. To pre-process such a spectrum in a moving filtering window of size $2k + 1$ centered at the point j ($j \in$ [305+kT, 1700-kT]), a classic SG smoothing filter uses a n-order polynomial to fit $2k$ +1 points in the window j is used and expressed by the following formula:

$$\hat{\boldsymbol{y}}_{\boldsymbol{j}} = \mathbf{B}(\mathbf{B}^{\mathbf{T}}\mathbf{B})^{-1}\mathbf{B}^{\mathbf{T}}\boldsymbol{y}_{\boldsymbol{j}} \quad (3)$$

where $\mathbf{B} = \begin{bmatrix} 1, & \cdots & ,(T)^n \\ \vdots & \ddots & \vdots \\ 1, & \cdots & ,(2k+1)^n T^n \end{bmatrix}$, $\boldsymbol{y}_{\boldsymbol{j}} = \left[y_{j-kT}, \cdots, y_{j+kT}\right]^T$ are spectral (reflectance, or absorbance) values at the points $(j - kT)$ nm, $\cdots$,$(j + kT)$nm, $\hat{y}_j$ in $\hat{\boldsymbol{y}}_{\boldsymbol{j}} = \left[\cdots, \hat{y}_j, \cdots\right]^T$ is the estimated value for

the point j. The above SG smoothing filter can also be regarded as zero-order SG differential filter. Under the same formulation, a d^{th}- order (d is an integer) SG differential filter can be represented by the following formula:

$$\widehat{\boldsymbol{y}_j^d} = \mathbf{M_d}(\mathbf{B^T B})^{-1}\mathbf{B^T}\boldsymbol{y}_j \quad (4)$$

where $\widehat{y_j^d}$ in $\widehat{\boldsymbol{y}_j^d} = \left[\cdots, \widehat{y_j^d}, \cdots\right]^T$ is the estimated SG d^{th} - order differential value at the point j nm, and $\mathbf{M_d}$ is denoted as:

$$\mathbf{M_d} = \begin{bmatrix} \overbrace{0, \cdots, 0}^{d\ columns}, \cdots, & P_{i-1}^d(T)^{i-1-d}, & \cdots, & P_n^d(T)^{n-d} \\ 0, \cdots, 0, \cdots, & P_{i-1}^d(2T)^{i-1-d}, & \cdots, & P_n^d(2T)^{n-d} \\ \vdots \quad \vdots \quad \vdots \cdots, & \vdots & \vdots & \vdots \\ 0, \cdots, 0, \cdots, & P_{i-1}^d((2k+1)T)^{i-1-d}, & \cdots, & P_n^d((2k+1)T)^{n-d} \end{bmatrix} \quad d+1 \le i \le n+1 \quad (5)$$

where P_i^d is a d-permutations of i, i.e. $P_i^d = i \times (i-1) \times \cdots \times (i-d+1),\ i \ge d$.

By combining the definition of R-L calculus (1) with the formulas (4) and (5), an analytical formula in the form of matrix is deduced to calculate both integer and any positive non-integer order SG differentiators. A α^{th}-order (α>0) derivate at the point j can be calculated in a moving filtering window by the following formula:

$$\widehat{\boldsymbol{Y}_j^\alpha} = \mathbf{M_\alpha}(\mathbf{B^T B})^{-1}\mathbf{B^T}\boldsymbol{y}_j \quad (6)$$

where $\widehat{y_j^\alpha}$ in $\widehat{\boldsymbol{y}_j^\alpha} = \left[\cdots, \widehat{y_j^\alpha}, \cdots\right]^T$ is the estimated SG α^{th}-order differential value at the point j, $\mathbf{M_\alpha}$ is an $(2k+1) \times (n+1)$ matrix, written as follows:

$$\mathbf{M_\alpha} = \begin{bmatrix} \overbrace{0 \quad \cdots \quad 0}^{\lfloor\alpha\rfloor\ columns}, \cdots, & P_{i-1}^{\lfloor\alpha\rfloor} \cdot \frac{\Gamma(i-\lfloor\alpha\rfloor)}{\Gamma(i-\alpha)} \cdot (T)^{i-1-\alpha} & \cdots, & P_n^{\lfloor\alpha\rfloor} \cdot \frac{\Gamma(n+1-\lfloor\alpha\rfloor)}{\Gamma(n+1-\alpha)} \cdot (T)^{n-\alpha} \\ \vdots \quad \vdots \quad \vdots, \cdots, & \vdots & \cdots, & \vdots \\ 0 \quad \cdots \quad 0, \cdots, & P_{i-1}^{\lfloor\alpha\rfloor} \cdot \frac{\Gamma(i-\lfloor\alpha\rfloor)}{\Gamma(i-\alpha)} \cdot \left((2k+1)T\right)^{i-1-\alpha} & \cdots, & P_n^{\lfloor\alpha\rfloor} \cdot \frac{\Gamma(n+1-\lfloor\alpha\rfloor)}{\Gamma(n+1-\alpha)} \cdot \left((2k+1)T\right)^{n-\alpha} \end{bmatrix} \quad (7)$$

and $\lfloor\cdot\rfloor$ denotes the rounding down to the next lower integer of α. Since $P_{i-1}^{\lfloor\alpha\rfloor} \cdot \frac{\Gamma(i-\lfloor\alpha\rfloor)}{\Gamma(i-\alpha)} \cdot (T)^{i-1-\alpha} = P_{i-1}^d(T)^{i-1-d}$ when $\alpha = d$, the above formula (7) covers the conventional integer-order filters in (5),

e.g., the SG smoothing filter ($\alpha = d = 0$), the first order SG derivative ($\alpha = d = 1$), and the second order SG derivative ($\alpha = d = 2$). This means that the integer order can be considered as the special cases of the formula (6) and (7) when α is of integer order. Therefore, the NISGDF is more flexible and generalized than the classical ISGDF. The zero values in the first-column in $\mathbf{M}_{\boldsymbol{\alpha}}$ would cause non-integer order derivatives of constant terms in the interpolation polynomial to have zero values if derivative orders are greater than 1. Similarly, both the constant and first terms of a polynomial would be zero, since the first two columns of $\mathbf{M}_{\boldsymbol{\alpha}}$ are zero, when using a non-integer order derivative larger than the second order. It is intuitively consistent with integer-order derivatives.

