



Tadesse Gashaw Asrat

DRY SPECTRAL DIAGNOSTIC TOOLS AND METHODS FOR
PRECISE FERTILIZER APPLICATION

School of Water, Energy and Environment (SWEE)
PhD degree of Environment & Agrifood

This thesis is submitted in partial fulfilment of the requirements for
the PhD degree of Environment & Agrifood
Academic Year: 2019 - 2024

Supervisor: Prof. Ruben Sakrabani (CU)
Associate Supervisors: Prof. Ronald Corstanje (CU), Dr. Stephan M.
Haefele (RRes), & Prof. Fassil Kebede (UM6P, Morocco)

March 2024



School of Water, Energy and Environment
Environment & Agrifood

PhD in Environment & Agrifood

Academic Year 2019 - 2024

Tadesse Gashaw Asrat

DRY SPECTRAL DIAGNOSTIC TOOLS AND METHODS FOR
PRECISE FERTILIZER APPLICATION

Supervisor: Prof. Ruben Sakrabani (CU)

Associate Supervisors: Prof. Ronald Corstanje (CU), Dr. Stephan M.
Haefele (RRes), & Prof. Fassil Kebede (UM6P, Morocco)

March 2024

This thesis is submitted in partial fulfilment of the requirements for
the PhD degree of Environment & Agrifood

© Cranfield University 2024. All rights reserved. No part of this
publication may be reproduced without the written permission of the
copyright owner.

Academic integrity declaration

I declare that:

- the thesis submitted has been written by me alone.
- the thesis submitted has not been previously submitted to this university or any other.
- that all content, including primary and/or secondary data, is true to the best of my knowledge.
- that all quotations and references have been duly acknowledged according to the requirements of academic research.

I understand that to knowingly submit work in violation of the above statement will be considered by examiners as academic misconduct.

Abstract

This study explores the potential of soil spectroscopy to enhance fertilizer decision-making by providing cost-effective, portable instruments for farm-level soil property prediction. The study focused on assessing the performance of various spectrometers, including low-cost near-infrared (NIR) devices, compared to mid-infrared (MIR) bench-top instruments, using a case study on maize productivity in East Africa and soil data from Morocco's semi-arid rainfed wheat-growing regions. The overall aim was to determine whether these spectroscopic methods could generate reliable predictions of key soil properties for nitrogen, phosphorus and potassium fertilizer recommendations.

The results indicate that NIR spectroscopy, despite being the most affordable and portable option, demonstrated sufficient accuracy for predicting key soil properties such as soil pH, organic carbon, and exchangeable potassium, with concordance correlation coefficients (CCCs) ranging from 0.77 to 0.96. However, the prediction of phosphorus (Olsen P) showed considerable uncertainty, particularly for values above 15 mg P kg⁻¹, where deviations from measured values increased. Comparatively, the MIR spectrometer showed better prediction accuracy for phosphorus, though its higher cost and complexity limit its applicability in resource-limited settings. The NIR spectrometer, with a prediction accuracy suitable for nitrogen fertilization (deviation between -8 to 8 kg N ha⁻¹), emerged as a promising tool for cost-effective and rapid nutrient recommendations in developing countries.

Furthermore, this research demonstrated that integrating spectroscopic data into crop models like QUEFTS for nutrient management enhances decision-making by considering both soil supply and crop response to nutrients. The findings underscore the necessity of developing region-specific calibration models to improve prediction reliability, with spatial autocorrelation analysis of soil spectra suggesting that proper calibration sample selection can improve prediction performance, especially for phosphorus and other key properties.

Ultimately, this thesis contributes to the ongoing development of soil spectral libraries and highlights the potential of low-cost, field-friendly spectrometers to improve nutrient management and crop productivity in regions with limited soil data.

Keywords: Soil spectroscopy, soil-crop-fertilizer models, spatial autocorrelation, soil spectral library.

Dedication

This thesis is dedicated to my family Mrs Amarech Tadesse (wife) and my children Adoniyas, Ruth and Rodas Tadesse who persevered during my absence from home for my studies!

Acknowledgements

My primary thanks go to my diverse supervisory team Dr Stephan Haefele, Prof Ruben Sakrabani, Prof Ron Corstanje and Prof Fassil Kebede. I appreciate all your encouragement and enthusiasm, as well as the time you committed to consistently support and guide me over the last four and a half years.

Special gratitude is extended to Dr Timo Breure for his endless support for my in-depth understanding of soil spectral data management and utilizing efficiently the R programming language to handle large data analysis. Dr Breure, you always managed to help me progress well in data management pitfalls, which would have taken a longer time to navigate without your support.

I would like to thank the OCP Group of Morocco sponsoring my PhD study in the UK under the Sustainable Agriculture for Africa (SAFA) studentship program.

I would like to thank my cohorts for your friendship and technical support at times. The monthly catch-up meetings and the three-month research progression presentations have given me the opportunity to improve my communication skills and extended my horizon in knowledge.

I express gratitude to The CESFRA team at UM6P Dr Leila Tajeddine, Mr Issam Aboud and Mrs Hind Belarbi who have been a cornerstone to exporting the research materials used in this study from Morocco.

And special thank goes to The Al Moutmir Agroengineer team members Mr Abdella Hama, Kamal Alahiane, Mr Achraf Nejoud, Mr Sifeddine Ouahdani who support and assisted me in carrying out sampling and experimentation in Morocco. Recognition goes beyond to whom I did not mentioned their names for their contributions and facilitation involved in this part of the study activity.

Additionally, I would like to acknowledge Prof Rachid Bouabid at the National School of Agriculture (ENA), Meknes, for his support and guidance regarding the nutrient application study for the Moroccan rainfed wheat system.

I am extremely grateful for the technical support I received from the Dry spectroscopy lab group at Rothamsted Research including Dr Sarah Dunham, Dr

Cathy Thomas, Dr Javier Hernandez, Ms Joanna Carter and Mr Christopher Herbin.

I would like also to extend my thanks to the Horticultural and Controlled Environment teams for facilitating and supporting to undertake the pot experimentation in the controlled glasshouse facility.

I also would like to acknowledge the analytical units at both Rothamsted Research and Cranfield University for their high quality and timely service for the soil and plant analysis carried out for my study.

Finally, I am thankful for Ian Truckell and the GIM team at Cranfield University who supported me in managing spatial datasets with ArcGIS Pro and allowed me to access the GIM modules to study and understand spatial analysis.

Table of Contents

Academic integrity declaration.....	i
Abstract	ii
Dedication	iv
Acknowledgements	v
Table of Contents	vii
List of Figures.....	xi
List of Tables	xvi
List of Equations.....	xviii
List of Abbreviations	xx
1 Introduction.....	1
1.1 Background.....	1
1.2 Overall Scientific Aim	3
1.3 Structure of the thesis	4
1.4 References	6
2 Infrared spectroscopy for prediction of soil properties input for crop models:	
a literature review	10
2.1 Introduction	11
2.2 Overview of IR soil spectroscopy towards soil analysis	12
2.2.1 General comparisons between MIR and vis-NIR spectroscopy	12
2.2.2 Advancement of tools and methods in Vis-NIR and MIR spectral ranges for assessing soil properties.....	14
2.2.3 Spectral data handling and processing	18
2.2.4 Soil spectral analysis and predictive models.....	20
2.3 Soil spectral libraries development and usage.....	23
2.3.1 Structure of soil spectral libraries and considerations	23
2.3.2 Global and regional initiatives for developing a soil spectral library .	24
2.3.3 Applicability of soil spectral libraries	25
2.4 Precision among spectrally predicted soil properties important in soil- crop-fertilizer model inputs by vis-NIR and MIR range.....	28
2.4.1 Soil organic carbon and total nitrogen.....	28
2.4.2 Plant available phosphorus, total phosphorus, and exchangeable potassium.....	31
2.4.3 Clay, CEC, and pH.....	31
2.5 Crop Fertiliser Recommendation Approaches with Soil Property Inputs	32
2.5.1 Overview of Approaches to fertilizer recommendations	32
2.5.2 Quantification of soil nutrient supply capacity and fertilizer recovery using IR soil spectroscopy.....	33
2.6 Conclusions and research gaps.....	34
2.7 Acknowledgment.....	35

2.8 References	35
3 Spectral soil analysis for fertilizer recommendations by coupling with QUEFTS for maize in East Africa: a sensitivity analysis.....	52
3.1 Abstract.....	52
3.2 Introduction	53
3.3 Materials and methods.....	56
3.3.1 Soil sample sources	56
3.3.2 Soil spectral data collection with a brief description of spectrometers	57
3.3.3 Conventional soil analysis	59
3.3.4 Soil property prediction from spectra.....	59
3.3.5 Derivation of soil test-based fertilizer estimation	61
3.3.6 Bootstrapping and error propagation.....	63
3.4 Results.....	65
3.4.1 Soil spectral characteristics and preprocessing across soil origins and instruments.....	65
3.4.2 Description of the soil properties	67
3.4.3 Soil property prediction accuracy and consistency for QUEFTS inputs	68
3.4.4 Comparison of soil testing methods for fertilizer decision making	70
3.5 Discussion	75
3.5.1 Prediction of soil property inputs for QUEFTS.....	75
3.5.2 Fertilizer determination through soil spectral tools by coupling with QUEFTS.....	77
3.6 Conclusions	79
3.7 Acknowledgment.....	80
3.8 Author contributions	80
3.9 References	81
4 A Moroccan soil spectral library use framework for improving soil property diagnostics: Evaluating a Geostatistical Approach.....	90
4.1 Abstract.....	90
4.2 Introduction	91
4.3 Material and methods	95
4.3.1 Soil sampling locations and sampling strategy.....	95
4.3.2 Baseline soil property analyses.....	97
4.3.3 Soil spectral collection and spectra preprocessing.....	98
4.3.4 Framework for calibration sample selection for modelling and performance evaluation.....	99
4.3.5 The entire MSSL with Partial Least Square regression (PLSR)	100
4.3.6 Spatial autocorrelation of spectra PC scores	101
4.3.7 MBL spectra similarity analysis	102

4.3.8 Cluster grouping using environmental covariates and FAO soil units	104
4.3.9 Accuracy assessment	106
4.4 Results.....	107
4.4.1 Baseline soil properties and their correlations with spectra preprocessing.....	107
4.4.2 Spatial autocorrelation and scale dependency of spectra PC scores	111
4.4.3 Prediction performance among sample selection processes	112
4.5 Discussion	118
4.5.1 Robustness of MSSL for soil property attribute predication?.....	118
4.5.2 Suitability of spatial autocorrelation of spectra for calibration sample selection	120
4.5.3 Spectra similarity and clustering performance for for calibration sample selection	121
4.6 Conclusion	122
4.7 Author contributions	123
4.8 Conflicts of Interest	123
4.9 Acknowledgment.....	124
4.10 References	124
5 Improving fertilizer application for Moroccan rainfed wheat growing regions using crop models and NIR soil spectroscopy.....	133
5.1 Abstract.....	133
5.2 Introduction	134
5.3 Materials and methods.....	138
5.3.1 Pot experimentation	138
5.3.2 Soil sample analyses.....	143
5.3.3 Actual nutrient uptake and recovery efficiency	144
5.3.4 The QUEFTS approach to quantify INS and internal nutrient use efficiency	145
5.3.5 Crop model fitting for parameter estimations.....	146
5.3.6 NIR applicability for indigenous nutrient supply estimation.....	147
5.3.7 Statistical analysis and model performance evaluation.....	148
5.4 Results.....	149
5.4.1 Crop yield response to applied nutrients and yield limiting nutrient per soil type.....	149
5.4.2 Actual uptake from applied and indigenous nutrient supply	152
5.4.3 Recovery efficiency estimates for applied nutrients across soil types	154
5.4.4 Total plant dry matter and grain yield with total nutrient uptake	155
5.4.5 Soil-crop-fertilizer model choice for Moroccan rainfed wheat	158

5.4.6 NIR soil spectroscopy use towards informed decision making in Morocco	165
5.5 Discussion	167
5.5.1 Yield limiting nutrients and nutrient uptake for rainfed wheat growing regions of Morocco	167
5.5.2 Crop models for rainfed wheat growing regions of Morocco	168
5.5.3 Soil spectroscopy for decision support in rainfed wheat growing regions of Morocco.....	170
5.6 Conclusions	171
5.7 Acknowledgment.....	172
5.8 Author Contributions	172
5.9 Conflict Of Interest	173
5.10 References	173
6 Implications of use of soil spectroscopy and its contribution towards informed decisions in agriculture	183
6.1 Can affordable spectrometers provide prediction accuracies adequate for meaningful fertilizer recommendations?	183
6.2 Soil spectral library spatial dependency can be useful to subset calibration sample selection for precise soil property prediction	186
6.3 Influence of soil-crop-fertilizer models to quantify indigenous nutrient supply and hence encourage use of soil spectroscopic tools and methods	189
6.4 References	191
7 Conclusion and future research directions	193
7.1 Overall Conclusion.....	193
7.2 Research Impact and Future Directions.....	194
Appendices.....	195
Appendix A : Ethical approval letter	195
Appendix B : Spectral soil analysis for fertilizer recommendations by coupling with QUEFTS for maize in East Africa: a sensitivity analysis	196
Appendix C : A Moroccan soil spectral library use-framework for improving soil property diagnostics: Evaluating a geostatistical approach	204
Appendix D : Improving fertilizer application using crop models and NIR soil spectroscopy for Moroccan rainfed wheat growing regions.....	218

List of Figures

- Figure 2.1 : Some of the widely used soil spectroscopy instruments for soil property assessment. a) FieldSpec with either contact probe or mug light (Source: Malvern Panalytical) - portable; b) Neospectra saucer (Source: Si-Ware systems) – handheld; c) Spectral Evolution with contact probe (Source: Pro-lite technology) – portable; d) Alpha II (Bruker) – benchtop; e) Tensor 27 (Source: Bruker) – benchtop and e) Mobile-IR II (Source: Bruker) – portable. 15
- Figure 2.2 : Wavelength regions for visNIR and MIR soil spectroscopy. The vertical dotted lines within the visNIR range indicate the wavelength region covered by the Neospectra NIR instrument (Si-Ware systems). Characteristic soil wavelength regions are adapted from Coblinski et.al. (2020), and Viscarra Rossel et.al. (2006). 17
- Figure 3.1 : Spectrometers compared in this study: a) FTIR Bruker Tensor 27 (Tensor-II), b) FieldSpec 4 mug light (FS4-ML) and contact probe (FS4-CP); and c) NeoSpectra Saucer (Neospec). 59
- Figure 3.2: Raw absorbance of the soil spectrometers (FS4-CP -> FieldSpec 4 contact probe, FS4-ML -> FieldSpec 4 mug light, Neospec -> NeoSpectra Saucer, Tensor-II -> FTIR Bruker Tensor 27) coloured by soil sample origin. 66
- Figure 3.3: Mean absorbance after Savitzky-Golay smoothing and first derivative preprocessing (SG + 1D) of spectra collected by soil spectrometers (FS4-CP -> FieldSpec 4 contact probe, FS4-ML -> FieldSpec 4 mug light, Neospec -> NeoSpectra Saucer, Tensor-II -> FTIR Bruker Tensor 27). 67
- Figure 3.4: Prediction accuracy and consistency for all soil properties (exchangeable K, total N, Olsen P, pH, and SOC, respectively) and spectrometers (CP -> FieldSpec 4 contact probe, ML -> FieldSpec 4 mug light, NeoSpec -> NeoSpectra Saucer, Tensor-II -> FTIR Bruker Tensor 27). The horizontal lines are the error bars which represented 95% confidence intervals. 70
- Figure 3.5: Nitrogen fertilizer rate predictions for a 7 t ha⁻¹ target maize yield based on spectral (x-axis) and conventional (y axis) soil analysis. The statistics shown are the root mean square error (RMSE) and Bias for each spectrometer (CP -> FieldSpec 4 contact probe, ML -> FieldSpec 4 mug light, NeoSpec -> NeoSpectra Saucer, Tensor-II -> FTIR Bruker Tensor 27). The horizontal lines are the marginal uncertainties represented as $\pm 1\sigma$ 72
- Figure 3.6: Fertilizer N difference among the conventional and spectrometers (FS4-CP -> FieldSpec 4 contact probe, FS4-ML -> FieldSpec 4 mug light, Neospec -> NeoSpectra Saucer, Tensor-II -> FTIR Bruker Tensor 27). The dashed horizontal lines are the mean difference $\pm z\alpha/2\sigma$ and the horizontal line is the mean difference. 72

- Figure 3.7: Phosphorus fertilizer rate recommendation for a 7 t ha⁻¹ maize target yield based on spectral (x-axis) and conventional (y axis) soil analysis. The statistics shown are the root mean square error (RMSE) and Bias for each spectrometer (FS4-CP -> FieldSpec 4 contact probe, FS4-ML -> FieldSpec 4 mug light, Neospec -> NeoSpectra Saucer, Tensor-II -> FTIR Bruker Tensor 27). The horizontal lines are the 95% confidence intervals. 74
- Figure 3.8: Fertilizer P difference among the conventional and spectrometers (FS4-CP -> FieldSpec 4 contact probe, FS4-ML -> FieldSpec 4 mug light, Neospec -> NeoSpectra Saucer, Tensor-II -> FTIR Bruker Tensor 27). The dashed horizontal lines are the mean difference $\pm z\alpha/2\sigma$. The horizontal line is the mean difference. 74
- Figure 4.1: Soil sample locations with crop cover map of 2019-2020 cropping season and sample subset. There were two geographic distinct sampling regions situated in the north and south of the rainfed wheat growing regions of Morocco. Note that soil sampling for the cropping area between the two geographic distinct sampling regions is ongoing. 96
- Figure 4.2: Mean with standard error (top row) and raw (bottom) absorbance spectra of the dominant soil types in the NIR and MIR region. Bk = Calcic Cambisols; Kk = Calcic Kastanozems; Lc = Chromic Luvisols; Je = Eutric Gleysols; Kl = Luvic Kastanozems; V = Vertisols. 99
- Figure 4.3: Schematic demonstrating the sample selection processes for soil property prediction from a soil spectral library of MIR and NIR region. Tensor-II; Neospec = Neospectra Saucer. StBCS design -> stratified balanced coverage sampling design. 100
- Figure 4.4: Correlation boxplot of each soil properties with spectra pre-processing techniques (SG -> Savitzsky-Golay smoothing) at each wavelength by the scanning instrumentation (Neospec -> Neospectra, Tensor-II -> Bruker Tensor 27). The lower and upper limits of each box are the 25th and 75th percentiles, the tails show the correlation range, the dots represent outliers and the horizontal line in each box represents the median. 110
- Figure 4.5: Semivariograms of PC vectors fitted with a spherical model (solid line) from the respective soil scanning instruments (Neospec -> Neospectra for NIR spectra; Tensor-II -> Bruker Tensor 27 for MIR spectra) with cut-off distances of 120 and 200 km. The dashed line is the sill. 112
- Figure 4.6: Prediction performance for the soil properties and soil scanning instruments (Neospec -> Neospectra for NIR spectra; Tensor-II -> Bruker Tensor 27 for MIR spectra) using the entire MSSL. The statistics shown are the R², concordance correlation coefficient (CCC), the root mean square error (RMSE), Bias, Ratio of standard error of Performance to standard Deviation (RPD) and the ratio of performance to interquartile range (RPIQ). 115
- Figure 5.1: Pictorial response of wheat crop for applied nitrogen (upper row) and phosphorus (bottom row) across the soil types. 142

Figure 5.2: Relationships between total plant dry matter or grain yield with total nutrient uptake (N, P or K). The solid lines are from Liu et al. (2006). The dashed lines are the 95 and 5 percentiles from the nutrient study from the three soil types.	158
Figure 5.3: A hyperbolic relationship between total nitrogen uptake and application from a nonlinear least squares regression fit. The parameters shown are as defined in section 5.3.4.2. The statistical shown are R^2 = Coefficient of determination; AICc = corrected version of Akaike information criterion; and RMSE = Root means square error.....	160
Figure 5.4: A hyperbolic relationship between total phosphorus uptake and application from a nonlinear least squares regression fit. The parameters shown are as defined in section 5.3.4.2. The statistical shown are R^2 = Coefficient of determination; AICc = corrected version of Akaike information criterion; and RMSE = Root means square error.....	161
Figure 5.5: A quadratic plateau relationship between grain yield and nitrogen application rates from a nonlinear least squares regression fit. The parameters shown are as defined in section 5.3.4.1. The statistical shown are R^2 = Coefficient of determination; AICc = corrected version of Akaike information criterion; and RMSE = Root means square error.	163
Figure 5.6: A quadratic plateau relationship between grain yield and phosphorus application rates from a nonlinear least squares regression fit. The parameters shown are as defined in section 5.3.4.1. The statistical shown are R^2 = Coefficient of determination; AICc = corrected version of Akaike information criterion; and RMSE = Root means square error.	164
Figure 5.7: Residual in soil supply capacity between conventional method and NIR spectrometer (NeoSpectra), relative to the conventionally determined soil supply capacity. The dashed horizontal lines present the mean residual with $\pm z\alpha/2\sigma$. The horizontal line presents the overall mean residual.....	166

Appendices

Figure B-1: Outlier detection from PC score plots of PC1 and PC2. The numbers are the sample codes.	196
Figure B-2: Model calibration and independent test sets sample distribution at the first two PC score space using cLHS sampling.	197
Figure B-3: Loadings of the first two PCs for pH across the instruments.	201
Figure B-4: Loadings of the first two PCs for exchangeable potassium across the instruments.	201
Figure B-5: Loadings of the first two PCs for OlsenP across instruments.	202
Figure B-6: Loadings of the first two PCs for SOC across instruments.	202

Figure B-7: Loadings of the first two components for Total N across instruments.	203
Figure C-1: Moving window size with the RMSE of calibration across soil properties and soil scanning instruments (Neospec -> Neospectra; Tensor-II -> Bruker Tensor 27).	207
Figure C-2: Soil textural triangle of the MSSL according to ISSS classification.	208
Figure C-3: Statistical distribution of soil properties in training sets (SSL).	209
Figure C-4: Statistical distribution of soil properties for independent test sets.	209
Figure C-5: The proportion of variance explained by first eleven PC's and their cumulative variance explained (99%) for Tensor-II (Bruker Tensor 27)...	210
Figure C-6: The proportion of variance explained by first four PC's and their cumulative variance explained (99%) for Neospec (Neospectra).	210
Figure C-7: Prediction performance for the soil properties (Neospec -> Neospectra) using a reflectance spectrum using the entire MSSL. The statistics shown are the R^2 , concordance correlation coefficient (CCC), the root mean square error (RMSE), Bias and the ratio of performance to interquartile range (RPIQ).	211
Figure C-8: Prediction performance for the soil properties and soil scanning instruments (Neospec -> Neospectra; TensorII -> Bruker Tensor 27) from calibration sample entry selection using the spatial non-stationarity analysis of the spectra PC vector. The statistics shown are the R^2 , concordance correlation coefficient (CCC), the root mean square error (RMSE), Bias and the ratio of performance to interquartile range (RPIQ).	213
Figure C-9: Prediction performance for the soil properties and soil scanning instruments (Neospec -> Neospectra; TensorII -> Bruker Tensor 27) using spectra similarity sample entry selection method. The statistics shown are the R^2 , concordance correlation coefficient (CCC), the root mean square error (RMSE), Bias and the ratio of performance to interquartile range (RPIQ).	215
Figure C-10: Prediction performance for the soil properties and soil scanning instruments (Neospec -> Neospectra; TensorII -> Bruker Tensor 27) from calibration sample selection using the cluster grouping of environmental covariates. The statistics shown are the R^2 , concordance correlation coefficient (CCC), the root mean square error (RMSE), Bias and the ratio of performance to interquartile range (RPIQ).	217
Figure D-1 : Split-plot experimental layout for pot nutrient study across Blocks and soil types.	218

List of Tables

Table 1.1 : Description of the chapters and their status.	5
Table 2.1 : Soil property prediction by vis-NIR and MIR ranges for the validation set.....	30
Table 3.1 : Characteristics of the spectrometers compared [FTIR Bruker Tensor 27 (Tensor-II), FieldSpec 4 mug light (FS4-ML) and contact probe (FS4-CP), and NeoSpectra Saucer (Neospec)].	58
Table 3.2: Descriptive statistics of soil chemical properties for model training and independent test set.	68
Table 3.3: Predictive metrics by soil property and soil sensor. The statistics shown are the root mean square error (RMSE), the concordance correlation coefficient (CCC), Bias and the ratio of performance to interquartile range (RPIQ).	69
Table 4.1: Descriptive statistics of soil property in the training and validation sets.	107
Table 4.2: Prediction performance for the soil properties using calibration sample selection approaches in the NIR and MIR spectral ranges.	117
Table 5.1: Soil characteristics of the soil types used for nutrient study using pot experimentation.	139
Table 5.2: Treatment set up and target response.....	140
Table 5.3 Analysis of variance for the split plot wheat grain and straw yield and grain nutrient concentration to treatments and soil types.....	151
Table 5.4 Effect of nutrient combinations and nutrient levels on grain yield and straw productivity of wheat across soil types.	151
Table 5.5: Uptake of each nutrient by the grain and straw of wheat on the untreated plots by soil types.	153
Table 5.6: Indigenous nutrient supply capacity (INS) using QUEFTS estimation approach with uptake at zero nutrient addition for each soil.	153
Table 5.7: Recovery efficiency among applied nitrogen levels across soil types.	154
Table 5.8: Recovery efficiency of applied phosphorus rates across soil types.	155
Table 5.9: Recovery efficiency of applied potassium rates across soil types.....	155
Table 5.10: Optimum fertilizer rates for predicted yields from response models for Cambisol, Kastanozems and Vertisol.	164

Table 5.11: Overall mean and standard error of indigenous nutrient supply capacity of soils estimated by NIR spectroscopy (p) and the conventional soil analysis (c) for the three soil types.	166
---	-----

Appendices

Table B.1: Descriptive statistics of the soil property attributes across countries, model calibration and independent test set subset by cLHS.	199
Table C.1: Area coverage of dominant FAO soil types from 2017/2018 cropping season crop cover map.	204
Table C.2: Cluster size of training and independent test sets based on nearest neighbours clustering from environmental covariates.....	208
Table D.1: Analysis of variance for nutrient study across soil types and treatments for the split plot design.....	218
Table D.2: Duncan multiple range mean separation for wheat grain concentration of zinc.	219
Table D.3: Duncan multiple range mean separation for wheat grain concentration of nitrogen.....	219
Table D.4: Duncan multiple range mean separation for wheat grain concentration of phosphorus.....	219
Table D.5: Duncan multiple range mean separation for wheat grain concentration of potassium.	220
Table D.5: Model summary - Formula: Total_N_Uptake ~ QUEFTS model N (n_fert, a_N, b_N).....	220
Table D.6: Model summary - Formula: Total_P_Uptake ~ QUEFTS model P (P_supply, a_P, b_P).....	221
Table D.7: Model summary - Formula: GY_ha ~ qp(n_fert, a, b, jp).....	221
Table D.8: Model summary - Formula: GY_ha ~ qp(p2o5_fert, a, b, jp).....	222

List of Equations

(3-1)..... 61

(3-2)..... 61

(3-3)..... 62

(3-4)..... 62

(3-5)..... 62

(3-6)..... 62

(3-7)..... 63

(3-8)..... 64

(3-9)..... 64

(3-10)..... 65

(4-1)..... 101

(4-2)..... 102

(4-3)..... 104

(4-4)..... 105

(4-5)..... 106

(4-6)..... 106

(4-7)..... 106

(4-8)..... 106

(5-1)..... 142

(5-2)..... 144

(5-3)..... 144

(5-4)..... 146

(5-5)..... 146

(5-6)..... **Error! Bookmark not defined.**

(5-7)..... 145

(5-8)..... 145

(5-9)..... 145

(5-10)..... 145

(5-11).....	145
(5-12).....	145
(5-13).....	147
(5-14).....	147
(5-15).....	Error! Bookmark not defined.

Appendices

(B.1)	197
(B.2)	198
(B.3)	198
(B.4)	198
(B.5)	198
(C.1)	204
(C.2)	204
(C.3)	205
(C.4)	205

List of Abbreviations

AfSIS	the Africa Soil Information Service
CAES	College of Agriculture and Environmental Science
CESFRA	Center of Excellence for Soil and Fertilizer Research in Africa
UNFAO	Food and Agricultural Organization of the United Nations
FS4-CL	FieldSpec-4 + Contact Probe
FS4-ML	FieldSpec-4 + Mug Light
ICRAF	International Centre for Research in Agroforestry
IR	Infrared
LDPA	Laser Diffraction Particle Size Analyser
MBL	Memory Based Learner
MIR	Mid Infrared
MSSL	Moroccan Soil Spectral Library
NIR	Near Infrared
Neospec	NeoSpectra Spectrometer
OCP	Office Chérifien des Phosphates
PLSR	Partial Least Squares Regression
pXRF	Portable X-ray fluorescence
QUEFTS	QUantitative Evaluation of the Fertility of Tropical Soils
SSL	Soil Spectral Library
Tensor-II	FTIR Bruker Tensor 27 Bench-Top
UM6P	University Mohammed VI Polytechnic
visNIR	Visible and Near Infrared

1 Introduction

1.1 Background

Sustainable soil management requires a thorough investigation of soil processes and underlying agricultural production systems for optimum productivity through precise fertilizer management. This can be accomplished with a qualitative and quantitative determination of spatial (Ge et al., 2014; Haefele et al., 2014; Bah et al., 2016; Oghaz et al., 2019; Amede et al., 2020) and temporal (Shanahan et al., 2008; Mueller et al., 2012) variability of soil properties and associated differential responses of crops to applied nutrients. Such studies might be accompanied by calibration (Greenwood et al., 1982; Rutkowska and Fotyma, 2009; Dhakal and Lange, 2021) and determination of critical concentrations of major nutrients in the soil and crop produce (Bell et al., 2013; Brennan and Bell, 2013). Furthermore, soil and plant chemical test results can be used as input parameters for further analyses using soil-crop-fertilizer simulations (Williams et al., 1989; Stöckle et al., 2003; Banterng et al., 2010; Wajid et al., 2021) and subsequent extrapolations using crop models can help to achieve optimal plant nutrition (Janssen et al., 1990; Sattari et al., 2014) under different growth conditions.

Conventional soil chemical testing has major shortcomings such as high costs and significant workload, limiting the feasible number of samples that can be feasibly collected and analysed, thus reducing the representation of field variability. Moreover, the diversity of testing methods used globally makes it challenging to compare results across different regions. Studies by Dobermann et al. (2003), Menzies et al. (2007) and Mason et al. (2010) have highlighted those chemical tests often showed deviation from crop nutrient status and / or yield response for major crop nutrients. Consequently, there can be yield penalties (Shanahan et al., 2008) accompanied by subsequent loss of underutilized fertilizer into the ecosystem (Andraski et al., 2000; Vitousek et al., 2009) and soil degradation (Rickson et al., 2015). Furthermore, the time delays, access to soil analysis service and potential errors in processing and waiting for conventional analysis can further compromise timely decision-making in nutrient management.

These problems could be addressed with the acquisition of rapid, low-cost, high-throughput, high-quality and detailed soil inventories using the infrared regions of the electromagnetic spectrum (Shepherd and Walsh, 2007; Soriano-Disla et al., 2014; Bah et al., 2016; Shepherd et al., 2022). Promising spectral sensing techniques which showed high potential in predicting soil variables include visible near-infrared (visNIR) and mid-infrared (MIR). The comparison between spectral analysis and conventional methods will confirm the validity of the former technique which is quicker and easier to deploy in field conditions. Once the validity of the spectra-based predictions soil characteristics is established, the respective nutrient values can be incorporated into suitable crop models that can predict crop requirements (Janssen and Guiking, 1990; Dobermann et al., 2003; Pearce et al., 2022) and thus predict optimal fertilizer rates to meet crop demand. There could also be an opportunity to directly relate soil spectral measurements to crop requirements/traits (Breure et al., 2021) which could be correlated with nutrient use efficiency to assess accuracy of model predictions.

Soriano-Disla et al. (2017) highlighted that handheld spectral tools for analysing soil (or plant) characteristics have the potential to significantly enhance fertilizer application in crop production. These tools enable instant site-specific adaptation of fertilizer recommendations and identification of soil characteristics important to diagnose other soil fertility and soil health parameters. These functionalities of soil spectroscopy could enhance the use of soil information for informed-decision making by increasing the opportunity access to soil information, specifically for developing countries and highly sensitive agro-ecological areas which demand precise management of nutrients such as the semi-arid region of Morocco as well as highly degraded regions of the East African rainfed regions. Sub-Saharan Africa has an average fertilizer application rate of 22 kg ha⁻¹, which is significantly below the global average of 146 kg ha⁻¹ (Malpass, 2022). Similarly, the overall fertiliser use trend in Africa is low, with a continental average of about 8 kg ha⁻¹, but there is a goal to increase this average to 50 kg ha⁻¹ (Dimkpa et al., 2023). Consequently, there is a huge yield gap for wheat in Morocco (Abail et al., 2014) as well as maize in East Africa (Leitner et al., 2020) and here is where dry spectral

can play a significant role in improving fertilizer use thereby increase sustainable crop productivity through informed decision-making.

However, proper calibration remains an essential component (Shepherd and Walsh, 2002; Guerrero et al., 2016; Breure et al., 2022), together with the site-specific estimation of crop response and nutrient availability model parameters. Spectral sensors need to be calibrated using conventional soil analysis to generate the soil information which significantly limits their use as it adds cost and complexity. Although calibration samples will never be removed entirely, there is scope to develop libraries in which the quantity, location and timing of entries is optimised based on statistical models improving the predictive power for new samples (Savvides et al., 2010). And as the soils vary in space, the spectra generated could be scale dependent and the calibration sample entries for a specific farmland could be selected based on a spatial analysis approach.

Hence, this project intended to fill the knowledge gap on the applicability of soil and plant spectral analysis tools and methods for precise fertiliser management, examining the potential of soil reflectance as an integral indicator of specific soil functions. To address these knowledge gaps, the study targeted case studies focused on the diverse soils and wheat production in Morocco and maize in East Africa. In both cases, application of spectral sensing techniques and crop modelling could enable better informed decisions on crop production and consequently improve appropriate input use.

1.2 Overall Scientific Aim

This project aimed at employing spectral techniques and crop modelling for precise fertilizer management. It also targeted a better understanding of the sensitivity and robustness of spectral analysis methods in the laboratory and in the field.

Therefore, it was hypothesized that non-destructive methods of analysis using spectral sensing could potentially determine available soil nutrients to develop better fertilizer recommendations. In addition, it was assumed that a targeted use strategy of soil spectral libraries would be necessary for optimized predictions of

soil variables at a range of scales. To test these hypotheses and realize the aims, the following objectives were set out:

Objective 1. To evaluate handheld spectral tools for in-field soil analysis and determine the characteristics which can be most reliably predicted compared to the MIR tools,

Objective 2. To develop an analytical framework for the best use of spectral libraries for prediction of soil variables at a range of scales,

Objective 3. To apply and test spectral methods for identification of crop available nutrients economically and quickly and,

Objective 4. To develop and test models that use spectral soil (and plant) data to determine fertilizer recommendations.

1.3 Structure of the thesis

This thesis is structured with chapters prepared as journal manuscripts (see Table 1.1). All the manuscripts of this thesis are written by Tadesse Gashaw Asrat with review and inputs by the supervisory team as indicated in the thesis cover page. Statistical and R coding support was obtained from Dr Kirsty Hassall (RRes) and Dr Timo Breure (FAO) for the data analysis in Chapters 3 and 4. All data generation and experimentation were designed, coordinated, and completed by Tadesse Gashaw Asrat with the support of the supervisory team. The soil sampling design was developed by Tadesse Gashaw Asrat whereas the soil sampling campaign and exportation to the UK was carried out by the OCP Al Moutmir collaborators and the UM6P CAES team. Soil scanning by various instrumentations and extraction of total soil elements by the Aqua regia method was completed by Tadesse with support from the soil spectral laboratory team at Rothamsted Research (RRes). All soil and crop sample processing and submission for analysis was completed by Tadesse Gashaw Asrat, and the analyses were done at RRes, Cranfield University (CU) and the NRM (part of Cawood Scientific) analytical laboratories at various times.

At the research project inception, a literature review (Chapter 2) was conducted, studying soil spectroscopy trends and soil-crop-fertilizer decision support tools towards understanding the efficiency of these approaches for quantifying the nutrient supply capacity of soils and response of crops to applied nutrients. This analysis identified key research gaps for which the following studies emanated and completed.

Following the critical review, soil scanning instrumentations with different wavelength ranges were evaluated for their efficiency and sensitivity for fertilizer decision support, using the crop model QUEFTS and existing East African soil samples from farmers' maize fields (Chapter 3: Asrat, T.G., Sakrabani, R., Corstanje, R., Breure, T., Hassall, K.L., Kebede, F. and Haefele, S.M., 2023. Spectral soil analysis for fertilizer recommendations by coupling with QUEFTS for maize in East Africa: A sensitivity analysis. *Geoderma*, 432, p.116397). Among other results, it was found that the spectral library used might have contributed considerably to the uncertainty of predictions due to dependency on global spectral libraries, which is not uncommon in this type of studies. Consequently, spectral library development and use-frameworks were evaluated for site-specific or spectral similarity analysis (Chapter 4). Further on, the main research target was addressing a comprehensive integration of soil spectroscopy with crop models. Therefore, a crop model was calibrated, and the input parameters were tested using a pot nutrient study targeted at estimating relationships of nutrient application, grain yield and total nutrient uptake for selected soils from Morocco, thereby developing a soil-site-crop specific fertilizer recommendation for rainfed wheat of Morocco (Chapter 5). Finally, in Chapter 6, the overall achievement of the research project with its contribution to knowledge and applicability in a wider context were discussed.

Table 1.1 : Description of the chapters and their status.

Chapter	Objective/s addressed	Focus	Journal	Status
2	1,2,3,4	Infrared spectroscopy for prediction of input soil properties	To be proposed	Completed

		for crop models: a literature review		
3	1,2,3	Spectral soil analysis for fertilizer recommendations by coupling with QUEFTS for maize in East Africa: a sensitivity analysis	Geoderma	Published
4	4	A Moroccan soil spectral library use-framework for improving soil property prediction: evaluating a Geostatistical Approach	Geoderma	Published
5	2	Optimizing fertilizer application for Moroccan rainfed wheat using crop models and NIR soil spectroscopy	To be proposed	Completed
6	1,2,3,4	Implication of use of soil spectroscopy and its efficiency towards informed decision in agriculture	----	----

1.4 References

- Abail, B.Z., Halima, O.I., Boulal, H., Gharous, M. El and Brouj, E. (2014), *Fertilizer Use Patterns in the Semi-Arid, Cereal Producing Region of Chaouia*, available at: <https://doi.org/10.13140/RG.2.1.2274.0886>.
- Amede, T., Gashaw, T., Legesse, G., Tamene, L., Mekonen, K., Thorne, P. and Schultz, S. (2020), "Landscape positions dictating crop fertilizer responses in wheat-based farming systems of East African Highlands".
- Andraski, T.W., Bundy, L.G. and Brye, K.R. (2000), "Crop Management and Corn Nitrogen Rate Effects on Nitrate Leaching", *Journal of Environmental Quality*, Vol. 29, pp. 1095–1103.
- Bah, A., Balasundram, S.K. and Husni, M.H.A. (2016), "Sensor Technologies for Precision Soil Nutrient Management and Monitoring", *American Journal of Agricultural and Biological Sciences*, Vol. 7 No. 1, pp. 43–48.
- Bantern, P., Hoogenboom, G., Patanotai, A., Singh, P., Wani, S.P., Pathak, P., Tongpoonpol, S., et al. (2010), "Application of the cropping system model (csm)-cropgro-soybean for determining optimum management strategies for soybean in tropical environments", *Journal of Agronomy and Crop Science*, Vol. 196 No. 3, pp. 231–242.
- Bell, R., Reuter, D., Scott, B., Sparrow, L., Strong, W. and Chen, W. (2013), "Soil phosphorus-crop response calibration relationships and criteria for winter cereal crops grown in Australia", *Crop and Pasture Science*, Vol. 64 No. 5, pp. 480–498.

