Contents lists available at SciVerse ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Energy dispersive X-ray fluorescence and scattering assessment of soil quality via partial least squares and artificial neural networks analytical modeling approaches

M.I. Kaniu $^{\mathrm{a},\mathrm{*}}$, K.H. Angeyo $^{\mathrm{b}}$, A.K. Mwala $^{\mathrm{c}}$, F.K. Mwangi $^{\mathrm{a}}$

^a Institute of Nuclear Science & Technology, University of Nairobi, P.O. Box 30197, 00100 Nairobi, Kenya

^b Department of Physics, University of Nairobi, P.O. Box 30197, 00100 Nairobi, Kenya

^c Department of Land Resource Management & Agricultural Technology, University of Nairobi, P.O. Box 30197, 00100 Nairobi, Kenya

article info

Article history: Received 21 May 2012 Received in revised form 27 June 2012 Accepted 30 June 2012 Available online 11 July 2012

Keywords: Artificial neural networks Energy dispersive X-ray fluorescence and scattering Partial least squares Soil quality assessment Spectra modeling

ABSTRACT

Soil quality assessment (SQA) calls for rapid, simple and affordable but accurate analysis of soil quality indicators (SQIs). Routine methods of soil analysis are tedious and expensive. Energy dispersive X-ray fluorescence and scattering (EDXRFS) spectrometry in conjunction with chemometrics is a potentially powerful method for rapid SQA. In this study, a 25 m Ci ¹⁰⁹Cd isotope source XRF spectrometer was used to realize EDXRFS spectrometry of soils. Glycerol (a simulate of ''organic'' soil solution) and kaolin (a model clay soil) doped with soil micro (Fe, Cu, Zn) and macro (NO $_3^-$, SO $_4^2^-$, H $_2$ PO $_4^-$) nutrients were used to train multivariate chemometric calibration models for direct (non-invasive) analysis of SQIs based on partial least squares (PLS) and artificial neural networks (ANN). The techniques were compared for each SQI with respect to speed, robustness, correction ability for matrix effects, and resolution of spectral overlap. The method was then applied to perform direct rapid analysis of SQIs in field soils. A one-way ANOVA test showed no statistical difference at 95% confidence interval between PLS and ANN results compared to reference soil nutrients. PLS was more accurate analyzing C, N, Na, P and Zn ($R^2 > 0.9$) and low SEP of (0.05%, 0.01%, 0.01%, and 1.98 µg g⁻¹respectively), while ANN was better suited for analysis of Mg, Cu and Fe ($R^2 > 0.9$ and SEP of 0.08%, 4.02 μ g g⁻¹, and 0.88 μ g g⁻¹ respectively).

 \odot 2012 Elsevier B.V. All rights reserved.

1. Introduction

The interest in SQA is currently stimulated by the increasing awareness that soil is critical for evaluation of environmental quality [\[1\]](#page-4-0). Lack of rapid and reliable SQA techniques is a critical barrier to promote (especially) precision agriculture [\[2\].](#page-4-0) SQA is routinely realized by evaluating basic SQIs which are known as the minimum data set [\[3\]](#page-4-0). Compared to wet analyses, spectroanalytical techniques for SQA are more timely, non-destructive, and can be applied in situ [\[4\]](#page-4-0).

* Corresponding author.

E-mail address: i.kaniu@gmail.com (M.I. Kaniu).

EDXRF spectroscopy is suited to SQA in these respects [\[5\]](#page-4-0). The technique is however limited by extreme matrix effects, high DL (ppm for solid samples) and it is only applicable to mostly elements with $Z \ge 19$ for most EDXRF set-ups [\[6\]](#page-4-0). Since plants absorb nutrients in their bioavailable form and thus trace levels, estimating the total element concentration is less important [\[7\].](#page-4-0) The challenge in SQA using XRF spectroscopy is how to realize accurate and simultaneous determination of micronutrients at low (trace) concentrations, and how to analyze low-Z elements (i.e. macronutrients) as well as the 'bio-available' nutrients [\[8\]](#page-4-0).