Curve fitting techniques applied to reflectance spectra are invaluable resources to spectral modelling, allowing both the mathematical description of the band shapes and the qualitative and quantitative estimation of the absorption components, at least in some cases [26]-[27]. The absorption bands can be fitted using Gaussian, modified Gaussians, Lorentz, or mixed Gaussian-Lorentz (Voight) shapes [28]-[29], when they are analysed as log reflectance versus photon energy (wavelength). Therefore, the reflectance spectra are converted to absorbance by taking the logarithm of the reflectance, as follow:

$$\boldsymbol{a_b} = log_{10}(1/\boldsymbol{r}) \quad (8)$$

If we consider the absorption bands to be Gaussian in shape, a series of absorption bands may be modelled by the following equation:

$$g(\omega) = l \cdot \left(\frac{\omega}{\omega_0}\right)^m \left(e^{-\frac{(\omega-\omega_0)^2}{\sigma^2}} + s\right) + p(\omega) \quad (9)$$

Where l is the magnitude of the Gaussian, ω_0 *nm* is the peak wavelength, σ *nm* is the width of the band, m and s are the power-law slope and constant offset of the baseline. The motivation for using the function $\left(\frac{\omega}{\omega_0}\right)^m$ to represent the wavelength dependence of the baseline stems from the observation that the spectral baselines of a wide variety of particulate media can be fitted well by power curves [29]. To investigate the benefits of NISGDF, a set of simulated Gaussian absorption bands were used to evaluate the performance of NISGDF, compared with conventional ISGDF.

The formulas (6) and (7), presented above, were incorporated in code packages used to implement the NISGDF including the SG smoothing, the SG first order derivative and the SG second order derivative. They were developed with three languages, i.e. MATLAB, R, and Python, all of which can be easily called by the following command line:

```
nisgdf(spectrum, difforder,windowwidth,polynomialorder)
```

where `spectrum` is the processed spectrum, `difforder` denotes the integer or non-integer order α of the derivative, `windowwidth` means the size of a moving filtering window, and `polynomialorder` is the polynomial order n. The Matlab code can be obtained from the corresponding author on request.

2.2 Manure sampling and spectroscopic measurements

To demonstrate the potential of modelling manure compositions based on vis-NIRS, 110 pig manure samples were collected from different farms across Flanders, Belgium in 2019 and 2020. Each sample was divided into two parts. One part was sent to Eurofins (Eurofins Agro Testing NV, Nazareth, Belgium) for chemical analysis by autoanalyzer spectrophotometry method, whereas the other part was stored in a cooling room at 5°C until measurement by the vis-NIR sensor in laboratory few weeks later. Experience has shown that samples can be stored in this way for more than one year with minor changes in the components [4]. These samples were scanned in the laboratory using a vis-NIRS (TEC5 CompactSpec® by Tec5 AG, Germany). The spectrophotometer has a wavelength range of 305-1700 nm and a light source of 5 Watt with 45° illumination angle. The manure samples of 170 ml each were placed in a home-made glass cups (60 mm in diameter, and 6 cm high), which were scanned from the bottom in reflectance mode. Before pouring the manure into the cup, it was manually stirred and stirring was resumed between successive measurements to prevent solid particles from settling on the bottom of the measurement cup. Five measurements of two scans each (4 seconds per scan) were taken and these ten measurements were averaged in one spectrum for each sample. The cups were carefully cleaned and dried before the next sample was measured.

Calibration model for P content was developed, after an integrated spectra pre-treatment, using a PLSR analysis carried out by MATLAB. The pre-treatment of vis-NIRS starts with removing the spectral jump at 1045 nm, where the two detectors of vis-NIRS are joined, following the correction algorithm suggested by Mouazen et al. (2009) [30]. The noisy part of the spectra at the two ends was then removed. The following successive pre-treatment steps were conducted: 1) averaging neighbouring wavelengths with the moving average algorithm, 2) normalizing spectra to a range of -1 and 1, 3) calculating the derivative spectra by ISGDF and NISGDF ($T = 1$), respectively, and 4) performing the SG smoothing. To compare the performances of ISGDF and NISGDF, all parameters,

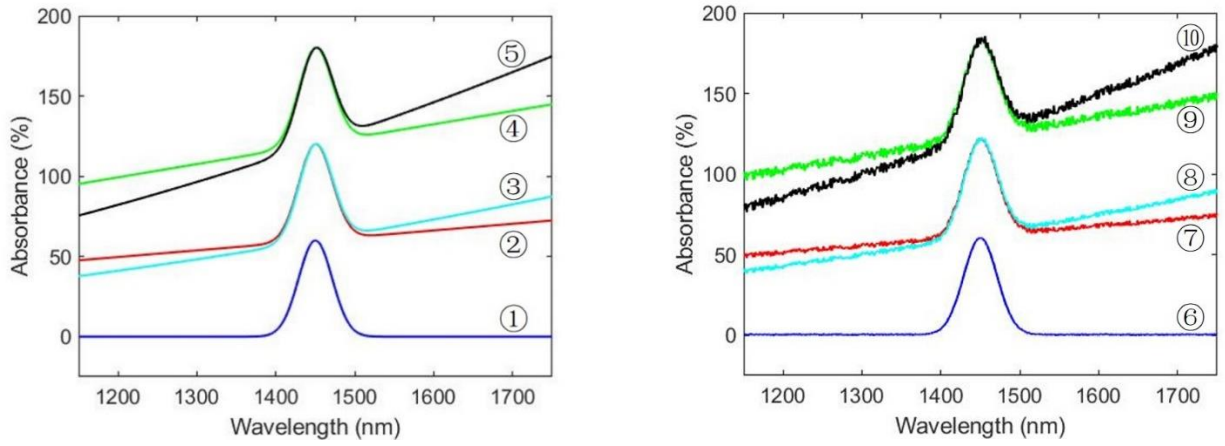
including the window size of moving average, the normalization range, and the number of latent variables of PLSR were set to the same values except the order of NISGDF and ISGDF.

PLSR cross-validation and independent validation were performed on samples obtained from the 110 manure samples, by dividing the entire dataset into calibration (70%) and validation (30%) sets. PLSR was carried out on the calibration set with a leave-one-out cross-validation method to develop the calibration model. The developed calibration models were validated using the independent validation set. The prediction performance of the developed models was evaluated using the coefficient of determination (R^2), root means square error (RMSE), ratio of performance to inter-quartile distance (RPIQ), and residual prediction deviation (RPD), which was calculated as the ratio of the standard deviation divided by RMSE.

3. Results and discussions

3.1 Performance of NISGDF in pre-processing vis-NIR spectra

Based on formula (9), an individual Gaussian-type absorption band can be simulated under different parameter settings. The simulation shows that the peak wavelength ω_0 was at 1450 nm, the width of the band σ was 30 nm, and l was 0.6. Fig. 1 depicts a set of simulated spectra in the range of 1150



(a) (b)

①and ⑥: m=0, d=0, ② and ⑦: m=1, d=1, ③ and ⑧: m=2, d=1, ④ and ⑨: m=1, d=2, ⑤ and ⑩: m=2, d=2.