- Brennan, R.F. and Bell, M.J. (2013), "Soil potassium—crop response calibration relationships and criteria for field crops grown in Australia", *Crop and Pasture Science*, Vol. 64 No. 5, pp. 523–530.
- Breure, T.S., Milne, A.E., Webster, R., Haefele, S.M., Hannam, J.A., Moreno-Rojas, S. and Corstanje, R. (2021), "Predicting the growth of lettuce from soil infrared reflectance spectra: the potential for crop management", *Precision Agriculture*, Springer US, Vol. 22 No. 1, pp. 226–248.
- Breure, T.S., Prout, J.M., Haefele, S.M., Milne, A.E., Hannam, J.A., Moreno-Rojas, S. and Corstanje, R. (2022), "Comparing the effect of different sample conditions and spectral libraries on the prediction accuracy of soil properties from near- and mid-infrared spectra at the field-scale", *Soil and Tillage Research*, Elsevier B.V., Vol. 215 No. July 2021, p. 105196.
- Dhakal, C. and Lange, K. (2021), "Crop yield response functions in nutrient application : A review", *Agronomy Journal*, Vol. 113 No. 6, pp. 5222–5234.
- Dimkpa, C., Adzawla, W., Pandey, R., Atakora, W.K., Kouame, A.K., Jemo, M. and Bindraban, P.S. (2023), "Fertilizers for food and nutrition security in sub-Saharan Africa: An overview of soil health implications", *Frontiers in Soil Science*, Vol. 3 No. March, pp. 1–18.
- Dobermann, A., Witt, C., Abdulrachman, S., Gines, H.C., Nagarajan, R., Son, T.T., Tan, P.S., *et al.* (2003), "Estimating indigenous nutrient supplies for site-specific nutrient management in irrigated rice", *Agronomy Journal*, Vol. 95 No. 4, pp. 924–935.
- Ge, Y., Thomasson, J.A. and Sui, R. (2014), "Remote Sensing of Soil Properties in Precision Agriculture : A Review Remote sensing of soil properties in precision agriculture : A review", *Frontiers of Earth Science*, Vol. 5, pp. 229–238.
- Greenwood, D.J., Floate, M.J.S., Greenwood, D.J., Nelder, J.A. and Frissel, M.J. (1982), "Modelling of Crop Response to Nitrogen Fertilizer", *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, Vol. 296 No. 1082, pp. 351–362.
- Guerrero, C., Wetterlind, J., Stenberg, B., Mouazen, A.M., Gabarrón-Galeote, M.A., Ruiz-Sinoga, J.D., Zornoza, R., *et al.* (2016), "Do we really need large spectral libraries for local scale SOC assessment with NIR spectroscopy?", *Soil and Tillage Research*, Elsevier B.V., Vol. 155, pp. 501–509.
- Haefele, S.M., Bhattachan, B.K., Adhikari, B.B., Jr, C.C.A. and Shrestha, S.M. (2014), "Agriculture , Ecosystems and Environment Spatial variability of fertilizer management and response in rainfed rice of Nepal", *Agriculture, Ecosystems and Environment*, Elsevier B.V., Vol. 189, pp. 190–198.
- Janssen, B.H. and Guiking, F.C.T. (1990), "Modelling the response of crops to fertilizers", *In Plant Nutrition—Physiology and Applications*, pp. 699–703.
- Janssen, B.H., Guiking, F.C.T., van der Eijk, D., Smaling, E.M.A., Wolf, J. and van Reuler, H. (1990), "A system for quantitative evaluation of the fertility of tropical soils (QUEFTS)", *Geoderma*, Vol. 46 No. 4, pp. 299–318.
- Leitner, S., Pelster, D.E., Werner, C., Merbold, L., Baggs, E.M., Mapanda, F. and Butterbach-Bahl, K. (2020), "Closing maize yield gaps in sub-Saharan Africa will boost soil N₂O emissions", *Current Opinion in Environmental Sustainability*, Elsevier B.V., Vol. 47, pp. 95–105.
- Malpass, D. (2022). A transformed fertilizer market is needed in response to the food crisis in

- Africa. Voices blog: Perspectives on Development. World Bank. <https://blogs.worldbank.org/en/voices/transformed-fertilizer-market-needed-response-food-crisis-africa>.
- Mason, S., McNeill, A., McLaughlin, M.J. and Zhang, H. (2010), "Prediction of wheat response to an application of phosphorus under field conditions using diffusive gradients in thin-films (DGT) and extraction methods", *Plant and Soil*, Vol. 337, pp. 243–258.
- Menzies, N.W., Donn, M.J. and Kopittke, P.M. (2007), "Evaluation of extractants for estimation of the phytoavailable trace metals in soils", *Environmental Pollution*, Vol. 145, pp. 121–130.
- Mueller, N.D., Gerber, J.S., Johnston, M., Ray, D.K., Ramankutty, N. and Foley, J.A. (2012), "Closing yield gaps through nutrient and water management", *Nature*, Nature Publishing Group, Vol. 490 No. 7419, pp. 254–257.
- Oghaz, M.M., Razaak, M., Kerdegari, H., Argyriou, V. and Remagnino, P. (2019), "Scene and Environment Monitoring Using Aerial Imagery and Deep Learning", In *2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*, No. IEEE, pp. 362–369.
- Pearce, A.W., Slaton, N.A., Lyons, S.E., Bolster, C.H., Bruulsema, T.W., Grove, J.H., Jones, J.D., *et al.* (2022), "Defining relative yield for soil test correlation and calibration trials in the fertilizer recommendation support tool", *Soil Science Society of America Journal*, No. March, pp. 1338–1353.
- Rickson, R.J., Deeks, L.K., Graves, A. and Harris, J.A.H. (2015), "Input constraints to food production : the impact of soil degradation", *Food Security*, Vol. 7, pp. 351–364.
- Rutkowska, A. and Fotyma, M. (2009), "Calibration of soil test for mineral nitrogen in Poland", *Communications in Soil Science and Plant Analysis*, Vol. 40 No. 1–6, pp. 987–998.
- Sattari, S.Z., van Ittersum, M.K., Bouwman, A.F., Smit, A.L. and Janssen, B.H. (2014), "Crop yield response to soil fertility and N, P, K inputs in different environments: Testing and improving the QUEFTS model", *Field Crops Research*.
- Savvides, A., Corstanje, R., Baxter, S.J., Rawlins, B.G. and Lark, R.M. (2010), "The relationship between diffuse spectral reflectance of the soil and its cation exchange capacity is scale-dependent", *Geoderma*, Elsevier B.V., Vol. 154 No. 3–4, pp. 353–358.
- Shanahan, J.F., Kitchen, N.R., Raun, W.R. and Schepers, J.S. (2008), "Responsive in-season nitrogen management for cereals", *Computers and Electronics in Agriculture*, Vol. 61 No. 1, pp. 51–62.
- Shepherd, K.D., Ferguson, R., Hoover, D., van Egmond, F., Sanderman, J. and Ge, Y. (2022), "A global soil spectral calibration library and estimation service", *Soil Security*, Elsevier Ltd, Vol. 7 No. April, p. 100061.
- Shepherd, K.D. and Walsh, M.G. (2002), "Development of Reflectance Spectral Libraries for Characterization of Soil Properties", *Soil Science Society of America Journal*, Vol. 66 No. 3, pp. 988–998.
- Shepherd, K.D. and Walsh, M.G. (2007), "Infrared spectroscopy - Enabling an evidence-based diagnostic surveillance approach to agricultural and environmental management in developing countries", *Journal of Near Infrared Spectroscopy*, Vol. 15 No. 1, pp. 1–19.

- Soriano-Disla, J.M., Janik, L.J., Allen, D.J. and McLaughlin, M.J. (2017), "Evaluation of the performance of portable visible-infrared instruments for the prediction of soil properties", *Biosystems Engineering*, Vol. 161, pp. 24–36.
- Soriano-Disla, J.M., Janik, L.J., Rossel, R.A.V., MacDonald, L.M. and McLaughlin, M.J. (2014), "The Performance of Visible , Near- , and Mid-Infrared Reflectance Spectroscopy for Prediction of Soil Physical , Chemical , and Biological Properties", *Applied Spectroscopy Reviews*, Vol. 49 No. 2, pp. 139–186.
- Stöckle, C.O., Donatelli, M. and Nelson, R. (2003), "CropSyst, a cropping systems simulation model", *European Journal of Agronomy*, Vol. 18 No. 3–4, pp. 289–307.
- Vitousek, P.M., Naylor, R., Crews, T., David, M.B., Drinkwater, L.E., Holland, E., Johnes, P.J., *et al.* (2009), "Nutrient imbalances in agricultural development", *Science: PolicyForum*, Vol. 324, pp. 1519–1520.
- Wajid, A., Hussain, K., Ilyas, A., Habib-Ur-rahman, M., Shakil, Q. and Hoogenboom, G. (1989), "Crop Models: Important Tools in Decision Support System to Manage Wheat Production under Vulnerable Environments", *Agriculture (Switzerland)*, Vol. 11 No. 1166, pp. 1–22.
- Williams, J.R., Jones, C.A., Kiniry, J.R. and Spanel, D.A. (1989), "EPIC crop growth model", *Transactions of the American Society of Agricultural Engineers*, Vol. 32 No. 2, pp. 497–511.

2 Infrared spectroscopy for prediction of soil properties input for crop models: a literature review

Tadesse Gashaw Asrat¹, Ruben Sakrabani¹, Ron Corstanje¹, Fassil Kebede³,
Stephan M. Haefele²

¹Cranfield University, Cranfield, UK; ²Rothamsted Research, Harpenden, UK;

³College of Agriculture and Environmental Science, AgroBioscience Program,
Mohammed VI Polytechnic University, Ben Guerir, Morocco.

Abstract

This literature review evaluates the efficacy of soil spectroscopic tools and methods for enhancing soil property predictions and the subsequent integration with crop models to facilitate informed decision-making regarding fertilizer management. Soil spectroscopy is a valuable tool for providing timely and cost-effective soil information, helping bridge the gap in access to data. It supports decision-making based on inherent soil productivity potential and facilitating the development of management strategies tailored to specific soil characteristics and crop-nutrient interactions, such as nutrient responses, imbalances, or toxicity, at the site level. Utilizing the visible-near infrared (vis-NIR) and mid-infrared (MIR) ranges of the electromagnetic spectrum, soil scanning instruments are being developed and commercialized, each with specific technical requirements, sample preparation protocols, and associated costs. Different soil properties exhibit varied responses across different wavelength regions and soil scanning instruments, impacting the precision of predictions. However, overall performance remains relatively consistent across these ranges. While most soil properties demonstrated satisfactory prediction performance using soil spectroscopic techniques, discrepancies existed, particularly concerning the utilization of soil spectral libraries and the absence of proper evaluation methods.

The review underscores the importance of assessing the efficiency and uncertainty of the low cost and handheld soil scanning instruments and the

wavelength regions they cover in relation to soil property predictions with expensive instruments. Additionally, it emphasized the need to translate the information generated by soil spectroscopy into actionable insights for decision-making, particularly in addressing variability in crop response to soil conditions and nutrient application. Crop models that incorporate a combination of soil information have the potential to leverage advancements in soil spectroscopy, facilitating the exploration and establishment of relationships between soil properties and crop performance. By improving access to soil information through soil spectroscopy, these models can enhance decision support capabilities, ultimately contributing to more informed fertilizer management practices.

Keywords: vis-NIR, MIR, soil spectral library, crop models, QUEFTS, informed decision-making

2.1 Introduction

Soil spectroscopy enables the rapid acquisition of soil information using multiple stages, devices and parameters in-field or in-lab. The technique employs most notably the visible near infrared (vis-NIR) (350 – 2500 nm wavelength) and mid infrared (MIR) (2500 - 16667 nm wavelength) regions of the electromagnetic spectrum. The spectra generated in the vis-NIR and MIR electromagnetic regions are based on the interaction of the radiation with soil molecules, resulting in absorption, reflection and scattering patterns that are characteristic of soil properties. These techniques allow estimation of multiple agriculturally important soil properties from acquired spectra for less time and cost than conventional soil analysis methods (Chang et al., 2001; Xu et al., 2018; Barra et al., 2021; Rodríguez-Pérez et al., 2021; Li et al., 2022).

The principle is that a material can reflect, absorb, scatter, and emit electromagnetic radiation in a characteristic manner that depends on its molecular composition and shape, resulting in a unique spectral fingerprint (Lemke and Quinlan, 2003). The spectral signatures of a soil sample contain information on the inherent composition, comprised of minerals, organic compounds, water, and soil microorganisms. The concealed and encoded information within the spectral signature necessitates extraction through the

construction of predictive models. The spectra, hence, are calibrated for each soil property with conventionally measured or estimated soil attributes. This can be achieved via multivariate regression techniques, such as the partial least square (PLS) and principal component regressions (PCR), which establish a linear relationship (Morellos et al., 2016; Lucà et al., 2017). Alternatively, machine learning algorithms, including Random Forest, Support Vector Machines and Deep Learning models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), are employed to capture non-linear patterns and interactions between spectra features and soil properties (Viscarra Rossel and Behrens, 2010; Shi et al., 2013; Kuang et al., 2015). All these models are developed using a paired database of soil spectral and conventional soil tests, the soil spectral library (SSL), forming the foundation of soil spectral analysis. The entries of this database should be as diverse, and representative of the target soil types and prediction scales as possible to ensure valid predictions.

2.2 Overview of IR soil spectroscopy towards soil analysis

2.2.1 General comparisons between MIR and vis-NIR spectroscopy

The primary emphasis of soil spectral sensing methods for assessing soil properties revolves mainly around the vis-NIR and MIR spectral ranges, which have been demonstrated to offer improved accuracy in estimating soil variables and are easier to apply than conventional soil analysis methods (Blitz, 1998; Vågen et al., 2010; Bellon-Maurel and McBratney, 2011; Towett et al., 2015). In the vis-NIR and MIR regions, the diffusively reflected light provides information not only on the chemical composition of the sample (through absorption) but also on its microstructure (via scattering) (Rinnan et al., 2009). The spectral sensing techniques in the vis-NIR cover the wavelength range from 400 to 700 nm in the visible spectrum, extending to 701-1100 nm in the NIR range, and further to 1101-2500 nm in the short-wave infrared (SWIR) region. This classification provided valuable insights for developing a specialized spectroscopic tool that incorporates key wavelength regions, helping to make the tool more compact and efficient for soil characterization. In contrast, the MIR analysis spans the wavelength range from 2500 to 25,000 nm, with soil scanning instrumentation typically recording up

to 16,667 nm. These two electromagnetic regions principally vary in terms of absorption mechanism, absorption coefficients, spectral complexity, and light diffusion characteristics. The main difference between the vis-NIR and MIR ranges is that absorption in MIR corresponds to fundamental bands of molecular vibrations, whereas absorptions in vis-NIR corresponds to overtones and combinations of these fundamental bands (Abbas et al., 2012; Gholizadeh et al., 2013).

Bellon-Maurel and McBratney (2011) point out that the significance of this variation is that (i) absorption coefficients (also called extinction coefficients) are much smaller in the vis-NIR range and (ii) the vis-NIR spectrum is much more overloaded because of the abundance or combination and overtone bands. This implies that the specificity of bands is less in the vis-NIR range due to broader and less specific absorption features which reduces the sensitivity in detecting subtle variations in soil properties, minimal interference from background noise and other factors. Better spectral resolution and sample size may not make any significant difference (Manley, 2014). In contrast, diffusion of light is less in the MIR range which enable to obtain more focused and precise measurements in terms of penetration depth and sample characterization (Viscarra Rossel et al., 2006).

Soriano-Disla et al. (2014) highlighted the distinct spectral characteristics of the vis-NIR and MIR regions, emphasizing the dominance of overtones and combination vibrations of lighter atoms, particularly hydrogen, in the vis-NIR, while heavier atomic groups contribute to absorption in the MIR (Viscarra Rossel et al., 2006). This observation is supported by Williams and Norris (1987), who noted that factors affecting light diffusion, such as physical soil structure and water content, have a greater impact on vis-NIR measurements. Equally, Reeves (2010) reported that the MIR region is strongly influenced by moisture content, which can degrade vis-NIR measurements as well. Hutengs et al. (2019) compared the effects of soil moisture variation and sample heterogeneity on vis-NIR and MIR measurements, reported that the MIR region outperformed vis-NIR in this regard. They also noted that sample pre-treatment had a significant effect

on MIR calibrations, with finely ground samples yielding more accurate results. However, variations in vis-NIR measurements were negligible in comparison. This finding was also supported by Li et al. (2022) who reported that calibrations made with samples of finer particle size ($\leq 0.05\text{mm}$) improved predictions with portable MIR spectrometers.

Such variations suggested the distinction for a given spectra acquisition circumstance by the two electromagnetic regions. These characteristic differences in spectral response inform the choice of spectrometers and applicability, with vis-NIR techniques favoured for in-field use due to their ease of application, while the MIR region is considered a gold standard in soil spectroscopy. As a result, vis-NIR spectroscopy has garnered more attention for advancement, particularity in field conditions and instrument configuration. This suggests a thorough consideration of both options when undertaking spectral analysis and identification of spectral regions for a specific soil property prediction and applicability for informed decision-making in agriculture (Soriano-Disla et al., 2014).

2.2.2 Advancement of tools and methods in Vis-NIR and MIR spectral ranges for assessing soil properties

Using a soil spectrometer, samples will be scanned to acquire the characteristic diffuse reflection or absorbance spectra which will then be studied to predict target variable properties using a wide array of classification and regression algorithms. Viscarra Rossel et al. (2016) elaborated that spectra can be recorded from different platforms; by proximal sensing in the field, in the laboratory using prepared samples, or from remote sensing platforms with multi- and hyperspectral capabilities. More efficient proximal soil reflectance spectroscopic tools have been developed as alternatives to support or, to some extent, replace the conventional soil analysis for quantification of soil properties and crop nutrient determination for the past few decades (e.g. Ben-Dor and Banin, 1995; Shepherd and Walsh, 2002; Stenberg et al., 2010; Gholizadeh et al., 2013; Nocita et al., 2015; O'Rourke et al., 2016).

At the beginning, measurements or scanning of samples using proximal methods have been done on bench-top in laboratory conditions with finely ground (usually that pass through less than 2mm sieve), compacted, smooth surface, and with lowest humidity possible. However, this constraint will hinder the widespread adoption of these methods in procuring detailed soil information at higher resolution, which is a key benefit of soil spectroscopy. This limitation arises from the need for sample processing and transportation to centralized laboratory facilities, which reduces the advantage of soil spectroscopy (Viscarra Rossel et al., 2022). Therefore, the advancement of compact and easily transportable spectrometers for both vis-NIR and MIR ranges, with minimal sample processing needs is ongoing (Figure 2.1).

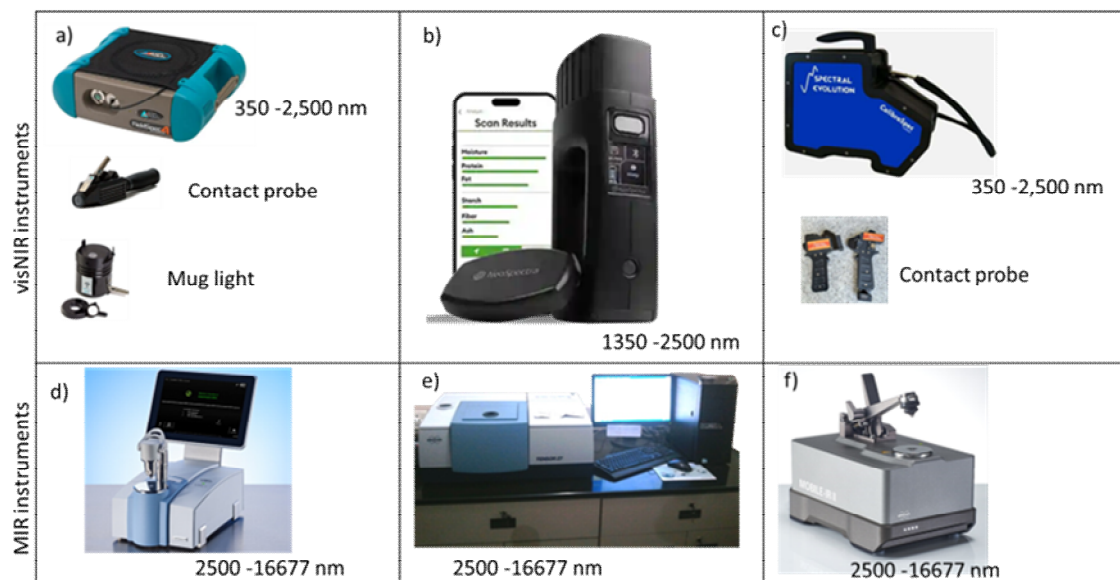


Figure 2.1 : Some of the widely used soil spectroscopy instruments for soil property assessment. a) FieldSpec with either contact probe or mug light (Source: Malvern Panalytical) - portable; b) Neospectra saucer (Source: Si-Ware systems) – handheld; c) Spectral Evolution with contact probe (Source: Pro-lite technology) – portable; d) Alpha II (Bruker) – benchtop; e) Tensor 27 (Source: Bruker) – benchtop and e) Mobile-IR II (Source: Bruker) – portable.

The trend towards miniaturization of spectrometers began over two decades ago (Sudduth and Hummel, 1993). In recent years, compact handheld spectrometers operating in the MIR and vis-NIR ranges have become commercially available

and have found extensive use in various analytical applications (Sorak et al., 2012; Guillemain et al., 2017; Soriano-Disla et al., 2018; Yan et al., 2019). Technological advancements leveraging micro-electromechanical structures (MEMSs), thin film filters, lasers, light emitting diodes (LEDs), fibreoptic assemblies, and high-performance detector arrays have led to the production of rugged and affordable handheld instruments (Viscarra and Bouma, 2016).

Studies have evaluated the applicability of these research-grade handheld instruments for soil property determination in both MIR and vis-NIR ranges (Gholizadeh et al., 2017; Hutengs et al., 2018; Hutengs et al., 2019). Recent developments also introduced portable and cost-effective miniature spectrometers for the NIR region, such as Neo SpectraTM, NIRScan Nano (Ng et al., 2020; Wolfrum et al., 2020), and the ASD systems Agrispec and FieldSpec (Tang et al., 2019). These studies reported that performance of miniature spectrometers varied due to technical limitations, including energy source type, detector quality, resolution, sampling accessories, and energy intensity. Maintaining measurement performance is crucial as the reduction in equipment size and cost should not compromise data quality.

Promising results have been reported with miniature spectrometers, though, for predicting certain properties with reasonable accuracy compared to bench-top measurements of well-processed samples (Sudduth and Hummel, 1993; Lee et al., 2009; Tang et al., 2014). The studies identified certain limitations particularly with portable NIR instruments covering a limited spectral range, hindering the acquisition of crucial regions for certain soil characteristics (Soriano-Disla et al., 2014). This led to the development of NIR instruments covering broader spectral ranges for improved prediction accuracy (Sudduth and Hummel, 1993; Lee et al., 2009; Figure 2.2). Recent studies demonstrated comparable accuracies between miniature NIR spectrometers (1300-2600 nm) ($R^2 = 0.70$) and full vis-NIR spectrometers (350-2500 nm) ($R^2 = 0.75$) for predicting total carbon and organic carbon in diverse soils (Sharififar, 2019; Tang et al., 2014). Similarly, portable MIR instruments also showed performance similar to bench-top equipment with

spectral ranges of 1650-5000 nm (Soriano-Disla et al., 2017; Hutengs et al., 2018).

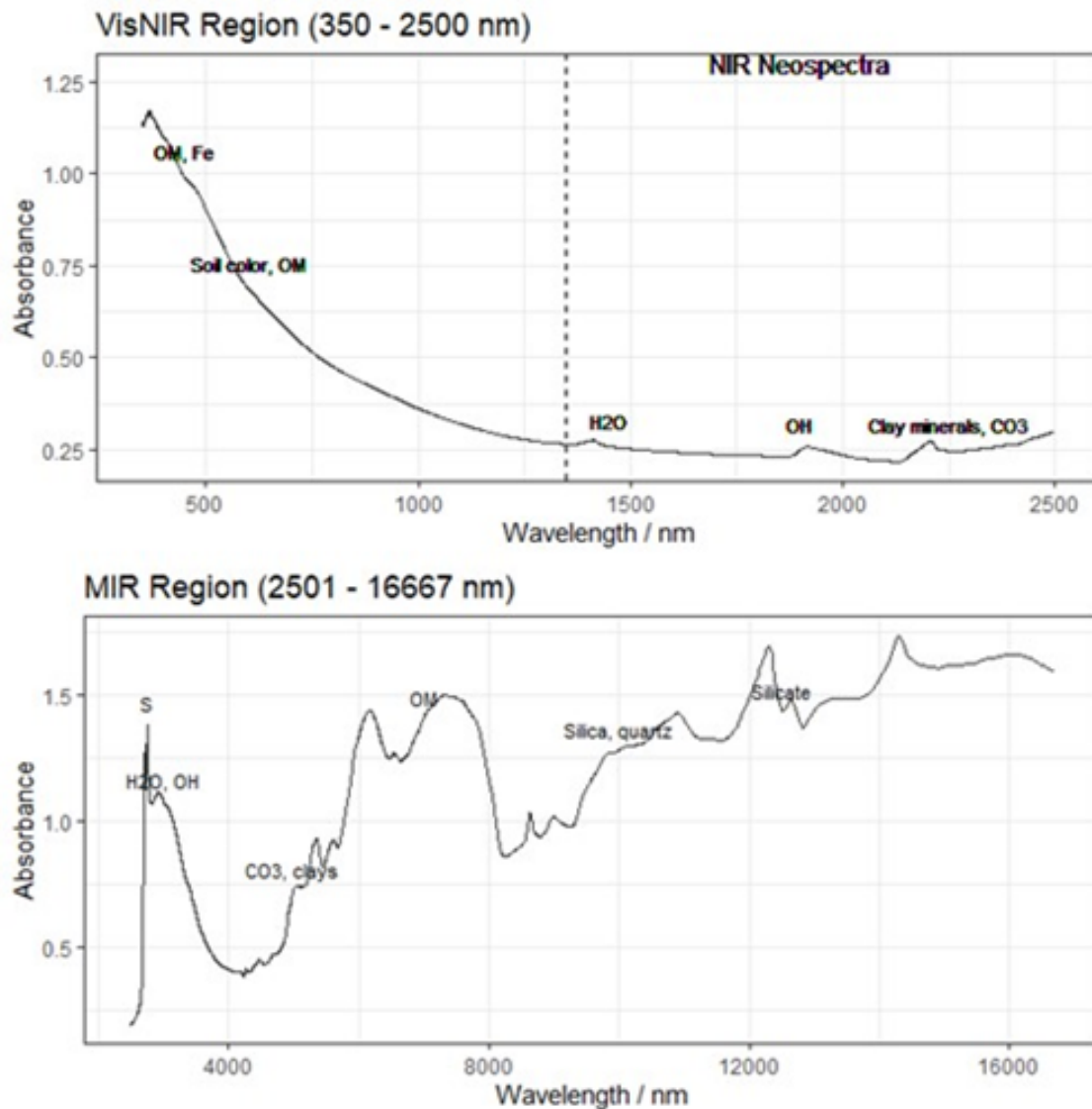


Figure 2.2 : Wavelength regions for visNIR and MIR soil spectroscopy. The vertical dotted lines within the visNIR range indicate the wavelength region covered by the Neospectra NIR instrument (Si-Ware systems). Characteristic soil wavelength regions are adapted from Coblinski et.al. (2020), and Viscarra Rossel et.al. (2006).

The miniaturization of handheld spectrometers has resulted in reduced weight, with some NIR instruments weighing less than 1 kg, and efforts are underway to integrate them into mobile phones (Pügner et al., 2016; Pätzold et al., 2019). This trend is expected to lower costs down to below US\$1,000, making the technology

even more accessible (Yan et al., 2019). While miniaturized NIR spectrometers are anticipated to find applications in agriculture and other sectors, MIR spectrometers are typically more expensive due to their design, which includes sensitive detectors and light sources to handle the longer wavelengths. These costs often restrict the MIR instruments to industrial, military, or homeland security applications (Yan et al., 2019). The disadvantage of MIR soil analysis under field conditions stems from its greater sensitivity to unprepared samples and reduced signal-to-noise ratio compared to the NIR instruments (Helfenstein et al., 2021). Furthermore, the components required for the MIR instruments, such as cooled detectors (requiring liquid nitrogen) and specialized optics, add significantly to their production costs. As a result, the MIR instruments can cost several thousands (>>£70,000), making them less accessible for general agricultural use.

Similarly, the ASD visNIR instruments, though highly accurate and widely used in research, can be prohibitively expensive (~£70,000). This price point limits their accessibility for routine agricultural applications. On the other hand, newer, more affordable NIR instruments like Neospectra are significantly less expensive and still offer robust performance for fieldwork. As these miniaturized NIR instruments become more cost-effective and efficient, they could inherently be integrated into agricultural workflows.

Portable miniature instruments overall present a promising solution for rapid, cost-effective, in-field testing. This development may broaden opportunities for growers, consultants, and soil sampling contractors to consider the use of handheld instruments for soil analysis in the field or the integration of smaller instruments into commercial soil testing services. However, their efficiency and effectiveness for use with existing decision support tools remains to be fully investigated.

2.2.3 Spectral data handling and processing

The application of infrared spectroscopy in soil fertility evaluation is largely dependent on spectra pre-treatment and multivariate calibration due to strong interferences in the spectra. Prior to any further spectra pre-processing methods,

the reflectance spectra (R) collected will be averaged from replicated readings and converted to absorption with an inverse logarithmic relationship (Maarschalkerweerd, 2014). Next, some regions will be trimmed to leave the appropriate wavelength range. The wavelength regions to be trimmed will have a low signal to noise ratio and are usually between 350–399 nm and 2451–2500 nm for vis-NIR (Lazaar et al., 2020) and below 2500 nm for MIR (Ng et al., 2022). Following this, a preprocessing procedure will be employed for the assessment of outliers to improve data quality, which consists of a variety of mathematical methods for correcting light scattering in reflectance measurements and data enhancement before the data is used in calibration models (Gholizadeh et al., 2015).

Given that, preprocessing techniques can be divided into four categories, namely smoothing (used for noise or random measurement error reduction), baseline removal (correcting background signal which influences measurements like peak height and peak area), scaling (to compare intensities so that every spectrum has a mean of zero and a standard derivation of 1) and normalization (used for absorption feature enhancement and identification) (Rinnan et al., 2009; Santos et al., 2013; Gholizadeh et al., 2015). These spectral preprocessing techniques have provided enhanced prediction performance for specific soil attributes which are closely related to specific infrared regions. Studies showcased important contributions of spectral preprocessing techniques, which improved prediction accuracy for some of the most important soil health indicators (Rinnan et al., 2009; Santos et al., 2013; Gautam et al., 2015; Gholizadeh et al., 2015; Vestergaard et al., 2021). Basic soil physical and chemical properties showed high correlation with the derivative reflectance values located around the principal absorption features in the vis-NIR regions (Shepherd and Walsh, 2001) and in MIR wavelength regions in the PLSR modelling (Terhoeven-Urselmans et al., 2010). Gholizadeh et al. (2015) compared the performance of different preprocessing algorithms for several soil chemical properties. Their study identified that the first derivative preprocessing was determined as the best method ($R^2 > 0.9$) for predicting Cu, Mn, Pb, and Zn concentration, whereas continuum removal preprocessing was chosen as the final method for predicting

Cd. Heil and Schmidhalter (2021) reviewed the effect of several preprocessing techniques and their contribution to prediction performance. They indicated that the choice of preprocessing techniques is somewhat arbitrary, and little is known about how it affects the final prediction of soil properties.

Carmon and Ben-dor (2017) also emphasized that the selection of preprocessing algorithms for a given dataset is often non-systematic and ends once the user obtains a favourable result based on a subjective impression. They recommended that novel pre-processing techniques and model-validation approaches are vital for the development of a systematic procedure for model development that considers many possible alternatives. To achieve this goal, software (e.g. R, Python and other built-in software) with appropriate packages could aid in handling and managing acquired spectra to test a wide array of procedural data processing, calibration, validation and prediction methods, for an improved level of precision and technicality.

2.2.4 Soil spectral analysis and predictive models

Besides developing soil scanning instruments and protocols for collecting and then preprocessing the soil spectra under various conditions, the selection of an appropriate modelling platform plays a crucial role in accurately predicting soil properties for decision-making purpose. This selection not only influences the predictive performance but also determines how the resulting soil information, along with its associated uncertainty, is effectively integrated into decision support tools. Soil spectral analysis is focused on extracting meaningful information from spectral data, which is increasingly available from remote sensing, laboratory analysis, and proximal field scanning. This approach goes beyond mere prediction; it aims to deepen our understanding of the underlying relationships between soil properties and their spectral signatures. By analysing these spectral patterns, researchers can gain insights into the mechanisms driving soil behaviour, which can then be used to improve soil management practices. The progress in this field is accelerated by the availability of open-source algorithms and application of advanced machine learning techniques, which allow for more sophisticated and accurate models (Padarian et al., 2020).

Given the complexity of infrared spectra and the large datasets involved, straightforward interpretations are often challenging. Therefore, machine learning methods have been devised to process spectral data and conduct big data analytics (Santos et al., 2013). Various multivariate calibration methods have been assessed for their relevance and precision in extracting pertinent information from extensive datasets in soil applications. They can be broadly classified as machine learning regression, deep learning neural networks and other decision trees, and they might respond differently to calibration sample numbers used. The choice of the modelling technique depends on factors such as the nature of the data, the complexity of the relationships between spectral features and soil properties, and the desired level of interpretability and prediction accuracy.

Linear modelling frameworks plays a crucial role in soil spectroscopy by enabling dimension reduction in datasets by extracting principal components (PCs). These PCs capture the majority of the variation in the soil variables and are then utilized to establish the relationships in the model. The PCs are derived either from the covariance matrix, which maintains the original scale of the variable, or from the correlation matrix, which standardizes the variables to ensure equal weighting across all dimensions. This choice depends on whether the focus is on preserving the data's original scale or normalizing it. As such, these frameworks offer simplicity, interpretability, and computational efficiency, while also providing insight into the direction and magnitude of relationships between spectral features and soil properties. Common linear modelling techniques include principal components regression (PCR) (Chang et al., 2001; Nocita et al., 2014), partial least squares regression (PLSR) (Breure et al., 2021), stepwise multiple linear regression (SMLR) (Vasques et al., 2010), and backward interval partial least squares (BIPLS) (Xiao et al., 2018; Kawamura et al., 2020). Among these, the PLSR has become the most widely adopted approach in soil spectroscopy due to its superior ability to handle multivariate data, including multicollinearity, and its strong prediction performance coupled with computational simplicity.

While methods like PCR, focus on explaining the variance within the predictor variables, PLSR goes a step further by maximizing the covariance between predictors and responses, making it specifically effective in cases where predictor variables (i.e., spectral data) are highly collinear or redundant. In contrast, SMLR and BIPLS offer more selective modelling approaches by isolating specific predictors or intervals within the spectrum, but they may lack the robustness and flexibility provided by PLSR in handling large, complex datasets. From a systematic review and meta-analysis, Ahmadi et al. (2021) reported that PLSR and its variations (e.g., MPLSR) accounted for about nearly 70% of the studies in soil spectroscopy, underscoring its dominance. This preference stems from PLSR's suitability for handling data with multiresolution and collinearity, which are common in soil spectral datasets, all while delivering strong predictive results with relatively straightforward computation.

The non-linear modelling frameworks such as deep learning neural networks and decision trees learn and detect patterns in a hierarchical way. The non-linear modelling frameworks mainly applied in soil spectroscopy include random forests (RF) (Viscarra Rossel and Behrens, 2010), artificial neural networks (ANN) (Ayoubi et al., 2011), and support vector machine regression (SVMR) (Stevens et al., 2013; Gholizadeh et al., 2015). These categories of soil spectroscopic modelling frameworks have advantages such as flexibility in capturing complex relationships, high predictive accuracy, and adaptability to different types of data. However, they require high computational processing and adequate training datasets.

Researchers studied these modelling platforms to identify their suitability, requirements, and adaptability in predicting various soil properties in both MIR and vis-NIR ranges. Ng et al. (2020) compared the number of sample entries among machine learning models and deep learning approaches. They found varying performance among modelling platforms depending on calibration sample entry numbers. The use of a smaller number of calibration samples caused better performance of machine learning regressions while deep learning platforms required large number of calibration samples. This could indicate that

the deep learning platforms may have the capability to select samples close to the unknown sample from the large spectral library, which requires intensive computation and expert inputs that complicate applicability in a wider context (Padarian et al., 2019; Albinet et al., 2022). Meanwhile, for machine learning regression, the calibration samples should only contain representative and enough samples to balance between over-estimation and under-estimation (Wetterlind and Stenberg, 2010; Luce et al., 2022) which can be useful in real-time soil property prediction.

2.3 Soil spectral libraries development and usage

2.3.1 Structure of soil spectral libraries and considerations

Calibration of spectral data using conventional soil analysis requires analysing adequate numbers of soil samples for a wide suite of properties representative for the region of interest. This imposes a limitation for the widespread use of soil spectroscopy due to added costs and complexity. The goal is to develop a robust predictive model from the database to estimate the soil characteristics of new samples without the need for further conventional soil analysis, thereby reducing workload, time needed, and use of hazardous chemicals. This led to the establishment of soil spectral libraries (SSL), which are the databases containing paired datasets of conventional soil analysis and spectral data, necessary for interpreting the soil spectral information of new soil samples (Brown et al., 2006). Defining, describing, and showcasing the structure and steps of development of the spectral library can support the reliable use and adoption among users to predict soil parameters for new samples at large.

Consequently, the SSL is expected to include as many sub-structures and descriptions as possible to be useful in wider applications which includes a proper database management system to organize it efficiently. Along with the paired spectral and conventional analysis datasets, metadata ought to be incorporated providing information regarding methods used, sample locations, sampling depth, sampling date, and soil type (Brodský et al., 2011). It is also advised that quality control information and standardization protocols are included in the soil spectral database which provide reliability and other information important for subsequent

data analytical procedures to be carried out (Viscarra Rossel et al., 2016; Baumann et al., 2017; Lobsey et al., 2017). Such approaches can minimize systematic errors while adopting or merging different databases and hence could improve the precision to be obtained (Liu et al., 2018). Accordingly, there are three basic requirements for developing a spectral library of soil: (a) the SSL must contain sufficient number of samples (usually recommended >30), representing the variability of soils found in the region to which the SSL refers; (b) samples must be carefully sub-sampled, handled, prepared, stored and scanned (any interference to the sample may affect the spectral signature); and (c) the analytical reference data from the samples to be used in calibrations must be acquired through recognized and trusted analytical procedures (Brown et al., 2006; Bellinaso et al., 2010).