Normally the X-ray photopeak (fluorescence) intensities are used to evaluate the elemental concentrations using mostly the FP or empirical coefficients methods [\[9\]](#page-4-0) based on the assumptions of sample homogeneity, plain surface, negligible particle size effects and a priori knowledge of the [sample] matrix composition. The ''complex'' nature of the soil matrix imposes further difficulties in the XRF analysis. ''Complex matrix'' with respect to XRF is here defined as that composed of mostly low-Z elements, refractory, inhomogeneous and those with irregular geometry. Such samples have severe matrix effects, as well as enhanced Compton scatter and elevated background [\[10\]](#page-4-0). Hence

Abbreviations: ANOVA, analysis of variance; ANN, artificial neural networks; CMD, coefficient of multiple determination; DL, detection limits; EDXRF, energy dispersive X-ray fluorescence; EDXRFS, energy dispersive X-ray fluorescence and scattering; FP, fundamental parameters; MCA, multi-channel analyzer; NARL, national agricultural research laboratory; N-PLS, non-linear partial least squares; PLS, partial least squares; SNR, signal to noise ratio; SQA, soil quality assessment; SQIs, soil quality indicators; SEP, standard error of prediction; XRF, X-ray fluorescence

^{0039-9140/\$ -} see front matter \circ 2012 Elsevier B.V. All rights reserved. http://dx.doi.org/10.1016/j.talanta.2012.06.081

fluorescence SNR and analyte DL are very poor for direct SQA. Although the scatter peaks are normally considered an inconvenience, this is changing after demonstration of the utility of scatter in the analysis of light element matrices [\[11](#page-4-0)–[13\]](#page-4-0).

Multivariate chemometric techniques such as PLS and ANN are applicable to quantitative spectroanalysis. The increased analytical capability that results from incorporating chemometrics in XRF spectroanalysis has been demonstrated in other studies [\[14–18\]](#page-4-0). Compared to the classical methods like FP, these techniques embody both spectrum evaluation and quantitative analysis in one step and have capacity to handle spectral nonlinearities. PLS has shown capabilities to handle non-linearities using techniques such as N-PLS [\[19\].](#page-4-0) The use of ANN for nonlinear spectroscopic calibration has also been demonstrated [\[20,21\]](#page-4-0). The benefits of ANN include flexibility, as they do not require an underlying mathematical model, they can work with noisy data or random variance, and are also robust. A major drawback with ANN is that they are prone to overfitting [\[22\]](#page-4-0) and they involve a great deal of computing time.

We have studied and exploited the applicability of PLS and ANN for rapid and direct analysis of SQIs utilizing EDXRFS spectroscopy. In EDXRFS, during calibration a model is built based on both fluorescence and scatter spectra patterns, describing the relation between the EDXRFS spectra and the analyte (for both low-Z and heavy elements, as well as anionic components) concentrations for a set of samples. During the prediction step, the actual analysis, the calibration model is applied to future EDXRFS spectra from unknown samples in order to predict the same analyte concentrations rapidly and directly (non-invasively).

2. Materials and methods

Calibration models were developed from glycerol, a liquid ''organic'' matrix simulate, and kaolin, a synthetic model clay soil, with doped compositions of macro- (NO $_3^-,$ H₂PO $_4^-$, SO $_4^{2-}$) and micro-nutrients (Fe, Cu, Zn). Soil samples from the NARL, Kenya with known nutrient levels were also used.

A relatively low activity (25 m Ci) 109 Cd isotope source XRF spectrometer was used for excitation of the samples. The spectrometer consisted of an external Canberra Model 2020 spectrometer amplifier, and external Model 459 power supply and ADC Canberra Model 8075 pulse height analyzer interfaced to an inbuilt S100 MCA card in a Desksaver 486 work station. The X-ray fluorescence and scatter radiation was detected by a 30 mm² \times 10 mm thick EG&G ORTEC Si(Li) detector located inside a cryostat, 5 mm under a 25 μ m beryllium entrance window. The detector crystal had a 200 Å thick gold contact. The detector resolution was 190 eV (Mn K_{α} X-ray line). Spectra were acquired for 200 s and 750 s live time(s) in order to simulate the influence of SNR on the performances of PLS and ANN.