Fig. 1 Simulated spectra for spectra with absorbance peaks shown. (a) Simulated Gaussian-shape bands without noise and (b) Simulated Gaussian with noise. Data are shown for different combinations of power-law slope (m) and offset (D)

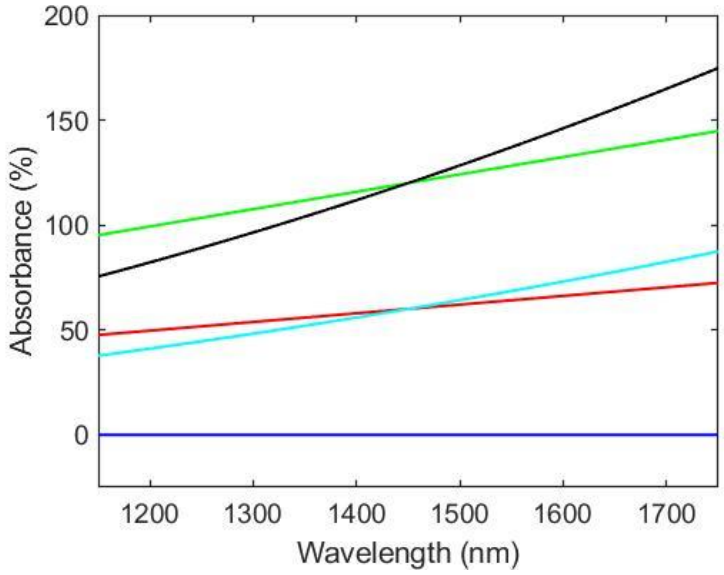
to 1750 *nm* that were obtained from different offsets under different power exponents without (Fig. 1 (a)) and with (Fig.1 (b)) random noise. Since white noise was present in almost all analytical

measurements and is often attributed to the photometric noise in the detector system, the white noise with a root mean squared (RMS) equal to 3% was added prior to pre-processing.

Since the offset *s* dominates the peak value of spectrum, the sensitivity of the peak value to the offset can be quantified as $\frac{\Delta\, Peak}{\Delta\, s}$. In the case of Fig.1 (a), this factor is approximately equal to 0.6, i.e. $\frac{\Delta\, Peak}{\Delta\, s} \approx l$. Unlike the peak feature, the baselines depend on both the offset *s* and the low-power exponent *m* as shown by the following equation:

$$f(\omega) = l \cdot s \cdot \left(\frac{\omega}{\omega_0}\right)^m \quad (10)$$

Based on Eq. (10), the baselines for the spectra ①-⑤ can be depicted in Fig. 2.



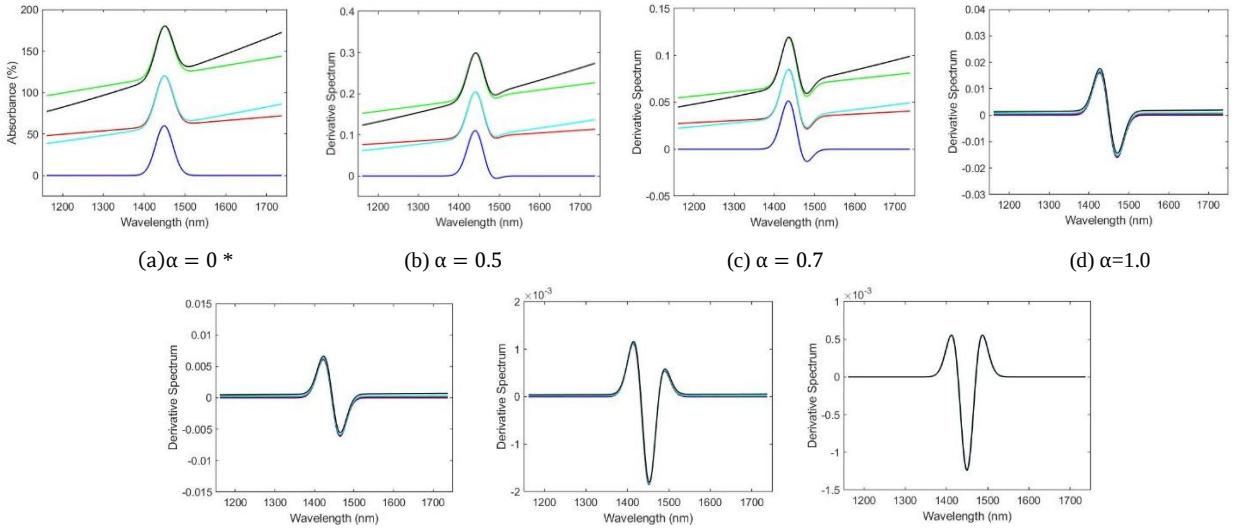
Blue: ① and ⑥, Red: ② and ⑦, Cyan: ③ and ⑧, Green: ④ and ⑨, Black: ⑤ and ⑩

Fig. 2 Baselines of simulated absorbance spectra for the same spectra as those of Fig.1.

By using the developed software package in this work, different orders of derivatives can be obtained for the curves ①-⑤ in the Fig.1 (a). The parameters set are: `window=25` and `polynomialorder=2`. The typical derivative curves are shown in Fig. 3 for $0 \leq \alpha \leq 2$ where the differentiated absorbance peaks are transformed smoothly and gradually from their original shape (Fig. 3 (a)) to a bipolar shape at α=1 (Fig. 3 - (d)) and finally into a sharp inverted peak with two smaller side lobes at α= 2 (Fig. 3 (g)). These results illustrate the ability of the NISGDF to interpolate between integer-order derivatives.

It can be observed from Fig. 3 that as the order α of the fractional derivative increases from 0 to 2, the offset and slope of the baseline progressively diminish. All of them are approximately reduced to 0 when $\alpha = 2$. In order to describe the ability of NISGDF to suppress the offsets, the offset sensitivity

$\frac{\Delta\, Peak}{\Delta\, s}$ was calculated as a function of the derivative order α and plotted in Fig. 4. Results show that the higher the order of the fractional derivative within the range of $0 \leq \alpha \leq 2$, the fewer effects result from the additional of an undesired offset onto a spectrum. For derivative orders greater than 0.5, the



(a)α = 0 * (b) α = 0.5 (c) α = 0.7 (d) α=1.0

(e) α = 1.3 (f) α = 1.8 (g) α = 2 **

① and ⑥ (Blue): *m*=0, *s*=0, ② and ⑦ (Red):*m*=1, *s*=1, ③ and ⑧ (Cyan): *m*=2, *s*=1, ④ and ⑨ (Green): *m*=1, *s*=2, ⑤ and ⑩ (Black): *m*=2, *s*=2.