2.3.2 Global and regional initiatives for developing a soil spectral library

Various initiatives and institutions worldwide are actively engaged in the development of a comprehensive global SSL, encouraging contributions from diverse institutions. This collective effort aims to enhance accessibility of the SSL, allowing users to access and utilize subsets or the entire dataset for various purposes and scales. One prominent global initiative, the Global Soil partnership (GloSIS), strives to assist countries in organizing and sharing their soil data online. By employing customized open-source software and adopting a distributed system design, GloSIS promotes data sharing and makes soil information readily available for decision-making at local, national, continental, and global levels. According to Shepherd et al. (2022), the infrastructure is intended to store and share data through national, regional, and institutional nodes or soil information systems (SIS), interconnected via Open Geospatial Consortium (OGC) webservices to a central portal. This approach facilitates customization at the national level, ensures control over data access, and enhance findability and accessibility through the different portals.

Additionally, several open-access SSLs have been developed and made available to users through various initiatives and contributors. Examples include

the Global Soil Spectral Library (GSSL), the Geocradle Regional Soil Spectral Library (GRSSL), and the NSSC-KSSL MIR Soil spectral library. These libraries contain diverse datasets covering a range of soil property attributes, thereby providing valuable resources for understanding soil spectral behaviour and predicting soil attributes from spectra. It is recommended and encouraged to contribute to the ongoing efforts to develop a dynamic and easily updateable database with broader global coverage (Viscarra Rossel et al., 2016; Wadoux et al., 2021). While there are debates regarding the representativeness of archived soil samples, both legacy and new soil sampling schemes offer valuable insights for data analysis, interpretation, and model development across various applications (Bellinaso et al., 2010; Shepherd et al., 2022). Efforts to establish and disseminate standards for sample preparation, spectral measurement, and data exchange aim to minimize systematic effects and ensure the interoperability of soil spectra from different libraries (Liu et al., 2018). Overall, ongoing efforts in developing and maintaining SSLs hold significant promise for advancing soil science research, facilitating informed decision-making, and supporting sustainable land management practices through the aid of soil spectroscopic approaches.

2.3.3 Applicability of soil spectral libraries

To take the soil spectroscopic approach towards wider application, there are still concerns about the source of uncertainties and the related selection of appropriate calibration sample entries from a large SSL for new samples to be predicted (Ramirez-Lopez et al., 2013; Sun et al., 2021). This is because the SSL might possess redundant information as well as noisy or non-informative samples (regarding the local spectral variability) hindering the development of a specific predictive model for a specific farmland (Ramirez-Lopez et al., 2013). The routine evaluation of soil predictions using soil spectroscopy has been to split the data into a calibration and a validation set, either randomly or with stratifying algorithms in the spectra space such as k-means clustering (Næs, 1987), Kennard-Stone sampling (Kennard and Stone, 1969; Nawar and Mouazen, 2018; Ng et al., 2018), conditioned hypercubic sampling (Minasny and McBratney,

2006; Moloney et al., 2023), a combination of these (Ogen et al., 2019); or by stratification using environmental covariates (Wijewardane et al., 2018; Dotto et al., 2020). Soil spectral library can be developed at various spatial extent such as local (Guerrero et al., 2016; Dotto et al., 2018; Moura-Bueno et al., 2019), national (Nocita et al., 2014; Wijewardane, 2016; Demattê et al., 2019), regional (Shepherd and Walsh, 2002; Vasques et al., 2010), continental (Stevens et al., 2013), or global (Viscarra Rossel et al., 2016). In all these scaled SSLs a spatial component is embedded depending on the variation in the sampling design, the number of samples included, and the spectra information generated.

Optimised spectral libraries are supposed to minimize and/or avoid further requirement of conventionally analysed soil samples to calibrate a spectroscopic model while predicting soil parameters for new samples. Moreover, large datasets might not be computationally effective and can cause problems when using routine calibration methods for regression by increasing noise and leading to model inadequacy (Sáiz-Abajo et al., 2005). By selecting representative, well-distributed data for calibration sets from abundant SSL data entries, fewer samples could be sufficient to build a robust and reliable model without losing the prediction accuracy.

Given that, several approaches have been explored to enhance the prediction performance for soil properties and develop strategies for informed decision-making using large and diverse SSL. One key approach involves enhancing model accuracy by incorporating local data into existing SSLs, a technique known as spiking (Padarian et al., 2019a). Spiking, combined with model calibration and similarity analysis, has proven effective in improving predictive performance with only a small number of additional samples, often surpassing the results achieved by conventional SSL methods (Angelopoulou et al., 2020). Another approach proposed by Ng, Husnain, et al. (2020) involves addressing the challenge of low prediction capability for potential available nutrients by developing categorical models. Using such an approach, they categorized numerical values into three classes (low, medium, and high) for various properties relevant to fertilizer recommendation doses, such as total N, potential P and K, available P (Olsen

and Bray), available K (Morgan) or exchangeable K, to use soil spectral predicted soil properties for decision-making. Notably, rapid discrimination between high, optimal, and low P classes has shown promise in improving common practice (Pätzold et al., 2019).

Recent advancements have opened the door to downscaling large spectral libraries for local application, reducing the need of new local calibrations while enhancing prediction performance and lowering the cost of soil information. One such method is geographically weighted partial least square regression (GW-PLS), which applies spatial weighting to the regression process, assigning greater importance to coefficients near the new sample location (Song et al., 2021). This spatial weighting adjusts the influence of calibration samples based on their geographic proximity to the target location, ensuring that local variations in soil properties are better captured. In this GW-PLS approach, the residual between the predicted values from the ordinary model and the observed soil properties are used to capture the spatial autocorrelation. This means that the model not only learns from the spectral information but also accounts for the spatial relationships between samples, improving prediction accuracy by recognizing patterns of spatial dependency and associated soil property variation due to varying moisture levels, organic matter content, or mineral compositions.

Another approach is continuous spectra similarity analysis, such as Memory-based Learner (MBL), which selects calibration samples from a large SSL based on spectral distance to the new sample (Dangal et al., 2019; Summerauer et al., 2021). Unlike traditional calibration models that use all available data, MBL narrows down the pool of calibration samples to those most relevant for predicting soil properties in a new sample. This is achieved by calculating the spectral distance—a measure of how similar the spectral signature of the new sample is to the other samples in the SSL. By prioritizing spectrally similar samples, MBL creates a more targeted and efficient calibration model, which not only improves prediction accuracy but also reduces computational complexity arising from models learning while learning from a large and diverse dataset.

Both the GW-PLS and MBL have demonstrated their ability to improve soil property predictions for new samples using fewer calibration samples than the full SSL, leveraging geographic and spectral proximity, respectively. These approaches offer alternatives to conventional groupings by better utilizing large SSLs to enhance local-scale soil property predictions. The spatial autocorrelation of spectral signatures can thus play a crucial role in selecting calibration samples, improving predictability as soil properties and their spectral behaviour often mirror spatial variability.

2.4 Precision among spectrally predicted soil properties important in soil-crop-fertilizer model inputs by vis-NIR and MIR range

2.4.1 Soil organic carbon and total nitrogen

Soil organic carbon (SOC) plays a crucial role in regulating soil nutrients by acting as reservoirs for plant nutrients, including nitrogen, phosphorus, sulphur, and many micronutrients. Through decomposition and mineralization processes, these nutrients are gradually released into the soil solution over time. Additionally, SOC contributes to several beneficial soil functions such as cation exchange capacity, soil aggregation, microbial activity and diversity, pH buffering, and water holding capacity. Consequently, many environmental and soil-crop-fertilizer models incorporate the role of SOC in optimizing empirical and/or deterministic models. SOC and total nitrogen have a consistent relationship in soil, where for every 10 units of SOC, 1 unit of total nitrogen is present. And although these two soil properties are not directly linked to plant nutrition, they significantly influence nutrient budgeting and availability to plants in a relational manner. Hence, there is high demand to assess soil organic carbon and total nitrogen in the soil easily and cheaply with high resolution information. Consequently, most of the studies in soil spectroscopy have evaluated the determination of soil organic carbon and total nitrogen (Stenberg et al., 2010; Dotto et al., 2019). Mostly, these were predicted with good to excellent agreement (Table 2.1). This is mainly because SOC have well-recognized absorption features in the vis-NIR region and MIR regions (Viscarra Rossel et al., 2006; Rodríguez-Pérez et al., 2021). But in most

of the studies, MIR was superior and better in predicting SOC and total N, although the results are comparable in the prediction category performance evaluation criteria.

Table 2.1 : Soil property prediction by vis-NIR and MIR ranges for the validation set.

Soil property	Vis-NIR			MIR			References
	Training set (N)	R2	RMSE	Training set (N)	R2	RMSE	
SOC (g/kg)	544; 217; 81; 45; 696	0.88; 0.95; 0.68; 0.90; 0.87	1.6; 1.75; 0.54; 1.5; 0.4	544; 217; 81; 45; 696	0.96; 0.96; 0.98; 0.94; 0.96	1.1; 1.4; 0.15; 1.1; 0.23	McCarty and Reeves (2006); Xie et al. (2011); Gates (2018); Araújo et al. (2015); Knox et al. (2015)
SOM (g/kg)	276	0.77	3.1	276	0.84	2.4	Cañasveras et al. (2012)
Total N (g/kg)	544; 217; 81; 285	0.85; 0.95; 0.69; 0.75	0.16; 0.12; 0.05; 0.06	544; 217; 81; 285	0.93; 0.96; 0.95; 0.74	0.1; 1.4; 0.02; 0.06	McCarty and Reeves (2006); Xie et al. (2011); Gates (2018); Johnson et al. (2019)
OlsenP (mg/kg)	50	0.71	1.6	50	0.55	1.9	Van Groenigen et al. (2003)
Available P (Bray-2) (mg/kg)	285	0.25	4.64	285	0.18	4.79	Johnson et al. (2019)
Mehlich I P (gm/kg)	374; 45	0.21; 0.05	89.1; 6.4	374; 45	0.15; 0.23	131; 5.6	McCarty and Reeves (2006); Araújo et al. (2015)
Mehlich III P (mg/kg)	273	0.16	14.56	273	0.27	13.69	Johnson et al. (2019)
Total P (mg/kg)	81	0.49	138.6	81	0.63	123.96	Gates (2018)
NH4-Ac Exch.K	81; 220	0.63; 0.8	0.51; 43.21	81; 220	0.77; 0.85	0.33; 37.78	Gates (2018); Shao and He (2011)
Mehlich I K	374	0.31	27.3	374	0.61	19.1	McCarty and Reeves (2006)
Mehlich III K	273	0.47	0.16	273	0.54	0.15	Johnson et al. (2019)
Clay (%)	529; 276; 285	0.69; 0.8; 0.70	22.2; 40.7; 9.96	529; 276; 285	0.77; 0.82; 0.72	19.1; 37.9; 9.16	McCarty and Reeves (2006); Cañasveras et al. (2012); Johnson et al. (2019)
CEC (cmol+/kg)	276; 45; 273	0.88; 0.80; 0.67	3.9; 10.6; 6.71	276; 45; 273	0.91; 0.88; 0.75	3.5; 8.5; 5.78	Cañasveras et al. (2012); Araújo et al. (2015); Johnson et al. (2019)
pH (1:1 soil suspension in water)	544; 81	0.53; 0.40	0.31; 0.31	544; 81	0.75; 0.42	0.23; 0.20	McCarty and Reeves (2006); Gates (2018);
pH (1:2.5 soil suspension in water)	276; 45; 285	0.47; 0.31; 0.59	0.2; 0.33; 0.55	276; 45; 285	0.61; 0.85; 0.73	0.2; 0.15; 0.45	Cañasveras et al. (2012); Araújo et al. (2015); Johnson et al. (2019)

2.4.2 Plant available phosphorus, total phosphorus, and exchangeable potassium

Phosphorus and potassium are crucial macro-nutrients that significantly impact agricultural productivity and ecosystem health. Various conventional methods have been employed to assess soil phosphorus and potassium levels, and many studies have evaluated predictions using soil spectroscopy. Generally, it was found that soil properties with indirect associations to spectra can be estimated with varying degrees of accuracy, such as extractable nutrients (Towett et al., 2015). Thus, prediction performance of available soil phosphorus and potassium using soil spectroscopy typically have not been very precise and highly variable (Table 2.1), and many studies reported the challenges in achieving precise quantitative prediction of phosphorus and potassium using infrared spectroscopy (Stenberg et al., 2010; Kuang et al., 2012; Soriano-Disla et al., 2014). Moreover, the choice of conventional methods for these soil properties and the variability addressed within the calibration database may significantly impact the performance of prediction using soil spectroscopic approaches. When compared among these two soil properties, soil potassium is predicted in most studies with satisfactory agreement based on the choice of the conventional method and spectral analysis followed. However, soil phosphorus only could be predicted satisfactory ($R^2 > 0.5$) at the most, if not poor. Across the infrared range, vis-NIR or MIR do not seem different in prediction performance for these two soil properties.

2.4.3 Clay, CEC, and pH

Clay, CEC and pH are important soil properties to understand the fertility status and health of the soil. Due to this, they are often used in pedo-transfer functions and models to predict important soil functional attributes for agricultural, environmental, and engineering applications, as these soil properties directly relate to other soil properties, including soil organic carbon (Shepherd and Walsh, 2001). Due to these functional attributes, a high prediction performance using soil spectroscopy can be achieved for these soil properties due to the relationship with soil chromophores. Cation exchange capacity (CEC) and clay content have

been predicted with sufficient accuracy ($R^2 > 0.70$) from soil reflectance spectra (Table 2.1; Terhoeven-Urselmans et al., 2010). This good results for clay and CEC could support the lack of information about these properties as their standard determination method is costly and time-consuming. Meanwhile, soil pH was mostly predicted satisfactorily ($R^2 \sim 0.5-0.7$) or poorly ($R^2 < 0.5$) especially in the vis-NIR range. However, soil pH has also been predicted with good to excellent agreement in other studies (McCarty and Reeves, 2006; Mclaughlin et al., 2014; Soriano-Disla et al., 2014; Breure et al., 2022).

2.5 Crop Fertiliser Recommendation Approaches with Soil Property Inputs

2.5.1 Overview of Approaches to fertilizer recommendations

Soil analysis and the subsequent application of fertilizers play a pivotal role in modern agriculture, with the aim of optimizing crop yield while minimizing the environmental impact. Various methods have been employed to achieve this goal, including conventional soil testing methods and the interpretation of results based on existing soil fertility guidelines. Additionally, soil fertility models such as the Mitscherlich-Bray model (Sonar and Babhulkar, 2002), quadratic models (Correndo et al., 2023) and the QUEFTS model (Janssen and Guiking, 1990), along with precision agriculture techniques (Bah et al., 2016) and Decision Support Systems (Sulaeman et al., 2013), have been utilized. To understand the relationship between crop yield and soil characteristics, it is crucial to assess the nutrient uptake in relation to yield and maintain the balance for the different nutrients in the soil.

Methods such as soil chemical testing, crop response experiments, long-term experiments, and observation of plant symptoms for specific nutrient deficiencies or toxicity have been extensively employed, enabling potential technological advancement and skill developments (Arshad and Martin, 2002; Sattari et al., 2014). In addition to identifying yield-limiting nutrients, determining key soil fertility indicators remains fundamental to science-based nutrient management (Chuan et al., 2013). Defining appropriate fertilizer rates is essential for optimising crop production and application of inputs. Soil testing, coupled with yield targets, forms

the basis for fertilizer recommendations, enhancing both crop yield and nutrient use efficiency. These recommendations are informed by soil property assessments, previous input additions, and crop responses, often integrated into empirical, theoretical or hybrid mathematical models.

2.5.2 Quantification of soil nutrient supply capacity and fertilizer recovery using IR soil spectroscopy

Among the multitude of factors influencing global food production, such as water and land, three pivotal nutrients, namely Nitrogen (N), Phosphorus (P), and Potassium (K), play significant roles. These nutrients can be introduced into the soil system through various forms of organic and inorganic fertilizers, undergoing biogeochemical reactions in the soil after application. Subsequently, a portion becomes available for plant uptake during the growing season, influenced by diverse environmental factors that also affects crop growth. Enhancing the ability to determine fertilizer recommendations that sustain crop productivity while safeguarding the environment requires improved soil testing and crop nutrient studies. These studies provide crucial insights into the availability of plant nutrients in the soil before planting and the quantity of nutrients absorbed throughout the crop production cycle (Leytem et al., 2020). Key components considered in developing soil-crop-fertilizer models include soil supply capacity, which determines the (potential) soil contribution of nutrients for plant uptake, and the fertilizer recovery fraction, representing the proportion of applied nutrients taken up by the plants.

Estimating indigenous nutrient supply capacity can be accomplished through various methods, including measuring plant uptake in aboveground dry matter on omission plots (Dobermann et al., 2003; Haefele et al., 2003) or utilizing empirical relationships based on important soil properties (Janssen and Guiking, 1990; Sattari et al., 2014). While the former provides precise estimates of soil nutrient provision, it is labour-intensive to implement across soil systems and agroecological belts. Emerging technologies such as infrared soil spectroscopy can offer feasible estimates of soil properties used to understand soil supply capacity with reasonable uncertainty. Given that the soil supply capacity varies

spatially, high-resolution data is necessary for fertilizer estimation models reliant on this parameter. Assessing the fraction of applied nutrients taken up by the plant may result in overestimation or underestimation, depending on the accuracy of soil nutrient supply quantification and target yield. The complexity of estimating soil nutrient availability is further compounded by soil-specific sorption characteristics and exchange dynamics between soil pools. These estimations are influenced by many environmental factors and can vary both spatially and temporally (Schut and Reymann, 2023). This strongly suggests that these soil-crop-fertilizer model parameters need to be parameterized or calibrated for a given agroecosystem and production.

2.6 Conclusions and research gaps

Infrared spectroscopy has emerged as an invaluable tool for soil prediction, facilitated by various, sometimes portable, soil scanning instruments that cover diverse wavelength regions. These instruments offer different levels of uncertainty, ease of applicability and varying cost while contributing to the fast generation and utilization of soil information for e.g., agricultural advisory services and soil health monitoring. While the vis-NIR spectroscopy is preferred for in-field applications due to its simplicity, the MIR region is considered as the gold standard in soil spectroscopy. Consequently, vis-NIR spectroscopy has received more attention for advancement, particularly in field conditions and various instrument configurations, and ever more economic instruments are becoming available.

Although comparisons of soil scanning instruments, and the wavelength regions the instruments covered, have primarily focused on a limited set of soil properties within the vis-NIR wavelength ranges, there has been minimal comparison between vis-NIR and MIR instruments. Furthermore, the question remains whether more affordable spectrometers can provide prediction accuracies adequate for meaningful fertilizer recommendations, highlighting the need for further investigation.

For spectral calibration, conventional soil analysis is essential, requiring the analysis of sufficient soil samples to capture the variability in soil properties within

the region of interest. This necessitates optimized development of soil spectral libraries to minimize the reliance on conventional soil analysis while preserving prediction accuracy. And as soil properties vary in space, their spectra interpretation might be scale dependent and prediction performance of a new sample could be dependent on its proximity to the spectral library entry locations, which is an area that has not yet been studied extensively.

Furthermore, as soil spectroscopy enables provision of more detailed soil information, there are opportunities to document soil property interactions and their effects on soil dynamics, influencing decision-making processes. Integrating soil spectroscopy with crop models allows for site-specific or recalibrated models, enhancing precision in fertilizer decision-making. This integration is crucial for maximizing the effectiveness of soil spectroscopy in agricultural applications, but again few examples of this integration have been studied.

2.7 Acknowledgment

Funding: This study was financially supported by OCP Group, as part of the FPO2 “New soil and plant diagnostics tools for better fertilizer recommendations” under collaborative project of Sustainable Agriculture for Africa (SAFA) between Cranfield University, Rothamsted Research and Mohammed VI Polytechnic University.

2.8 References

Abbas, O., Dardenne, P. and Baeten, V. (2012), Near-Infrared, Mid-Infrared, and Raman Spectroscopy, Chemical Analysis of Food: Techniques and Applications, available at :<https://doi.org/10.1016/B978-0-12-384862-8.00003-0>.

Ahmadi, A., Emami, M., Daccache, A. and He, L. (2021), “Soil properties prediction for precision agriculture using visible and near-infrared spectroscopy: A systematic review and meta-analysis”, *Agronomy*, Vol. 11 No. 3, available at: <https://doi.org/10.3390/agronomy11030433>.

Albinet, F., Peng, Y., Eguchi, T., Smolders, E. and Dercon, G. (2022), “Prediction of exchangeable potassium in soil through mid-infrared spectroscopy and deep

learning: From prediction to explainability”, *Artificial Intelligence in Agriculture*, The Authors, Vol. 6, pp. 230–241.

Angelopoulou, T., Balafoutis, A., Zalidis, G. and Bochtis, D. (2020), “From Laboratory to Proximal Sensing Spectroscopy for Soil Organic Carbon Estimation — A Review”, *Sustainability*, Vol. 12 No. 2, pp. 1–24.

Araújo, S.R., Söderström, M., Eriksson, J., Isendahl, C., Stenborg, P. and Demattê, J.A.M. (2015), “Determining soil properties in Amazonian Dark Earths by reflectance spectroscopy”, *Geoderma*, Vol. 237, pp. 308–317.

Arshad, M.A. and Martin, S. (2002), “Identifying critical limits for soil quality indicators in agro-ecosystems”, *Agriculture, Ecosystems and Environment*, Vol. 88 No. 2, pp. 153–160.

Ayoubi, S., Pilehvar, A., Mokhtari, P. and L., K. (2011), “Application of Artificial Neural Network (ANN) to Predict Soil Organic Matter Using Remote Sensing Data in Two Ecosystems”, *Biomass and Remote Sensing of Biomass*, Vol. 10, pp. 181–196.

Bah, A., Balasundram, S.K. and Husni, M.H.A. (2016), “Sensor Technologies for Precision Soil Nutrient Management and Monitoring”, *American Journal of Agricultural and Biological Sciences*, Vol. 7 No. 1, pp. 43–48.

Barra, I., Haefele, S.M., Sakrabani, R. and Kebede, F. (2021), “Soil spectroscopy with the use of chemometrics, machine learning and pre-processing techniques in soil diagnosis: Recent advances—A review”, *TrAC - Trends in Analytical Chemistry*, Elsevier Ltd, Vol. 135 No. 116166, pp. 1–13.

Baumann, P., Helfenstein, A., Gubler, A., Keller, A., Meuli, R.G., Wächter, D., Lee, J., et al. (2017), “Developing the Swiss mid-infrared soil spectral library for local estimation and monitoring”, *European Journal of Soil Science*, Vol. 68 No. 6, pp. 840–852.

Bellinaso, H., Demattê, J.A.M. and Romeiro, S.A. (2010), “Soil spectral library and its use in soil classification”, *Revista Brasileira de Ciência Do Solo*, Vol. 34 No. 3, pp. 861–870.

Bellon-Maurel, V. and McBratney, A. (2011), "Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils - Critical review and research perspectives", *Soil Biology and Biochemistry*, Elsevier Ltd, Vol. 43 No. 7, pp. 1398–1410.

Ben-Dor, E. and Banin, A. (1995), "Near-Infrared Analysis as a Rapid Method to Simultaneously Evaluate Several Soil Properties", *Soil Science Society of America Journal*, Vol. 59 No. 2, pp. 364–372.

Blitz, J.B. (1998), "Diffuse Reflectance Spectroscopy", *Modern Techniques in Applied Molecular Spectroscopy: Techniques in Analytical Chemistry Series*, Vol. 14, pp. 185–219.

Breure, T.S., Milne, A.E., Webster, R., Haefele, S.M., Hannam, J.A., Moreno-Rojas, S. and Corstanje, R. (2021), "Predicting the growth of lettuce from soil infrared reflectance spectra: the potential for crop management", *Precision Agriculture*, Springer US, Vol. 22 No. 1, pp. 226–248.

Breure, T.S., Prout, J.M., Haefele, S.M., Milne, A.E., Hannam, J.A., Moreno-Rojas, S. and Corstanje, R. (2022), "Comparing the effect of different sample conditions and spectral libraries on the prediction accuracy of soil properties from near- and mid-infrared spectra at the field-scale", *Soil and Tillage Research*, Elsevier B.V., Vol. 215 No. July 2021, p. 105196.

Brodský, L., Klement, A., Penížek, V., Kodešová, R. and Borůvka, L. (2011), "Building Soil Spectral Library of the Czech Soils for Quantitative Digital Soil Mapping", *Soil and Water Research*, Vol. 6 No. 4, pp. 165–172.

Brown, D.J., Shepherd, K.D., Walsh, M.G., Dewayne Mays, M. and Reinsch, T.G. (2006), "Global soil characterization with VNIR diffuse reflectance spectroscopy", *Geoderma*, Vol. 132 No. 3–4, pp. 273–290.

Cañasveras, J.C., Barrón, V., del Campillo, M.C. and Viscarra Rossel, R.A. (2012), "Reflectance spectroscopy: a tool for predicting soil properties related to the incidence of Fe chlorosis", *Spanish Journal of Agricultural Research*, Vol. 10 No. 4, pp. 1133–1142.

Carmon, N. and Ben-dor, E. (2017), "An Advanced Analytical Approach for Spectral - Based Modelling of Soil Properties", *International Journal of Emerging Technology and Advanced Engineering*, Vol. 7 No. 3, available at: <https://doi.org/10.1002/9781119043553.CITATIONS>.

Chang, C., Laird, D.A., Mausbach, M.J. and Hurburgh, C.R. (2001), "Near-Infrared Reflectance Spectroscopy–Principal Components Regression Analyses of Soil Properties", *Soil Sci. Soc. Am.*, Vol. 65, pp. 480–490.

Chuan, L., He, P., Pampolino, M.F., Johnston, A.M., Jin, J., Xu, X., Zhao, S., et al. (2013), "Establishing a scientific basis for fertilizer recommendations for wheat in China: Yield response and agronomic efficiency", *Field Crops Research*, Elsevier B.V., Vol. 140, pp. 1–8.

Coblinski, J.A., Giasson, É., Demattê, J.A., Dotto, A.C., Costa, J.J.F. and Vašát, R., 2020. Prediction of soil texture classes through different wavelength regions of reflectance spectroscopy at various soil depths. *Catena*, 189, p.104485.

Correndo, A.A., Pearce, A., Bolster, C.H., Spargo, J.T., Osmond, D. and Ciampitti, I.A. (2023), "The soiltestcorr R package: An accessible framework for reproducible correlation analysis of crop yield and soil test data", *SoftwareX*, Elsevier B.V., Vol. 21, p. 101275.

Dangal, S.R.S., Sanderman, J., Wills, S. and Ramirez-lopez, L. (2019), "Accurate and Precise Prediction of Soil Properties from a Large Mid-Infrared Spectral Library", pp. 1–23.

Demattê, J.A.M., Dotto, A.C., Paiva, A.F.S., Sato, M. V., Dalmolin, R.S.D., de Araújo, M. do S.B., da Silva, E.B., et al. (2019), "The Brazilian Soil Spectral Library (BSSL): A general view, application and challenges", *Geoderma*, Elsevier, Vol. 354 No. October, p. 113793.

Dobermann, A., Witt, C., Abdulrachman, S., Gines, H.C., Nagarajan, R., Son, T.T., Tan, P.S., et al. (2003), "Estimating indigenous nutrient supplies for site-specific nutrient management in irrigated rice", *Agronomy Journal*, Vol. 95 No. 4, pp. 924–935.

Dotto, A.C., Dalmolin, R.S.D., Caten, A. ten, Gris, D.J. and Ruiz, L.F.C. (2019), "AlradSpectra : a Quantification Tool for Soil Properties Using Spectroscopic Data in R", pp. 1–15.

Dotto, A.C., Dalmolin, R.S.D., ten Caten, A., Grunwald, S., Grunwald, S. and Albrecht, A. (2018), "A systematic study on the application of scatter-corrective and spectral-derivative preprocessing for multivariate prediction of soil organic carbon by Vis-NIR spectra", *Geoderma*, Elsevier, Vol. 314 No. May 2017, pp. 262–274.

Dotto, A.C., Demattê, J.A.M., Viscarra Rossel, R.A. and Rizzo, R. (2020), "Soil environment grouping system based on spectral, climate, and terrain data: a quantitative branch of soil series", *Soil*, Vol. 6 No. 1, pp. 163–177.

Gates, J.R. (2018), "A Comparison of VNIR and MIR Spectroscopy for Predicting Various Soil Properties", *Dissertations & Theses in Natural Resources*, No. 264.

Gautam, R., Vanga, S., Ariese, F. and Umapathy, S. (2015), "Review of multidimensional data processing approaches for Raman and infrared spectroscopy", *EPJ Techniques and Instrumentation*, EPJ Techniques and Instrumentation, Vol. 2 No. 8, pp. 1–38.

Gholizadeh, A., Boruvka, L., Saberioon, M.M., Kozák, J., Vašát, R. and Nemecek, K. (2015), "Comparing different data preprocessing methods for monitoring soil heavy metals based on soil spectral features", *Soil and Water Research*, Vol. 10 No. 4, pp. 218–227.

Gholizadeh, A., Carmon, N., Klement, A., Ben-Dor, E. and Boruvka, L.B. (2017), "Agricultural Soil Spectral Response and Properties Assessment: Effects of Measurement Protocol and Data Mining Technique", *Remote Sensing*, Vol. 9 No. 10, p. 1079.

Gholizadeh, A., Luboš, B., Saberioon, M. and Vašát, R. (2013), "Visible, near-infrared, and mid-infrared spectroscopy applications for soil assessment with emphasis on soil organic matter content and quality: State-of-the-art and key issues", *Applied Spectroscopy*, Vol. 67 No. 12, pp. 1349–1362.

Van Groenigen, J.W., Mutters, C.S., Horwath, W.R. and Van Kessel, C. (2003), "NIR and DRIFT-MIR spectrometry of soils for predicting soil and crop parameters in a flooded field", *Plant and Soil*, Vol. 250 No. 1, pp. 155–165.

Guerrero, C., Wetterlind, J., Stenberg, B., Mouazen, A.M., Gabarrón-Galeote, M.A., Ruiz-Sinoga, J.D., Zornoza, R., et al. (2016), "Do we really need large spectral libraries for local scale SOC assessment with NIR spectroscopy?", *Soil and Tillage Research*, Elsevier B.V., Vol. 155, pp. 501–509.

Guillemain, A., Dégardin, K. and Roggo, Y. (2017), "Performance of NIR handheld spectrometers for the detection of counterfeit tablets", *Talanta*, Elsevier, Vol. 165, pp. 632–640.

Haefele, S.M., Wopereis, M.C.S., Ndiaye, M.K., Barro, S.E. and Ould Isselmou, M. (2003), "Internal nutrient efficiencies, fertilizer recovery rates and indigenous nutrient supply of irrigated lowland rice in Sahelian West Africa", *Field Crops Research*, Vol. 80 No. 1, pp. 19–32.

Heil, K. and Schmidhalter, U. (2021), "An evaluation of different nir-spectral pre-treatments to derive the soil parameters c and n of a humus-clay-rich soil", *Sensors*, Vol. 21 No. 4, pp. 1–24.

Helfenstein, A., Baumann, P., Viscarra Rossel, R., Gubler, A., Oechslin, S. and Six, J. (2021), "Quantifying soil carbon in temperate peatlands using a mid-IR soil spectral library", *Soil*, Vol. 7 No. 1, pp. 193–215.

Hutengs, C., Ludwig, B., Eisele, A. and Vohland, M. (2018), "Comparison of Portable and Bench-Top Spectrometers for Mid-Infrared Diffuse Reflectance Measurements of Soils", *Sensors*, Vol. 18 No. 4, p. 998.

Hutengs, C., Seidel, M., Oertel, F., Ludwig, B. and Vohland, M. (2019), "In situ and laboratory soil spectroscopy with portable visible-to-near-infrared and mid-infrared instruments for the assessment of organic carbon in soils", *Geoderma*, Elsevier, Vol. 355 No. July, p. 113900.

Janssen, B.H. and Guiking, F.C.T. (1990), "Modelling the response of crops to fertilizers", In *Plant Nutrition—Physiology and Applications*, pp. 699–703.

Johnson, J.M., Vandamme, E., Senthilkumar, K., Sila, A., Shepherd, K.D. and Saito, K. (2019), "Near-infrared, mid-infrared or combined diffuse reflectance spectroscopy for assessing soil fertility in rice fields in sub-Saharan Africa", *Geoderma*, Vol. 354 No. May, available at: <https://doi.org/10.1016/j.geoderma.2019.06.043>.

Kawamura, K., Nishigaki, T., Tsujimoto, Y., Andriamananjara, A., Rabenaribo, M., Asai, H., Rakotoson, T., et al. (2020), "Exploring relevant wavelength regions for estimating soil total carbon contents of rice fields in Madagascar from Vis-NIR spectra with sequential application of backward interval PLS", *Plant Production Science*, Taylor & Francis, Vol. 24 No. 1, pp. 1–14.

Kennard, R.W. and Stone, L.A. (1969), "Computer Aided Design of Experiments", *Technometrics*, Vol. 11 No. 1, pp. 137–148.

Knox, N.M., Grunwald, S., McDowell, M.L., Bruland, G.L., Myers, D.B. and Harris, W.G. (2015), "Modelling soil carbon fractions with visible near-infrared (VNIR) and mid-infrared (MIR) spectroscopy", *Geoderma*, Elsevier B.V., Vol. 239–240, pp. 229–239.

Kuang, B., Mahmood, H.S., Quraishi, M.Z., Hoogmoed, W.B., Mouazen, A.M. and Henten, E.J. Van. (2012), *Sensing Soil Properties in the Laboratory , In Situ , and On-Line :*, *Advances in Agronomy*, 1st ed., Vol. 114, Elsevier Inc., available at: <https://doi.org/10.1016/B978-0-12-394275-3.00003-1>.

Kuang, B., Tekin, Y. and Mouazen, A.M. (2015), "Comparison between artificial neural network and partial least squares for on-line visible and near infrared spectroscopy measurement of soil organic carbon, pH and clay content", *Soil and Tillage Research*, Elsevier B.V., Vol. 146 No. PB, pp. 243–252.

Lazaar, A., Mouazen, A.M., EL Hammouti, K., Fullen, M., Pradhan, B., Memon, M.S., Andich, K., et al. (2020), "The application of proximal visible and near-infrared spectroscopy to estimate soil organic matter on the Triffa Plain of Morocco", *International Soil and Water Conservation Research*, Elsevier Ltd, Vol. 8 No. 2, pp. 195–204.

- Lee, K.S., Lee, D.H., Sudduth, K.A., S. O. Chung, N. R. Kitchen and S. T. Drummond. (2009), "Wavelength Identification and Diffuse Reflectance Estimation for Surface and Profile Soil Properties", Transactions of the ASABE, Vol. 52 No. 3, pp. 683–695.
- Lemke, M. and Quinlan, D. (2003), "Spectral Imaging for Remote Sensing", Lincoln Laboratory Journal, Vol. 14 No. 1, pp. 121–126.
- Leytem, A.B., Rogers, C.W., Tarkalson, D., Dungan, R.S., Haney, R.L. and Moore, A.D. (2020), "Comparison of nutrient management recommendations and soil health indicators in southern Idaho", Agrosystems, Geosciences and Environment, Vol. 3 No. 1, pp. 1–16.
- Li, S., Rossel, R.A.V. and Webster, R. (2022), "The cost-effectiveness of reflectance spectroscopy for estimating soil organic carbon", European Journal of Soil Science, Vol. 73 No. 1, pp. 1–16.
- Liu, L., Ji, M. and Buchroithner, M. (2018), "Transfer learning for soil spectroscopy based on convolutional neural networks and its application in soil clay content mapping using hyperspectral imagery", Sensors (Switzerland), Vol. 18 No. 9, available at: <https://doi.org/10.3390/s18093169>.
- Lobsey, C.R., Viscarra Rossel, R.A., Roudier, P. and Hedley, C.B. (2017), "Rs-Local Data-Mines Information From Spectral Libraries To Improve Local Calibrations", European Journal of Soil Science, Vol. 68 No. 6, pp. 840–852.
- Lucà, F., Conforti, M., Castrignanò, A., Matteucci, G. and Buttafuoco, G. (2017), "Effect of calibration set size on prediction at local scale of soil carbon by Vis-NIR spectroscopy", Geoderma, Elsevier B.V., Vol. 288, pp. 175–183.
- Luce, S.M., Ziadi, N. and Viscarra Rossel, R.A. (2022), "GLOBAL-LOCAL : A new approach for local predictions of soil organic carbon content using large soil spectral libraries", Geoderma, Elsevier B.V., Vol. 425, p. 116048.
- Maarschalkerweerd, M. Van. (2014), New Ways to Determine Plant Nutrient Deficiencies Using Fast Spectroscopy., University of Copenhagen.

Manley, M. (2014), "Near-infrared spectroscopy and hyperspectral imaging: Non-destructive analysis of biological materials", *Chemical Society Reviews*, Royal Society of Chemistry, Vol. 43 No. 24, pp. 8200–8214.

McCarty, G.W. and Reeves, J.B. (2006), "Comparison of near infrared and mid infrared diffuse reflectance spectroscopy for field-scale measurement of soil fertility parameters", *Soil Science*, Vol. 171 No. 2, pp. 94–102.

Mclaughlin, A.M., Disla, J.M.S., Janik, L., Früzangohar, M., Land, C., Campus, W. and Zerner, M. (2014), "Real time soil tests in the field ? Science fiction or just over the horizon".

Moloney, J.P., Malone, B.P., Karunaratne, S. and Stockmann, U. (2023), "Leveraging large soil spectral libraries for sensor-agnostic field condition predictions of several agronomically important soil properties", *Geoderma*, Elsevier B.V., Vol. 439 No. March, p. 116651.

Morellos, A., Pantazi, X.E., Moshou, D., Alexandridis, T., Whetton, R., Tziotziou, G., Wiebenson, J., et al. (2016), "Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy", *Biosystems Engineering*, Elsevier Ltd, Vol. 152 No. December, pp. 104–116.

Moura-Bueno, J.M., Dalmolin, R.S.D., ten Caten, A., Dotto, A.C. and Demattê, J.A.M. (2019), "Stratification of a local VIS-NIR-SWIR spectral library by homogeneity criteria yields more accurate soil organic carbon predictions", *Geoderma*, Elsevier, Vol. 337 No. November 2017, pp. 565–581.

Næs, T. (1987), "The design of calibration in near infra-red reflectance analysis by clustering", *Journal of Chemometrics*.

Nawar, S. and Mouazen, A.M. (2018), "Optimal sample selection for measurement of soil organic carbon using on-line vis-NIR spectroscopy", *Computers and Electronics in Agriculture*, Elsevier, Vol. 151 No. June, pp. 469–477.

Ng, W., Husnain, Anggria, L., Siregar, A.F., Hartatik, W., Sulaeman, Y., Jones, E., et al. (2020), "Developing a soil spectral library using a low-cost NIR

spectrometer for precision fertilization in Indonesia”, *Geoderma Regional*, Elsevier B.V., Vol. 22, p. e00319.

Ng, W., Minasny, B., Jeon, S.H. and McBratney, A. (2022), “Mid-infrared spectroscopy for accurate measurement of an extensive set of soil properties for assessing soil functions”, *Soil Security*, Elsevier Ltd, Vol. 6, p. 100043.

Ng, W., Minasny, B., Malone, B. and Filippi, P. (2018), “In search of an optimum sampling algorithm for prediction of soil properties from infrared spectra”, *PeerJ*, Vol. 6 No. 5722, pp. 1–27.