2.1. Sample preparation

General requirements for preparation of calibration standards materials are that (i) they should have a matrix that as closely as possible matches the samples to be analyzed, both chemically and physically, and (ii) they are homogeneous. The reason for working with glycerol was that it is a complex organic liquid matrix that mimics the solution (in soil) that carries the bio-available SQIs. Glycerol was meant to investigate the possibility of direct SQA from soil liquid filtrates, currently a challenge in XRF analysis. Moreover, a direct EDXRFS method developed on a complex organic liquid such as glycerol would easily find utility for heavily organic solids prepared as pellets. Kaolin mimics a solid clay-type

Fig. 1. Arrangement of sample, excitation source and detector for EDXRFS analysis.

soil, for the case where direct SQA is to be based on soil pellet analysis.

2.1.1. Preparation of calibration standards in a glycerol base matrix

The pH of glycerol was adjusted to 4.1, corresponding to acidic environment typical for most soils. For the selected macro (NO $_3^-$, $H_2PO_4^-$, SO_4^{2-}) and micro (Fe, Cu, Zn) nutrients of interest, calibration standards were prepared by spiking various analyte concentrations in the glycerol base in solution. Micro (trace) nutrients were spiked in the range $0-100 \,\mathrm{\mu g\,ml^{-1}}$. For macronutrients, appropriate chemical compounds in the range 0–2% were weighed and dissolved in double distilled water and volume made to 5 ml using glycerol. Contents in the vials were then shaken thoroughly to ensure homogeneity. 1 ml of the prepared samples was pipetted on a mylar film for EDXRFS analysis (Fig. 1).

2.1.2. Preparation of simulate soil samples

A calibration set design with uniform distribution in the representative concentration space (μ g g⁻¹ level regime for micronutrients to percent levels for macronutrients) was used [\[23\]](#page-4-0). To avoid extrapolation, the nutrient concentrations in the training set samples spanned the full concentration ranges (0–300 μ g g⁻ 1 for micronutrients and 0–3.0% for macronutrients) as expected in typical (real) soils. The selected nutrients (Cu, Fe, Zn, NO_3^- , SO_4^{2-} , H_2 PO $_4^-$) were milled simultaneously but randomly via measurements of weights of chemical compounds using pestle and mortar to fine powder and mixed with kaolin matrix base to the required concentration. Fifteen samples were prepared in triplicates by weighing approximately 0.5 g from each sample to make pellets using a hydraulic press.

2.1.3. Preparation of field soil samples

The field soil samples were of ferrosol type, obtained from an agricultural research field station in western Kenya. Bioavailable analyses of the soils had been done at NARL using the standard laboratory methods [\[24\].](#page-4-0) The soil samples were dried at 45 \degree C to expel moisture, then ground thoroughly using pestle and mortar and sieved with a $75 \mu m$ mesh. Approximately $0.5 g$ of each sample was used to prepare three replicate pellets.

2.2. Chemometric analyses of EDXRFS spectra

ANN and PLS were used to build multivariate calibration models between two matrices **X** and **Y**. The **X** $(n \times p)$ matrix represents the EDXRFS spectra, with n measured spectra each having p X-ray line intensities (channels). The **Y** ($n \times m$) matrix holds the concentrations of m analytes in the n samples. The macronutrient analyses simulate soil dataset for PLS and ANN consisted of a matrix $(X_{45 \times 110})$ in which the rows corresponded to soil sample spectra i.e. 38 calibration samples and 7 validation samples; and the columns corresponded to the Compton scatter

peak energy/channel counts. The micronutrient analyses dataset consisted of a matrix $(X_{45 \times 98})$ in which the rows corresponded to soil sample spectra i.e. 38 calibration samples and 7 validation samples; and the columns corresponded to the Fe–Cu–Zn fluorescence region energy/channel counts. The validation samples were used to optimize the performance of the calibration models viz. the coefficient of multiple determination (CMD) (R^2) and standard error of prediction (SEP).

For ANN, a back propagation neural network (BP-ANN) was designed with two-layers i.e. input and output with a hidden layer as illustrated in Fig. 2 [\[25\]](#page-4-0).