Fig. 3. Non-integer order derivatives of the simulated Gaussian-shape bands, shown for different order of the fractional derivative (α) for same spectra in Fig. 1. *α = 0 means the Savitzky-Golay (SG) smoothing filter. **Five curves coincide with each other. Data are shown for different combinations of power-law slope (m) and offset (s)

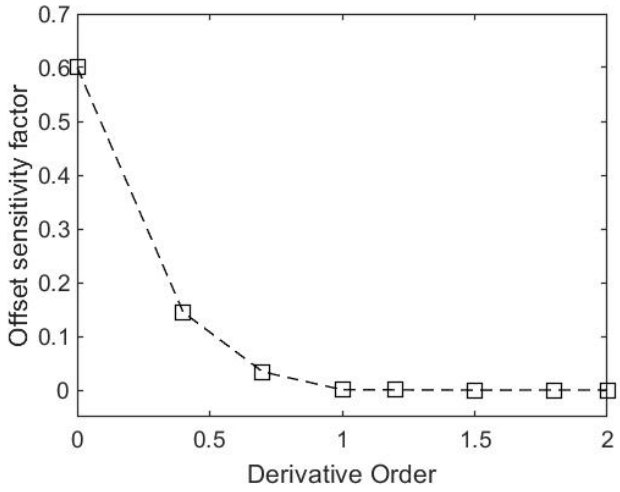
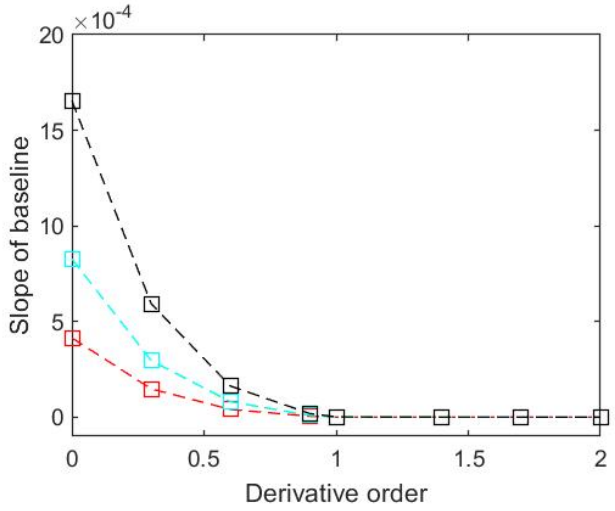


Fig. 4. Sensitivity of peak magnitudes to offsets



Red (② and ⑦): m=0, s=0, Cyan (③ and ⑧): m=1, s=1, Green (④ and ⑨)*: m=1, s=2, Black (⑤ and ⑩): m=2, s=2.

Fig 5. Residual slopes of baseline resulted from implementation of the non-integer Savitzky-Golay derivative function (NISGDF). Data are shown for different combinations of power-law slope (m) and offset (s). *The cyan and green curves coincide with each other.

offset sensitivity falls to less than 10% of that of the original spectrum. In addition to their reduced sensitivity to a constant offset, non-integer derivative spectra also have, in general, flatter baselines because differentiation reduces spectral slopes gradually. Since the SG differentiation is a derivative of a least-squares fitting polynomial, the residual baseline slope of a baseline by differentiation is not only dependent on the low-power exponent *m* but also on the polynomial order used for fitting in moving windows. This is different from non-SG type integer- and even non-integer order differentiation that does not use a moving window and a local least-square fitting polynomial [20]-[21]. Since a non-SG type differentiation imposes directly on spectral data instead of a fitting polynomial, the target baseline to be reduced would be dependent on raw data only. Regardless of the order of the polynomial chosen, conventional ISGDF are either first or second order derivatives to reduce the slope of baseline. By contrast, the NISGDF offers more options by selecting any order of derivative to handle different polynomial orders. Fig. 5 is a plot of the residual slopes of the baselines obtained by the SG differentiators of different orders. It shows that the residual baseline slopes decrease approximately exponentially to near zero as the derivative order increases. This is the reason why it can be concluded that NISGDF provides more flexibility under more order options than ISGDF to approximately remove slope baselines.

Another characteristic of NISGDF is that it enables users to shift the positions of the peaks and zero-crossings of the derivative spectrum in the vicinity of peaks. Table 1 illustrates an increase in the order of the differentiation produces a systematic shift in the peak of the spectrum ①. It shows that the shift of the peak varies almost linearly with the derivative order and the wavelength at which the

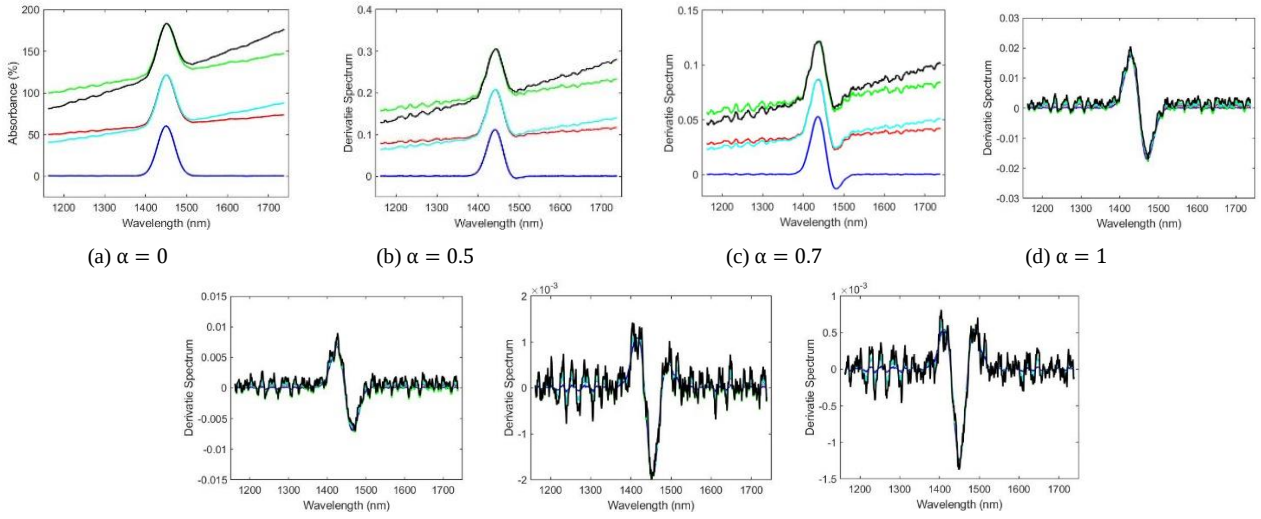
Table 1. The peak shift systematically with respect to the non-integer Savitzky-Golay derivative function (NISGDF) order α.

Derivative order	0	0.2	0.4	0.6	0.8	1	1.2	1.4	1.6	1.8	2
Wavelength of peak (nm)	1450	1447	1444	1439	1434	1428	1425	1422	1418	1415	1412

peak of the derivative spectrum can be adjusted anywhere between the zero-order spectral peak and the second order spectral peak.

To demonstrate the performance that the NISGDF can offer in the analysis of noisy spectra, Fig. 6 illustrates the derivatives of the same example spectra used in Fig.1 (b), where the white noise is added to the ones in Fig.1 (a) so that the noise-to-signal ratio (NSR) is 3%. It can be observed that the white noise has little effect on the quality of the spectra, but with increasing noise, respectively, for ⑥ ($\alpha \leq 1.3$), for ⑦ and ⑧ ($\alpha \leq 0.7$), and for ⑨ and ⑩ ($\alpha \leq 0.5$), however, it destructs all the spectra when $\alpha \geq 1.8$ significantly.