Ng, W., Minasny, B., de Sousa Mendes, W. and Melo Demattê, J.A. (2020), “The influence of training sample size on the accuracy of deep learning models for the prediction of soil properties with near-infrared spectroscopy data”, *Soil*, Vol. 6 No. 2, pp. 565–578.

Nocita, M., Stevens, A., Toth, G., Panagos, P., van Wesemael, B. and Montanarella, L. (2014), “Prediction of soil organic carbon content by diffuse reflectance spectroscopy using a local partial least square regression approach”, *Soil Biology and Biochemistry*, Elsevier Ltd, Vol. 68, pp. 337–347.

Nocita, M., Stevens, A., Wesemael, B. Van, Aitkenhead, M., Bachmann, M., Barth, B., Csorba, A., et al. (2015), “Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring”, *Advances in Agronomy*, Vol. 132, pp. 139–159.

O’Rourke, S.M., Minasny, B., Holden, N.M. and McBratney, A.B. (2016), “Synergistic Use of Vis-NIR, MIR, and XRF Spectroscopy for the Determination of Soil Geochemistry”, *Soil Science Society of America Journal*, Vol. 80 No. 4, pp. 888–899.

Ogen, Y., Zaluda, J., Francos, N., Goldshleger, N. and Ben-Dor, E. (2019), “Cluster-based spectral models for a robust assessment of soil properties”, *Geoderma*, Elsevier, Vol. 340 No. August 2018, pp. 175–184.

Padarian, J., Minasny, B. and McBratney, A.B. (2019), “Transfer learning to localise a continental soil vis-NIR calibration model”, *Geoderma*, Elsevier, Vol. 340 No. January, pp. 279–288.

Padarian, J., Minasny, B. and McBratney, A.B. (2020), "Machine learning and soil sciences: A review aided by machine learning tools", *Soil*, Vol. 6 No. 1, pp. 35–52.

Pätzold, S., Leenen, M., Frizen, P., Heggemann, T., Wagner, P. and Rodionov, A. (2019), "Predicting plant available phosphorus using infrared spectroscopy with consideration for future mobile sensing applications in precision farming", No. 0123456789, available at: <https://doi.org/10.1007/s11119-019-09693-3>.

Pügner, T., Knobbe, J. and Grüger, H. (2016), "Near-Infrared Grating Spectrometer for Mobile Phone Applications", *Applied Spectroscopy*, Vol. 70 No. 5, pp. 734–745.

Ramirez-Lopez, L., Behrens, T., Schmidt, K., Rossel, R.A.V., Demattê, J.A.M. and Scholten, T. (2013), "Distance and similarity-search metrics for use with soil vis-NIR spectra", *Geoderma*, Elsevier B.V., Vol. 199, pp. 43–53.

Ramirez-Lopez, L., Behrens, T., Schmidt, K., Stevens, A., Demattê, J.A.M. and Scholten, T. (2013), "The spectrum-based learner: A new local approach for modeling soil vis-NIR spectra of complex datasets", *Geoderma*, Vol. 195–196, pp. 268–279.

Reeves, J.B. (2010), "Near- versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: Where are we and what needs to be done?", *Geoderma*, Elsevier B.V., Vol. 158 No. 1–2, pp. 3–14.

Rinnan, A., Berg, F. van den and Engelsen, S.B. (2009), "Review of the most common pre-processing techniques for near-infrared spectra", *Trends in Analytical Chemistry*, Vol. 28 No. 10, available at: <https://doi.org/10.1016/j.trac.2009.07.007>.

Rodríguez-Pérez, J.R., Marcelo, V., Pereira-obaya, D., Garc, M. and Sanz-Ablanedo, E. (2021), "Estimating Soil Properties and Nutrients by Visible and Infrared Diffuse Reflectance Spectroscopy to Characterize Vineyards", *Agronomy*, Vol. 11 No. 1895.

Sáiz-Abajo, M.J., Mevik, B.H., Segtnan, V.H. and Næs, T. (2005), “Ensemble methods and data augmentation by noise addition applied to the analysis of spectroscopic data”, *Analytica Chimica Acta*, Vol. 533 No. 2, pp. 147–159.

Santos, C., Páscoa, R. and Lopes, J.A. (2013), “A Review on the Applications of Portable Near-Infrared Spectrometers in the Agro-Food Industry”, No. March 2015, available at:<https://doi.org/10.1366/13-07228>.

Sattari, S.Z., van Ittersum, M.K., Bouwman, A.F., Smit, A.L. and Janssen, B.H. (2014), “Crop yield response to soil fertility and N, P, K inputs in different environments: Testing and improving the QUEFTS model”, *Field Crops Research*.

Schut, A.G.T. and Reymann, W. (2023), “Towards a better understanding of soil nutrient dynamics and P and K uptake”, *Plant and Soil*, Springer International Publishing, Vol. 492 No. 1–2, pp. 687–707.

Shao, Y. and He, Y. (2011), “Nitrogen, phosphorus, and potassium prediction in soils, using infrared spectroscopy”, *Soil Research*, Vol. 49 No. 2, pp. 166–172.

Sharififar, A. (2019), “Evaluating a low- - cost portable NIR spectrometer for the prediction of soil organic and total carbon using different calibration models”, *Soil Use and Management*, Vol. 35 No. 4, pp. 607–616.

Shepherd, K.D. (2010), “Soil spectral diagnostics – infrared , x-ray and laser diffraction spectroscopy for rapid soil characterization in the Africa Soil Information Service”, In 19th World Congress of Soil Science, Soil Solutions for a Changing World, No. August, pp. 1–6.

Shepherd, K.D., Ferguson, R., Hoover, D., van Egmond, F., Sanderman, J. and Ge, Y. (2022), “A global soil spectral calibration library and estimation service”, *Soil Security*, Elsevier Ltd, Vol. 7 No. April, p. 100061.

Shepherd, K.D. and Walsh, M. (2001), “Light reflectance provides rapid assessment of soil quality”, *International Centre for Research in Agroforestry*, Vol. 1 No. June, pp. 1–11.

Shepherd, K.D. and Walsh, M.G. (2002), "Development of Reflectance Spectral Libraries for Characterization of Soil Properties", *Soil Science Society of America Journal*, Vol. 66 No. 3, pp. 988–998.

Shi, T., Cui, L., Wang, J., Fei, T., Chen, Y. and Wu, G. (2013), "Comparison of multivariate methods for estimating soil total nitrogen with visible/near-infrared spectroscopy", *Plant and Soil*, Vol. 366 No. 1–2, pp. 363–375.

Sonar, K.R. and Babhulkar, V.P. (2002), "Application of Mitscherlich-Bray equation for fertilizer use in wheat", *Communications in Soil Science and Plant Analysis*, Vol. 33 No. 15–18, pp. 3241–3249.

Song, Y., Shen, Z., Wu, P. and Rossel, R.A.V. (2021), "Wavelet geographically weighted regression for spectroscopic modelling of soil properties", *Scientific Reports*, Nature Publishing Group UK, No. 0123456789, pp. 1–11.

Sorak, D., Herberholz, L., Iwascek, S., Altinpinar, S., Siesler, H.W. and Pfeifer, F. (2012), "New developments and applications of handheld Raman, mid-infrared, and near-infrared spectrometers", *Taylor & Francis*, Vol. 47 No. 2, pp. 83–115.

Soriano-Disla, J.M., Janik, L.J., Allen, D.J. and McLaughlin, M.J. (2017), "Evaluation of the performance of portable visible-infrared instruments for the prediction of soil properties", *Biosystems Engineering*, Vol. 161, pp. 24–36.

Soriano-Disla, J.M., Janik, L.J. and McLaughlin, M.J. (2018), "Assessment of cyanide contamination in soils with a handheld mid-infrared spectrometer", *Talanta*, Elsevier B.V., Vol. 178, pp. 400–409.

Soriano-Disla, J.M., Janik, L.J., Rossel, R.A.V., MacDonald, L.M. and McLaughlin, M.J. (2014), "The Performance of Visible , Near- , and Mid-Infrared Reflectance Spectroscopy for Prediction of Soil Physical , Chemical , and Biological Properties", *Applied Spectroscopy Reviews*, Vol. 49 No. 2, pp. 139–186.

Stenberg, B., Rossel, R.A.V., Mouazen, A.M. and Wetterlind, J. (2010), "Visible and near infrared spectroscopy in soil science", *Advances in Agronomy*, Vol. 107, pp. 163–215.

Stevens, A., Nocita, M., Tóth, G., Montanarella, L. and van Wesemael, B. (2013), "Prediction of Soil Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy", *PLoS ONE*, Vol. 8 No. 6, available at: <https://doi.org/10.1371/journal.pone.0066409>.

Sudduth, K.A. and Hummel, J.W. (1993), "PORTABLE, NEAR-INFRARED SPECTROPHOTOMETER FOR RAPID SOIL ANALYSIS", *Transactions of the ASAE*, Vol. 36 No. 1, pp. 185–193.

Sulaeman, Y., Nursyamsi, D., Widowati, L.R., Muhrizal Sarwani Husnain, Karet, J.K., Loktabat Utara and Banjarbaru., S.B. (2013), "Phosphorus and Potassium Decision Support System: Bridging Soil Database and Fertilizer Application", *Proceedings of the International Workshop on Soil Information System-Oriented Nutrient Management for Major Asian Crops*, pp. 1–15.

Summerauer, L., Baumann, P., Ramirez-Lopez, L., Barthel, M., Bauters, M., Bukombe, B., Reichenbach, M., et al. (2021), "The central African soil spectral library: A new soil infrared repository and a geographical prediction analysis", *Soil*, Vol. 7 No. 2, pp. 693–715.

Sun, Y., Yuan, M., Liu, X., Su, M., Wang, L., Zeng, Y., Zang, H., et al. (2021), "A sample selection method specific to unknown test samples for calibration and validation sets based on spectra similarity", *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, Elsevier B.V., Vol. 258 No. 119870, pp. 1–8.

Tang, G., Huang, Y., Tian, K., Song, X., Yan, H., Hu, J., Xiong, Y., et al. (2014), "A new spectral variable selection pattern using competitive adaptive reweighted sampling combined with successive projections algorithm", *The Analyst*, Royal Society of Chemistry, Vol. 00 No. June, pp. 1–9.

Tang, Y., Jones, E. and Minasny, B. (2019), "Evaluating low-cost portable near infrared sensors for rapid analysis of soils from South Eastern Australia", *Geoderma Regional*, Elsevier B.V., Vol. 20 No. e00240, pp. 1–11.

Terhoeven-Urselmans, T., Vagen, T.-G., Spaargaren, O. and Shepherd, K.D. (2010), "Prediction of Soil Fertility Properties from a Globally Distributed Soil Mid-Infrared Spectral Library", *Soil Science Society of America Journal*, Vol. 74 No. 5, pp. 1792–1799.

Thomas, C.L., Hernandez-Allica, J., Dunham, S.J., McGrath, S.P. and Haefele, S.M. (2021), "A comparison of soil texture measurements using mid-infrared spectroscopy (MIRS) and laser diffraction analysis (LDA) in diverse soils", *Scientific Reports*, Nature Publishing Group UK, Vol. 11 No. 1, pp. 1–12.

Towett, E.K., Shepherd, D. and Drake, B.L. (2016), "Plant elemental composition and portable X-ray fluorescence (pXRF) spectroscopy : quantification under different analytical parameters", *X-Ray Spectrometry*, Vol. 45 No. 2, pp. 117–124.

Towett, E.K., Shepherd, K.D., Sila, A., Aynekulu, E. and Cadisch, G. (2015), "Mid-Infrared and Total X-Ray Fluorescence Spectroscopy Complementarity for Assessment of Soil Properties", *Soil Science Society of America Journal*, Vol. 79 No. 5, pp. 1375–1385.

Vågen, T.-G., Shepherd, K.D., Walsh, M.G., Winowiecki, L., Desta, L.T. and Tondoh, J.E. (2010), "AfSIS Technical Specifications: Soil Health Surveillance", *Africa Soil Information Service*, Vol. 0, available at: <http://africasoils.net>.

Vasques, G.M., Grunwald, S. and Harris, W.G. (2010), "Spectroscopic Models of Soil Organic Carbon in Florida, USA", *Journal of Environmental Quality*, Vol. 39 No. 3, pp. 923–934.

Vestergaard, R.-J., Vasava, H.B., Aspinall, D., Chen, S., Gillespie, A., Adamchuk, V. and Biswas, A. (2021), "Evaluation of Optimized Preprocessing and Modeling", *Sensors*, Vol. 21 No. 20, p. 6745.

Viscarra, R.A. and Bouma, J. (2016), "Soil sensing : A new paradigm for agriculture", *Agricultural Systems*, Elsevier B.V., Vol. 148, pp. 71–74.

Viscarra Rossel, R.A. and Behrens, T. (2010), "Using data mining to model and interpret soil diffuse reflectance spectra", *Geoderma*, Elsevier B.V., Vol. 158 No. 1–2, pp. 46–54.

Viscarra Rossel, R.A., Behrens, T., Ben-dor, E., Brown, D.J., Demattê, J.A.M., Shepherd, K.D., Shi, Z., et al. (2016), "A global spectral library to characterize the world's soil", *Earth-Science Reviews*, Vol. 155, pp. 198–230.

Viscarra Rossel, R.A., Behrens, T., Ben-Dor, E., Chabrilat, S., Demattê, J.A.M., Ge, Y., Gomez, C., et al. (2022), "Diffuse reflectance spectroscopy for estimating soil properties: A technology for the 21st century", *European Journal of Soil Science*, Vol. 73 No. 4, pp. 1–9.

Viscarra Rossel, R.A., Walvoort, D.J.J., Mcbratney, A.B., Janik, L.J. and Skjemstad, J.O. (2006), "Visible , near infrared , mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties", *Geoderma*, Vol. 131 No. 2, pp. 59–75.

Wadoux, A., Malone, B., Minasny, B., Fajardo, M. and McBratney, A. (2021), *Soil Spectral Inference with R: Analysing Digital Soil Spectra Using the R Programming Environment*, Springer International Publishing AG, Cham., Vol. 1908, available at:https://doi.org/10.1007/978-3-030-64896-1_1.

Wetterlind, J. and Stenberg, B. (2010), "Near-infrared spectroscopy for within-field soil characterization : small local calibrations compared with national libraries spiked with local samples", No. December, pp. 823–843.

Wijewardane, N. k., Ge, Y., Wills, S. and Libohova, Z. (2018), "Predicting Physical and Chemical Properties of US Soils with a Mid-Infrared Reflectance Spectral Library", *Soil Science Society of America Journal*, Vol. 82 No. 3, pp. 722–731.

Wijewardane, N.K. (2016), "Using a VNIR Spectral Library to Model Soil Carbon and Total Nitrogen Content", *Biological Systems Engineering--Dissertations, Theses, and Student Research*, Vol. Biological No. University of Nebraska-Lincoln.

Williams, P. and Norris, K. (1987), *Near-Infrared Technology in the Agricultural and Food Industries*, American Association of Cereal Chemists, Inc.

Wolfrum, E.J., Payne, C., Schwartz, A., Jacobs, J. and Kressin, R.W. (2020), “A Performance Comparison of Low-Cost Near-Infrared (NIR) Spectrometers to a Conventional Laboratory Spectrometer for Rapid Biomass Compositional Analysis”, *Bioenergy Research*, BioEnergy Research, Vol. 13, pp. 1121–1129.

Xiao, S., He, Y., Dong, T. and Nie, P. (2018), “Spectral analysis and sensitive waveband determination based on nitrogen detection of different soil types using near infrared sensors”, *Sensors (Switzerland)*, Vol. 18 No. 2, pp. 1–16.

Xie, H.T., Yang, X.M., Drury, C.F., Yang, J.Y. and Zhang, X.D. (2011), “Predicting soil organic carbon and total nitrogen using mid- and near-infrared spectra for Brookston clay loam soil in Southwestern Ontario, Canada”, *Canadian Journal of Soil Science*, Vol. 91 No. 1, pp. 53–63.

Xu, D., Ma, W., Chen, S., Jiang, Q., He, K. and Shi, Z. (2018), “Assessment of important soil properties related to Chinese Soil Taxonomy based on vis–NIR reflectance spectroscopy”, *Computers and Electronics in Agriculture*, Elsevier, Vol. 144 No. November 2017, pp. 1–8.

Yan, H., Xu, Y.C., Siesler, H.W., Han, B.X. and Zhang, G.Z. (2019), “Hand-Held Near-Infrared Spectroscopy for Authentication of Fengdous and Quantitative Analysis of Mulberry Fruits”, *Frontiers in Plant Science*, Vol. 10 No. November, pp. 1–15.

3 Spectral soil analysis for fertilizer recommendations by coupling with QUEFTS for maize in East Africa: a sensitivity analysis

Tadesse Gashaw Asrat^{1*}, Ruben Sakrabani¹, Ron Corstanje¹, Timo Breure⁴, Kirsty L. Hassall², Fassil Kebede³, Stephan M. Haefele²

¹Cranfield University, Cranfield, UK; ²Rothamsted Research, Harpenden, UK;

³College of Agriculture and Environmental Science, AgroBioscience Program, Mohammed VI Polytechnic University, Ben Guerir, Morocco; ⁴European Commission, Joint Research Centre (JRC), Ispra, Italy.

3.1 Abstract

Laboratory analysis of soil properties is prohibitively expensive and difficult to scale across the soils in sub-Saharan Africa. This results in a lack of soil-specific fertilizer recommendations, where recommendation can only be provided at a regional scale. This study aims to assess the feasibility of using spectral soil analysis to provide soil-specific fertilizer recommendations.

Using a range of spectrometers [NeoSpectra Saucer (NIR), FieldSpec 4 (vis-NIR) with contact probe or mug light interface, FTIR Bruker Tensor 27 (MIR)], 346 archived soil samples (0-20 cm) with known soil chemical properties collected from Ethiopia, Kenya and Tanzania were scanned. Partial least square regression (PLSR) was used to develop prediction models for selected soil properties including pH, soil organic carbon (SOC), total nitrogen, Olsen P, and exchangeable K. These predicted properties, and associated uncertainty, were used to derive fertilizer recommendations for maize using the Quantitative Evaluation of the Fertility of Tropical Soils (QUEFTS) model parameters for sub-Saharan Africa.

Most soil properties (pH, SOC, total nitrogen, and exchangeable K) were well predicted (Concordance Correlation Coefficient values between 0.88-0.96 and Ratio of Performance to Interquartile values between 1.4-5.9) by all the spectrometers but there were performance variations between soil properties and

spectrometers. Use of the predicted soil data for the development of fertilizer recommendations gave promising results when compared to the recommendations obtained with the conventional soil analysis. For example, the least performing NeoSpectra Saucer over/under-estimated up to 8 and 24 kg ha⁻¹ N and P, respectively, though there was insignificant variation in estimation of P fertilizer among spectrometers. We conclude that spectral technology can be used to determine major soil properties with satisfactory precision, sufficient for specific fertilizer decision making in East Africa, possibly even with portable equipment in the field.

Keywords: NIR; visNIR; MIR; QUEFTS model; PLSR model; East Africa

3.2 Introduction

Intensified farming with increased productivity is necessary in many regions of the world to match the growing food and energy demand. Among many management factors, mineral fertilizer use is regarded as an essential component in agricultural intensification. Fertilizers play a key role in providing global food supplies and will continue to be needed even with climate change (Wortmann and Sones, 2017; Krasilnikov et al., 2022). To optimize their use, soil tests can help to identify whether it would be productive to grow a particular crop in an area, what fertilizer to apply and how much of it. However, an important constraint in the judicious use of fertilizers in Africa arises from inadequate, often slow and expensive soil testing facilities with limited capacities (Shepherd and Walsh, 2002). Historically, soil laboratories have been under-supported in many developing countries resulting in poor quality diagnosis, which quickly leads to inappropriate fertilizer recommendations, reduced productivity, and higher costs.

Furthermore, conventional soil analysis methods often suffer from a lack of standardization across laboratories, primarily due to differences in protocols, equipment calibration, and analytical techniques. These inconsistencies create challenges when trying to compare or synthesize data from various countries, regions, time periods, and studies. Without uniform methods, results can vary significantly, leading to unreliable comparisons and hindering large-scale data integration. This lack of consistency complicates the development of region-

specific insights and strategies, often resulting in the use of generalized fertilizer recommendations for entire regions. Consequently, this approach perpetuates imbalanced fertilizer use and low nutrient efficiency (Tittonell et al., 2008; Ichami et al., 2019). However, blanket recommendations fall short because soil conditions fluctuate widely across different areas and over time, directly affecting resource use efficiency in agricultural systems (Tittonell et al., 2008). To address this, there is a growing emphasis on developing cost-effective, standardized, repeatable soil analysis methods in contemporary soil science (Grunwald et al., 2011), particularly to create harmonized spatial soil datasets that support precise, localized soil management.

The trade-off between the growing need for large scale soil information and its high cost could be resolved by a widespread use of reflectance soil spectroscopy (RSS) in the visible-near infrared (vis-NIR) and the mid infrared (MIR) regions. The diffusely reflected/absorbed spectral signatures have been proven to reliably predict important soil properties (Viscarra Rossel et al., 2016; Lausch et al., 2013; Bah et al., 2016; Shepherd and Walsh, 2007). They were also applied in digital soil mapping (Paul et al., 2019; Morais et al., 2018; Hengl et al., 2021; Brodský et al., 2011), evaluated for crop productivity and performance estimation (Breure et al., 2021), and for fertilizer decision making using a soil fertility categorical approach (Ng et al., 2020). Therefore, these tools could be useful to improve fertilizer decision making in low-income countries for higher agricultural productivity through informed decisions which can be continuously improved and updated.

One of the advancements in soil spectroscopy is the development of diverse spectrometers which can be suitable for various applications with a range of costs, precision, and flexibility while in use. The soil scanning spectrometers employed in soil science are mostly operating in the vis-NIR (350-2500 nm), NIR (1350 – 2500nm) or MIR (2500-16700nm) spectrum with diverse equipment setups such as bench-top, portable or handheld spectrometers. These spectrometers further differ in their technical specifications which encompasses the type of light source and detector, measurement principles, energy intensity

and resolution (Mouazen et al., 2007; Dos Santos et al., 2013; Rodionov et al., 2016; Xu et al., 2016; Fajardo et al., 2017; Ng et al., 2019). These differences may result in varied precision and consistency across soil properties, soil types and scanning conditions. The spectrometers also differ in weight, size, and sample preparation protocols (Tang et al., 2019) which may affect the speed of analysis, sample processing duration, cost of soil analysis and ability to perform measurements in the field. Particularly, the MIR spectroscopy provides clear and distinct absorbance peaks, however, the technology is more complex, demands better facilities and training and often liquid nitrogen to cool down the light source which may not be accessible easily in developing countries. It is also more costly compared with vis-NIR and NIR spectroscopy (Li et al., 2022; Reeves, 2010; Viscarra Rossel et al., 2006). In contrast, vis-NIR and NIR spectroscopy are robust methods that do not require much soil preparation (Sharififar, 2019) and can be integrated into simpler scanning instrumentations for reduction in cost and field applicability.

A number of studies have compared and evaluated various soil scanning instrumentations and spectral ranges for their precision, consistency, cost and complexity in obtaining spectral recordings (Van Groenigen et al., 2003; Mouazen et al., 2005; Bellon-Maurel & McBratney, 2011; Ge et al., 2011; Xie et al., 2011; Soriano-Disla et al., 2014; Piikki et al., 2016; Gates, 2018; Hutengs et al., 2018; Pätzold et al., 2020; Sharififar, 2019; Tang et al., 2019). However, most comparisons of soil scanning instrumentations were carried out on a small number of soil properties, particularly soil organic carbon (SOC) and clay content (Knadel et al., 2013); soil moisture content and some chemical properties of manures (Mouazen et al., 2005); and SOC and total carbon estimation (Sharififar, 2019). Further comparisons on the predictive accuracy and instrument-specific trade-offs are needed to understand relationships among the scanning outcomes and prediction performance for a specific soil property (Knadel et al., 2013; Cécillon et al., 2009; Breure et al., 2022; Clingensmith et al., 2019).

Furthermore, the soil information generated should be interpreted in relation to its use, in our case fertilizer recommendations. Among other approaches to

interpret the soil information, mechanistic or empirical soil-crop-fertilizer models such as the QUEFTS model (**QU**antitative **E**valuation of the **F**ertility of **T**ropical **S**oils) and soil test-based (STB) fertilizer recommendation models have been used to determine the amount of nutrients to be applied (Janssen et al., 1990; Tiftonnell et al., 2008; Sattari et al., 2014).

But it remains unclear if cheaper spectrometers can provide sufficient prediction accuracy to provide meaningful fertilizer recommendations (Tang et al., 2019; Sharififar, 2019). Hence, this study aims to i) compare the performance of desktop and portable soil scanning infrared instruments with conventional reference soil analysis techniques and in predicting multiple soil properties and ii) to use these predictions to derive and evaluate fertilizer recommendations for maize grown in selected soils of East Africa.

3.3 Materials and methods

3.3.1 Soil sample sources

The soil samples used in this study were archived at the Africa Soil Information Service (AfSIS) spectral laboratory at Rothamsted Research, UK. These samples were collected from maize experimental plots during various seasons between 2017 and 2019 across Ethiopia, Kenya and Tanzania, as part of several research projects. Specifically, they were sourced from the Geo-Nutrition project (BB/P023126/1) in Ethiopia, the iCGRAF project (also BB/P023126/1) in Kenya and Tanzania, and a second Geo-Nutrition project funded by the Bill and Melinda Gates Foundation (INV-009129). A total of 346 topsoil (i.e., 0-20 cm depth) samples were used for this study, representing a wide-ranging agroecology and many soil types. Geographically, there were 15, 259 and 76 soil samples from Ethiopia, Kenya and Tanzania, respectively. Samples were air dried, crushed with a pestle and mortar and sieved to pass through a 2 mm stainless steel sieve. Then, sub-samples were further fine milled ($< 50 \mu\text{m}$) and stored in glass vials with a screw cap for scanning.

3.3.2 Soil spectral data collection with a brief description of spectrometers

The soil samples were scanned using four different commercially available spectrometers (Figure 3.1) whose characteristics are summarized in Table 3.1. Technical descriptions of each infrared spectrometers are briefly presented below including the labelling used in brackets for each spectrometer in the graphs and discussions afterwards.

- *FTIR Bruker Tensor 27 bench-top spectrometer (Tensor-II) (Figure 3.1 a) was used to collect the spectral data in the MIR region (FTIR Bruker Tensor 27; Bruker Optik GmbH, Ettlingen, Germany) using a nitrogen purged integrating sphere to cool the MCT (mercury cadmium telluride) mid-band detector. The high throughput screening accessory (HTS-XT), which scans 95 samples in one plate, was used with a spectral resolution of 4 cm⁻¹ and scan time of 32 s per sample. Absorbance data in the spectral range 2500–16,666 nm were obtained in two replicates.*
- *The portable ASD Contact Probe (FS4-CP) (Figure 3.1 b) together with the FieldSpec 4 (2019 Malvern Panalytical) spectrometer, is used to measure samples through contact which minimizes measurement errors associated with stray light. The light source for this spectrometer is a Halogen bulb/1500 hour with a colour temperature of 2900 k that requires 12-18 VDC and 6.5 W. It had a 3 nm vis, 8 nm short-wave infrared (SWIR) spectral resolution with wavelength accuracy of 0.5 nm. Three repeated scans were taken in the same sample holder/Petri dish by moving the 12 mm diameter visNIR sensor each time.*
- *The portable ASD Mug light (FS4-ML) (Figure 3.1 c) together with the FieldSpec 4 (2019 Malvern Panalytical) spectrometer, is employed to acquire reflectance and absorbance measurements of samples using the ASD sampling tray adapter. This scanning interface has the advantage of minimizing measurement errors related to stray light and specular reflected components. The light source for this scanning interface is a Tungsten Quartz Halogen b/1500hrs with a bulb colour temperature of 2900 k. It had a 3 nm vis, 8 nm short-wave infrared (SWIR) spectral*

resolution with wavelength accuracy of 0.5 nm. Three replicate measurements of each soil sample were acquired with independent subsamples placed on the ASD sampling tray adapter.

- *NeoSpectra Saucer (Neospec) spectrometer (Figure 3.1 d) is a hand-held Near InfraRed (NIR) spectral sensor manufactured by Si-ware containing a single-chip Michelson interferometer with monolithic opto-electro-mechanical (OEM) structure (Si-ware, 2017). It has a wavelength resolution of 16 nm and a wavelength accuracy of ± 1.5 nm. The sample scanning was performed with the settings set to boxcar, 3N zero padding and a 10 second integration resulting in 1024 data points per spectrum using linear interpolation. Three replicate measurements per sample were recorded by moving the scanning spot over the soil sample in a petri dish.*

Table 3.1 : Characteristics of the spectrometers compared [FTIR Bruker Tensor 27 (Tensor-II), FieldSpec 4 mug light (FS4-ML) and contact probe (FS4-CP), and NeoSpectra Saucer (Neospec)].

Spectrometer	Spectral region (nm)		Spot size		Unit price	Weight	Power source	Spectra data points
			(mm)		(£) *	(kg)		
Tensor-II	MIR/NIR	1340 - 16669	6		71,163.00	>30.0	DC	2376
FS4-CP	Vis-NIR	350 - 2500	10		75,140.00	6.1	AC	2151
FS4-ML	Vis-NIR	350 - 2500	12		72,950.00	6.7	AC	2151
Neospec	NIR	1339 - 2500	10		3,850.00	1.0	AC	1023

*Prices were collected in July 2021.

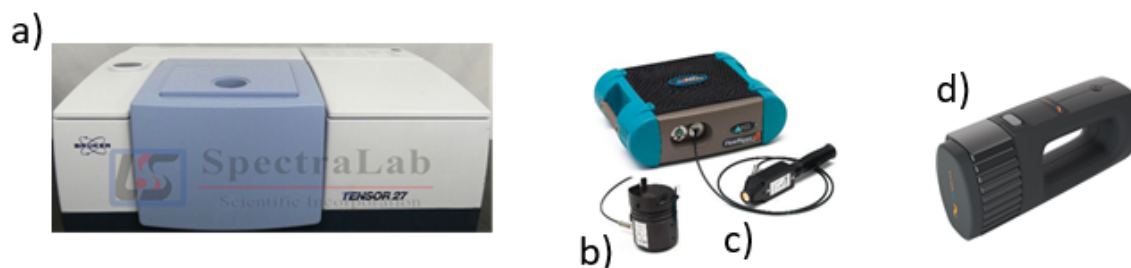


Figure 3.1 : Spectrometers compared in this study: a) FTIR Bruker Tensor 27 (Tensor-II), b) FieldSpec 4 mug light (FS4-ML) and contact probe (FS4-CP); and c) NeoSpectra Saucer (Neospec).

3.3.3 Conventional soil analysis

The dry combustion method was used to determine total carbon (Nelson and Sommers, 1996) and total nitrogen (Bremner, 1996) with a Leco dry combustion analyzer (Stockport, UK). Inorganic carbon was analyzed using wet acidification with a Skalar Primacs AIC 100 (Skalar Analytical BV, Breda, Netherlands). Soil total organic carbon (SOC) was then calculated by subtracting inorganic carbon from total soil carbon. The Cobalthexamine extract (ISO 23470) of exchangeable K was analyzed with inductively coupled plasma optical emission spectrometry (ICP-OES) (Ciesielski and Sterckeman, 1997). Soil pH was measured in a 1:2.5 soil: water suspension (ISO 10390: 2005) one hour after mixing, using a thin semi-micro sealed combined pH electrode from Fisher scientific (Loughborough, UK). Olsen P was measured in the sodium bicarbonate extract (Olsen, 1954) using the phospho-molybdenum blue method on a continuous colorimetric flow analyzer.

3.3.4 Soil property prediction from spectra

3.3.4.1 Spectral pre-processing

The MIR scans of the Tensor II were converted to wavelengths (nm). The reflectance (R) measurements recoded by all spectrometers were transformed to the logarithmic apparent absorbance using $A = \log(1/R)$. Splice correction (de-stepping) on FS4-ML and FS4-CP at 1000 and 1800 nm was performed using the “spliceCorrection” function in the “prospectr” package (Stevens and Lopez, 2014) in R (R Core Team, 2017). The spectral records of FS4-ML and FS4-CP

outside the 500–2450 nm range and of Neospec beyond 2450 nm were removed because of the low signal-to-noise ratio often caused by light scattering effects of quartz sand or by instrument drift. This was followed by performing a combination of spectral pre-processing techniques that improved absorption features by reducing the noise (the Savitsky-Golay Smoothing with spectral first derivative) (Vestergaard et al., 2021). The smoothing gap sizes were a window of 11 nm for vis-NIR and MIR and a window of 5 nm for NIR. Thereafter, the spectral data was cleaned from water absorption regions (H₂O band 1 between 1350 and 1460 nm; H₂O band 2 between 1790 and 1960 nm) in the vis-NIR, and CO₂ peaks in the MIR absorption regions (between 4274 and 4464 nm).

3.3.4.2 Principal component analysis

A principal component (PC) analysis was performed to compute using the “resemble” R package (Ramirez-lopez et al., 2022) to generate the PC scores of the respective infrared regions recorded from each spectrometer. The PCA was conducted by centering the spectral data and projecting it onto a new set of orthogonal components, known as the principal components. This transformation reduces the dimensionality of the spectral dataset by converting the original variables into a smaller set of uncorrelated variables that capture the most variance in the data. In this study, we retained 99% of the total variance, ensuring that most of the spectral information was preserved while minimizing less significant variations. The primary goal was to use these principal components to split the dataset into representative calibration and validation sets for developing and evaluating the predictive PLSR model for soil properties. By retaining the most informative components, the dataset becomes more manageable and efficient for subsequent predictive modelling in the spectral space.

3.3.4.3 Dividing samples into train and test sets

The dataset was split into a calibration (training) set, comprising 75% of the data, and a validation set, comprising of the remaining 25%. This division was achieved using the conditioned Latin Hypercube sampling (cLHS) based on the PC scores. After excluding the outliers (as illustrated in Figure **B-1**), the final dataset included 260 samples for the training and 86 samples for the validation. The distribution of

these samples is depicted in Figure B-2, which shows the spread of the training and validation sets in the score plot of the first two PCs for each spectrometer.

3.3.4.4 Model Calibration and performance evaluation

Partial least squares regression (PLSR) was used to develop predictive models of each soil property from the calibration spectra using the “pls” package (Wehrens, 2007). The number of components to retain were chosen based on the minimum root-mean squared error estimated from a ten-fold cross validation. Models for each spectrometer were then evaluated on the independent test set. The performance of predictive models was evaluated using concordance correlation coefficients (CCC), the Ratio of Performance to Inter Quartile distance (RPIQ), the Root Mean Square Error (RMSE) and the Bias, quantified using the independent test set. CCC measures both the accuracy and precision (Lin, 1989) with a range of values between -1 and +1, where a value of +1 denotes a perfect agreement.

3.3.5 Derivation of soil test-based fertilizer estimation

3.3.5.1 Estimation of indigenous soil nutrient supply (NPK)

Based on the Quantitative Evaluation of the Fertility of Tropical Soils (QUEFTS) model, the indigenous soil supply of N, P and K were derived from the empirical relationship of values of SOC, Olsen P and exchangeable K with pH and SOC according to equation ((3-1)(3-2)(3-3), respectively (Janssen et al., 1990; Tittonell et al., 2008; Sattari et al., 2014). We utilized potassium measured with the Cohex chloride K as a substitute for ammonium acetate extraction (NH₄-Ac) (for which QUEFTS parameters were calibrated for K). Ciesielski and Sterckeman (1997) found that the slope between the two methods was 0.978 when comparing them using percolation techniques. However, they noted that NH₄-Ac method generally extracts more potassium than the Cohex chloride method.

$$SN_j = fN_j \times 6.8 \times SOC_j \quad (3-1)$$

$$SP_j = fP_j \times 0.35 \times SOC_j + 0.5 \times Olsen P_j \quad (3-2)$$

$$SK_j = \frac{fK_j \times 500 \times \text{Exchangeable } K_j}{2 + (0.9 \times \text{SOC}_j)} \quad (3-3)$$

where SN, SP and SK are the indigenous soil supply of N, P and K, respectively, from soil mineralization and expressed in kg ha⁻¹; SOC is organic carbon content of the soil expressed in g kg⁻¹ of soil; P and Exchangeable K are expressed in mg kg⁻¹ and cmol kg⁻¹ soil, respectively; and j is the soil testing (conventional as well as spectral soil analysis) methods. Soil pH correction factors were used according to equations (3-4), (3-5) and (3-6) because pH considerably influences N mineralization, P dissolution and K exchangeability from soils (Sattari et al., 2014).

$$fN_j = f(x) = \begin{cases} 0.4, & pH_j < 4.7 \\ 0.25 \times (pH_j - 3), & 4.7 < pH_j < 7 \\ 1, & pH_j > 7 \end{cases} \quad (3-4)$$

$$fP_j = f(x) = \begin{cases} 0.02, & pH_j < 4.7 \\ 1 - 0.5 \times (pH_j - 6)^2, & 4.7 < pH_j < 6 \\ 1, & 6 < pH_j < 6.7 \\ 1 - 0.25 \times (pH_j - 6.7)^2, & 6.7 < pH_j < 8 \\ 0.57, & pH_j > 8 \end{cases} \quad (3-5)$$

$$fK_j = f(x) = \begin{cases} 1, & pH_j < 4.5 \\ 6.1 \times (pH_j)^{-1.2}, & 4.5 < pH_j < 6.8 \\ 0.6, & pH_j > 6.8 \end{cases} \quad (3-6)$$

where fN, fP, and fK are pH-H₂O correction factors for N, P and K, respectively, and j is the soil testing method.

3.3.5.2 Estimation of nutrient requirement for maize

We estimated fertilizer recommendations based on Rurinda et al. (2020), equation (3-7) for the Sub-Saharan region in this study. In addition, some key assumptions were considered for estimating the nutrient requirements of maize. These include, firstly, the maize target yield is 7 t ha⁻¹, which is a median of the yield range from 6 - 13 t ha⁻¹ retrieved from the Global Yield Gap Atlas protocol (i.e., <https://www.yieldgap.org/>) for the East Africa region, assuming that the

attainable yield is 70% of the water limited yield potential. Secondly, we assumed that the optimum requirement of each nutrient in kg per tonne of maize grain is between the inverse of the maximum accumulation and dilution of the crop specific internal efficiencies (Smaling and Janssen, 1993). Hence, we took the average of the inverse internal efficiencies reported by Rurinda et al. (2020) for East Africa for each nutrient (for N ~ 16.74; for P ~ 3.02; and for K ~ 15.44). Thirdly, the fertilizer recovery efficiency used in this study for N and K was 50% as reported by Sattari et al. (2014), and 30% for P as reported by Rurinda et al. (2020). Thus, the NPK requirements are calculated according to:

$$NC_{ij} = \frac{(NR_i \times Y_{at}) - (S_{ij})}{RE_i} \quad (3-7)$$

where NC is the crop nutrient input requirement in kg ha⁻¹; NR is the N, P or K kg requirement per tonne of maize grain; Y_{at} is the attainable maize grain yield target in t ha⁻¹; S is the indigenous nutrient supply of N, P or K derived from equation (3-2), (3-3) and (3-4); and RE is the recovery efficiency for applied nutrients (%; ratio of crop nutrient uptake to nutrients applied) ; i is N, P or K; j is the soil testing method.