Training functions i.e. 'tan-sigmoid' non-linear transfer function in the hidden layer and 'purelin' linear transfer function in the output layer were used to train the feed forward networks for function approximation (non-linear regression). The number of neurons in the hidden layer was chosen based on the least value of minimum square error (MSE) and network training performance. The ANN training process consisted of four steps namely; (i) assembling training data, (ii) creating the network object, (iii)

Fig. 2. BP-ANN design for SQI analysis. Source: [\[25\]](#page-4-0).

training the network, and (iv) simulating the network response to the validation samples. The training process was repeated several times (i.e. by initializing the network and adjusting the number of hidden neurons) and the trained network that provided the best performance was retained [\[26,27](#page-4-0)]. Training stopped automatically after the validation error increased, to minimize model overfitting [\[26,28\]](#page-4-0). After the neural network model had been selected, the model was validated by the validation samples.

The Unscrambler 9.7 (CAMO, Oslo) and MATLAB 7.1 (Mathworks, Natick) software were used for PLS and ANN modeling respectively. Mean-centering and auto-scaling techniques were utilized for data pre-processing. Scaling reduced collinearity among the variables and allowed the variables to contribute to the model.

2.3. Application of the chemometric EDXRFS method

Owing to the differences in matrix composition between simulate and real soils, the simulate soil calibration models could not be used to predict SQIs in real soils. PLS and ANN models were therefore developed for application in real 'field' soils utilizing the developed conceptual framework. Eight soil nutrients were utilized i.e. (C, N, Na, Mg, P, Fe, Cu, Zn) and their concentrations utilized as reference concentrations.

The macronutrient dataset consisted of a matrix $(X_{64 \times 110})$ in which the rows corresponded to samples spectra i.e. 40 calibration, 12 validation, and 12 independent test samples; and columns corresponded to the Compton scatter peak energy/channel counts. The micronutrient analyses consisted of a matrix $(X_{64 \times 98})$ in which the rows corresponded to samples spectra i.e. 40 calibration, 12 validation, and 12 independent test samples; and columns corresponded to the Fe–Cu–Zn fluorescence region energy/channel counts.

3. Results and discussion

3.1. Analysis of glycerol samples

Representative spectra of three of [non-digested] glycerol samples for spiked and pure glycerol samples irradiated directly at 750 s are shown in Fig. 3.

Fig. 3. Typical EDXRF spectra of non-digested glycerol spiked with trace elements. (Analysis time: 750 s).

Fig. 4. Example of EDXRF scatter spectra of non-digested glycerol spiked with $NO₃⁻$. (Analysis time: 200 s).

Table 1

Summary of PLS and ANN nutrient (SQIs) prediction performances for simulate soil validation samples.

^a Effective Z (Zeff).

One of the features immediately apparent in [Fig. 3](#page-2-0) is the absence of micronutrient (trace) element (Fe, Cu, Zn) fluorescence peaks even at high concentrations (100 μ g ml⁻¹). This can be attributed to the complex organic nature of glycerol which causes high background, masking the fluorescence peaks during rapid analysis ($t=750$ s). The Compton peak is observed to be more intense instead. A method had to be found with which to recover SQI information from such subtle peaks in a typical organic liquid matrix. Multivariate chemometric techniques viz PLS and ANN were found to be capable of performing this task.

The scatter signal response to different $\mathsf{NO_3^-}$ concentration levels is shown in Fig. 4 (H_2 PO₄⁻ and SO₄⁻ exhibited similar profiles). As can be observed from Fig. 4, the ''blank'' i.e. pure glycerol matrix does not correspond to the minimum scatter profile (i.e. peak amplitude, peak width) indicating that the scatter-analyte (low Z) concentration relationship is non-linear.

This relationship was exploited to develop PLS and ANN quantitative models for the analysis of macronutrients in the presence of spectral non-linearities.

3.2. Calibration performance for simulate samples

The results (Table 1) show that PLS performed well for Zn but was not accurate (compared to ANN) for the estimation of Fe, Cu, NO_3^- , $H_2PO_4^-$ and SO_4^{2-} , as the SEP values were high. The weak performance could be attributed to non-linearities (spectral overlap and low fluorescence yield of ''low-Z'' elements at low concentration) of the SQI models to which the PLS method was applied. ANN performed better for non-linear cases, demonstrating its capability for modeling non-linearities and hence effectively correcting for matrix effects in XRF analysis. However for Zn i.e. the most linear case, the PLS method performed better than ANN due to the inapplicability of the non-linear 'sigmoid' training functions in ANN to the analysis linear cases [\[23\]](#page-4-0).