Since the NSR was kept as 3%, the raw spectra with larger absorption amplitudes were coupled with larger noise as depicted in Fig. 1 (b). This explains why spectra ⑨ and ⑩ contain relatively greater



(a) α = 0 (b) α = 0.5 (c) α = 0.7 (d) α = 1

(e) α = 1.3 (f) α = 1.8 (g) α = 2

Blue (⑥): *m*=0, *s*=0, Red (⑦): *m*=1, *s*=1, Cyan (⑧): *m*=2, *s*=1, Green (⑨): *m*=1, *s*=2, Black (⑩): *m*=2, *s*=2.
Polynomial order: 3, window width: 25.

Fig. 6. The non-integer Savitzky-Golay derivative function (NISGDF) of spectra with white noise added, shown for the same spectra in Fig. 1. Data are shown for different combinations of power-law slope (*m*) and offset (*s*)

Table 2. Noise-to-signal ratio (NSR* calculated with the non-integer Savitzky-Golay derivative function (NISGDF) for different orders α between 0 and 2, shown for the same spectra in Fig. 1.

	Raw	α = 0	α = 0.5	α = 0.7	α = 1	α = 1.3	α = 1.8	α = 2
Spectrum ⑥	3%	2.57%	2.48%	2.42%	2.76%	4.91%	6.90%	6.28%
Spectrum ⑦	3%	2.61%	2.77%	3.42 %	12.32%	21.92%	30.88%	28.11%
Spectrum ⑧	3%	2.62%	2.79%	3.48 %	23.05%	41.07%	58.26%	53.10%
Spectrum ⑨	3%	2.62%	2.76%	3.38%	12.56%	22.37%	31.71%	28.90%
Spectrum ⑩	3%	2.62%	2.77%	3.43%	22.95%	41.09%	59.82%	54.76%

* NSR= (root mean square of noise signal after NIGSF implementation) / (root mean square of noise-free signal after NIGSF)

noise than spectra ⑥, ⑦, and ⑧ under the same order of derivative (Fig. 6). On the other hand, Table 2 depicts the variations in NSR with respect to the derivative order. It can be observed that the NISGDF reduces the NSR for all the noisy spectra when α is in the range of 0 to 0.5, presenting more smoothing effects, which is similar to the effect of the SG smoothing filter at $\alpha = 0$. At $\alpha = 0.7$, the spectrum ⑥ without slope, has the minimum value of NSR. The baseline has a negative influence, which causes a larger increase in the NSR of the spectra ⑦-⑩ along with the derivative order. For example, both ⑧ and ⑩ are almost drowned in noise when $\alpha \geq 1.8$ to the extent that it is impossible to separate useful information from the derivative results. Mathematically, this can be explained by the fact that the NISGDF has a similar filtering effect on noise to that of the SG smoothing filter when α is close to zero and that the NISGDF with higher derivative orders amplifies the RMS when $\alpha \geq 1$. Since the NISGDF lowers the spectral slope as discussed above, the RMS of the noise-free signal decreases as the slope decreased with the increased order of the NISGDF. This results in a larger increase in the NSR for spectra ⑦-⑩ than for the spectrum ⑥ without the baseline.

In the choice of the order of the derivative to use, it follows from the above analysis that noise immunity must be traded off with a reduction in the offset and NSR of the derivative spectrum. As the resulting spectrum varies continuously with the order of the derivative, a better order can be searched between a lower order corresponding to a less noisy spectrum and a higher one corresponding to a spectrum with free baseline.

3.2 Manure spectral and chemical characteristics

Fig.7 (a) and Fig.7 (b) show both the raw reflectance spectra and the transformed absorption spectra respectively. The derivative spectra at different orders α are presented in Fig. 7 (c) to Fig.7 (i). As reported by Saeys et al. (2004) [17], some manure properties e.g., P, Mg, and calcium (Ca) are not spectrally active in the NIR range. Their NIR prediction relies on correlations they have with other manure properties