3.3.6 Bootstrapping and error propagation

Our dataset was divided into calibration and validation sets using PC scores derived from a cLHS approach as outlined in section 3.3.4.3. This method ensures a representative distribution of the spectral data across the entire data space by stratifying samples based on their PC scores, balancing the need for robust training set while maintaining an adequate validation samples. The systematic selection of the calibration set based on PC scores and cLHS effectively controlled the design and minimized sampling bias (Wehrens et al., 2002).

To estimate prediction uncertainty, we used a non-parametric bootstrapping procedure with 10,000 iterations. This method involved resampling residuals from the PLSR model to derive leverage and residual for the calibration set, as outlined in equations (B.1 and (B.2 (Zhang, 2014). Bootstrapping was used to generate

a distribution of model performance metrics, allowing us to assess variability and uncertainty in the predictions without assuming a specific residual distribution.

Following bootstrapping, we estimated confidence intervals using a bias-corrected method. First, the bias was calculated as the proportion of individual bootstrap estimates exceeding the overall bootstrap mean (p_b). The proportional bias was then transformed into standard normal deviates using the inverse (quantile) normal distribution function ($b = \Phi^{-1}(p_b)$). To adjust the 95% confidence intervals, twice the calculated bias was subtracted from the standard normal limits (z), and adjusted limits were converted to proportions using of the normal probability distribution function. This approach provided a more accurate representation of model uncertainty by accounting for systematic biases in the bootstrap estimates. The percentiles were thus given by equation (3-8) and (3-9):

$$\alpha_1 = \Phi \left(b + \frac{b + z^{(\alpha/2)}}{1 - b + z^{(\alpha/2)}} \right) \quad (3-8)$$

$$\alpha_2 = \Phi \left(b + \frac{b + z^{(1-\alpha/2)}}{1 - b + z^{(1-\alpha/2)}} \right) \quad (3-9)$$

where, $z^{(\alpha/2)}$ was the $100 \times \left(\frac{\alpha}{2}\right)^{th}$ percentile point of the normal probability distribution function (e.g., $z^{(.05/2)} = -1.96$); Φ represents the cumulative distribution function (CDF) of the standard normal distribution.

The standard deviation of each individual prediction was then computed as the difference between the confidence limits (CI) divided by twice the 95% normal deviate: $\sigma = \frac{CI_u - CI_l}{2z}$, where: $2z = 3.92$.

The standard deviations associated with predictions for the validation set were used to propagate marginal prediction errors in the estimation of indigenous soil N and K supply (Equations (B.3, (B.4, and (B.5). Finally, uncertainty in the indigenous soil N and K supply was propagated into the nutrient requirement for a given yield target (equation (3-7) from the soil nutrient supply of N and K as follows:

$$\sigma_{N_{ij}} = \sqrt{\left(\frac{\sigma_{S_{ij}}^2}{S_{ij}^2}\right)} NC_{ij} \quad (3-10)$$

Where $\sigma_{S_{ij}}^2$ represents the standard deviation of the crop nutrient requirement estimate for nutrient i and soil scanning instrument j ; S_{ij}^2 is the estimated soil nutrient supply which is the central value around which the $\sigma_{S_{ij}}^2$ is calculated.

Given that Olsen P was log-transformed for the PLS regression, we propagated the $100(1-\alpha/2)\%$ confidence interval (CI) limits through in the equations for the soil P supply (Equation (3-2), the correction factor (Equation (3-5), and P requirement (Equation (3-7). Olsen P was exponentiated during the confidence interval calculation to revert the log transformation. The uncertainty in SOC predictions was not included in the error propagation for the nutrient requirement, but Olsen P uncertainty, as part of the soil P supply (SP_j), was propagated through to the nutrient requirement (NC_{Pj}).

3.4 Results

3.4.1 Soil spectral characteristics and preprocessing across soil origins and instruments

Figure 3.2 and Figure 3.3 illustrate the raw absorbance for individual soil sample and the mean absorbance of the preprocessed data using the Savitzky-Golay and first derivative (SG + 1D), categorized by origin of soil sample, respectively. The MIR spectral (Tensor-II) exhibited multiple, narrow absorbance peaks, which are characteristics of the region. Notably, the spectral curves varied among the source of the soil samples, likely due to distinct variation in soil types between the countries. Specifically, the spectral patterns of soils from Ethiopia and Tanzania were similar across all instruments, suggesting that soil type variation is more a significant source of differences than the scanning instrumentation itself, although some variation among instruments may still exist. Both the visNIR and NIR instruments produced similar spectral curves from 1000 nm onwards, displaying broad absorption features centred around approximately 1400, 1900 and 2200 nm. However, the spectra obtained using fieldspec4 scanner with

contact probe differed from those acquired with the mug light in the visible range, extending from 500 - 700 nm.

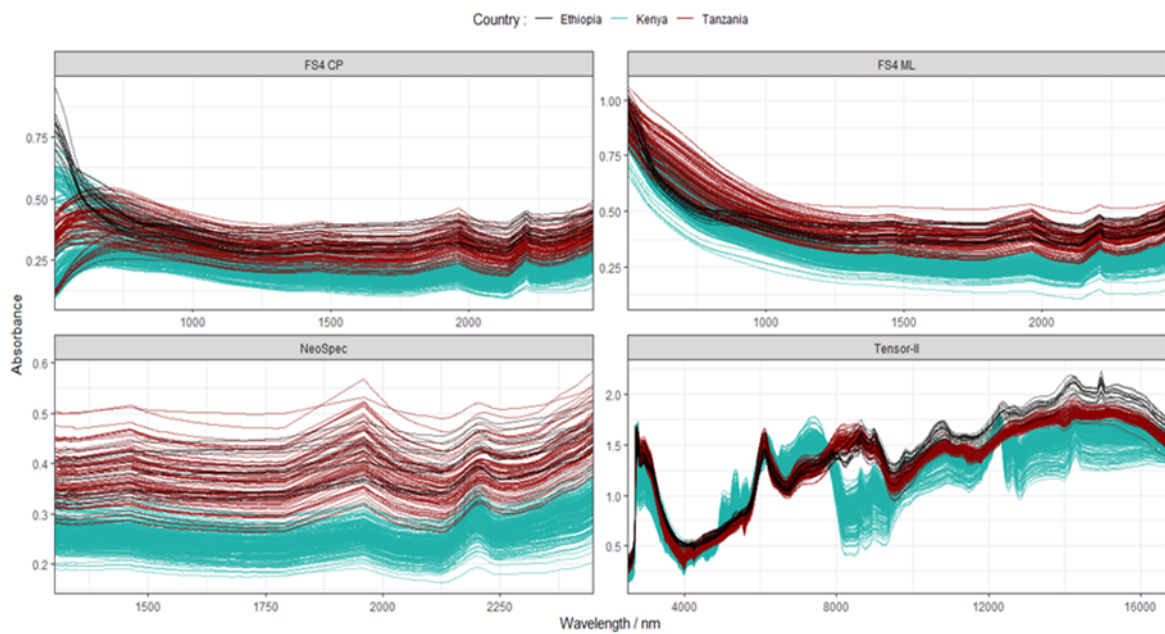


Figure 3.2: Raw absorbance of the soil spectrometers (FS4-CP -> FieldSpec 4 contact probe, FS4-ML -> FieldSpec 4 mug light, Neospec -> NeoSpectra Saucer, Tensor-II -> FTIR Bruker Tensor 27) coloured by soil sample origin.

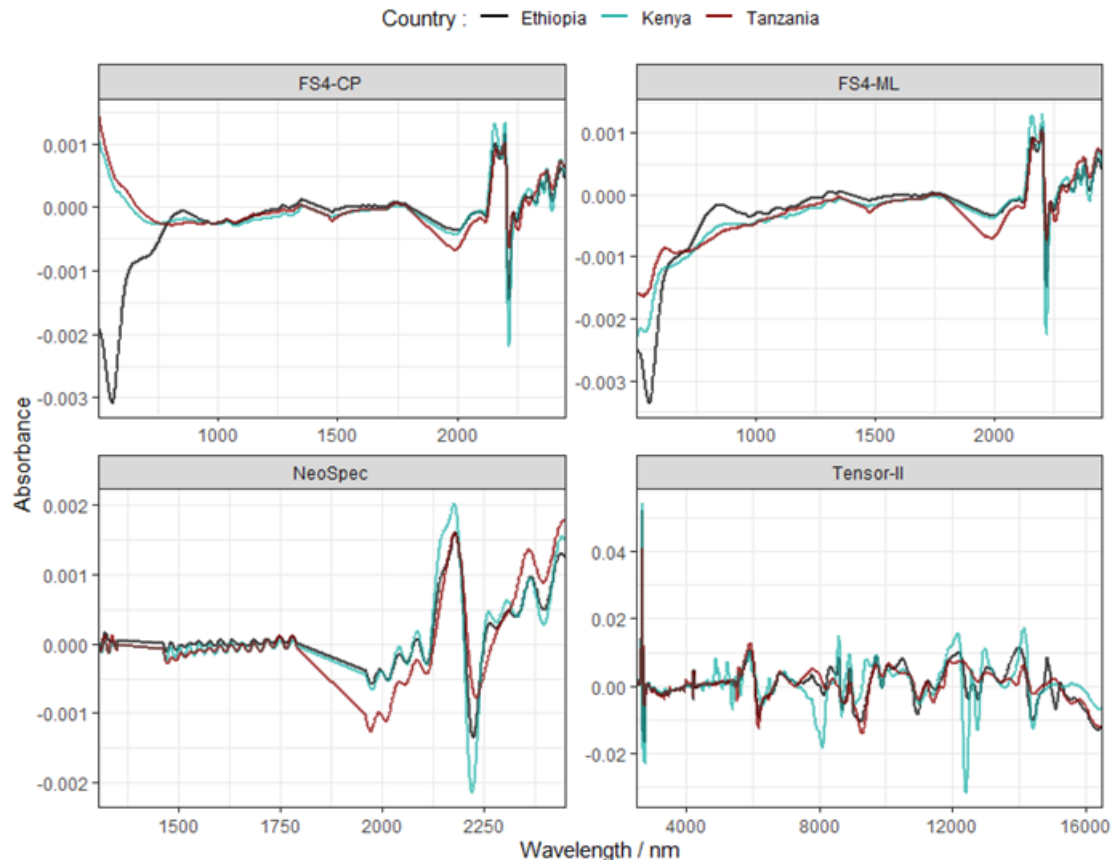


Figure 3.3: Mean absorbance after Savitzky-Golay smoothing and first derivative preprocessing (SG + 1D) of spectra collected by soil spectrometers (FS4-CP -> FieldSpec 4 contact probe, FS4-ML -> FieldSpec 4 mug light, NeoSpec -> NeoSpectra Saucer, Tensor-II -> FTIR Bruker Tensor 27).

3.4.2 Description of the soil properties

Table 3.2 presents descriptive statistics of the soil properties obtained by conventional analysis separately for the calibration and independent test sets. Soil pH ranged from 4.37 to 8.89 well representing the diverse pH values in agricultural soils of the region. The average, as measured by the mean (for most properties) or median (for the skewed Olsen P), of the soil attributes for all properties, were comparable in the calibration and independent test sets. Olsen P data was highly skewed, and a log transformation was carried out before calibrating the predictive models from the spectra based on an investigation of the PLSR residuals for model calibration and back-transformed for the prediction exercise. Table B.1 further separates these statistics by country.

Table 3.2: Descriptive statistics of soil chemical properties for model training and independent test set.

Soil property	Sample set	N	min	q1	median	mean	q3	max
pH	Cal	260	4.48	5.09	5.50	5.99	6.42	8.89
	val	86	4.37	5.06	5.48	5.96	6.37	8.76
SOC	Cal	260	0.50	1.12	1.42	1.42	1.67	2.45
	val	86	0.57	1.15	1.40	1.43	1.70	2.51
TotalN	Cal	260	0.04	0.09	0.11	0.11	0.13	0.20
	val	86	0.04	0.09	0.11	0.11	0.13	0.20
Exch.K	Cal	260	1.02	2.28	3.81	6.69	7.81	45.83
	val	86	0.57	2.62	3.55	6.35	8.54	36.87
OlsenP	Cal	260	0.42	3.53	6.41	10.17	11.30	89.96
	val	86	0.94	3.98	6.40	8.99	10.87	35.62

3.4.3 Soil property prediction accuracy and consistency for QUEFTS inputs

Table 3.3 shows the prediction performance and associated errors for the validation set across all spectrometers, while Figure 3.4 illustrates the distribution of predicted versus observed values in scatter plot. As indicated by the table and figure, all soil properties except Olsen P were adequately predicted from spectra obtained from all spectrometers, as indicated by CCC values ranging from 0.77 to 0.96 and RPIQ values between 1.7 to 4.7. In particular, soil pH was predicted with higher precision and consistency, when compared to other soil properties considered, by the vis-NIR, NIR and the MIR spectrometers. Furthermore, SOC, exchangeable K and total N were slightly better predicted by the vis-NIR and MIR spectrometer than the NIR spectrometer. Predictions from NIR spectrometer resulted in the lowest RPIQ values for SOC, exchangeable K and total N (2.1, 1.7 and 2.2, respectively) though the values are still in the range of model stability and reproducibility. The CCC and RPIQ values for vis-NIR and MIR spectrometers are in the category of good model performance and consistency whereas the respective values for NIR would be categorized as moderate according to Viscarra Rossel et al. (2016). This could have arisen from the

missing absorbance ranges between 500 - 1350 which are clearly informative as demonstrated in the loadings plots shown in Figure B-3, Figure B-4, Figure B-5, Figure B-6 and Figure B-7. Across all four spectrometers, the spectra generated from the MIR spectrometer excelled in predicting SOC and total N, showing higher precision and consistency than any of the other spectrometers.

Across soil properties, Olsen P had the poorest agreement with larger errors than all other soil properties. This indicated that Olsen P prediction using the spectroscopic analysis was more uncertain than for the other soil properties considered. As can be seen from the scatter plot (Figure 3.4), the deviation between the predicted and measured Olsen P increased above 15 mg P kg⁻¹ irrespective of the spectrometer. Overall, there was a tendency to underestimate higher values of Olsen P (>15 mg P kg⁻¹) and exchangeable K (>25 cmol kg⁻¹ soil) by all soil spectrometers.

Table 3.3: Predictive metrics by soil property and soil sensor. The statistics shown are the root mean square error (RMSE), the concordance correlation coefficient (CCC), Bias and the ratio of performance to interquartile range (RPIQ).

Soil property	Sensor	CCC	RMSE	Bias	RPIQ
K[exch]	FS4 CP	0.91	3.20	0.41	2.10
K[exch]	FS4 ML	0.93	2.80	0.20	2.30
K[exch]	NeoSpec	0.82	4.20	0.65	1.50
K[exch]	Tensor-II	0.94	2.60	0.24	2.50
N[tot]	FS4 CP	0.88	0.01	0.00	2.90
N[tot]	FS4 ML	0.86	0.02	0.00	2.70
N[tot]	NeoSpec	0.74	0.02	0.00	2.10
N[tot]	Tensor-II	0.95	0.01	0.00	4.30
P[exch]	FS4 CP	0.42	11.00	2.80	0.86
P[exch]	FS4 ML	0.44	11.00	2.20	0.86
P[exch]	NeoSpec	0.11	13.00	3.80	0.72
P[exch]	Tensor-II	0.77	9.40	0.99	1.00
pH	FS4 CP	0.97	0.30	-0.04	4.90
pH	FS4 ML	0.97	0.29	-0.04	5.00
pH	NeoSpec	0.95	0.41	0.01	3.60
pH	Tensor-II	0.99	0.21	-0.03	7.20
SOC	FS4 CP	0.88	0.19	0.01	2.80
SOC	FS4 ML	0.87	0.21	0.03	2.60
SOC	NeoSpec	0.77	0.26	0.05	2.10
SOC	Tensor-II	0.95	0.13	0.02	4.20

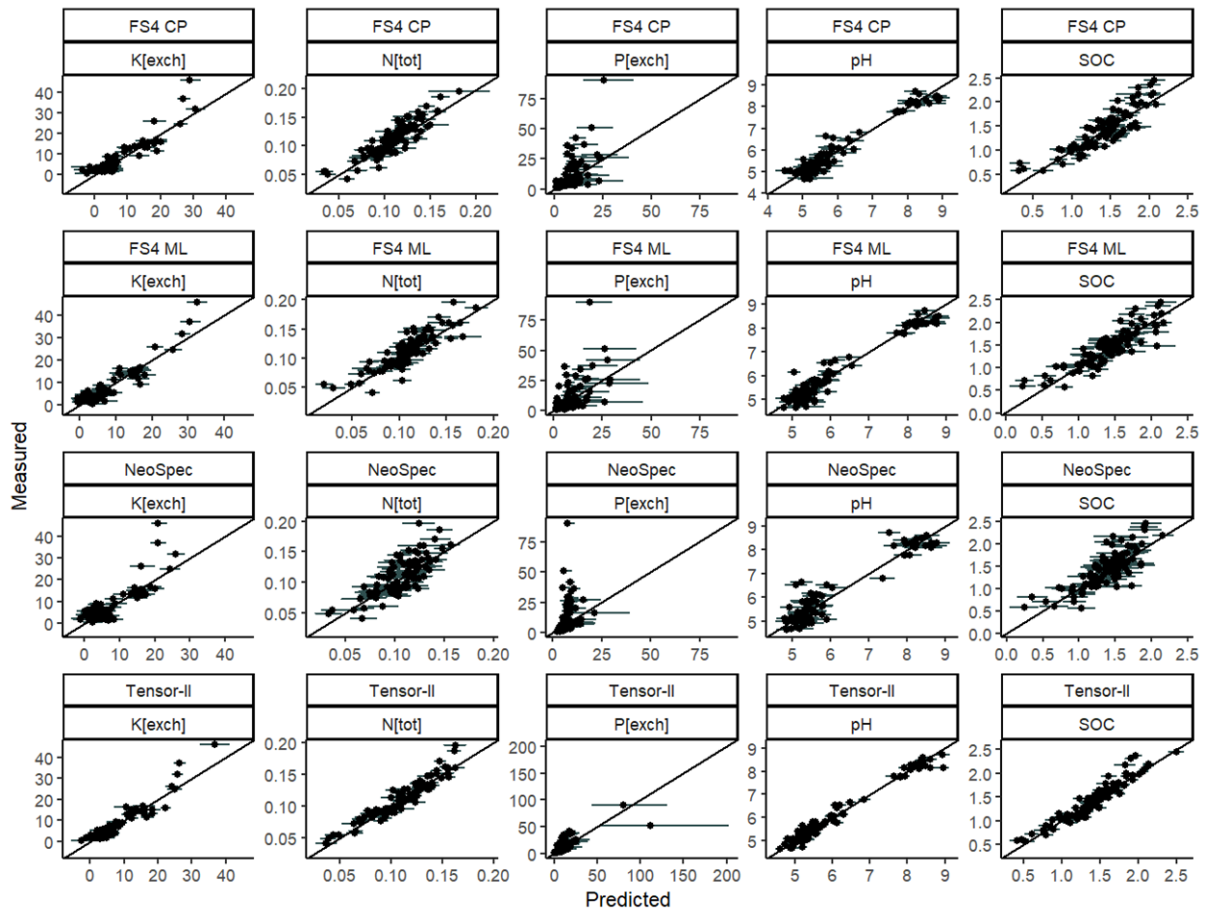


Figure 3.4: Prediction accuracy and consistency for all soil properties (exchangeable K, total N, Olsen P, pH, and SOC, respectively) and spectrometers (CP -> FieldSpec 4 contact probe, ML -> FieldSpec 4 mug light, NeoSpec -> NeoSpectra Saucer, Tensor-II -> FTIR Bruker Tensor 27). The horizontal lines are the error bars which represented 95% confidence intervals.

3.4.4 Comparison of soil testing methods for fertilizer decision making

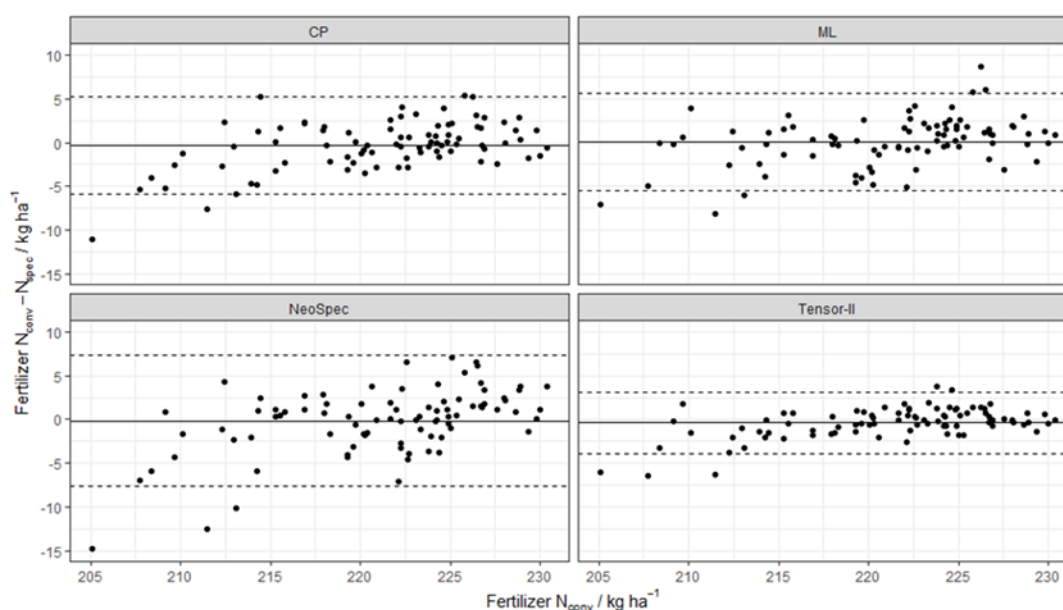
There was no potassium fertilizer recommended for a 7 t ha⁻¹ maize grain yield target because all studied soil samples had sufficient plant available K supply (Table 3.2). Hence, we only included nitrogen and phosphorus recommendation results.

3.4.4.1 Nitrogen fertilizer recommendation

Figure 3.5 shows the scatter plot of the estimated nitrogen requirements using the spectra from the spectrometers versus the conventional soil analysis methods

for a 7 t ha⁻¹ yield target with a 95% confidence interval (CI). The nitrogen requirement ranged between 200 to 240 kg ha⁻¹ for this yield target, with SOC levels varying from 0.57% to 2.51% and pH values spanning from 4.37 to 8.76, as determined using the QUEFTS model. The MIR spectrometer had the lowest RMSE (1.8 kg ha⁻¹) for recommended nitrogen fertilizer rates while the RMSE were 2.8, 2.8 and 3.8 kg ha⁻¹ for the vis-NIR contact probe sample interface, the mug light sample interface and the NIR, respectively. However, the vis-NIR FS4 contact probe and MIR spectrometer resulted in a higher bias than the other two spectrometers. The recommended nitrogen fertilizer rates from all spectrometers were comparably close to each other and fell close to the 1:1 line with the fertilizer rate determined based on conventional soil analysis.

There was a tendency for wider confidence intervals at higher rates of estimated nitrogen fertilizer from predictions based on the contact probe spectrometer and in particular for the Tensor II (Figure 3.5; Figure 3.6



While the confidence intervals are consistent for the other two spectrometers. Between spectrometers, the confidence intervals and the distance between the lower and upper bounds of the 95% fertilizer N estimation difference between the spectrometers and conventional method were narrower with MIR followed by vis-NIR and then NIR, confirming a better performance of the former than the rest of the spectrometers (Figure 3.6). The NIR predictions resulted in the widest

confidence interval and 95% fertilizer N estimation difference (-7.61 to 7.29 kg N ha⁻¹), however the difference was not prominent. Meanwhile the 95% fertilizer N estimation differences lied in a range of -3.86 to 3.07, 5.43 to 5.59, and -5.84 to 5.17 kg ha⁻¹ for Tensor-II MIR, vis-NIR FS4-CP and vis-NIR FS4-ML, respectively. This confirmed again that the MIR spectrometer was best for nitrogen rate decisions. And the NIR and vis-NIR spectrometer-based recommendations performed slightly worse.

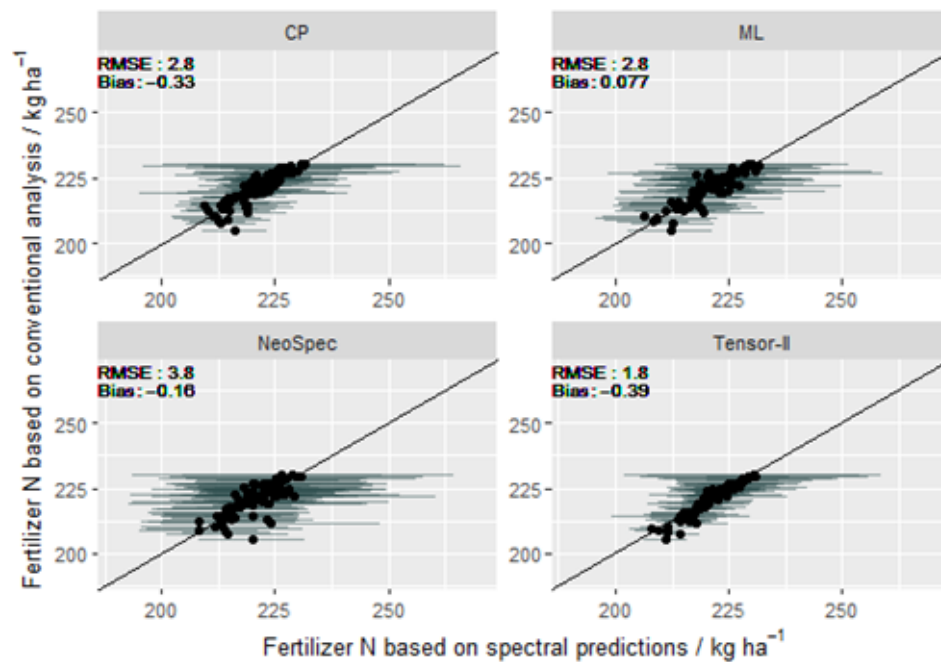


Figure 3.5: Nitrogen fertilizer rate predictions for a 7 t ha⁻¹ target maize yield based on spectral (x-axis) and conventional (y axis) soil analysis. The statistics shown are the root mean square error (RMSE) and Bias for each spectrometer (CP -> FieldSpec 4 contact probe, ML -> FieldSpec 4 mug light, NeoSpec -> NeoSpectra Saucer, Tensor-II -> FTIR Bruker Tensor 27). The horizontal lines are the marginal uncertainties represented as $\pm 1\sigma$.

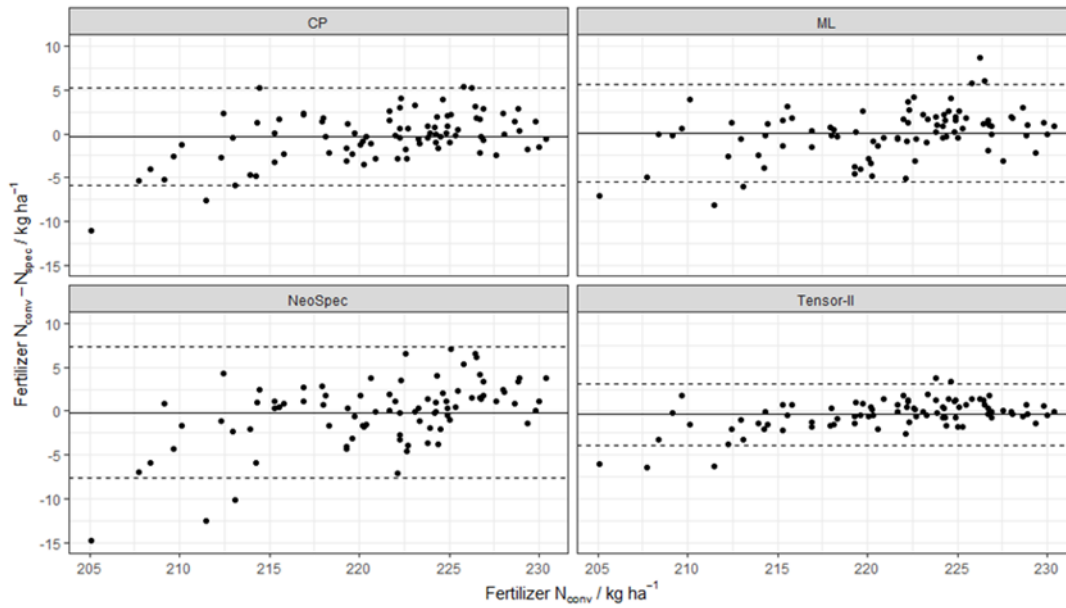


Figure 3.6: Fertilizer N difference among the conventional and spectrometers (FS4-CP -> FieldSpec 4 contact probe, FS4-ML -> FieldSpec 4 mug light, Neospec -> NeoSpectra Saucer, Tensor-II -> FTIR Bruker Tensor 27). The dashed horizontal lines are the mean difference $\pm z^{(\alpha/2)}\sigma$ and the horizontal line is the mean difference.

3.4.4.2 Phosphorus fertilizer recommendation

Figure 3.7 shows the estimated phosphorus fertilizer rate (kg P ha^{-1}) required for a 7 t ha^{-1} yield target with the associated confidence interval estimated based on the error propagation from the prediction of Olsen P. As with, the phosphorus fertilizer recommendations ranged between 10 to 70 kg P ha^{-1} for this yield target, with SOC levels varying from 0.57% to 2.51% and pH values spanning from 4.37 to 8.76 . Higher P recommendations ($>50 \text{ kg P ha}^{-1}$) were more accurately estimated, as there was a substantial agreement between the recommendations based on conventional Olsen P across the spectrometers. However, the disagreement substantially increased for recommended P rates below 50 kg P ha^{-1} , indicated by a considerable scatter of points away from the 1:1 line. This trend agreed well with the prediction pattern in Figure 3.7 which had good predictions at low Olsen P values (and therefore high P fertilizer rates) but had longer error bars with deviation from the 1:1 line at higher Olsen P values (and therefore low P fertilizer rates). However, the overall RMSE of P recommendation was ranging from 10 to 12 kg ha^{-1} with very high negative bias (from -3.0 to -0.91)

of estimation across the spectrometers, that could question the credibility of P fertilizer rate estimations using soil spectroscopic techniques. Besides, there were larger confidence intervals, irrespective of the spectrometers, at higher rates of phosphorus recommendations, indicating higher uncertainties and inefficiencies across the infrared spectrometers. When compared among spectrometers, the distance between the lower and upper bounds of the 95% fertilizer P estimation difference between the spectrometers and conventional method appeared insignificant among the spectrometers. Besides, there was a trend of positive relationships among which may be attributed to the tendency to over-estimate higher values of Olsen P by the spectrometers (Figure 3.8). The 95% P fertilizer estimation difference were -23.56 to 17.92, -21.23 to 18.98, -25.43 to 19.47 and -21.47 to -19.66 kg P ha⁻¹ for vis-NIR FS4-CP, vis-NIR FS4-ML, NIR Neospec and MIR Tensor-II, respectively.

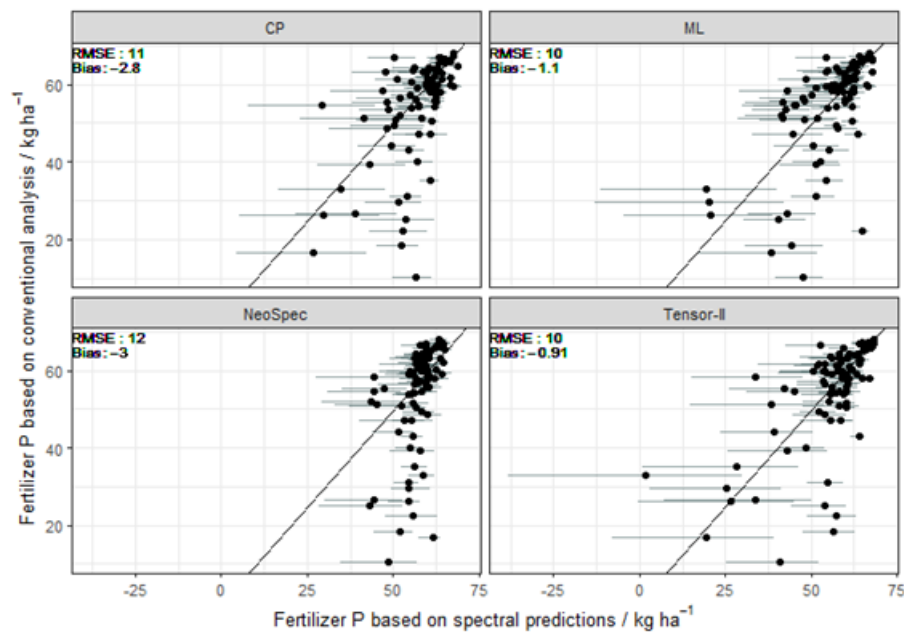


Figure 3.7: Phosphorus fertilizer rate recommendation for a 7 t ha⁻¹ maize target yield based on spectral (x-axis) and conventional (y axis) soil analysis. The statistics shown are the root mean square error (RMSE) and Bias for each spectrometer (FS4-CP -> FieldSpec 4 contact probe, FS4-ML -> FieldSpec 4 mug light, Neospec -> NeoSpectra Saucer, Tensor-II -> FTIR Bruker Tensor 27). The horizontal lines are the 95% confidence intervals.

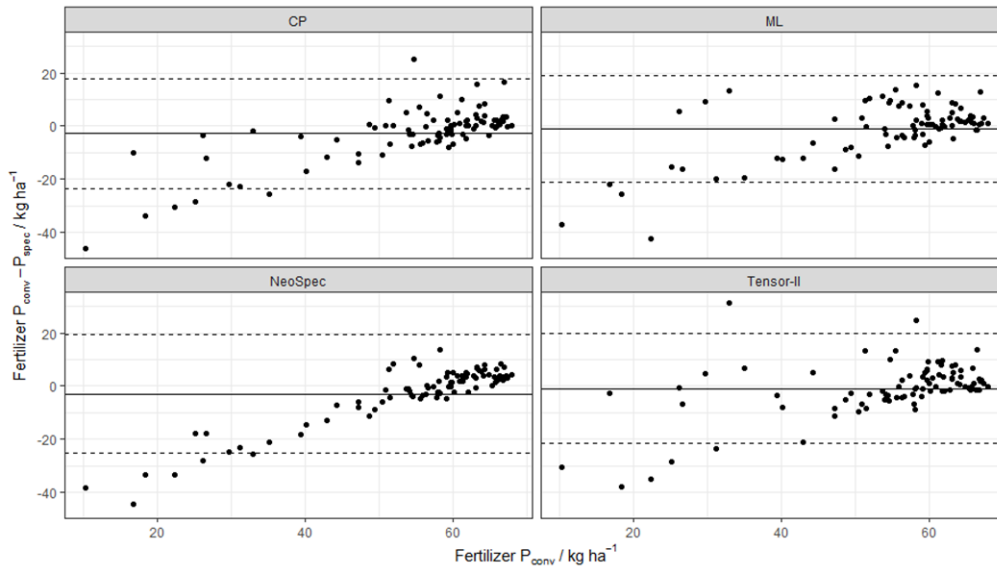


Figure 3.8: Fertilizer P difference among the conventional and spectrometers (FS4-CP -> FieldSpec 4 contact probe, FS4-ML -> FieldSpec 4 mug light, Neospec -> NeoSpectra Saucer, Tensor-II -> FTIR Bruker Tensor 27). The dashed horizontal lines are the mean difference $\pm z^{(\alpha/2)}\sigma$. The horizontal line is the mean difference.

3.5 Discussion

3.5.1 Prediction of soil property inputs for QUEFTS

One of the sources of prediction imprecisions of soil properties by all spectrometers can be associated with the representativeness of the values in the training set for the specific soil properties (Brodský et al., 2013; Viscarra Rossel et al., 2016; Ng et al., 2020). This is due to the skewed distribution of the conventional analysis of some of the soil properties whereby the attribute ranges were not always well represented in training set with an adequate sample number. This can be observed in the scatter plot of Figure **B-2** whereby samples were not evenly distributed across the PC score space. Hence, this might have created a smaller number of occasions for the model to train for those samples which are found in less dense space in the scatter plot. This result agrees with Viscarra Rossel et al., (2016) who reported prediction inconsistency due to the poor soil attribute value representation in the training set. In addition, Guerrero *et al.*, (2016) reported that the inaccuracy of predictive models can be caused by

the inadequacy of the soil libraries and a shortage of sufficient representative soil data. Brodský et al. (2013) reported a high deviation of measured versus predicted SOC at smaller values due to less representation in the training set. Hence, a representative calibration set is a vital component in soil spectroscopy, and some of the prediction discrepancy might be solved with an optimized sampling strategy for the training set. Ng et al. (2020) also mentioned that large variations within the measured values without enough representation might yield an unstable model, which affects the soil nutrients predictions.

In general, the model prediction parameters for most of the QUEFTS inputs did not vary much between the spectrometers, though the prediction performance varied among soil properties whereby the prediction performance of Olsen P was the poorest of all. Several studies concluded that infrared spectroscopy is not satisfactory to predict plant available P when testing the predictive models with an independent test set (e.g., Pätzold et al., 2020). And the prediction performances in all the visible and infrared regions for the soil properties considered were comparable or better than reported by other researchers (Gates, 2018; Ng et al., 2020; Rodríguez-Pérez et al., 2021; Vestergaard et al., 2021; Wijewardane et al., 2018). Despite this, differences were observed between the spectrometers, with prediction accuracies for most of the soil properties being highest in the MIR and vis-NIR range as compared to the NIR. This was particularly true for Olsen P and SOC, likely due to the absence of certain wavelength ranges (Bellon-Maurel and McBratney, 2011). Our result is in consent with the findings of Zelikman and Carmina (2013) who reported lower performance of the NIR range to predict SOC Barra et al. (2021) identified in their review paper that most comparisons favoured the MIR spectroscopy over the vis-NIR or NIR spectroscopy with regards to prediction accuracy and consistency although some studies also found better performance of NIR for some soil properties. However, Li et al. (2022) advised that NIR is the most cost-effective method for SOC analysis, compared with MIR spectroscopy and dry combustion, and assuming > 250 samples per day had to be analyzed. They recommended that vis-NIR spectroscopy was cheaper, accurate and had a large capacity for quick measurements.

Besides these factors, the complexity of acquiring the spectra could be added to the evaluation of different spectrometers, especially for the application in developing countries. Portable NIR spectrometers could allow field-based application whereas bench-top MIR spectrometers require proper laboratories and liquid nitrogen to cool the light source. MIR spectrometers also need fine milled samples whereas NIR and Vis-NIR spectrometers give good results with sieved soil or even in-field scanning (this study and Breure et al., 2021). And finally, the initial equipment cost is another argument, clearly favouring the low cost NIR equipment used in our study. Therefore, our findings suggested that the small NIR spectral range covered by the NeoSpec had little negative effects on the prediction accuracy for the QUEFTS inputs even though the overall prediction performance was in the order of MIR > vis-NIR > NIR spectrometers. Hence, NIR seems to have a comparative advantage in generating relevant soil information regarding prediction precision, cost effectiveness and ease of application.