Optimal performance for both PLS and ANN models was achieved at low SNR for the macronutrients. This is because at low SNR i.e. 200 s, the most dominant spectral features (with high count statistics) in the full spectrum are the scatter peaks; heavy elements have minimum contribution. Predictions of NO_3^- , $\rm H_2PO_4^-$ and $\rm SO_4^{2-}$ using ANN achieved the highest overall accuracy with high CMD ($R^2 > 0.9$) and low SEP (0.09% for NO₃, 0.05% for H_2 PO₄ and 0.16% for SO₄² compared to PLS. The calibration performance of PLS and ANN in SQA was observed to depend on training (generalizing) capability.

The utility of spectrum pre-processing techniques i.e. meancentering and smoothing using the moving averages technique (3 points) were found to reduce collinearities in PLS data matrix. Utility of nonlinear training functions i.e. tan-sigmoid in ANN was found to correct for nonlinearities better.

3.3. Method performance for field soil samples

The results (Table 2) show that PLS realized better accuracy for C, N, Na, P and Zn with high CMD ($R^2 > 0.9$) and low SEP (0.05% for C, 0.01% for N and Na, 1.98 μ g g $^{-1}$ for P, and 0.7 μ g g $^{-1}$ for Zn). This shows that PLS was able to model the calibration samples relatively well for these SQIs compared to ANN. However for Mg, Fe and Cu, ANN performed better with SEP of 0.08% for Mg, 4.02 μ g g $^{-1}$ for Fe and 0.88 μ g g $^{-1}$ for Cu. Thus non-linear relationships (matrix effects) between nutrient concentration and EDXRF spectra are significantly overcome using the EDXRFS method in conjunction with PLS and ANN.

A one-way analysis of variance (ANOVA) test was conducted to compare the estimated SQIs using the EDXRFS method in conjunction with both PLS and ANN, and the reference SQI concentrations. The *p*-values for all SQIs for both techniques were (>0.174) i.e., the minimum level of significance for which a difference in mean concentration of the SQIs could be rejected. The mean concentrations from PLS and ANN models compared to reference SQIs, were not statistically significant at 95% confidence level ($p > 0.05$).

As can also be noted in [Table 3,](#page-4-0) for six of the test soil samples analyzed, most SQIs were estimated with considerable accuracy compared to the reference concentrations; the higher accuracies between PLS and ANN highlighted in bold. For C, N, Na, Cu and Zn both PLS and ANN showed mostly similar performance. However, for Mg, P and Fe, ANN performance was better.

Table 2

Summary of PLS and ANN nutrient (SQI) prediction performances for field soil test samples.

SQI	Atomic number (Z)	Chemometric technique used	SEP	R^2
C^{a} (%)	6	PLS	0.05	0.928
		ANN	0.83	0.847
N(%)	$\overline{7}$	PLS	0.01	0.969
		ANN	0.02	0.800
Na(%)	11	PLS	0.01	0.977
		ANN	0.02	0.796
Mg(%)	12	PLS	0.21	0.913
		ANN	0.08	0.991
P (μ g g ⁻¹)	15	PLS	1.98	0.982
		ANN	6.70	0.920
Fe (μ g g ⁻¹)	26	PLS	5.85	0.937
		ANN	4.02	0.956
Cu $(\mu g g^{-1})$	29	PLS	0.93	0.847
		ANN	0.88	0.874
Zn $(\mu g g^{-1})$	30	PLS	0.70	0.955
		ANN	0.66	0.534

^a Organic form.

Key: the bolds stand for the concentrations with highest accuracy achieved.

4. Conclusion

We have utilized a new approach in XRF analysis, EDXRFS spectroscopy enabled by multivariate chemometric techniques namely PLS and ANN to perform rapid and direct SQA. It was found that matrix effects corrections in EDXRFS analysis to perform accurate quantitative analysis was equivalent to solving the nonlinear spectral effects in direct analysis of soil using PLS and ANN. Both techniques performed well (with mostly high $R^2 > 0.9$ and low SEP) in the rapid prediction of SQIs i.e. both macro (low Z) and micronutrients (high Z). ANN was more accurate for prediction of NO $_3^-$, H $_2$ PO $_4^-$, SO $_4^{2-}$, Mg, Fe and Cu while PLS was more accurate for C, N, Na, P and Zn. Spectra modeling by PLS and ANN thus corrects matrix effects and affords rapid detection and quantification of high- and low-Z elements, and direct transformation of EDXRFS spectra to SQIs. The capabilities of the method should be investigated further and applied to test for quality assurance (QA) characterization of other equally complex matrices.