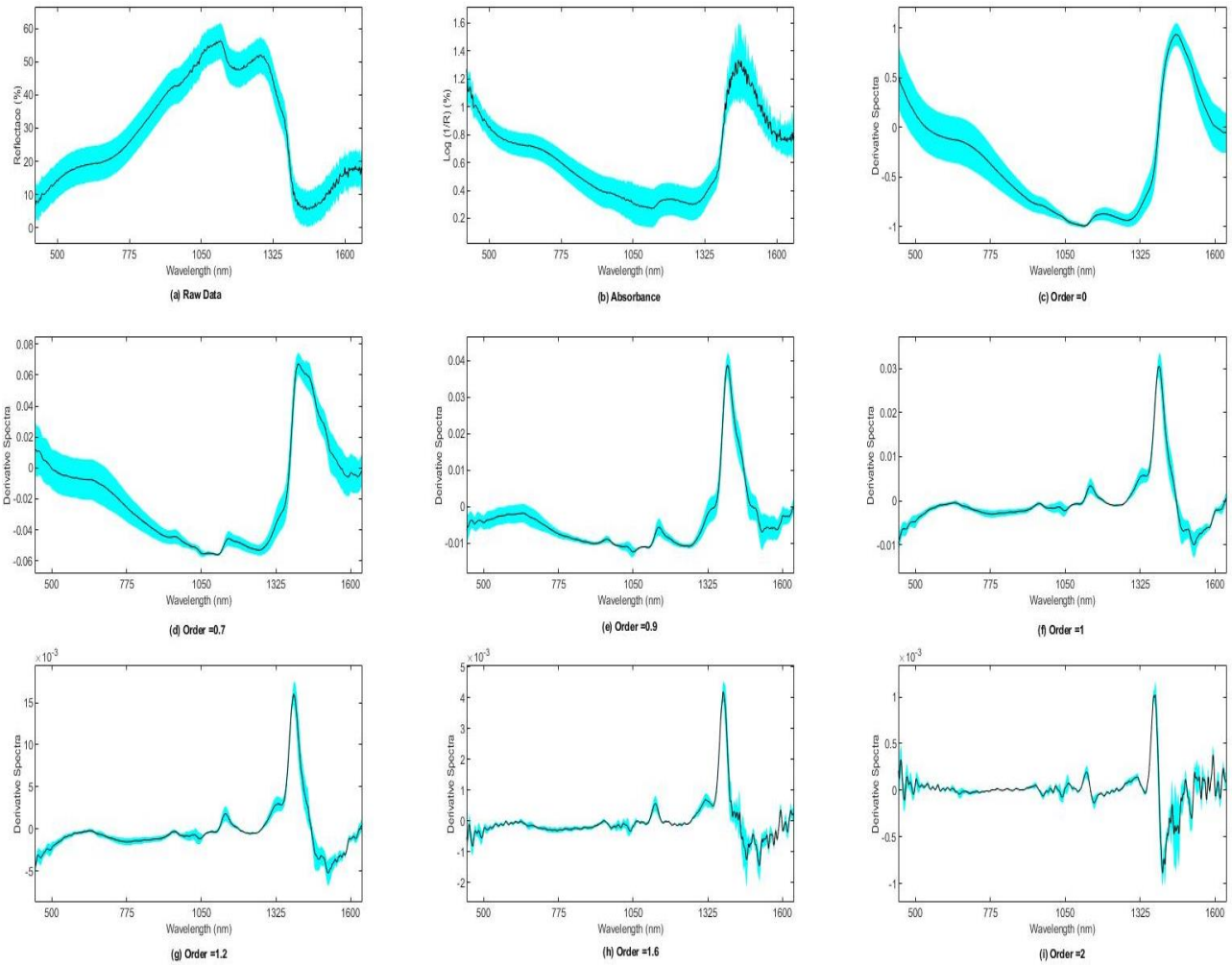


Fig.7 Non-integer order derivative spectra of 110 manure samples used in this study. Cyan areas represent the standard deviations of the spectra and the black curves are the mean of spectra of the 110 samples.

such as DM, total N and organic matter (OM), which are spectrally active, and this is due the direct molecular vibrations of e.g., C–H, O–H or N–H bonds. Researchers (Saeys et al. [17], Marino et al. [31], and Martinez-Suller et al. [32]) reported good linear relationships between P and DM in animal manure. Since the DM content in manure can usually be accurately predicted, the prediction of constituents correlated with DM content is expected to rely mostly on O–H and C–H bonds. The coefficients of determination between the different constituents measured by the wet chemical analysis for the manure samples used in this work are listed in Table 3, which confirms that P is highly correlated with DM. A plausible explanation of the high correlation among P with DM is that these are constituents of phytate, a significant component of grains in the feed, which is indigestible for pigs unless the enzyme phytase is supplemented. As a consequence, P is supplemented to pig feed.

Table 3. Pearson correlation coefficients among the constituents as determined by wet chemical analysis

	Dry matter (kg/1000 kg)	Total N (kg/1000 kg)	P (g/kg)
Dry matter (kg/1000 kg)	1	0.79	0.88
Total N (kg/1000 kg)	0.79	1	0.65
P (g/kg)	0.88	0.65	1

Because of its dark brown colour, liquid pig manures absorb light in the whole region of visible wavelengths in the range of 400-780 nm, resulting in an increase in reflectance (Fig.7 (a)), due to a decrease in absorption (Fig. 7 (b)). The significant peak around 1450 nm is the result of the strong absorption of radiation by water molecules (Fig.7), which is attributed to the first overtone of the fundamental vibration of the O-H molecule bond [33].

When the derivative order increased from 0 to 1, the absorbance profiles exhibited continuous variations. The baseline offset effects were gradually decreased as the derivative order increased. An absorption feature at 1200 nm, became increasingly evident that can be attributed to the CH2 second overtone vibrations [34]. The 0.7-order NISGDF derivative sharpened the peaks and the valleys, and a relatively large offset and slope were observed, while the magnitude of the high-frequency noise in the derivative spectrum was still low. The 0.9-order derivative reduced the offset and slope substantially, without increasing the effects of the noise. With increasing the order from 0.7 to 2, the peaks and valleys sharpened further. For the second order spectra, the noise overwhelmed the spectral features and few smaller peaks appeared, which explains that the higher order derivative spectra were more susceptible to the interference of spectral noises, a conclusion that is in line with the discussion above. When the derivative order increased from 0.7 to 1, a weak peak at 950 nm was observed, which increased in size with the increase in the derivative order, until it is transformed into a small sharp peak in the first order derivative spectrum. It is associated by the third overtone of fundamental vibration of the O-H molecule bond [34]. This indicates that the non-integer derivative spectra can capture changes in spectra details and improve spectral curve resolution. However, an aggressive derivative with high order of 2 have to be used with caution, due to the introduction of noise. These results suggest that NISGDF derivative spectra have advantages over conventional ISGDF derivative spectra, since the former permits users to optimize the performance of the following regression model by adjusting the order of the derivative, with small intervals. The optimization can be achieved by trial-and-error, hence, it does not require detailed knowledge of the separate effects of interfering substances, background scattering, or high-frequency noise.

3.3 Effect of NISGDF on PLSR prediction of P

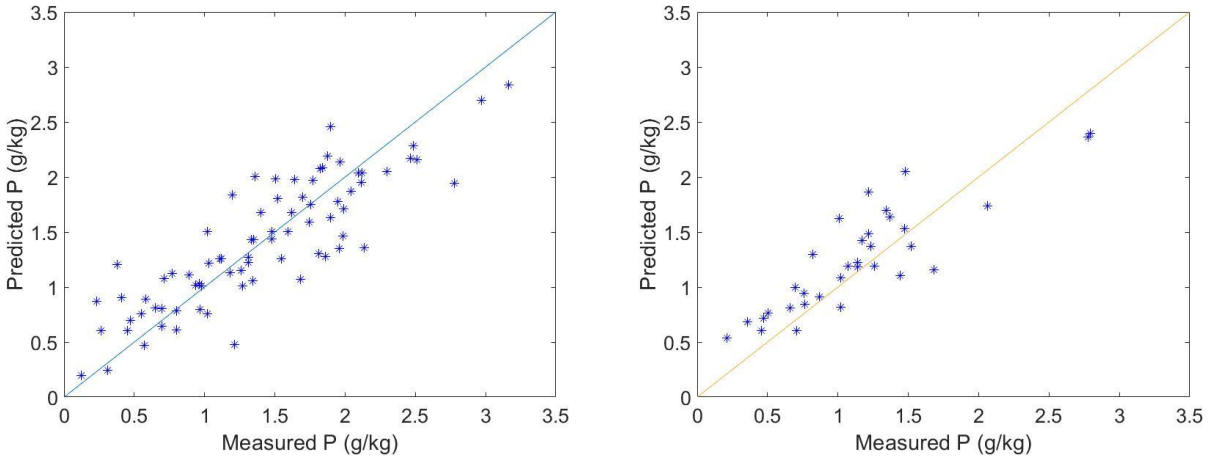
To investigate the influence of the NISGDF on model's prediction performance, PLSR models were developed for different derivative orders after pre-processing described in the Section 2.2. Table 4 shows the prediction results indicated as R^2, RMSE, and RPIQ. Since all of the parameters were set to the same values in the spectra pre-processing and PLSR modelling, the differences in model performance were entirely attributed to the different derivative order. Overall, the 1.2-order derivative spectra-based model outperformed the other models with R^2 values of 0.74 and 0.72, which is an increase of 0.04 for the calibration and validation sets respectively, compared with the first ISGDF derivative. These results prove that the NISGDF has a positive effect on model performance. The 1.1 order derivative exhibited similar performance to 1.2 order and is also better than the first order performance for the validation set. By contrast, the second order derivative produced the worst prediction performance for the validation set.