3.5.2 Fertilizer determination through soil spectral tools by coupling with QUEFTS

Knowledge of relevant soil characteristics is a vital component to advise precise nutrient management adjusted to local conditions and different crop types. Consequently, soil testing has been widely accepted and used as a tool for rational fertilizer use, helping to achieve optimal yields with high nutrient use efficiency and minimized losses to the environment. Moreover, soil-test and crop response based empirical models such as the QUEFTS model (Janssen and Guiking, 1990; Haefele and Wopereis, 2005; Sattari et al., 2014; Yang et al., 2017) or yield response functions (Dhakal and Lange, 2021) have been developed for many cropping systems, using general soil characteristics, nutrient availability tests and/or observed crop responses to applied nutrients as input parameters. These approaches progressed overtime to assess the soil fertility status and recommend suitable and economic nutrient rates through inorganic fertilizer and organic manure for different crops and cropping systems.

However, there are considerable challenges in many countries in getting soil test results in time and at the required scale with conventional soil testing methods.

Hence, applicability of easy soil testing methods and harmonized approaches of fertilizer recommendations could enable instant site-specific fertilizer management, enhancing overall productivity and improved nutrient use efficiency. Employing soil spectroscopic techniques with empirical fertilizer determination could be a breakthrough to alleviate the existing low productivity and nutrient use efficiency through informed decision in developing countries where technologies and information are scarce. Though some researchers argued the non-existence of a standard functional form to evaluate optimal fertilizer recommendation (Dhakal and Lange, 2021), we considered that the QUEFTS model could be a candidate platform to harmonize efforts towards fertilizer recommendations with low costs, especially when coupled with efficient soil testing. The QUEFTS model is a versatile soil-crop-fertilizer model that demands very few inputs (four soil properties and calibrated soil-crop-fertilizer parameters) to derive site specific nutrient rates for optimized nutrient use of various crops (maize, wheat, rice, oil palm, manioc, and cotton) and for both tropic and subtropic environments.

In this research, the resulting recommendation for NP fertilizer for the attainable yield (70% of water limited potential yield) using soil spectroscopic techniques coupled with the QUEFTS model agreed well with the nutrient rates reported by Kenea et al. (2021) and ten Berge et al. (2019) for maize in the soil sample regions. Our findings revealed that the error propagated from the spectral training models was not further exacerbated by the QUEFTS model and parameters although the relative marginal uncertainties for the recommendations increased. This signified the feasibility of predicted soil inputs to be used in the QUEFTS approach. Our results indicated that spectroscopic methods predicted the QUEFTS-relevant soil properties with reasonable accuracy, confirming that a site-specific fertilizer recommendation could be derived efficiently and much faster if compared to the conventional method. Ng et al. (2022) had also proposed a categorical nutrient recommendation based on the vis-NIR or NIR as well as MIR region of the electromagnetic spectrum to alleviate the challenges of cost, complexity, and accessibility of soil testing in developing countries.

When we compared the spectrometers for their efficiency, longer confidence interval bars, revealing higher uncertainties associated with the recommendation, were observed for total N predictions when using NIR and vis-NIR ranges rather than MIR spectra, though most points fall on or close to the 1:1 line when compared with the conventional nitrogen fertilizer determination. Besides, the fertilizer N difference was in a range of acceptable estimation which ranged between -8 to 8 kg N ha⁻¹ for the NIR which was the least performing spectrometer. In contrast, longer confidence intervals, which were comparably similar for all soil spectrometers, were observed for P fertilizer recommendations and the points tended to be underestimated for P rates below 50 kg ha⁻¹ with wider estimation differences with the conventional method. The MIR spectrometer was, in general, more efficient, and effective spectrometer when compared to the vis-NIR and NIR spectrometers for use in fertilizer recommendation. However, its current complexity and cost to use for such purposes for developing countries could be impractical and the vis-NIR spectrometers, though portable, are still costly with a minimum or no variation in prediction performance when compared to the NIR spectrometer. In contrary, the NIR spectrometer is cheaper, portable and has very low running cost which can be used in the field and in low-income countries. This result ascertained that the cheaper, flexible to use and handheld NIR can contribute a lot in access to soil information for informed decision to improve productivity while decreasing fertilizer cost which ultimately improve livelihood of the farmers.

To summarize, the spectroscopic methods followed were efficient and effective for N fertilizer recommendations, and we could not evaluate their use for K fertilizer estimation because all soils analyzed had a high K supply (but the prediction of exchangeable K was promising). The estimation of phosphorus fertilizer rates with soil spectroscopic techniques was unsatisfactory in precision, often overestimating the P fertilizer rate on soils with a good P supply. Its application for this purpose needs further investigation and consideration of larger and better-balanced calibration sets and/or improving the predictive model approach.

3.6 Conclusions

This study evaluated commercially available spectrometers in the vis-NIR, NIR, and MIR ranges for their effectiveness, efficiency, and applicability in analyzing soil properties critical for fertilizer decision-making tools. Subsequently, the predicted soil data were integrated with the QUEFTS model to assess the impact of uncertainty on potential fertilizer recommendations.

Among the spectrometers, the MIR and portable vis-NIR were more costly and required more time-intensive sample preparation compared to the handheld NIR, which is more accessible and practical, especially for low-income countries. Despite the cost and preparation differences, our findings demonstrated that nitrogen and phosphorus fertilizer decisions can be effectively guided by spectra from NIR, vis-NIR, and MIR spectrometers when coupled with QUEFTS and soil-test-based models for site-specific fertilizer recommendations. The confidence intervals indicated that the uncertainty in these predictions is lower than that of blanket or uninformed fertilizer recommendations, offering advantages from environmental, productivity, and cost-efficiency perspectives.

Moreover, the spectra could potentially classify soils into productivity levels through well-trained models, as they represent an aggregate signature of various soil properties. Since there is a direct relationship between spectral patterns and soil properties, this approach could provide a robust foundation for decision-making. However, predicting nutrient requirements directly from spectra may introduce uncertainty due to complex interactions between soil properties and crop nutrient demands. Integrating target yield data could help reduce this uncertainty, enhancing the precision of fertilizer recommendations.

Finally, our study did not compare fertilizer recommendations with actual crop response. Future research should evaluate the accuracy of nitrogen and phosphorus fertilizer estimates generated by soil spectroscopic techniques using the QUEFTS model against crop performance data to further validate these methods.

3.7 Acknowledgment

This work was supported by GeoNutrition projects funded by the Bill & Melinda Gates Foundation (INV-009129) and the UKRI Biotechnology and Biological Sciences Research Council (BBSRC)/Global Challenges Research Fund (GCRF) (BB/P023126/1). The funders were not involved in the study design, the collection, management, analysis, and interpretation of data. We gratefully acknowledge the AfSIS lab of Rothamsted Research for maintaining the soil samples for use in this study. We also acknowledge the OCP Groups of Morocco for the financial support of this research.

3.8 Author contributions

T.G.A., S.M.H, R.S., F.K., R.C. and K.H. conceptualized the study, and acquired and administered project funding. T.G.A. and T.B. contributed to data analyses, including method development, data visualizations and data interpretation. T.G.A wrote the primary draft of the paper with editing and reviewing inputs from other authors.

3.9 References

- Bah, A., Balasundram, S.K. and Husni, M.H.A. (2016), "Sensor Technologies for Precision Soil Nutrient Management and Monitoring", *American Journal of Agricultural and Biological Sciences*, Vol. 7 No. 1, pp. 43–48.
- Barra, I., Haefele, S.M., Sakrabani, R. and Kebede, F. (2021), "Soil spectroscopy with the use of chemometrics, machine learning and pre-processing techniques in soil diagnosis: Recent advances—A review", *TrAC - Trends in Analytical Chemistry*, Elsevier Ltd, Vol. 135 No. 116166, pp. 1–13.
- Bellon-Maurel, V. and McBratney, A. (2011), "Near-infrared (NIR) and mid-infrared (MIR) spectroscopic techniques for assessing the amount of carbon stock in soils - Critical review and research perspectives", *Soil Biology and Biochemistry*, Elsevier Ltd, Vol. 43 No. 7, pp. 1398–1410.
- ten Berge, H.F.M., Hijbeek, R., van Loon, M.P., Rurinda, J., Tesfaye, K., Zingore, S., Craufurd, P., et al. (2019), "Maize crop nutrient input requirements for food

security in sub-Saharan Africa”, *Global Food Security*, Elsevier B.V., Vol. 23 No. November 2018, pp. 9–21.

Bremner, J.M. (1996), “Nitrogen Total”, *Methods of Soil Analysis: Part 3 Chemical Methods*, pp. 1085–1121.

Breure, T.S., Milne, A.E., Webster, R., Haefele, S.M., Hannam, J.A., Moreno-Rojas, S. and Corstanje, R. (2021), “Predicting the growth of lettuce from soil infrared reflectance spectra: the potential for crop management”, *Precision Agriculture*, Springer US, Vol. 22 No. 1, pp. 226–248.

Breure, T.S., Prout, J.M., Haefele, S.M., Milne, A.E., Hannam, J.A., Moreno-Rojas, S. and Corstanje, R. (2022), “Comparing the effect of different sample conditions and spectral libraries on the prediction accuracy of soil properties from near- and mid-infrared spectra at the field-scale”, *Soil and Tillage Research*, Elsevier B.V., Vol. 215 No. July 2021, p. 105196.

Brodský, L., Klement, A., Penížek, V., Kodešová, R. and Borůvka, L. (2011), “Building Soil Spectral Library of the Czech Soils for Quantitative Digital Soil Mapping”, *Soil and Water Research*, Vol. 6 No. 4, pp. 165–172.

Brodský, L., Vašát, R., Klement, A., Zádorová, T. and Jakšík, O. (2013), “Uncertainty propagation in VNIR reflectance spectroscopy soil organic carbon mapping”, *Geoderma*, Vol. 199, pp. 54–63.

Cécillon, L., Barthès, B.G., Gomez, C., Ertlen, D., Génot, V., Hedde, M. and Brun, J.J. (2009), “Assessment and monitoring of soil quality using near-infrared reflectance spectroscopy (NIRS)”, *European Journal of Soil Science*, Vol. 60 No. January, pp. 770–784.

Ciesielski, H. and Sterckeman, T. (1997), “Determination of cation exchange capacity and exchangeable cations in soils by means of cobalt hexamine trichloride”, *Agronomie*, Vol. 17 No. 1, pp. 1–7.

Clingensmith, C.M., Grunwald, S. and Wani, S.P. (2019), “Evaluation of calibration subsetting and new chemometric methods on the spectral prediction

of key soil properties in a data-limited environment”, *European Journal of Soil Science*, Vol. 70 No. 1, pp. 107–126.

Dhakal, C. and Lange, K. (2021), “Crop yield response functions in nutrient application : A review”, *Agronomy Journal*, Vol. 113 No. 6, pp. 5222–5234.

Fajardo, M.P., McBratney, A.B. and Minasny, B. (2017), “Measuring functional pedodiversity using spectroscopic information”, *Catena*, Elsevier B.V., Vol. 152, pp. 103–114.

Gates, J.R. (2018), “A Comparison of VNIR and MIR Spectroscopy for Predicting Various Soil Properties”, *Dissertations & Theses in Natural Resources*, No. 264.

Ge, Y., Morgan, C.L.S., Grunwald, S., Brown, D.J. and Sarkhot, D. V. (2011), “Comparison of soil reflectance spectra and calibration models obtained using multiple spectrometers”, *Geoderma*, Elsevier B.V., Vol. 161 No. 3–4, pp. 202–211.

Van Groenigen, J.W., Mutters, C.S., Horwath, W.R. and Van Kessel, C. (2003), “NIR and DRIFT-MIR spectrometry of soils for predicting soil and crop parameters in a flooded field”, *Plant and Soil*, Vol. 250 No. 1, pp. 155–165.

Grunwald, S., Thompson, J.A. and Boettinger, J.L. (2011), “Digital Soil Mapping and Modeling at Continental Scales : Finding Solutions for Global Issues”, *Soil Science Society of America Journal*, Vol. 75 No. 4, pp. 1201–1213.

Guerrero, C., Wetterlind, J., Stenberg, B., Mouazen, A.M., Gabarrón-Galeote, M.A., Ruiz-Sinoga, J.D., Zornoza, R., et al. (2016), “Do we really need large spectral libraries for local scale SOC assessment with NIR spectroscopy?”, *Soil and Tillage Research*, Elsevier B.V., Vol. 155, pp. 501–509.

Haefele, S.M. and Wopereis, M.C.S. (2005), “Spatial variability of indigenous supplies for N, P and K and its impact on fertilizer strategies for irrigated rice in West Africa”, *Plant and Soil*, Vol. 270 No. 1, pp. 57–72.

Hengl, T., Miller, M.A.E., Križan, J., Shepherd, K.D., Sila, A., Kilibarda, M., Antonijević, O., et al. (2021), “African soil properties and nutrients mapped at 30

m spatial resolution using two - scale ensemble machine learning”, Scientific Reports, Nature Publishing Group UK, Vol. 11 No. 6130, pp. 1–18.

Hutengs, C., Ludwig, B., Eisele, A. and Vohland, M. (2018), “Comparison of Portable and Bench-Top Spectrometers for Mid-Infrared Diffuse Reflectance Measurements of Soils”, *Sensors*, Vol. 18 No. 4, p. 998.

Ichami, S.M., Shepherd, K.D., Sila, A.M., Stoorvogel, J.J. and Hoffland, E. (2019), “Fertilizer response and nitrogen use efficiency in African smallholder maize farms”, *Nutrient Cycling in Agroecosystems*, Springer Netherlands, Vol. 113 No. 1, pp. 1–19.

Janssen, B.H. and Guiking, F.C.T. (1990), “Modelling the response of crops to fertilizers”, In *Plant Nutrition—Physiology and Applications*, pp. 699–703.

Janssen, B.H., Guiking, F.C.T., van der Eijk, D., Smaling, E.M.A., Wolf, J. and van Reuler, H. (1990), “A system for quantitative evaluation of the fertility of tropical soils (QUEFTS)”, *Geoderma*, Vol. 46 No. 4, pp. 299–318.

Kenea, W.B., Reidsma, P., Descheemaeker, K., Rurinda, J., Balemi, T. and van Ittersum, M.K. (2021), “Variability in yield responses, physiological use efficiencies and recovery fractions of fertilizer use in maize in Ethiopia”, *European Journal of Agronomy*, Elsevier B.V., Vol. 124 No. January, p. 126228.

Knadel, M., Stenberg, B., Deng, F., Thomsen, A. and Greve, M.H. (2013), “Comparing predictive abilities of three visible-near infrared spectrophotometers for soil organic carbon and clay determination”, *Journal of Near Infrared Spectroscopy*, Vol. 21 No. 1, pp. 67–80.

Krasilnikov, P., Taboada, M.A. and Amanullah, 2022. Fertilizer use, soil health and agricultural sustainability. *Agriculture*, 12(4), p.462.

Lausch, A., Dierke, C., Doktor, D. and Werban, U. (2013), “Analysis of Vegetation and Soil Patterns using Hyperspectral Remote Sensing , EMI , and Gamma- Ray Measurements”, No. October 2015, available at:<https://doi.org/10.2136/vzj2012.0217>.

Li, S., Rossel, R.A.V. and Webster, R. (2022), "The cost-effectiveness of reflectance spectroscopy for estimating soil organic carbon", *European Journal of Soil Science*, Vol. 73 No. 1, pp. 1–16.

Lin, L.I.-K. (1989), "A Concordance Correlation Coefficient to Evaluate Reproducibility", *Biometrics*, Vol. 45 No. 1, p. 255.

Malpass, D. (2022), "A transformed fertilizer market is needed in response to the food crisis in Africa", *Voices Blog: Perspectives on Development*, available at: <https://blogs.worldbank.org/en/voices/transformed-fertilizer-market-needed-response-food-crisis-africa>.

Morais, P., Inácio, E., Filho, F. and Rocha, M. (2018), "Digital Soil Mapping of Soil Properties in the ' Mar de Morros ' Environment Using Spectral Data", pp. 1–19.

Mouazen, A.M., Maleki, M.R., De Baerdemaeker, J. and Ramon, H. (2007), "On-line measurement of some selected soil properties using a VIS-NIR sensor", *Soil and Tillage Research*, Vol. 93 No. 1, pp. 13–27.

Mouazen, A.M., Saey, W., Xing, J., Baerdemaeker, De, J. and H., R. (2005), "Near Infrared Spectroscopy for Agricultural Materials: An Instrument Comparison", *Near Infrared Spectrosc*, Vol. 13 No. October 2018, pp. 87–97.

Nelson, D.W. and Sommers, L.E. (1996), "Total Carbon , Organic Carbon , and Organic Matter", *Methods of Soil Analysis: Part 3 Chemical Methods*, pp. 961–1010.

Ng, W., Husnain, Anggria, L., Siregar, A.F., Hartatik, W., Sulaeman, Y., Jones, E., et al. (2020), "Developing a soil spectral library using a low-cost NIR spectrometer for precision fertilization in Indonesia", *Geoderma Regional*, Elsevier B.V., Vol. 22, p. e00319.

Ng, W., Minasny, B., Jeon, S.H. and McBratney, A. (2022), "Mid-infrared spectroscopy for accurate measurement of an extensive set of soil properties for assessing soil functions", *Soil Security*, Elsevier Ltd, Vol. 6, p. 100043.

Ng, W., Minasny, B., Montazerolghaem, M., Padarian, J., Ferguson, R., Bailey, S. and McBratney, A.B. (2019), "Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra", *Geoderma*, Elsevier, Vol. 352 No. June, pp. 251–267.

Olsen, S.. (1954), "Estimation of available phosphorus in soils by extraction with sodium bicarbonate", US Department of Agriculture., Vol. 939, pp. 1–20.

Pätzold, S., Leenen, M., Frizen, P., Heggemann, T., Wagner, P. and Rodionov, A. (2020), "Predicting plant available phosphorus using infrared spectroscopy with consideration for future mobile sensing applications in precision farming", *Precision Agriculture*, Vol. 21 No. 4, pp. 737–761.

Paul, S.S., Coops, N.C., Johnson, M.S., Krzic, M. and Smukler, S.M. (2019), "Evaluating sampling efforts of standard laboratory analysis and mid-infrared spectroscopy for cost effective digital soil mapping at field scale", *Geoderma*, Vol. 356 No. April, available at:<https://doi.org/10.1016/j.geoderma.2019.113925>.

Pebesma, E.J. and Wesseling, C.G. (1998), "Gstat: A program for geostatistical modelling, prediction and simulation", *Computers and Geosciences*, Vol. 24 No. 1, pp. 17–31.

Piikki, K., Söderström, M., Eriksson, J., John, J.M., Muthee, P.I., Wetterlind, J. and Lund, E. (2016), "Performance Evaluation of Proximal Sensors for Soil Assessment in Smallholder Farms in Embu", *Sensors*, Vol. 16 No. 11, pp. 1950–1972.

R Core Team. (2017), "R Language Definition", Vol. 1.

Ramirez-lopez, L., Stevens, A., Orellano, C., Viscarra, R., Lobsey, C. and Wadoux, A. (2022), "resemble: Regression and similarity evaluation for memory-based learning in spectral chemometrics.", *R Package Version 1.2.2.*, Vol. 1 No. 2, pp. 1–47.

Reeves, J.B. (2010), "Near- versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis:

Where are we and what needs to be done?”, *Geoderma*, Elsevier B.V., Vol. 158 No. 1–2, pp. 3–14.

Rodionov, A., Pätzold, S., Welp, G., Pude, R. and Amelung, W. (2016), “Proximal field Vis-NIR spectroscopy of soil organic carbon: A solution to clear obstacles related to vegetation and straw cover”, *Soil and Tillage Research*, Elsevier B.V., Vol. 163, pp. 89–98.

Rodríguez-Pérez, J.R., Marcelo, V., Pereira-obaya, D., Garc, M. and Sanz-Ablanedo, E. (2021), “Estimating Soil Properties and Nutrients by Visible and Infrared Diffuse Reflectance Spectroscopy to Characterize Vineyards”, *Agronomy*, Vol. 11 No. 1895.

Rurinda, J., Zingore, S., Jibrin, J.M., Balemi, T., Masuki, K., Andersson, J.A., Pampolino, M.F., et al. (2020), “Science-based decision support for formulating crop fertilizer recommendations in sub-Saharan Africa”, *Agricultural Systems*, Elsevier, Vol. 180 No. December 2019, p. 102790.

Dos Santos, C.A.T., Lopo, M., Páscoa, R.N. and Lopes, J.A. (2013), “A review on the applications of portable near-infrared spectrometers in the agro-food industry”, *Applied Spectroscopy*, Vol. 67 No. 11, pp. 1215–1233.

Sattari, S.Z., van Ittersum, M.K., Bouwman, A.F., Smit, A.L. and Janssen, B.H. (2014), “Crop yield response to soil fertility and N, P, K inputs in different environments: Testing and improving the QUEFTS model”, *Field Crops Research*.

Sharififar, A. (2019), “Evaluating a low- - cost portable NIR spectrometer for the prediction of soil organic and total carbon using different calibration models”, *Soil Use and Management*, Vol. 35 No. 4, pp. 607–616.

Shepherd, K.D. and Walsh, M.G. (2002), “Development of Reflectance Spectral Libraries for Characterization of Soil Properties”, *Soil Science Society of America Journal*, Vol. 66 No. 3, pp. 988–998.

Shepherd, K.D. and Walsh, M.G. (2007), “Infrared spectroscopy - Enabling an evidence-based diagnostic surveillance approach to agricultural and

environmental management in developing countries”, *Journal of Near Infrared Spectroscopy*, Vol. 15 No. 1, pp. 1–19.

Si-ware. (2017), *NeoSpectra Micro – Development Kit: Evaluation Brief Guide V1.0*.

Smaling, E.M.A. and Janssen, B.H. (1993), “Calibration of QUEFTS, a model predicting nutrient uptake and yields from chemical soil fertility indices”, *Geoderma*, Vol. 59 No. 1–4, pp. 21–44.

Soriano-Disla, J.M., Janik, L.J., Rossel, R.A.V., MacDonald, L.M. and McLaughlin, M.J. (2014), “The Performance of Visible , Near- , and Mid-Infrared Reflectance Spectroscopy for Prediction of Soil Physical , Chemical , and Biological Properties”, *Applied Spectroscopy Reviews*, Vol. 49 No. 2, pp. 139–186.

Stevens, A. and Lopez, L.R. (2022), *An Introduction to the Prospectr Package*.

Tang, Y., Jones, E. and Minasny, B. (2019), “Evaluating low-cost portable near infrared sensors for rapid analysis of soils from South Eastern Australia”, *Geoderma Regional*, Elsevier B.V., Vol. 20 No. e00240, pp. 1–11.

Tittonell, P., Vanlauwe, B., Corbeels, M. and Giller, K.E. (2008), “Yield gaps, nutrient use efficiencies and response to fertilisers by maize across heterogeneous smallholder farms of western Kenya”, *Plant and Soil*, Vol. 313 No. 1–2, pp. 19–37.

Vestergaard, R.-J., Vasava, H.B., Aspinall, D., Chen, S., Gillespie, A., Adamchuk, V. and Biswas, A. (2021), “Evaluation of Optimized Preprocessing and Modeling”, *Sensors*, Vol. 21 No. 20, p. 6745.

Viscarra Rossel, R.A., Behrens, T., Ben-dor, E., Brown, D.J., Demattê, J.A.M., Shepherd, K.D., Shi, Z., et al. (2016), “A global spectral library to characterize the world’s soil”, *Earth-Science Reviews*, Vol. 155, pp. 198–230.

Viscarra Rossel, R.A., Walvoort, D.J.J., Mcbratney, A.B., Janik, L.J. and Skjemstad, J.O. (2006), “Visible , near infrared , mid infrared or combined diffuse

reflectance spectroscopy for simultaneous assessment of various soil properties”, *Geoderma*, Vol. 131 No. 2, pp. 59–75.

Wehrens, R. (2007), “The pls Package: Principal Component and Partial Least Squares Regression in R”, *Journal of Statistical Software*, Vol. 18 No. 2, pp. 1–23.

Wehrens, R., Simonetti, A.W. and Buydens, L.M.C. (2002), “Mixture modelling of medical magnetic resonance data”, *Journal of Chemometrics*, Vol. 16 No. 6, pp. 274–282.

Wijewardane, N. k., Ge, Y., Wills, S. and Libohova, Z. (2018), “Predicting Physical and Chemical Properties of US Soils with a Mid-Infrared Reflectance Spectral Library”, *Soil Science Society of America Journal*, Vol. 82 No. 3, pp. 722–731.

Wortmann, C.S. and Sones, K.R. eds., 2017. Fertilizer use optimization in sub-Saharan Africa (pp. 227-pp). Wallingford, UK: CABI.

Xie, H.T., Yang, X.M., Drury, C.F., Yang, J.Y. and Zhang, X.D. (2011), “Predicting soil organic carbon and total nitrogen using mid- and near-infrared spectra for Brookston clay loam soil in Southwestern Ontario, Canada”, *Canadian Journal of Soil Science*, Vol. 91 No. 1, pp. 53–63.

Xu, C., Zeng, W., Huang, J., Wu, J. and Van Leeuwen, W.J.D. (2016), “Prediction of soil moisture content and soil salt concentration from hyperspectral laboratory and field data”, *Remote Sensing*, Vol. 8 No. 1, available at:<https://doi.org/10.3390/rs8010042>.

Yang, F., Xu, X., Wang, W., Wang, J., Wei, D., He, P., Pampolino, M.F., et al. (2017), “Estimating nutrient uptake requirements for soybean using QUEFTS model in China”, *PLoS ONE*, Vol. 12 No. 5, pp. 1–12.

Zelikman, E. and Carmina, E. (2013), “The spectral response characteristics of the soils and their possible estimation by using partial least square regression (PLSR) analysis”, *International Journal of Geomatics and Geosciences*, Vol. 3 No. 3, pp. 438–453.

Zhang, Y. (2014), Quantification of Prediction Uncertainty for Principal Components Regression and Partial Least Squares Regression, University College London.

4 A Moroccan soil spectral library use framework for improving soil property diagnostics: Evaluating a Geostatistical Approach

Tadesse Gashaw Asrat^{1*}, Timo Breure¹, Ruben Sakrabani¹, Ron Corstanje¹, Kirsty L. Hassall², Abdellah Hamma⁴, Fassil Kebede³, Stephan M. Haefele²

¹Cranfield University, Cranfield, UK;

²Rothamsted Research, Sustainable Soils and Crops, Harpenden, UK;

³Centre for Soil and Fertilizer Research in Africa, College of Agriculture and Environmental Science, Mohammed VI Polytechnic University, Ben Guerir, Morocco.

⁴BU Al Moutmir, College of Agriculture and Environmental Sciences, Mohammed VI Polytechnic University, Ben Guerir, Morocco.

4.1 Abstract

A soil spectrum generated by any spectrometer requires a calibration model to estimate soil properties from it. To achieve best results, the assumption is that locally calibrated models offer more accurate predictions. However, achieving this higher accuracy comes with associated costs, complexity, and resource requirements, thus limiting widespread adoption. Furthermore, there is a lack of comprehensive frameworks for developing and utilizing soil spectral libraries (SSLs) to make predictions for specific samples. While calibration samples are necessary, there is the need to optimize SSL development through strategically determining the quantity, location, and timing of these samples based on the quality of the information in the library. This research aimed to develop a spatially optimized SSL and propose a use-framework tailored for predicting soil properties for a specific farmland context. Consequently, the Moroccan SSL (MSSL) was established utilizing a stratified spatially balanced sampling design, using six environmental covariates and FAO soil units. Subsequently, various criteria for calibration sample selection were explored, including a spatial autocorrelation of spectra principal component (PC) scores (spatial calibration

sample selection), spectra similarity memory-based learner (MBL), and selection based on environmental covariate clustering. Twelve soil properties were used to evaluate these calibration sample selections to predict soil properties using the near infrared (NIR) and mid infrared (MIR) ranges. Among the methods assessed, we observed distinct precision improvements resulting from spatial sample selection and MBL compared to the use of the entire MSSL. Notably, the Lin's Concordance Correlation Coefficient (CCC) values using the spatial calibration sample selection was improved for Olsen extractable phosphorus (OlsenP) by 41.3% and Mehlich III extractable phosphorus (P_M3) by 8.5% for the MIR spectra and for CEC by 25.6%, pH by 13.0% and total nitrogen (Tot_N) by 10.6% for the NIR spectra in reference to use of the entire MSSL. Utilizing the spatial autocorrelation of the spectra PC scores proved beneficial in identifying appropriate calibration samples for a new sample location, thereby enhancing prediction performance comparable to, or surpassing that of the use of the entire MSSL. This study signifies notable advancement in crafting targeted models tailored for specific samples within a vast and diverse SSL.

Keywords: Soil properties, IR spectroscopy, spatial non-stationarity, geostatistics, MBL, covariate clustering, Morocco

4.2 Introduction

Proximal spectral signatures acquired from soil samples in the near infrared (NIR) to mid infrared (MIR) ranges of the electromagnetic spectrum need to be calibrated, mostly using conventional wet chemistry analysis. A soil spectral library (SSL), which is fundamentally a paired dataset consisting of soil spectral records and conventional wet chemistry data of the same samples, is built to calibrate the predictive models for estimating soil properties for new samples (Knadel et al., 2012; Demattê et al., 2019; Baumann et al., 2017; Summerauer et al., 2021). Based on this approach, soil spectroscopy has been evaluated for diverse applications with the aim of minimising the need for future expensive and time-consuming conventional soil analysis such as for digital soil mapping (Brodský et al., 2011), soil classification and groupings (Viscarra Rossel et al., 2016; Zeng et al., 2017), and precision agriculture (Ng, et al., 2020; Breure et al.,

2021; Asrat et al., 2023). However, for the practical application of soil spectroscopy, it is essential to establish a comprehensive reference pattern in the SSL database. Addressing this need involves ensuring that the calibration samples used in the models are the most representative of the criteria such as the target region, the spatial scale, sample conditions, the range of soil variability, and the spectral space in which the models are to be applied (Savvides et al., 2010; Grunwald et al., 2018).

Therefore, it could be possible to achieve the required high resolution soil information easily and more economically (although not equally well for all soil characteristics) using soil spectroscopy to support informed decisions and monitoring of sustainable soil functions and ecological services obtained from soil systems compared to the conventional soil analysis methods. However, to take the approach towards wider application, there are still concerns over the source of uncertainties related to SSL development and selection of appropriate calibration sample entries from a large SSL for new samples to be predicted (Ramirez-Lopez et al., 2013; Sun et al., 2021). This is because the large SSL might possess non-informative samples (regarding the local spectral and soil variability) to develop a specific predictive model for a specific farmland. The routine evaluation of soil predictions using soil spectroscopy is to split the data into calibration and a validation sets, either randomly or with stratifying algorithms in the spectra space such as k-means clustering (Næs, 1987), Kennard-Stone sampling (Kennard and Stone, 1969; Nawar and Mouazen, 2018), conditioned hypercubic sampling (Minasny and McBratney, 2006; Moloney et al., 2023), a combination of these (Ogen et al., 2019), or by stratification using environmental covariates (Wijewardane et al., 2018; Dotto et al., 2020).

The spectral libraries might cover differing spatial areas, ranging from local (Guerrero et al., 2016) and national (Demattê et al., 2019) to regional (Shepherd and Walsh, 2002), continental (Stevens et al., 2013), or even global scales (Viscarra Rossel et al., 2016). In all these scaled SSLs, apart from variation in the sampling design, number of samples required, and spectra information generated, a spatial component is embedded which the current evaluation of

prediction performances may have not considered. Further on, the validation set used to evaluate the predictive models may not be similar in spectral and soil characteristics with the new samples to be analysed (Stenberg et al., 2010). Consequently, this may result in model overfitting which is characterized by smaller calibration errors and large prediction deviation causing a reliability issue (Summerauer et al., 2021).

Considering this, while calibration samples are essential, there is significant potential to optimize the quantity, spatial distribution, and type of samples used to enhance the quality of the spectral library (Shepherd et al., 2022). Such optimised spectral libraries are supposed to minimize and/or avoid further requirement of conventionally analysed soil samples to calibrate a spectroscopic model while predicting soil parameters for new samples. Besides, large datasets might not be computationally effective and can cause problems when using routine calibration methods for regression by increasing noise and leading to model inadequacy (Sáiz-Abajo et al., 2005). By selecting representative (spatially, spectrally, or other improving criteria and combinations of these) and well-distributed calibration sets from abundant SSL data entries, fewer samples could be sufficient to build a robust and reliable model without losing the prediction accuracy.

Thus, the challenge is to determine the best calibration samples from a large and diverse SSL to predict soil properties for any unknown sample which only possess geographic coordinates and soil spectra information. A few approaches have been tried and assessed to downscale large spectral libraries for local scale, avoiding the need of new local calibration sample wet chemistry analysis while improving prediction performance to minimizing the cost of soil information. Among other, geographically weighted partial least square regression Kriging (PLS-GWRK) implements a spatial weighted regression and gives larger weights to the coefficients closest to the new sample location (Song et al., 2021). Another approach is the implementation of a continuous spectra similarity analysis (Memory-based Learner [MBL]) to narrow down calibration sample entries based on the spectra distance to the new samples from a large SSL (Dangal et al., 2019;

Summerauer et al., 2021). Both approaches have been evaluated for their contribution to improved soil property prediction for new samples from fewer calibration samples (<the entire SSL) in geographic and spectra space, respectively.

However, these studies did not consider the spatial autocorrelation of spectra when selecting calibration samples for predicting a new sample. Spatial autocorrelation refers to the principle that spatially close objects are more likely to exhibit similar properties than those further apart. In the context of soil spectra, it means that the spectral signatures of soil samples located near each other are more likely to be similar in most structures. By accounting for the spatial dependency through variogram analysis, it can be more effective in selecting calibration samples that represent the spatial variability within a study area, ultimately improving the predictive accuracy of soil property models from spectra.

Additionally, previous studies limited their evaluation to a small number of soil properties and a single spectral source. In order to obtain a robust assessment of the effectiveness of subsetting entire SSLs, it is essential to examine and model the spatial relationship of spectra within a SSL generated from various soil scanning instruments. In this paper we consider both MIR and NIR spectral sources.

We targeted the Moroccan rainfed wheat growing areas as the research domain. To date, there are only a few samples collected from a Mediterranean climate included in the global (GSSL) (<https://explorer.soilspectroscopy.org/>) (SoilSpec4GG, 2020) as well as the Geocradle Regional Soil Spectral Libraries (GRSSL) (<http://datahub.geocradle.eu/dataset/regional-soil-spectral-library>) (Geocradle, 2018), with no soil samples included from Morocco. To evaluate different approaches for selecting calibration samples, the Moroccan soil spectral library (MSSL) was developed. The compared methods included selection based on spectra PC spatial autocorrelation (spatial sample selection), soil spectra similarity (MBL), and environmental covariate clusters, in comparison to the use of the entire MSSL. The specific objectives of this study were: - 1) to develop a representative SSL for the rainfed wheat-growing areas of Morocco, 2) to study

the spatial autocorrelation of the spectra PC scores in MIR and NIR spectra, 3) to evaluate the prediction performance of the calibration models built with spatial, spectral and covariate clustering based calibration sample selection for various soil properties and soil spectral range combinations, and 4) to propose a use-framework on how best to use the MSSL for prediction of soil properties for a specific farmland.

4.3 Material and methods

4.3.1 Soil sampling locations and sampling strategy

This research targeted the rainfed wheat-growing areas of Morocco and the soil samples were collected from regions of agricultural importance and their dominant soil types. Morocco is a country in the Maghreb region of North-western Africa with a Mediterranean climate. Morocco is mostly situated in the arid to semi-arid climate region, and the rainfed agriculture accounts for around 80% of the utilized agricultural area (Mamassi et al., 2023). Based on the 2017/2018 crop cover map (<https://lcviewer.vito.be/2018>) (Copernicus:Europe's eyes on Earth, 2020) and the FAO 1974 soil classification (Spaargaren and Batjes, 1995), Calcic Kastanozems (Kk), Calcic Cambisols (Bk) and Chromic Luvisols (Lc) dominate the cropped areas of Morocco in this order of importance (C.1.1).

We selected a subset of 599 soil sampling locations from the Al Moutmir BU (College of Agriculture and Environmental Sciences, Mohammed VI Polytechnic University, Ben Guerir, Morocco) dataset that classified as rainfed wheat farms in the 2019-2020 cropping season. Sampling sites were selected with a stratified balanced coverage sampling (StrBCS) design to have a balanced representation of the soil types, geographic coordinates, and environmental covariates. Stratified balanced coverage sampling or the doubly balanced sampling eliminates the poor spread of covariates on the axis in their spatial space. Details of the sampling algorithms are provided in the C.1.2. For the stratification, we used the `lcubestratified` function from the `BalancedSampling` package of R (Grafström, 2018).

Mean annual rainfall, mean annual temperature, slope percent and elevation, which are the most influential drivers of soil formation at local scale, were used as balancing covariates, longitude and latitude were used as spreading covariates, and the FAO soil types were used as stratifying variables. Details of the environmental covariate data types and sources are elaborated in the C.1.3. Furthermore, the StrBCS design was used to subset the database into a calibration database (further on named as the Moroccan Soil Spectral Library (MSSL) and new samples (validation set) to evaluate different methodologies of sample selection for calibrating soil spectra to laboratory data (Figure 4.1).

The soil samples were collected from the 0–20cm soil depth with a soil auger, each representing the field of a collaborating farmer. At each location, soil samples were collected in a ‘W’ pattern across the field, mixed thoroughly and the composite sample was taken for air drying and further sample processing for analysis. All sampling locations had linked data of farming practices, coordinates, and crop productivity. An overview of the sampling sites is given in Figure 4.1.

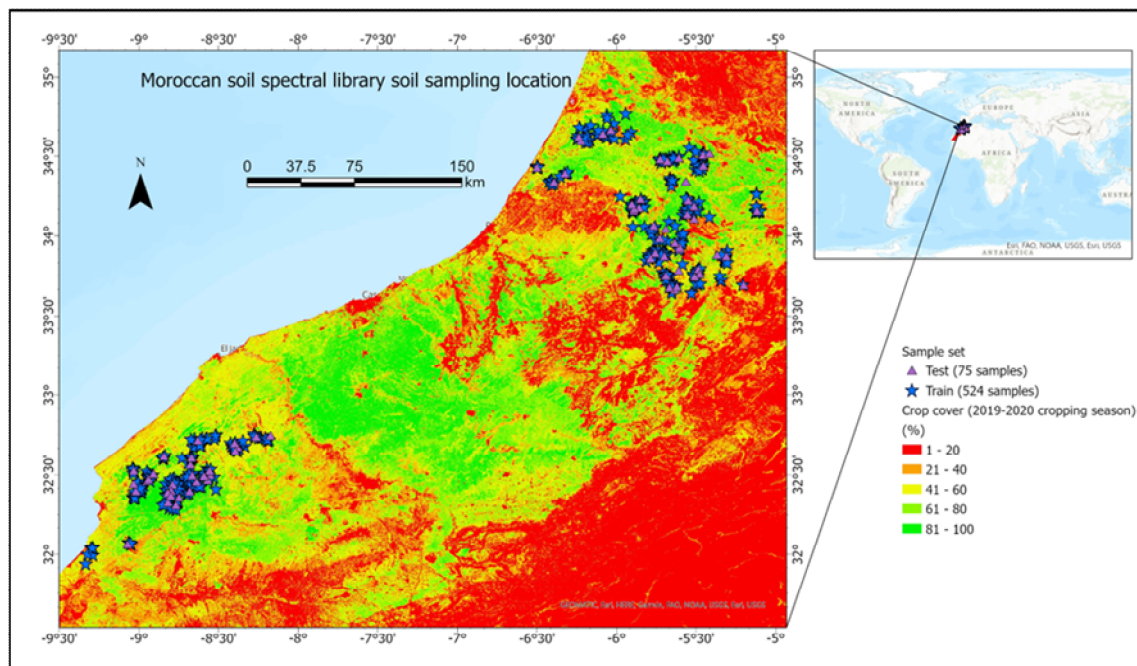


Figure 4.1: Soil sample locations with crop cover map of 2019-2020 cropping season and sample subset. There were two geographic distinct sampling regions situated in the north and south of the rainfed wheat growing regions of Morocco.