Acknowledgments

The authors wish to thank the University of Nairobi for the financial support extended to this work in the form of an M.Sc. fellowship (to the first author). They express their gratitude to Mr. Sijali, I.V. of the National Agricultural Research Laboratory (NARL), Kenya for providing the reference soil samples used in this work.

References

- [1] J. Glanz, Saving Our Soil: Solutions for Sustaining Earth's Vital Resource,
- Johnson Books, New York, 1995.
- [2] P. Robert, Plant Soil 247 (2002) 143–149.
- [3] D. Karlen, C. Ditzler, S. Andrews, Geoderma 114 (2003) 145–156.
- [4] M. Vohland, C. Bossung, H. Frund, J. Plant Nutr. Soil Sci. 172 (2009) 201–209. ¨
- [5] R. Grieken, A. Markowicz, Handbook of X-ray Spectrometry, CRC, 2002.
- [6] R. Jenkins, X-ray fluorescence spectrometry, Wiley, New York, 1999.
- [7] H. Foth, Fundamentals of Soil Science, 8th ed., John Willey & Sons, New York, 1990.
- [8] M. Kaniu, K. Angeyo, A. Mwala, M. Mangala, Anal. Chim. Acta 729 (2012) 21–25.
- [9] J. Criss, L. Birks, Anal. Chem. 40 (1968) 1080-1086.
- [10] F.M. Verbi, E.R. Pereira-Filho, M.I.M.S. Bueno, Microchim. Acta 150 (2005) 131–136.
- [11] A. Henrich, P. Hoffmann, H. Ortner, T. Greve, H. Itzel, J. Anal. Chem. 368 (2000) 130–138. [12] T. Kessler, P. Hoffmann, T. Greve, H. Ortner, X-Ray Spectrom. 31 (2002)
- 383–390.
- [13] R. Schramm, Anal. Chim. Acta 420 (2000) 197–203.
- [14] T. Alexandre, M. Bueno, X-Ray Spectrom. 35 (2006) 257–260.
- [15] E. Bakraji, X-Ray Spectrom. 35 (2006) 190–194.
- [16] C. Enrich, S. Boeykens, N. Caracciolo, G. Custo, C. Vazquez, X-Ray Spectrom. 36 (2007) 215–220.
- [17] K. Goraieb, T. Alexandre, M. Bueno, X-Ray Spectrom. 36 (2007) 241–246.
- [18] K. Goraieb, A.S. Lopes, C.A. Sato, M.G. Segatelli, J.C. Verzoto, M.I.M.S. Bueno, J.
- Chemom. 20 (2006) 455–463. [19] M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, J. Pages, Chemom. Intell. Lab.
- Syst. 50 (2000) 75–82.
- [20] M. Bos, A. Bos, W.E. van der Linden, Analyst 118 (1993) 323–328.
- [21] Y. Liu, B. Upadhyaya, M. Naghedolfeizi, Appl. Spectrosc. 47 (1993) 12–23.
- [22] F. Despagne, D.L. Massart, Analyst 123 (1998) 157R–178R.
- [23] M.I. Kaniu, K.H. Angeyo, M.J. Mangala, A.K. Mwala, S.K. Bartilol, X-Ray Spectrom. 40 (2011) 432–440.
- [24] J.M. Bremner, C.S. Mulvaney, Methods of Soil Analysis, Agronomy, 1982, pp. 595–625.
- [25] M.I. Kaniu, M.Sc. Thesis, University of Nairobi, 2011.
- [26] H. Demuth, M. Beale, M. Hagan, Neural Network Toolbox TM 6, User's Guide, The Math Works, Massachusetts, 2008.
- [27] A. Weigend, Proceedings of the 1993 Connectionist Models Summer School, Lawrence Erlbaum, 1994, pp. 335–342.
- [28] A. Zaknich, Neural Networks for Intelligent Signal Processing, World Scientific Pub Co. Inc., Toh Tuck Link, 2003.