Table 4. Predictive results for phosphorous (P) estimation in manure using partial least squares regression (PLSR) established with seven derivative orders. Results are shown for calibration and validation sets.

	Calibration set				Validation set			
Order	R^2	RPD	RPIQ	RMSE (g/kg)	R^2	RPD	RPIQ	RMSE (g/kg)
0	0.56	1.52	2.15	0.44	0.51	1.45	1.58	0.41
0.6	0.65	1.70	2.41	0.40	0.63	1.67	1.82	0.35
0.9	0.68	1.77	2.51	0.38	0.63	1.66	1.81	0.35
1	0.70	1.83	2.60	0.37	0.68	1.81	1.97	0.33
1.1	0.70	1.85	2.62	0.36	0.70	1.84	2.01	0.32
1.2	0.74	1.97	2.80	0.34	0.72	1.90	2.08	0.31
1.6	0.74	1.97	2.79	0.34	0.57	1.55	1.69	0.38
2	0.72	1.90	2.69	0.35	0.55	1.51	1.65	0.39

Coefficient of determination (R^2), root means square error (RMSE), ratio of performance to inter-quartile distance (RPIQ), and residual prediction deviation (RPD). Very poor: RPD < 1.0, Poor: RPD = 1.0-1.4, Fair: RPD = 1.4-1.8, Good: RPD = 1.8-2.0, Very good: RPD = 2.0-2.5, Excellent: RPD>2.5, Very poor: RPIQ < 1.4, Fair: RPIQ = 1.4-1.7, Good: RPIQ = 1.7-2.0, Very good: RPIQ = 2.0-2.5, Excellent: RPIQ > 2.5.

As can be seen from Table 4, there is no linear dependency or qualitative relationship between derivative order and model performance. Increasing or decreasing the derivative order does not necessarily lead to an increase or decrease in model prediction performance, and this seems to be true for the same dataset. Perhaps this is dataset, material and parameter specific, meaning that an optimal derivative order may be dependent on dataset itself. Since the implementation of the NISGDF



(a) Calibration (b) Validation

Fig.8. Scatter plots of the measured phosphorous (P) content vs. P content predicted using the best performing 1.2-order derivative spectra-based PLSR model

algorithm is nearly as fast and simple as that of the conventional ISGDF, the order tuning can be optimized by repeated trial-and-error, and the selection of best order will rely on the order that leads to the largest R^2, RPD and RPIQ and the smallest RMSE values. This trend is similar to that of other pre-processing algorithms, e.g., the moving average (MA). A small window size of MA may be better for highly variable datasets to increase sensitivity to changes, while for noisy datasets a large window size should be considered to decrease noise. However, there is also no theoretical argument that permits the determination or calculation of an optimal moving window width.

The scatterplots of the measured P content compared with the P content predicted by the PLSR model based on the 1.2-order derivative spectra are illustrated in Fig. 8. The regression points for the results of the PLSR model based on the 1.2-order derivative spectra were well grouped along the 1:1 line. This distribution indicates a strong relationship between the measured and the predicted P. Compared to conventional first SG differentiation under the same parameter settings, 1.2-order derivative spectra improve the model accuracy by 5-8%, measured as R^2, RPD and RPIQ values for the calibration and validation sets. The NISGDF exhibits good potential in spectral pre-processing for the prediction of P in manure, and for future real-time site-specific manure management.

Although the prediction model has shown an acceptable prediction performance with R^2 values above 0.7 and RMSE reaching a value of 0.31 g/kg, more accurate prediction would be ideal to obtain. However, this is not a simple hurdle to overcome, since P has no spectral response in the NIR spectroscopy [11]. Since manure is usually nonhomogeneous containing different ratios of solid and liquid fractions, samples should be mixed thoroughly to ensure homogeneity during scanning. However, obtaining the same level of homogeneity for all samples is out of reach, which may also negatively affect the prediction accuracy. For the reflectance spectrometer used, the phenomenon of

light scattering resulted from small particles or droplets in manure has additional negative impacts, as it can result in different spectra for equal chemical compositions, leading to a further deterioration of the vis-NIR measurement accuracy. Despite these factors affecting accuracy, the accuracy of P prediction in manure obtained in the current study is considered good. This is true, as the number of measured samples could be increased significantly over a short period of time, compared to the limited number of samples measured by the costly and time-consuming traditional laboratory analyses.

4. Conclusions

A more generalized formula was proposed to enable the calculation of NISGDF with an arbitrary order not limited to less than the first order. Based on the developed software package, the effects of non-integer order SG filter on the pre-processing of vis-NIR spectra were systematically investigated. The NISGDF was further applied to manure spectra before they were used as input for PLSR modelling to estimate the P content. Results allowed the following conclusions to be drawn:

1) The NISGDF offers superior flexibility compared to conventional integer order derivatives because it can interpolate between integer-order derivatives, reduce baseline offsets and tilts through the fine adjustment of the derivative order.

2) The NISGDF provides a better predictive model than the first and second order SG derivatives for P content in manure in terms of accuracy estimated as R^2, RPD and RPIQ.

3) The tuning of differential order NISGDF needs to be made by trial and error, and the selection of the best order can be made based on the one providing the best accuracy. It is expected that the optimal derivative order is property and dataset specific.

Future work will consider the NISGDF along with other advanced pre-processing algorithms and modelling methods such as the weighted moving average, the multivariate adaptive regression splines, and machine learning to improve the prediction of P in manure further, as to enable better manure management.

Acknowledgements

This research was funded by the Research Foundation—Flanders (FWO) for the Odysseus I SiTeMan Project (Nr. G0F9216N).

Reference

[1] B. M. Pilar, et.al, Evaluation of manure management systems in Europe, SARGA. URL: https://upcommons.upc.edu/handle/2117/88745, 2015. (Accessed 15.12.22)

[2] W.J. Powers, H.H. Van Horn, Nutritional implications for manure nutrient management planning, Applied Engineering in Agriculture 17(1) (2001) 27-39.

[3] S. Majee, G. Halder, T. Mandal, Formulating nitrogen-phosphorous-potassium enriched organic manure from solid waste: A novel approach of waste valorization, Process Safety and Environmental Protection 132 (2019) 160-168.

[4] W. Saeys, A.M. Mouazen, H. Ramon, Potential for onsite and online analysis of pig manure using variable and near infrared reflectance spectroscopy, Biosystems Engineering 91 (4) (2005) 393-402.