Note that soil sampling for the cropping area between the two geographic distinct sampling regions is ongoing.

4.3.2 Baseline soil property analyses

The soil samples were air-dried, crushed, sieved to less than 2mm, and further processed (milling and weighing) for laboratory analysis as required. Twelve soil properties were considered for this study based on their importance for plant nutrition, soil health and soil productivity. Soil particle size distribution was determined by a laser diffraction on a L-960 particle-size analyser (Horiba scientific Ltd.) in the Dry Spectral Laboratory at Rothamsted Research. The instrument allows for the measurement of particle sizes (volume, %) within the size range of 0.02-2000µm using a diode laser of 650nm wavelength and a blue LED light source of 405nm wavelength in wet mode. The intervals used were clay diameter < 5µm, silt 5 < d < 63µm, sand 63µm < d < 2000µm (Thomas et al., 2021). Total carbon and nitrogen were measured using the dry combustion method with a Leco analyser (Nelson and Sommers, 1996; Bremner, 1996). Inorganic carbon was determined via wet acidification using a Skalar Primacs AIC 100 (Skalar Analytical BV, Breda, Netherlands). Soil samples were introduced into the reactor, cleared with oxygen, then treated with Phosphoric acid and heated to 135°C. The released CO₂ was measured by an infrared detector. Total soil organic carbon (SOC) was then calculated by subtracting inorganic carbon from total soil carbon.

The ammonium acetate extracts (ISO 13536:1995) of exchangeable potassium (Exch.K) and magnesium (Exch.Mg) were determined by atomic emission and atomic absorbance, respectively, and CEC was determined with the ammonium acetate method (Ciesielski and Sterckeman, 1997). Soil pH-H₂O was measured in a 1:2.5 soil: water suspension (ISO 10390: 2005) one hour after mixing, using a semi-micro sealed pH electrode from Fisher scientific. Sodium bicarbonate extracted soil phosphorus (OlsenP) was determined following the Olsen method (Olsen, 1954), and Mehlich III extractable phosphorus (P_M3) was determined with an inductively coupled plasma membrane analyser (Mehlich, 1984). Available zinc (DTPA Zn) and manganese (DTPA Mn) were determined by atomic

absorption spectrophotometer after complexing with DTPA (Lindsay and Norvell, 1972).

4.3.3 Soil spectral collection and spectra preprocessing

Absorbance data in the MIR range (2500–16,666nm) were obtained in two replicates, and the average of these was used for subsequent spectral modelling. The spectral signatures were acquired from finely milled samples (< 50µm following overnight drying at 40°C. Measurements in the NIR range (1350-2600nm) were obtained in three replicates per sample. Each replicate was obtained from the same Petri dish by shifting the scanning spot across the sample holder. These samples were air-dried and sieved through 2mm stainless steel sieve. For more detailed information on each instrument's specifications, including the labelling used in subsequent graphs and discussions, please refer to the C.1.4. Spectra collected by each instrument is depicted in Figure 4.2.

The reflectance (R) measurements recoded by Neospectra was transformed to the logarithmic apparent absorbance using $A = \log(1/R)$ after which spectra beyond 2450nm were trimmed off due to low signal-to-noise ratio often caused by light scattering effects of quartz sand or instrument drift. For both spectra signatures, a combination of spectral pre-processing techniques were applied that improved absorption features by reducing or eliminating noise and thus enhancing the correlation with soil properties (Vestergaard et al., 2021). A correlation of these pre-processing algorithms with the conventional soil properties' attribute was performed. The smoothing was done with a window size of 11 for MIR and 3 for NIR and a polynomial order of 1. Afterwards, the spectral data was cleaned from water absorption regions (H₂O band 1 between 1350 and 1460nm; H₂O band 2 between 1790 and 1960nm) in the NIR, and CO₂ peaks in the MIR absorption regions (between 4274 and 4464nm). Spectral pre-processing was implemented in R with the `prospectr` package (Stevens and Ramirez-Lopez, 2022).

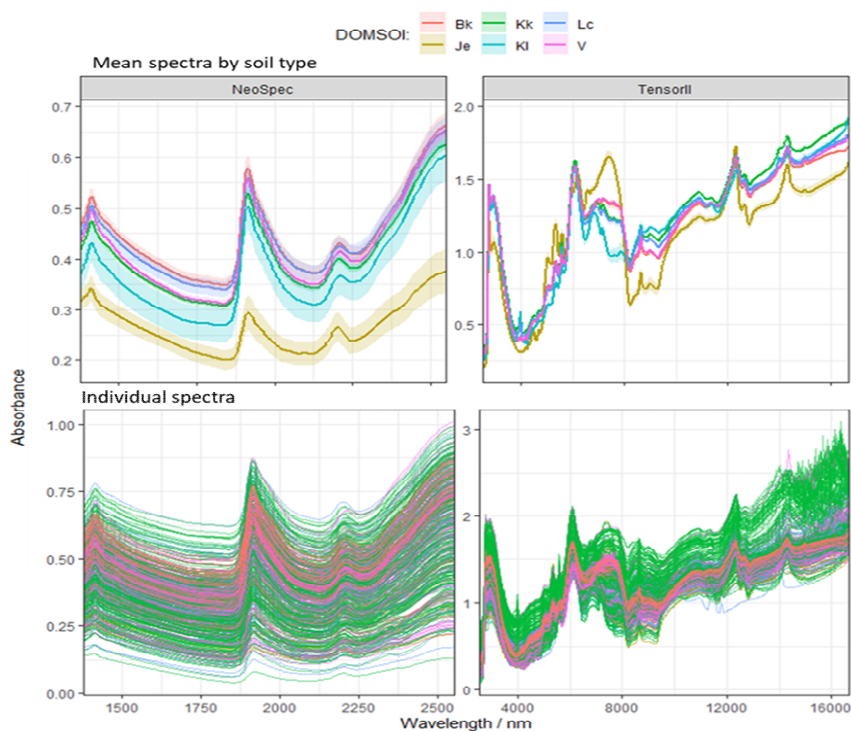


Figure 4.2: Mean with standard error (top row) and raw (bottom) absorbance spectra of the dominant soil types in the NIR and MIR region. Bk = Calcic Cambisols; Kk = Calcic Kastanozems; Lc = Chromic Luvisols; Je = Eutric Gleysols; Kl = Luvic Kastanozems; V = Vertisols.

4.3.4 Framework for calibration sample selection for modelling and performance evaluation

We evaluated four different methodologies for calibration sample selections and analysed their impact on the prediction performance, following the workflow illustrated in Figure 4.3. These were use of 1) the entire MSSL sample collection (MSSL), 2) subset by first four PC scores (weighted by their variance explained) variogram (Spatial selection), 3) based on spectra similarity (MBL) and 4) based on clustering using environmental covariates and FAO soil units (covariate clustering). The details of each methodology are described in section 4.3.5, 4.3.6, 4.3.7, and 4.3.8, respectively. We used 524 calibration sample entries for the library to develop the predictive model and 75 validation set samples. This was done for two spectral sources, i) MIR and ii) NIR. We also considered prediction of soil properties using the reflectance spectra in the NIR using the entire MSSL dataset.

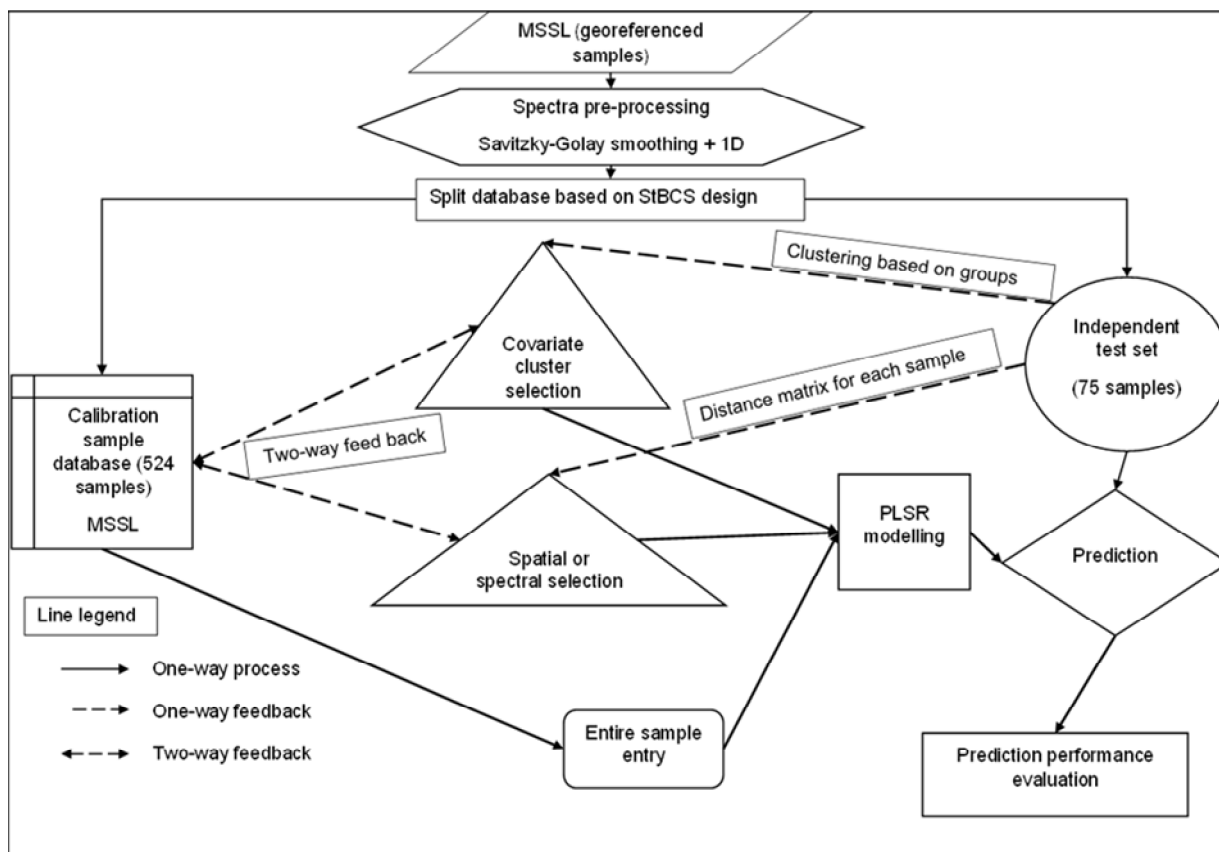


Figure 4.3: Schematic demonstrating the sample selection processes for soil property prediction from a soil spectral library of MIR and NIR region. Tensor-II; Neospec = Neospectra Saucer. StBCS design -> stratified balanced coverage sampling design.

4.3.5 The entire MSSL with Partial Least Square regression (PLSR)

The first methodology is the reference method in soil spectroscopy, generalising and combining features from principal component analysis and multiple regression to develop the 'global' PLSR predictive model. From a systematic review and meta-analysis, Ahmadi et al., (2021) reported that the most employed machine learning methods was the PLSR family which accounted for around 70% of the studies in soil spectroscopy. It establishes a linear relationship between a set of latent variables extracted from the spectral matrix (X) while maximizing the covariance with the soil property (y). As such, it eliminates the multicollinearity problem that would arise by regressing y on X . For more details see Forina et al. (2007). We selected the number of latent variables, or components to retain, by

the minimum RMSE from a 10-fold cross-validation and considered a maximum of 15 components. We used the `pls` function in the `pls` R package (Liland et al., 2023). This process generates a single model (per soil property) that was used to predict each sample in the validation set.

4.3.6 Spatial autocorrelation of spectra PC scores

The second methodology selected calibration samples under the assumption that soil variation is spatially autocorrelated. A principal component analysis (PCA) was computed on the soil spectra. A single vector representing the majority of the variation in the spectra space was calculated by weighting the first four principal components according to the variance they explained. We calculated the experimental variograms by the method of moments with (4-1), considering a cutoff distance of 120 km and a lag distance of 10 km. Additionally, we computed the experimental variograms for a longer cut-off distance (200 km) to investigate any further variogram characteristics that might be present in the MSSL dataset. Elevation, slope percentage, mean rainfall and temperature were used as covariates to account for long-distance trends.

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2m(\mathbf{h})} \sum_{i=1}^{m(\mathbf{h})} [Z(\mathbf{s}_i) - Z(\mathbf{s}_i + \mathbf{h})]^2 \quad (4-1)$$

where $\hat{\gamma}(\mathbf{h})$ is the empirical semivariogram value at lag $\mathbf{h}$, $m(\mathbf{h})$ is the number of paired comparisons separated by $\mathbf{h}$, $Z(\mathbf{s}_i)$ and $Z(\mathbf{s}_i + \mathbf{h})$ are the values of the weighted PC vectors at locations $\mathbf{s}_i$ and $\mathbf{s}_i + \mathbf{h}$, respectively.

After evaluating the residuals of various covariance models, including Exponential and Matern, we fitted a spherical covariance model ($\gamma(\mathbf{h})$) for characterising spatial autocorrelation and parameter estimation using equation (4-2).

$$\gamma(\mathbf{h}) = \begin{cases} 0 & \text{if } h = 0 \\ c_0 + c \left[\frac{3h}{2a} - \left(\frac{h^3}{2a^3} \right) \right] & \text{if } 0 < h \leq a \\ c_0 + c, & \text{if } h > a \end{cases} \quad (4-2)$$

Where c_0 is the nugget, c is the sill, and a is the effective range.

Restricted maximum likelihood fitting led to non-unique solutions depending on the initial range parameter. Hence, we used the '*fit.variogram*' function from the gstat package in R (Pebesma and Graeler, 2023), which uses weights based on the number of point pairs and the lag distance, *i.e.* smaller lag distances and bins with more data get more weight. Thus, the spatial dependency analysis determined the distance for which the weighted PC was autocorrelated (the effective range). We employed subsampling with spatial constraints to select a set of the MSSL dataset for each prediction location, based on a local neighbourhood determined by the effective range estimate (Pebesma and Wesseling, 1998).

Following this, the predictions of soil properties $\hat{z}(s_n)$ from spectral dataset obtained by a specific soil scanning instrument j for new sample at any given location s_n were derived using the following equation (4-3):

$$\hat{z}(s_n)_j = \text{PLSR}\{z(s_i)_j, \gamma(h)_j, s_n \mid s \in \mathbb{A}\} \quad (4-3)$$

Where the PLSR denotes the Partial Least Square Regression function, $z(s_i)$ are the input point data values within area $\mathbb{A}$, $\gamma(h)$ is the semivariogram model defining the spatial autocorrelation of the weighted PC vector at a defined cut-off distance, s_n is the location of the new sample, $s \in \mathbb{A}$ indicates that the points are within the spatial area of $\mathbb{A}$.

This process generates a separate PLSR model (using a specific subset of calibration samples) for every sample in the validation set.

4.3.7 MBL spectra similarity analysis

In the third methodology, we followed the approach of Summerauer et al., (2021) with slight modification to enhance calibration sample selection based on spectral similarity. In this approach, calibration samples were selected for each target sample based on spectral similarity. The process involved the following steps:

Moving Windows Correlation Dissimilarity: We calculated the dissimilarity between spectra using a moving window correlation method. This was done for a range of window sizes from 11 to 151 in steps of 10. For each window size w , the correlation dissimilarity matrix was computed using equation (4-4).

$$dissimilarity_{ij} = 1 - cor(X_i^w, X_j^w) \quad (4-4)$$

Where $cor(X_i^w, X_j^w)$ is the correlation between the spectral windows of size w for sample i and j . The optimal window size was determined by calculating the root mean squared errors (RMSE) between the nearest neighbours for each target soil property, selecting the window size that yielded the lowest RMSE (Figure C-1) using equation (4-5).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4-5)$$

Where y_i is the actual value, and $\hat{y}_i$ is the predicted value.

Selection of Nearest Neighbours and weighted Average PLSR Model: For each target sample, we selected the nearest neighbours (NN_i) with a dissimilarity below a threshold of 0.1 with equation (4-6).

$$NN_i = \{j : dissimilarity_{ij} < 0.1\} \quad (4-6)$$

And a weighted average PLSR model was fitted (Shenk et al., 1997) for each target sample. The weights for each component k of the PLSR model were calculated based on the reconstruction error E_k of the PLS loadings for the target samples using equation (4-7).

$$w_k = \frac{1}{E_k} \quad (4-7)$$

These weights were in turn used to average all PLSR models across the range of components considered (minimum 5 and maximum of 15 components) using equation (4-8).

$$\hat{y} = \sum_{k=1}^k w_k \hat{y}_k \quad (4-8)$$

Where $\hat{y}_k$ is the prediction from the $k - th$ component.

Optimum Number of Neighbours and Execution of MBL Routine: Lastly, the final number of nearest neighbours was determined by nearest neighbour cross-validation (Ramirez-Lopez et al., 2013) testing a range from 20 to 100 in interval of 10. The number that minimized the cross-validation RMSE was selected. To execute the MBL (Moving Block Local) routine, the ‘*resemble*’ package in R (Ramirez-Lopez et al., 2022) was used.

4.3.8 Cluster grouping using environmental covariates and FAO soil units

For the fourth method, calibration samples were selected based on the similarity in environmental conditions and FAO soil types. We used five environmental variables (mean annual rainfall, mean annual temperature, slope percent, elevation, and FAO soil types) to cluster the entire dataset ($n = 599$). All numerical variables were scaled to unit variance, and the FAO soil types were converted into binary variables using one-hot encoding. The one-hot encoding process converted each category of the FAO soil types into a separate binary (0 or 1) variable. If there are n unique categories $\{c_1, c_2 \dots c_n\}$ in the FAO soil types (FAO_{Soil}), the one-hot encoding process created n new binary variables (FAO_{c_i}) as follow (4-9):

$$FAO_{c_i} = \begin{cases} 1 & \text{if } FAO_{Soil} = c_i \\ 0 & \text{otherwise} \end{cases} \quad \text{for } i = 1, 2, \dots, n \quad (4-9)$$

where x_i is a data point belonging to the cluster C_k , and μ_k is the mean value of the points assigned to the cluster C_k .

After converting the FAO soil types, the dataset was expanded to include these binary variables. We checked for multicollinearity among the numerical variables using a correlation matrix and identify highly correlated variables (correlation >

0.9), which were removed to avoid multicollinearity issues in the clustering process.

We then applied the k-prototypes clustering algorithm to the scaled dataset with the original numerical variables and newly created binary variables. The k-prototype algorithm, which is suitable for datasets with mixed numerical and categorical variables, was employed to minimize the total within-cluster variation. The total within-cluster variation (total within-cluster sum of squares) was calculated from the sum of squared Euclidean distances between variables and the corresponding centroid following (4-10).

$$W(C_k) = \sum_{x_i \in C_k} (x_i - \mu_k)^2 \quad (4-10)$$

where x_i is a data point belonging to the cluster C_k , and μ_k is the mean value of the points assigned to the cluster C_k .

To determine the optimum number of clusters, we ran the k -prototype algorithm for a range of clusters and repeated this process 10 times for each cluster count to ensure robustness. The choice of the optimum number of clusters was based on evaluating the total within-cluster sum squares, selecting the number of clusters that minimised this measure while maintaining the minimum number of calibration sample for a PLSR model. Each observation x_i was then assigned to the cluster such that the sum of squares (SS) distance of the observation to the cluster centers μ_k was a minimum. The total within-cluster sum of squares, which measures the compactness (i.e. goodness) of the clustering for the entire dataset, is given by equation (4-11).

$$tot.withinss = \sum_{k=1}^k W(C_k) = \sum_{k=1}^k \sum_{x_i \in C_k} (x_i - \mu_k)^2 \quad (4-11)$$

Where $tot.withinss$ is the sum of the within-cluster variations for all clusters.

We used the '*kproto*' function from the '*clusteMixType*' R package and considered tuning the lambda parameter to adjust the weighting between numerical and categorical distances as needed. In addition, we evaluated the stability of the

clusters using silhouette scores to confirm the robustness of the chosen cluster number. The clustering resulted in 5 distinct clusters, and training and validation sets were assigned accordingly (Table C.2). Calibration regression was then based on cluster associations and the PLSR as described above (4.3.5). For each validation sample, one of 5 possible PLSR models was used according to the cluster assignment.

4.3.9 Accuracy assessment

Ten-fold-leave-group out cross validation was used to optimise PLSR model. The models' performance was further assessed using the measures a) Lin's concordance correlation coefficient (CCC) (Lin, 1989), b) the coefficient of determination (R^2), c) the root mean square error of prediction, d) the ratio of performance to the interquartile range (RPIQ) (Bellon-Maurel et al., 2010), and e) bias according to the equations below using the test set:

$$CCC = \frac{2rs_x s_y}{s_x^2 + s_y^2 + (\bar{x} - \bar{y})^2} \quad (4-12)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum_{i=1}^n (x_i - \bar{x}_i)^2} \quad (4-13)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2} \quad (4-14)$$

$$RPIQ = \frac{(Q_3 - Q_1)}{RMSE} \quad (4-15)$$

$$Bias = \frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2 \quad (4-16)$$

Where x_i and y_i are the observed and predicted value at point i ; $\bar{x}$ and $\bar{y}$ are the mean of the observed and predicted values; r is the correlation coefficient between the observed and predicted values, s_x and s_y are the variance of observed and predicted values, Q_1 and Q_3 are the 25th and 75th percentiles, respectively.

4.4 Results

4.4.1 Baseline soil properties and their correlations with spectra preprocessing

Descriptive statistics are presented in Table 4.1 for the entire MSSL soil samples for both the training and validation sets. The mean and median values of SOC across all samples are moderate, though the lowest and highest values fell into extremely low and very high categories, respectively (Agvise Laboratories, 2019). Higher SOC is typically linked to better soil health and nutrient availability. Most of the soil samples exhibited a mildly to moderately alkaline pH, with some reaching the strongly alkaline range (Hazelton and Murphy, 2019), a characteristic often found in semiarid climates with limited leaching. The CEC of the soils was generally high, indicating good availability of exchangeable cations, which helps regulate soil pH and its response to inputs (Agvise Laboratories, 2019). In contrast, low CEC is usually associated with lower resistance to change in soil chemistry. Similarly, exchangeable K and Mg were found in moderate to high availability categories, likely due to the high CEC of soils and low precipitation levels. The distribution of clay content across all samples followed a normal distribution in the soil textural diagram, according to the International Union of Soil Sciences (IUSS) soil textural classification system (Figure C-2).

The distribution of soil properties was largely similar between the MSSL and validation sets, though the range was generally higher for the former. While Olsen P, P_M3, Exch.K, DTPA.Zn, DTPA.Mn and Tot_IC showed right-skewed distribution, CEC and pH were left-skewed. In contrast, Exch.Mg, TotN and SOC exhibited normal distributions (Figure C-3 and Figure C-4). Data transformations were applied due to the skewed distribution of the data for some of the soil properties. Log transformation was applied for those positively skewed datasets viz., OlsenP, DTPA.Zn, DTPA.Mn, P-m3, and Exch.Mg, while a square root transformation was used for Exch.K. For negatively skewed data of pH and CEC, reflection and log transformation were carried out.

Table 4.1: Descriptive statistics of soil property in the training and validation sets.

Soil property	Sample set	N	Min	Median	Mean	Max	Std	Skewness
CEC (cmol (+) kg-1 of soil)	Validation	75	4.2	31.1	30.19	41.8	6.51	-0.42
	Train	524	0.6	31	29.46	41.9	7.6	-0.61
Clay (%)	Validation	75	8.46	42.4	40.6	65.6	14.6	-0.3
	Train	524	2.61	40.9	39.9	71.7	15.1	-0.19
DTPA.Mn (mg kg-1)	Validation	75	0.5	5.5	8.92	87.2	11.85	0.87
	Train	524	0.5	5.2	8.74	87.7	10.15	1.05
DTPA.Zn (mg kg-1)	Validation	75	0.1	0.7	0.75	2.6	0.52	0.29
	Train	524	0.1	0.6	0.78	7.2	0.74	0.73
Exch.K (cmol kg-1 of soil)	Validation	75	0.38	0.94	1.13	6.86	0.88	0.65
	Train	524	0.04	0.95	1.12	10.23	0.73	0.7
Exch.Mg (cmol kg-1 soil)	Validation	75	0.63	8.99	10.25	49.87	8.09	0.47
	Train	524	0.3	9.05	9.3	56.28	6.51	0.12
OlsenP (ppm)	Validation	75	0.72	9.37	14.41	96.22	14.55	1.04
	Train	524	0.52	10.08	15.49	130.27	16.93	0.96
P_M3 (ppm)	Validation	75	3	29	47.91	301	54.37	1.04
	Train	524	0	27	49.3	381	60.33	1.11
pH	Validation	75	6.01	8.06	7.98	8.68	0.42	-0.57
	Train	524	5.45	8.08	7.97	8.94	0.48	-0.69
SOC (%)	Validation	75	0.35	1.19	1.3	2.74	0.48	0.69
	Train	524	0.09	1.16	1.22	4.29	0.46	0.39
Tot_IC (%)	Validation	75	0	0.5	0.82	3.87	0.95	1.01
	Train	524	0	0.21	0.99	9.16	1.39	1.68
Tot_N (%)	Validation	75	0.03	0.13	0.14	0.29	0.05	0.6
	Train	524	0.03	0.13	0.13	0.42	0.05	0

The correlation boxplot of spectral pre-processing methods against soil property attributes (Figure 4.4; Wang et al., 2018) highlighted the critical role of spectral pre-processing in refining the selection of explanatory variables for modelling via dimension reduction algorithms. The box plots effectively illustrated how different pre-processing techniques influence the identification of relevant spectral

variables, which are shown at the tails of the plots, while irrelevant variables were concentrated around the centre, closer to zero correlation values.

The use of pre-processing methods such as SG + 1D and SG + gap 1D pre-processing demonstrated superior performance in generating clearer correlations between the spectral data and soil properties across both soil scanning instruments. This suggested that these pre-processing techniques enhanced the ability to distinguish relevant spectral features for modelling purposes. However, the analysis also revealed that certain soil properties, such as Exch.K exhibited the least correlation values on either side of the boxplot, indicating that such soil properties might be less responsive to spectral pre-processing or that its spectral signature is more challenging to capture accurately. This insight is crucial for understanding the limitations and strengths of different pre-processing methods in relation to specific soil properties and can guide the selection of appropriate spectral pre-processing techniques to improve model accuracy and predictive capability.

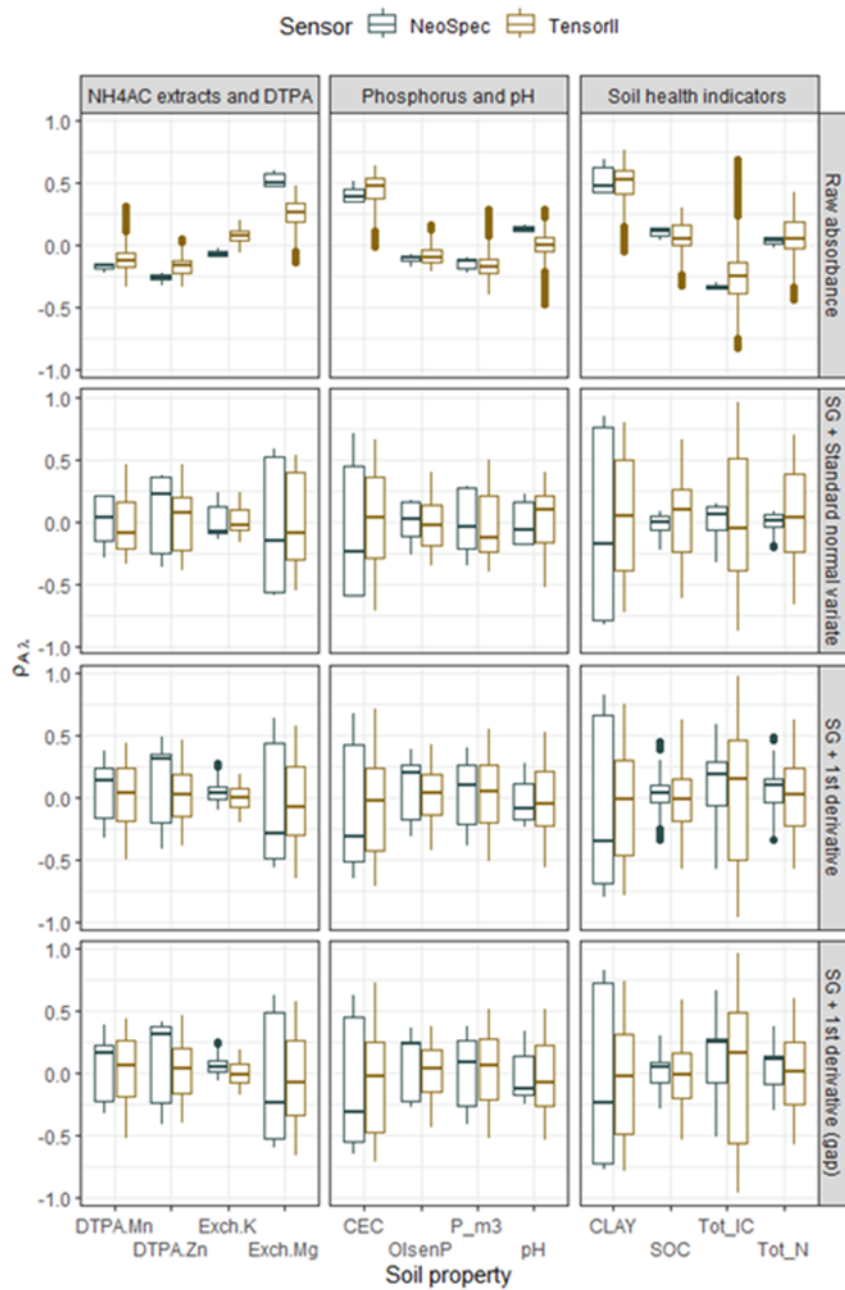


Figure 4.4: Correlation boxplot of each soil properties with spectra pre-processing techniques (SG -> Savitzsky-Golay smoothing) at each wavelength by the scanning instrumentation (Neospec -> Neospectra, Tensor-II -> Bruker Tensor 27). The lower and upper limits of each box are the 25th and 75th percentiles, the tails show the correlation range, the dots represent outliers and the horizontal line in each box represents the median.

4.4.2 Spatial autocorrelation and scale dependency of spectra PC scores

The first four PC components, used as a PC vector, explained 80% and 99% of the variation in the spectra for the MIR and NIR spectra, respectively (Figure C-5 and Figure C-6). We assumed that these PC components could reliably explain most of the spatial autocorrelation for the spectra generated. The number of PC components explaining 99% of the variation for the MIR were eleven. The spatial autocorrelation of the PC vector is depicted in Figure 4.5 with cut-off distances of 120 and 200km. The distances were chosen to reflect the width and length of the distinct sampling areas (Figure 4.1). The separation distance between the two distinct sampling areas was around 300km.

The variograms exhibited distinct patterns of spatial autocorrelation among the PC vectors of the NIR and MIR spectral ranges. However, the general pattern of the variograms for each spectral range tended to be similar between the cut-off distances used. The nugget effect, random variation which could be caused by mainly undetectable experimental error and field variation within the minimum sampling space (Sun et al., 2003), remained similar among these cut-off distances. The distinct variations notably were 1) a shorter cut-off distance of 120km yielded better fit for spherical variograms and 2) the longer cut-off distance of 200km brought up a new developing variogram characteristics for both NIR and MIR spectra. This divergence could stem from the additional spatial autocorrelation of the PC vectors, given that soil spectra constitute a multiresolution dataset (Lark and Webster, 1999; Song et al., 2021) which is influenced by various soil chromophores exhibiting different spatial scales (Savvides et al., 2010).

With the shorter cut-off distance, the spatial autocorrelation for the PC vectors reached a sill parameter at a range of 117 and 67km for the MIR and NIR, respectively. These distance values for each spectra source indicated the bounding spatial extent within which the first four spectra PC scores of the soil samples were spatially autocorrelated. The spherical variogram fit showed the presence of a strong spatial autocorrelation for both NIR and MIR spectra PC

vectors with a steep rise to the sill. This means that the soil sample spectra are strongly correlated over the effective range.

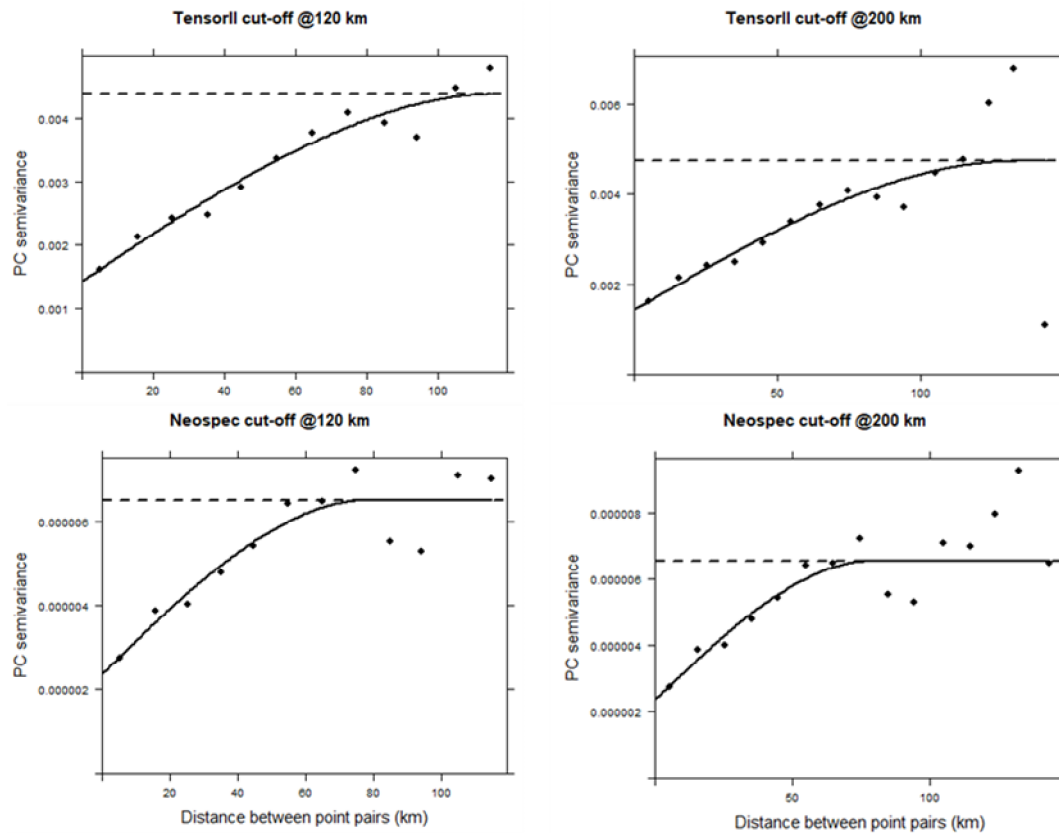


Figure 4.5: Semivariograms of PC vectors fitted with a spherical model (solid line) from the respective soil scanning instruments (Neospec -> Neospectra for NIR spectra; Tensor-II -> Bruker Tensor 27 for MIR spectra) with cut-off distances of 120 and 200 km. The dashed line is the sill.

4.4.3 Prediction performance among sample selection processes

4.4.3.1 Use of entire MSSL sample entry (MSSL)

The PLSR models, created for each soil property using the entire MSSL dataset (524 samples), was assessed based on their effectiveness in predicting the respective soil properties from the NIR and MIR spectra. The models demonstrated excellent predictive accuracy for clay and Tot_IC, as evidenced by high CCC (>0.8) and high RPIQ (>2) values for both NIR and MIR spectra (Figure 4.6). This suggested that these soil properties have strong, consistent spectral

signatures across the samples of the entire MSSL library, making them well-suited for prediction using the entire dataset with a PLSR models.

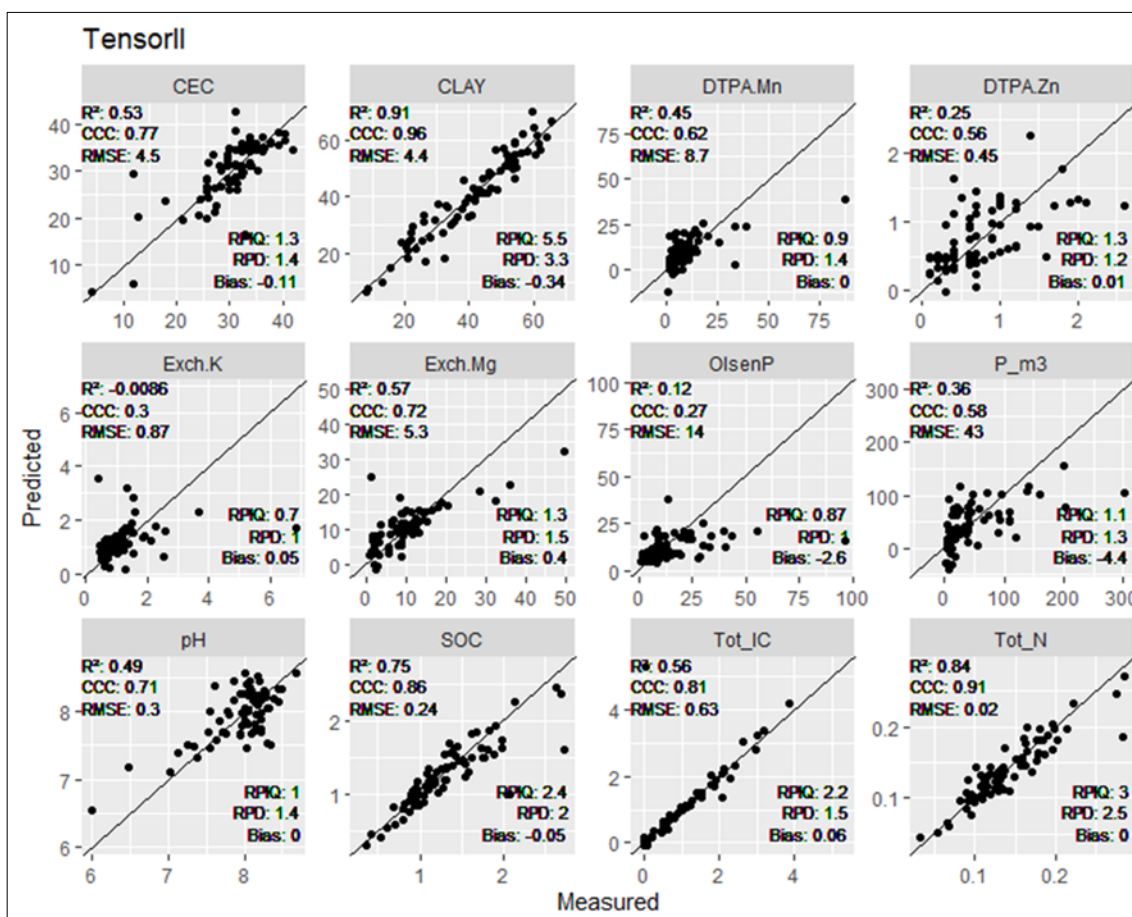
For SOC and Tot_N, the MIR spectra provided excellent predictive agreement (CCC = 0.86 and 0.91, respectively) while the NIR spectra showed good but lower agreement (CCC = 0.69 and 0.73, respectively). The discrepancy between the MIR and NIR performance could be attributed to the absence of critical absorbance ranges (500 – 1350nm) in the NIR spectra, which includes important C-H and O-H absorption regions (Stenberg, Rossel, Mouazen, Wetterlind, et al., 2010). These regions are essential for accurately capturing the spectral characteristics associated with organic carbon and nitrogen, thus explaining the superior performance of MIR in these cases. We also included the prediction performance using the reflectance spectra of NIR for further comparison, applying SG-1D preprocessing (see Figure C-7). However, the results showed similar predictive accuracy to those obtained with absorbance spectra in most soil properties.