[5] Z. Zou, D.T. Galligan, R.D. Allshouse, J.D. Toth, C.F. Ramberg, J.D. Ferguson, Manure sampling for nutrient analysis: variability and sampling efficacy, Journal Environmental Quality 30 (4) (2001) 1432-1437.

[6] P. M. Ndegwa, J. Zhu, Sampling procedures for piggery slurry in deep pits for estimation of nutrient content, Biosystems Engineering 85(2) (2003) 239-248.

[7] J.D. O'Dell, M.E. Essington, D.D. Howard, Surface application of liquid swine manure: chemical variability. Communications in Soil Science and Plant Analysis 26 (19-20)(1995)3113-3120.

[8] O.F. Schoumans, F. Bouraoui, C. Kabbe, O. Oenema, K.C. van Dijk, Phosphorus management in Europe in a changing world, AMBIO 44 (Suppl 2)(2015)180–192.

[9] G.M. Chescheir III, P.W. Westerman, L.M. Safley Jr, Laboratory methods for estimating available nitrogen in manures and sludges, Agricultural Wastes 18 (3) (1986) 175-195.

[10] J.O. Azeez, W.V. Averbeke, Fate of manure phosphorus in a weathered sandy clay loam soil amended with three animal manures, Bioresource Technology 101 (16) (2010) 6584-6588.

[11] B.Y. Kuang, H.S. Mahmood, M.Z. Quraishi, W.B. Hoogmoed, A.M. Mouazen, E.J. van Henten, Sensing soil properties in the laboratory, In Situ, and On-Line: A Review, Advanced in Agronomy 114 (2012) 155-223.

[12] L.J. Chen, L. Xing, L.J. Han, Review of the application of near-Infrared spectroscopy technology to determine the chemical composition of animal manure, Journal of Environmental Quality 42 (4) (2013) 1015-1028.

[13] S. Sankaran, A. Mishra, R. Ehsani, C. Davis, A review of advanced techniques for detecting plant diseases, Computers and Electronics in Agriculture 72 (1) (2010) 1-13.

[14] B.M. Nicolai, K. Beullens, E. Bobelyn, A. Peirs, W. Saeys, K.I. Theron, J. Lammertyn, Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review, Postharvest Biology and Technology, 46 (2) (2007) 99-108.

[15] W. Saeys, J. Xing, J. De Baerdemaeker, H. Ramon, Comparison of transflectance and reflectance to analyse hog manures, Journal Near Infrared Spectroscopy 13 (2) (2005) 99-107.

[16] A.M. Mouazen, W. Saeys, J. Xing, J. De Baerdemaeker, H. Ramon, Near infrared spectroscopy for agricultural materials: an instrument comparison, J. Near Infrared Spectroscopy, 13 (2) (2005) 87-97.

[17] W. Saeys, P. Darius, H. Ramon, Potential for on-site analysis of hog manure using a visual and near infrared diode array reflectance spectrometer, Journal of Near Infrared Spectroscopy 12 (5) (2004) 299-309.

[18] M. D. Dagnew, T.G. Crowe, J.J. Schoenau, Measurement of nutrients in Saskatchewan hog manures using near-infrared spectroscopy, Canadian Biosystems Engineering, 46 (6) (2004) 33-37.

[19] R.W. Schafer, What is a Savitzky - Golay filter, IEEE Signal Processing Magazine 28 (4) (2011) 111–117.

[20] Y.S. Hong, Y.L. Liu, Y.Y. Chen, Y.F. Liu, L. Yu, Y. Liu, H. Cheng, Application of fractional-order derivative in the quantitative estimation of soil organic matter content through visible and near-infrared spectroscopy, Geoderma 337 (2019) 758-769.

[21] Y.S. Hong, L. Guo, S.C. Chen, M. Linderman, A.M. Mouazen, L. Yu, Y.Y. Chen, Y.L. Liu, Y.F. Liu, H. Cheng, Y. Liu, Exploring the potential of airborne hyperspectral image for estimating topsoil organic carbon: Effects of fractional-order derivative and optimal band combination algorithm, Geoderma 365 (2020) 114228.

[22] J.M. Schmitt, Fractional derivative analysis of diffuse reflectance spectra, Applied Spectroscopy 52 (6) (1998) 840-846.

[23] D. L. Chen, Y.Q. Chen, D.Y. Xue, Digital fractional order Savitzky-Golay differentiator, IEEE Transactions on Circuits and Systems II: Express Briefs 58 (11) (2011) 758-762.

[24] P. J. Tong, Y. P. Du, K.Y. Zheng, T. Wu, J.J. Wang, Improvement of NIR model by fractional order Savitzky–Golay derivation (FOSGD) coupled with wavelength selection, Chemometrics and Intelligent Laboratory Systems 143 (2015) 40-48.

[25] H. A. Fallahgoul, S.M. Focardi, F.J. Fabozzi, Fractional calculus and fractional processes with applications to financial economics: Theory and application, Academic Press, (2016).

[26] T. L. Roush, R.B. Singer, Gaussian analysis of temperature effects on the reflectance spectra of mafic minerals in the 1-μm region, Journal of Geophysical Research: Solid Earth, 91(B10) (1986) 10301-10308.

[27] J. M. Sunshine, C. M. Pieters, S. F. Pratt, Deconvolution of mineral absorption bands: an improved approach, Journal of Geophysical Research: Soil Earch, 95 (B5)(1990) 6955–6966.

[28] A. J. Brown, Spectral curve fitting for automatic hyperspectral data analysis, IEEE Transactions on Geoscience and Remote Sensing 44 (6) (2006) 1601-1608.

[29] J. M. Schmitt, G. Kumar, Spectral distortions in near-Infrared spectroscopy of turbid materials, Applied Spectroscopy 50 (8) (1996) 1066-1073.

[30] A. M. Mouazen, M.R. Maleki, L. Cockx, M. Van Meirvenne, L.H.J. Van Holm, R. Merckx, J. De Baerdemaeker, H. Ramon, Optimum three-point linkage set up for improving the quality of soil spectra and the accuracy of soil phosphorous measured using an on-line visible and near infrared sensor, Soil and Tillage Research 103 (1) (2009) 144-152.

[31] P. Marino, G. De Ferrari, L. Bechini, Description of a sample of liquid dairy manures and relationships between analytical variables, Biosystems Engineering 100 (2) (2008) 256-265.

[32] L. Martínez-Suller, A. Azzellino, G. Provolo, Analysis of livestock slurries from farms across Northern Italy: Relationship between indicators and nutrient content, Biosystems Engineering 99 (4)(2008) 540-552.

[33] M. A. Arnold, G.W. Small, Determination of physiological levels of glucose in an aqueous matrix with digitally filtered Fourier transform near-infrared spectra, Analytical Chemistry, 62 (14)(1990) 1457-1464.

[34] I. Ben-Gera, K. H. Horris, Direct spectrophotometric determination of fat and moisture in meat products, Journal of Food Science 33 (1) (1968) 64-67.