The predictive performance for soil properties like CEC, DTPA.Zn, Exch.Mg, and P_M3 ranged from moderate to satisfactory when using both MIR and NIR spectra. However, the models showed low prediction accuracy for OlsenP and Exch.K across both spectral regions, as well as DTPA.Mn and pH specifically in the NIR spectra, with CCC values below 0.5 and RPIQ values under 1. Notably, NIR spectra outperformed MIR in predicting OlsenP and Exch.K. These low performance indicators suggest that these properties may have weaker or more variable spectral signatures, limiting accurate prediction with PLSR models (Dangal et al., 2019).

Interestingly, the poor prediction performance for certain soil properties correlated with the characteristics observed in the correlation boxplot between spectral pre-processing and conventional soil analysis values (Figure 4.4). The boxplot, which showed the distribution of correlations between spectral data and soil properties, could serve as a qualitative indicator of prediction performance. Specifically, the variability in box and tail length for properties like CEC, DTPA.Mn, pH, Exch.K, and OlsenP highlights the challenges in predicting these

parameters accurately across different soil spectra ranges using the MSSL dataset.

Despite these challenges, the study demonstrated that PLSR models, particularly those using MIR spectra, can effectively predict a range of soil properties. The higher RPIQ values (>1.4) observed for Clay, SOC, Tot_IC and Tot_N in the MIR region underscores the potential of both spectra regions in delivering more accurate predictions of these soil properties. The stronger performance of MIR compared to NIR for certain soil properties can be attributed to the richer spectral information in the MIR range. These findings are consistent with existing literature, which highlights the superior predictive ability of MIR (Johnson et al., 2019).



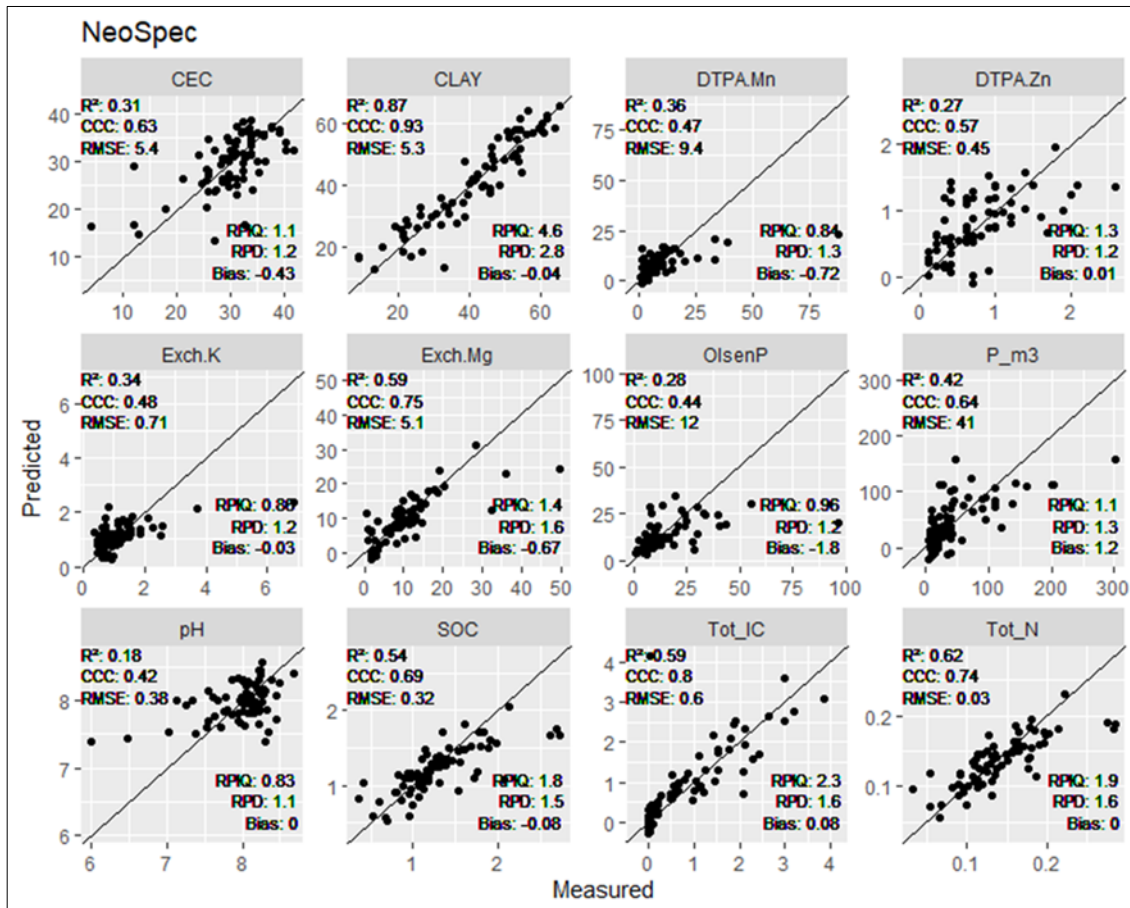


Figure 4.6: Prediction performance for the soil properties and soil scanning instruments (Neospec -> Neospectra for NIR spectra; Tensor-II -> Bruker Tensor 27 for MIR spectra) using the entire MSSL. The statistics shown are the R^2 , concordance correlation coefficient (CCC), the root mean square error (RMSE), Bias, Ratio of standard error of Performance to standard Deviation (RPD) and the ratio of performance to interquartile range (RPIQ).

4.4.3.2 Performance comparison between sample entry selection methods and the entire MSSL

The selection of calibration samples for the PLSR model produced varying prediction performances across different soil property-spectral region combinations, with both positive and negative impacts (Table 4.2). The scatter plots illustrating the spread of predicted values against the 1:1 line for each sample selection methods are also provided in Figure C-8, Figure C-9 and Figure C-10. For the MIR spectra, the spatial calibration sample selection improved the CCC values notably for OlsenP (by 41.3%) and P_M3 (by 8.5%). Similarly, for

the NIR spectra, improvements were obtained in CEC (by 25.6%), pH (by 13.0%), Tot_N (by 10.6%), Exch.Mg (by 9.6%) and SOC (5.6%) compared to using the entire MSSSL dataset. However, the CCC values decreased for DTPA.Mn (-21.4), Exch.K (-13.7%), pH (by -8.3%) and DTPA.Zn (-7.1) for the MIR spectra, and for Exch.K (by -23.2%), OlsenP (by -13.4%), DTPA.Mn (by -9.5%) DTPA.Zn (by -7.3%) and P_m3 (-6.7%) for the NIR spectra. Despite these variations, the spatial calibration sample selection generally provided comparable prediction performance for other soil property-spectral range combinations, even with smaller sample size (<175 samples) compared to the entire MSSSL dataset.

On the other hand, the MBL approach enhanced the CCC value for OlsenP (by 85.2%), P_m3 (by 21.2%) pH (by 12.2%), and Exch.Mg (by 7.7%) in the MIR spectra, and for pH (by 18.7%), Tot_N (by 13.5%), OlsenP (by 9.9%), SOC (by 8.5%), DTPA.Zn (by 8.1%), and Exch.Mg (by 6.3%) in the NIR spectra. However, it also resulted in decreased CCC values for DTPA.Zn (by -15.1%), Exch.K (by -10.9%), CEC (by -7.6%), and DTPA.Mn (by -7.0%) in the MIR spectra, as well as for Exch.K (by -18.9%) and DTPA.Mn (by -6.2%) in the NIR spectra. Overall, the spectra similarity selection demonstrated an improvement in prediction performance compared to using the entire MSSSL dataset aligning with findings from Gholizadeh et al. (2018) and Moloney et al. (2023). Conversely, the cluster-based calibration sample selection improved the CCC values only for OlsenP (by 9.1%); but led to reduced prediction performance for DTPA.Mn (by -27.5%), DTPA.Zn (by -21.9%), Exch.K (-20.9%), pH (by -15.8%), CEC (by -7.0), and P_m3 (by -6.3%) in the MIR spectra. Similarly, in the NIR spectra, it improved prediction performance for CEC (by 28.2%) and Exch.Mg (by 7.3%) while reducing it for DTPA.Mn (by -34.4%), OlsenP (by -29.1%), Exch.K (by -19.0%), pH (by -17.7%) and P_m3 (by -5.5%). These findings partially concur with those of Wijewardane et al. (2018), who reported that clustering with environmental covariates did not improve prediction performance for Exch.K and P_m3 using MIR spectra. Nonetheless, the cluster-based selection yielded prediction outcomes comparable to using the entire MSSSL dataset for other soil property – spectral range combinations with a PLSR model developed with only 48 samples.

Table 4.2: Prediction performance for the soil properties using calibration sample selection approaches in the NIR and MIR spectral ranges.

Soil properties	Wavelength Region	Spatial					MBL					Cluster				
		CCC	RMSE	RPIQ	Bias	*CCC deviation (%)	CCC	RMSE	RPIQ	Bias	*CCC deviation (%)	CCC	RMSE	RPIQ	Bias	*CCC deviation (%)
CEC	MIR	0.77	4.50	1.30	0.0	0.8%	0.71	5.00	1.20	-0.9	-7.6%	0.71	4.90	1.20	0.1	-7.0%
	NIR	0.79	4.10	1.50	0.1	25.6%	0.61	5.90	1.00	-1.2	-3.1%	0.80	3.90	1.50	0.0	28.2%
Clay	MIR	0.95	4.60	5.30	-0.1	-0.6%	0.95	4.70	5.20	0.2	-0.5%	0.94	5.20	4.70	-0.2	-1.8%
	NIR	0.94	5.10	4.80	0.2	0.8%	0.91	6.10	4.00	-0.4	-2.1%	0.92	5.70	4.30	-0.4	-1.2%
Mn	MIR	0.49	9.90	0.80	-0.1	-21.4%	0.58	9.50	0.83	-1.1	-7.0%	0.45	10.00	0.76	-0.8	-27.5%
	NIR	0.43	9.90	0.80	-0.8	-9.5%	0.44	11.00	0.75	-0.3	-6.2%	0.31	11.00	0.72	-0.9	-34.4%
Zn	MIR	0.52	0.49	1.20	0.0	-7.1%	0.48	0.56	1.10	0.1	-15.1%	0.44	0.53	1.10	0.0	-21.9%
	NIR	0.53	0.46	1.30	-0.1	-7.3%	0.62	0.52	1.10	0.1	8.1%	0.59	0.45	1.30	0.0	3.2%
K	MIR	0.26	0.87	0.70	0.1	-13.7%	0.27	0.88	0.69	0.0	-10.9%	0.24	0.89	0.69	0.0	-20.9%
	NIR	0.37	0.79	0.77	-0.1	-23.2%	0.39	0.79	0.77	0.0	-18.9%	0.39	0.75	0.81	0.0	-19.0%
Mg	MIR	0.73	5.40	1.30	0.5	1.3%	0.77	4.70	1.50	0.0	7.7%	0.74	5.20	1.30	0.2	2.9%
	NIR	0.82	4.60	1.50	-0.4	9.6%	0.79	4.70	1.50	-0.5	6.3%	0.80	4.60	1.50	-0.8	7.3%
OlsenP	MIR	0.39	13.00	0.93	-2.1	41.3%	0.51	12.00	1.00	-1.5	85.2%	0.30	16.00	0.76	-1.5	9.1%
	NIR	0.38	13.00	0.92	-1.8	-13.4%	0.48	13.00	0.93	-1.2	9.9%	0.31	14.00	0.84	-1.9	-29.1%
P_m3	MIR	0.63	42.00	1.10	1.1	8.5%	0.70	38.00	1.20	1.3	21.2%	0.54	48.00	0.94	-1.6	-6.3%
	NIR	0.59	43.00	1.10	2.8	-6.7%	0.62	45.00	1.00	1.1	-2.7%	0.60	43.00	1.10	-4.3	-5.5%
pH	MIR	0.65	0.33	0.96	0.0	-8.3%	0.80	0.26	1.20	0.0	12.2%	0.60	0.35	0.90	0.0	-15.8%
	NIR	0.47	0.38	0.83	0.0	13.0%	0.49	0.39	0.82	0.0	18.7%	0.34	0.41	0.76	0.0	-17.7%
SOC	MIR	0.87	0.24	2.40	-0.1	0.7%	0.88	0.22	2.60	0.0	2.6%	0.84	0.26	2.20	-0.1	-2.3%
	NIR	0.72	0.31	1.80	-0.1	5.6%	0.74	0.31	1.80	-0.1	8.5%	0.68	0.33	1.70	0.0	-1.3%
Tot_IC	MIR	0.81	0.63	2.30	0.1	0.1%	0.82	0.61	2.30	0.1	1.3%	0.79	0.66	2.20	0.1	-2.4%
	NIR	0.80	0.61	2.30	0.0	0.0%	0.80	0.64	2.20	0.0	-0.9%	0.80	0.62	2.30	0.1	-0.6%
Tot_N	MIR	0.92	0.02	3.20	0.0	0.8%	0.91	0.02	3.00	0.0	-0.1%	0.88	0.02	2.70	0.0	-3.0%
	NIR	0.81	0.03	2.20	0.0	10.6%	0.83	0.03	2.10	0.0	13.5%	0.74	0.03	1.90	0.0	1.0%

*Percent CCC deviation for the respective sample entry selection methods in reference to the entire MSSL sample entry.

4.5 Discussion

4.5.1 Robustness of MSSL for soil property attribute predication?

The Moroccan rainfed wheat belt soils mostly had calcic diagnostic characteristics with a pH in alkaline to strongly alkaline ranges which is mostly due to the low annual precipitation, $<600\text{mm}$ per annum, for much of this region. A further decrease in precipitation amounts of 20% by 2050 and of 40% by 2080 is projected because of the expected climate change which could further the severity of recurring and severe drought combined with increasing demand for water (Mokhtar et al., 2022). Hence, soil input applications and soil management for agricultural productivity requires informed decision making with the support of reliable tools and methods. The database established comprised 599 soil samples collected across an area spanning approximately $12,200\text{ km}^2$, with 524 of these samples constituting the entire calibration set. This equated to an average of about 23 km^2 per sample.

Our sampling strategy and sampling points were determined and selected to acquire representative samples to develop the MSSL for soil property prediction in rainfed wheat growing regions of Morocco, optimized by a stratified balanced coverage sampling design. This approach enabled prioritization for those soil types which had the largest area coverage while addressing the spatial distribution and soil variation, and hence optimized the number of samples required for soil property prediction using soil spectroscopy (Potash et al., 2023). Consequently, we targeted the relevant geographic coverage of rainfed wheat productivity and spread the sampling locations according to the spatial variation expected from the environmental covariates (Janik et al., 1998).

Using the MSSL established within the outlined framework, the most significant agricultural soil properties were accurately predicted with satisfactory to excellent agreement, accompanied by low RMSEs for the respective soil variables. However, certain soil properties, which are spectrally inactive, exhibited unsatisfactory predictions in both the NIR and MIR spectral regions (Shepherd et al., 2022). Sarathjith et al. (2014) classified SOC and clay as spectrally active,

while pH, P, K, and Zn fall under spectrally inactive soil properties. These associations were also explained for conventional extraction procedures by e.g., Ciesielski and Sterckeman (1997b) who noted that the ammonium acetate determination method, a percolation extraction method buffering the pH of the extracts, led to significant variations in the proportion of negatively charged sites and particularly those bonded to organic matter. For instance, the relatively poor prediction performance for Exch.K in this study mirrored findings by Hu et al. (2013) for Missouri soils, and attributed to low SOC levels. Conversely, Jin et al. (2020) reported better performance for Anhui soils in China, although detailed soil types and property information were lacking for comparison. This may also indicate that variations in prediction performance for certain soil properties can emanate from differing conventional extraction methods followed, besides general soil variations. Notably, Exch.K determined with cobalt hexamine (Cohex) trichloride showed excellent agreement (CCC of 0.81-0.94) with soil spectra in both vis-NIR and MIR for East African soils (Asrat et al., 2023). These disparities in prediction performance were also observed for phosphorus, i.e., for OlsenP versus P-M3, reinforcing the notion that the choice of conventional soil extraction methods substantially affect the estimation of specific soil property attributes using soil spectroscopy methods. Therefore, it is hypothesized that discrepancies in quantifying soil properties contribute to variation in prediction performance when utilizing spectroscopic methods, underscoring the importance of selecting the appropriate conventional analysis method for specific purposes and research objectives for a defined soil system.

A general observation was that the prediction performance within this database also depended on the two spectral regions used. The MIR spectra demonstrated better prediction accuracies for soil properties associated with organic matter (SOC, Tot_N and CEC), while the NIR spectra yielded superior predictions for soil properties related to geological materials and/or land use (DTPA extracts of Zn & Mn, Exch.K, Exch.Mg, OlsenP and P_M3). Hence, this study and the database created will support access to spatially explicit soil information easily and in time to support the soil information gap for soil spectroscopic methods, especially for the studied region and represented soil types.

4.5.2 Suitability of spatial autocorrelation of spectra for calibration sample selection

The scales of spatial variation is dependent on the sampling design and sampling intensity in a given area for specific soil properties, and each of the soil properties can be influenced at varying extent by the combination of soil forming factors and soil forming processes in space and time (Bogunovic et al., 2017). With that in mind, the parameter selection in variogram modelling and fitting can influence the analysis and interpretation of the outcome (Vasu et al., 2017). This may impose challenges to compare results across such research fields as the data used might have varying spatial density (Ye et al., 2017). To better fit a variogram in any geographic extent, a minimum of about a 100 sampling points is required with a set interval for the scale of variation needed (Iqbal et al., 2005; Kerry and Oliver, 2007). This study used the spectral PC scores to understand the overall spatial autocorrelation as it was noted that the general soil characteristics were encoded in the soil spectral signature (Viscarra Rossel et al., 2016; Demattê et al., 2019). Hence, the spatial variability will dictate the extent of the general soil variation within the sampling extent and scale (Hengl, 2009). We anticipated that the spatial dependency analysis in the soil spectra space can be performed to inform calibration sample entry selection with consideration of the databases' spatial extent and subsequent representation of the domain variability within it. Consequently, the semi-variogram of the spectra PC vectors delineated the spatial variability and extent of the spatial autocorrelation across both spectral ranges. Longer spatial autocorrelation was observed for the MIR region (range = 117km) than the NIR region (range = 67km) which supports the findings reported by Qiu et al. (2013) as the relatively longer wavelengths tended to result in higher spatial variability. In parallel, the spatial autocorrelation of the PC vectors for both spectral ranges were strong which indicates the geostatistical approach could reliably be used to inform calibration sample entry selection for local sample prediction using a soil spectral database.

The spatial autocorrelations of the spectra PC scores were useful in identifying proper and reliable calibration samples per target sample location, improving prediction performances or equivalent to using the entire MSSL. The observed

spectral spatial autocorrelation likely reflects the influence of factors such as parent material, agroclimatic, soil management practices, and crop production patterns, which contribute to the spatial variability in specific soil property across the study region. Using the spatial selection approach, mainly the prediction performance was improved for soil property - spectral region combinations which had poor prediction results when using the entire MSSL samples, suggesting that this approach was more reliable. We therefore recommend further research on spatial dependency analysis of the specific responsible wavelength regions for a soil property and its spatial variation in any geographic extent.

4.5.3 Spectra similarity and clustering performance for for calibration sample selection

Spectra similarity analysis using MBL had been evaluated and anticipated to be a better way to subset calibration sample entries for local scales with samples within the domain area of a large SSL. Summerauer et al. (2021) found that despite the spectral similarity of a new sample to those in the large SSL database, reasonable predictions of soil properties were not guaranteed. Their analysis revealed that prediction performance using MBL was subpar when calibration sample entries were selected for the new samples which were geographically distant from the SSL database location. Conversely, when the new samples fell within the geographic coverage of the SSL database, the MBL yielded improved prediction performance. Our findings and others (i.e., Gholizadeh et al., 2016) suggested that the MBL approach could be preferred for improved soil property prediction of new samples when a calibration model was developed within the geographic area covered by the SSL, as compared with the use of the entire database.

On the other hand, it was suggested that predictions of soil properties based on spectra might show enhanced performance within a database organized into homogeneous units using environmental covariates or readily accessible soil characteristics (Ogen et al., 2019; Angelopoulou et al., 2020). Such approaches have been applied in other studies such as digital soil mapping (Dunkl and Ließ, 2022). We implemented clustering of the entire MSSL entries using five

environmental covariates, as elaborated in 4.3.8, into their respective homogeneous groupings. The prediction performances were only better or similar to the use of the entire MSSL for soil properties predicted satisfactorily or better with the use of the entire MSSL. It is vital to note that a wide range of soil property attributes within the homogenous clusters that address the variability within the local and/or homogeneous clusters might be required for a better outcome (Morais et al., 2018). Hence, the question which needs to be addressed in future research is to optimize homogenous clusters with the range and distribution of soil property attributes.

4.6 Conclusion

Ordinary predictive models, such as PLSR, developed from a large soil spectral library might have high uncertainty because the models learn the general variation to explain the target soil property from the diverse spectral database. Hence, appropriate sub-setting methods could help to use such huge and important databases for the purpose of generating soil information in real-time and at the required spatial resolution from the proximal spectra information acquired. As soil properties vary in space due to varying combination and extent of soil forming factors, their multiresolution spectral signature will vary in space parallelly. We characterized this spatial variation of soil spectral PC vectors and used it for sub-setting calibration samples from a large SSL to predict soil properties at farm level in Morocco. We compared this approach with the use of the entire MSSL, spectral similarity sub-setting (MBL) and clustering with environmental covariates for twelve agricultural important soil properties and two soil scanning instruments. The spatial autocorrelation of the PC vectors varied among soil spectral ranges and gave rise to varying lag distances. Our findings suggested that the spatial and spectra similarity-based calibration sample entry selection improved prediction for those soil properties (OlsenP, P_{m3}, Exch.K and DTPA.Mn) with low prediction performances when using the entire MSSL database. Otherwise, these sub-setting methods resulted in equivalent prediction performance for soil properties predicted well with the entire MSSL samples. In general, the spatial selection and MBL approaches demonstrated superior

performance in predicting most soil properties that were inadequately predicted by using the entire MSSL, indicating the effectiveness of these targeted calibration sample selections methods. This improvement likely stems from the fine-tuning of calibration samples, which effectively addresses local variations arising from the site-specific soil-forming factors and land management practices. The observed spatial variation in soil property and spectral characteristics may be influenced by the scale of observation, landscape heterogeneity and other contextual factors. On the other hand, this study suggests that utilizing clustering based on environmental covariates and soil units can be a viable strategy for sub-setting a large SSL for predicting local samples, particularly for soil properties that achieve satisfactory or better predictions with the entire MSSL. Hence, this study brought up use-case frameworks for large SSLs to predict soil properties per location which have geographic coordinates and spectral recordings. Our study employed PCA to explore the general spatial autocorrelation of spectra data, providing a comprehensive overview of spatial variability across samples. This may generalize the specific variation per soil property whereby PLSR latent variables could offer more targeted understanding of spatial autocorrelations related to specific soil attributes. We anticipated future research investigating the application of PLSR latent variables to enhance the spatial analysis of soil properties, addressing this gap and potentially improving the precision of soil property predictions.

4.7 Author contributions

T.G.A., S.M.H, R.S., F.K., R.C. and K.H. conceptualized the study, and acquired and administered project funding. A.H, and A.K carried out soil sampling, processing, and data generation processes. T.G.A., K.H. and T.B. contributed to data analyses, including method development, data visualizations and data interpretation. T.G.A wrote the primary draft of the paper with editing and reviewing inputs from all other authors.

4.8 Conflicts of Interest

The authors declare that they have no conflict of interest.

4.9 Acknowledgment

The authors want to thank Mohammed VI Polytechnic University (UM6P) and OCP group, for the technical and financial support.

4.10 References

Agvise Laboratories. (2019), Interpreting a Soil Test Report, available at: <https://www.agvise.com/wp-content/uploads/2020/09/guide-soil-interpretation-report-2019.pdf>.

Ahmadi, A., Emami, M., Daccache, A. and He, L. (2021), “Soil properties prediction for precision agriculture using visible and near-infrared spectroscopy: A systematic review and meta-analysis”, *Agronomy*, Vol. 11 No. 3, available at: <https://doi.org/10.3390/agronomy11030433>.

Angelopoulou, T., Balafoutis, A., Zalidis, G. and Bochtis, D. (2020), “From Laboratory to Proximal Sensing Spectroscopy for Soil Organic Carbon Estimation — A Review”, *Sustainability*, Vol. 12 No. 2, pp. 1–24.

Asrat, T.G., Sakrabani, R., Corstanje, R., Breure, T., Hassall, K.L., Kebede, F. and Haefele, S.M. (2023), “Spectral soil analysis for fertilizer recommendations by coupling with QUEFTS for maize in East Africa: A sensitivity analysis”, *Geoderma*, Elsevier B.V., Vol. 432 No. February, p. 116397.

Baumann, P., Helfenstein, A., Gubler, A., Keller, A., Meuli, R.G., Wächter, D., Lee, J., et al. (2017), “Developing the Swiss mid-infrared soil spectral library for local estimation and monitoring”, *European Journal of Soil Science*, Vol. 68 No. 6, pp. 840–852.

Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.M. and McBratney, A. (2010), “Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy”, *TrAC - Trends in Analytical Chemistry*, Elsevier Ltd, Vol. 29 No. 9, pp. 1073–1081.

- Bogunovic, I., Trevisani, S., Seput, M., Juzbasic, D. and Durdevic, B. (2017), "Short-range and regional spatial variability of soil chemical properties in an agroecosystem in eastern Croatia", *Catena*, Elsevier B.V., Vol. 154, pp. 50–62.
- Bremner, J.M. (1996), "Nitrogen Total", *Methods of Soil Analysis: Part 3 Chemical Methods*, pp. 1085–1121.
- Breure, T.S., Milne, A.E., Webster, R., Haefele, S.M., Hannam, J.A., Moreno-Rojas, S. and Corstanje, R. (2021), "Predicting the growth of lettuce from soil infrared reflectance spectra: the potential for crop management", *Precision Agriculture*, Springer US, Vol. 22 No. 1, pp. 226–248.
- Brodský, L., Klement, A., Penížek, V., Kodešová, R. and Borůvka, L. (2011), "Building Soil Spectral Library of the Czech Soils for Quantitative Digital Soil Mapping", *Soil and Water Research*, Vol. 6 No. 4, pp. 165–172.
- Ciesielski, H. and Sterckeman, T. (1997), "A comparison between three methods for the determination of cation exchange capacity and exchangeable cations in soils", *Agronomie*, Vol. 17 No. 1, pp. 9–16.
- Copernicus:Europe's eyes on Earth. (2020), "Land Cover Classification:Global Land Cover", Land Monitoring Service, available at: <https://lcviewer.vito.be/2018>.
- Dangal, S.R.S., Sanderman, J., Wills, S. and Ramirez-lopez, L. (2019), "Accurate and Precise Prediction of Soil Properties from a Large Mid-Infrared Spectral Library", pp. 1–23.
- Demattê, J.A.M., Dotto, A.C., Paiva, A.F.S., Sato, M. V., Dalmolin, R.S.D., de Araújo, M. do S.B., da Silva, E.B., et al. (2019), "The Brazilian Soil Spectral Library (BSSL): A general view, application and challenges", *Geoderma*, Elsevier, Vol. 354 No. October, p. 113793.
- Dotto, A.C., Demattê, J.A.M., Viscarra Rossel, R.A. and Rizzo, R. (2020), "Soil environment grouping system based on spectral, climate, and terrain data: a quantitative branch of soil series", *Soil*, Vol. 6 No. 1, pp. 163–177.

Dunkl, I. and Ließ, M. (2022), "On the benefits of clustering approaches in digital soil mapping: an application example concerning soil texture regionalization", *Soil*, Vol. 8 No. 2, pp. 541–558.

Forina, M., Lanteri, S. and Casale, M. (2007), "Multivariate calibration", *Journal of Chromatography A*, Vol. 1158 No. 1–2, pp. 61–93.

Geocradle. (2018), "Regional Soil Spectral Library", PILOT 2: Improved Food Security – Water Extremes Management (IFS), available at: <http://datahub.geocradle.eu/dataset/regional-soil-spectral-library>.

Gholizadeh, A., Borůvka, L., Saberioon, M. and Vašát, R. (2016), "A memory-based learning approach as compared to other data mining algorithms for the prediction of soil texture using diffuse reflectance spectra", *Remote Sensing*, Vol. 8 No. 4, available at: <https://doi.org/10.3390/rs8040341>.

Gholizadeh, A., Saberioon, M., Carmon, N., Boruvka, L. and Ben-Dor, E. (2018), "Examining the performance of PARACUDA-II data-mining engine versus selected techniques to model soil carbon from reflectance spectra", *Remote Sensing*, Vol. 10 No. 8, available at: <https://doi.org/10.3390/rs10081172>.

Grafström, A.L. (2018), "Balanced and Spatially Balanced Sampling", R Package Version 1.5.4. [Http://Www.Antongrafstrom.Se/Balancedsampling](http://Www.Antongrafstrom.Se/Balancedsampling).

Grunwald, S., Yu, C. and Xiong, X. (2018), "Transferability and Scalability of Soil Total Carbon Prediction Models in Florida, USA", *Pedosphere, Soil Science Society of China*, Vol. 28 No. 6, pp. 856–872.

Guerrero, C., Wetterlind, J., Stenberg, B., Mouazen, A.M., Gabarrón-Galeote, M.A., Ruiz-Sinoga, J.D., Zornoza, R., et al. (2016), "Do we really need large spectral libraries for local scale SOC assessment with NIR spectroscopy?", *Soil and Tillage Research, Elsevier B.V.*, Vol. 155, pp. 501–509.

Hazelton, P. and Murphy, B. (2016), *Interpreting Soil Test Results: What Do All the Numbers Mean?.*, CSIRO Publishing, available at: <https://doi.org/10.1071/9781486303977>.

Hengl, T. (2009), A Practical Guide to the Superintendent of Documents Classification System, Government Publications Review, Vol. 13, available at:[https://doi.org/10.1016/0277-9390\(86\)90082-8](https://doi.org/10.1016/0277-9390(86)90082-8).

Iqbal, J., Thomasson, J.A., Jenkins, J.N., Owens, P.R. and Whisler, F.D. (2005), "Spatial Variability Analysis of Soil Physical Properties of Alluvial Soils", Soil Science Society of America Journal, Vol. 69 No. 4, pp. 1338–1350.

Johnson, J.M., Vandamme, E., Senthilkumar, K., Sila, A., Shepherd, K.D. and Saito, K. (2019), "Near-infrared, mid-infrared or combined diffuse reflectance spectroscopy for assessing soil fertility in rice fields in sub-Saharan Africa", Geoderma, Vol. 354 No. May, available at:<https://doi.org/10.1016/j.geoderma.2019.06.043>.

Kennard, R.W. and Stone, L.A. (1969), "Computer Aided Design of Experiments", Technometrics, Vol. 11 No. 1, pp. 137–148.

Kerry, R. and Oliver, M.A. (2007), "Comparing sampling needs for variograms of soil properties computed by the method of moments and residual maximum likelihood", Geoderma, Vol. 140 No. 4, pp. 383–396.

Knadel, M., Deng, F., Thomsen, A. and Greve, M.H. (2012), "Development of a Danish national vis-NIR soil spectral library for soil organic carbon determination", Digital Soil Assessments and Beyond, pp. 403–408.

Lark, R.M. and Webster, R. (1999), "Analysis and elucidation of soil variation using wavelets", European Journal of Soil Science, Vol. 50 No. 2, pp. 185–206.

Liland, H., Mevik, B.-H., Wehrens, R. and Hiemstra, P. (2023), Package 'Pls'.

Lin, L.I.-K. (1989), "A Concordance Correlation Coefficient to Evaluate Reproducibility", Biometrics, Vol. 45 No. 1, p. 255.

Lindsay, W.L. and Norvell, W.A. (1972), "Development of a DTPA soil test for zinc, iron, manganese and copper.", Soil Sci.Soc.Am.Proc., Vol. 36, pp. 778–783.

Mamassi, A., Balaghi, R., Devkota, K.P., Bouras, H., El-Gharous, M. and Tychon, B. (2023), "Modeling genotype × environment × management interactions for a

sustainable intensification under rainfed wheat cropping system in Morocco”, Agriculture and Food Security, BioMed Central, Vol. 12 No. 1, pp. 1–23.

Mehlich, A. (1984), “Mehlich 3 soil test extractant: A modification of Mehlich 2 extractant.”, Communications in Soil Science and Plant Analysis, Vol. 15 No. 12, pp. 1409–1416.

Moloney, J.P., Malone, B.P., Karunaratne, S. and Stockmann, U. (2023), “Leveraging large soil spectral libraries for sensor-agnostic field condition predictions of several agronomically important soil properties”, Geoderma, Elsevier B.V., Vol. 439 No. March, p. 116651.

Morais, P., Inácio, E., Filho, F. and Rocha, M. (2018), “Digital Soil Mapping of Soil Properties in the ‘ Mar de Morros ’ Environment Using Spectral Data”, pp. 1–19.

Næs, T. (1987), “The design of calibration in near infra-red reflectance analysis by clustering”, Journal of Chemometrics.

Nawar, S. and Mouazen, A.M. (2018), “Optimal sample selection for measurement of soil organic carbon using on-line vis-NIR spectroscopy”, Computers and Electronics in Agriculture, Elsevier, Vol. 151 No. June, pp. 469–477.

Nelson, D.W. and Sommers, L.E. (1996), “Total Carbon , Organic Carbon , and Organic Matter”, Methods of Soil Analysis: Part 3 Chemical Methods, pp. 961–1010.

Ng, W., Husnain, Anggria, L., Siregar, A.F., Hartatik, W., Sulaeman, Y., Jones, E., et al. (2020), “Developing a soil spectral library using a low-cost NIR spectrometer for precision fertilization in Indonesia”, Geoderma Regional, Elsevier B.V., Vol. 22, p. e00319.

Ogen, Y., Zaluda, J., Francos, N., Goldshleger, N. and Ben-Dor, E. (2019), “Cluster-based spectral models for a robust assessment of soil properties”, Geoderma, Elsevier, Vol. 340 No. August 2018, pp. 175–184.

Olsen, S.. (1954), “Estimation of available phosphorus in soils by extraction with sodium bicarbonate”, US Department of Agriculture., Vol. 939, pp. 1–20.

Pebesma, E. and Graeler, B. (2023), Spatial and Spatio-Temporal Geostatistical Modelling, Prediction and Simulation: Package ‘Gstat’, available at: <https://github.com/r-spatial/gstat/issues/>.

Pebesma, E.J. and Wesseling, C.G. (1998), “Gstat: A program for geostatistical modelling, prediction and simulation”, Computers and Geosciences, Vol. 24 No. 1, pp. 17–31.

Qiu, B., Zeng, C. and Chen, C. (2013), “Comparative spatio-spectral heterogeneity analysis using multispectral and hyperspectral airborne images”, Geo-Spatial Information Science, Vol. 16 No. 2, pp. 83–90.

Ramirez-Lopez, L., Behrens, T., Schmidt, K., Rossel, R.A.V., Demattê, J.A.M. and Scholten, T. (2013), “Distance and similarity-search metrics for use with soil vis-NIR spectra”, Geoderma, Elsevier B.V., Vol. 199, pp. 43–53.

Ramirez-Lopez, L., Behrens, T., Schmidt, K., Stevens, A., Demattê, J.A.M. and Scholten, T. (2013), “The spectrum-based learner: A new local approach for modeling soil vis-NIR spectra of complex datasets”, Geoderma, Vol. 195–196, pp. 268–279.

Ramirez-lopez, L., Stevens, A., Orellano, C., Viscarra, R., Lobsey, C. and Wadoux, A. (2022), “resemble: Regression and similarity evaluation for memory-based learning in spectral chemometrics.”, R Package Version 1.2.2., Vol. 1 No. 2, pp. 1–47.

Sáiz-Abajo, M.J., Mevik, B.H., Segtnan, V.H. and Næs, T. (2005), “Ensemble methods and data augmentation by noise addition applied to the analysis of spectroscopic data”, Analytica Chimica Acta, Vol. 533 No. 2, pp. 147–159.

Savvides, A., Corstanje, R., Baxter, S.J., Rawlins, B.G. and Lark, R.M. (2010), “The relationship between diffuse spectral reflectance of the soil and its cation exchange capacity is scale-dependent”, Geoderma, Elsevier B.V., Vol. 154 No. 3–4, pp. 353–358.

Shenk, J.S., Westerhaus, M.O. and Berzaghi, P. (1997), "Investigation of a LOCAL calibration procedure for near infrared instruments", *Journal of Near Infrared Spectroscopy*, Vol. 5 No. 4, pp. 223–232.

Shepherd, K.D., Ferguson, R., Hoover, D., van Egmond, F., Sanderman, J. and Ge, Y. (2022), "A global soil spectral calibration library and estimation service", *Soil Security*, Elsevier Ltd, Vol. 7 No. April, p. 100061.

Shepherd, K.D. and Walsh, M.G. (2002), "Development of Reflectance Spectral Libraries for Characterization of Soil Properties", *Soil Science Society of America Journal*, Vol. 66 No. 3, pp. 988–998.

SoilSpec4GG. (2020), "Open Soil Spectral Library (OSSL)", available at: <https://doi.org/10.7560/320624-010>.

Song, Y., Shen, Z., Wu, P. and Rossel, R.A.V. (2021), "Wavelet geographically weighted regression for spectroscopic modelling of soil properties", *Scientific Reports*, Nature Publishing Group UK, No. 0123456789, pp. 1–11.

Spaargaren, O.C. and Batjes, N.H. (1995), Report on the Classification Into Fao-Unesco Soil Units of Profiles Selected From the Nrcs Pedon Data Base for Igbp-Dis.

Stenberg, B., Rossel, R.A.V., Mouazen, A.M., Wetterlind, J., Viscarra Rossel, R.A., Mouazen, A.M. and Wetterlind, J. (2010), "Visible and Near Infrared Spectroscopy in Soil Science", *Advances in Agronomy*, Vol. 107 No. C, pp. 163–215.

Stevens, A. and Lopez, L.R. (2022), An Introduction to the Prospectr Package.

Stevens, A., Nocita, M., Tóth, G., Montanarella, L. and van Wesemael, B. (2013), "Prediction of Soil Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy", *PLoS ONE*, Vol. 8 No. 6, available at: <https://doi.org/10.1371/journal.pone.0066409>.

Summerauer, L., Baumann, P., Ramirez-Lopez, L., Barthel, M., Bauters, M., Bukombe, B., Reichenbach, M., et al. (2021), "The central African soil spectral

library: A new soil infrared repository and a geographical prediction analysis”, *Soil*, Vol. 7 No. 2, pp. 693–715.

Sun, B., Zhou, S. and Zhao, Q. (2003), “Evaluation of spatial and temporal changes of soil quality based on geostatistical analysis in the hill region of subtropical China”, *Geoderma*, Vol. 115 No. 1–2, pp. 85–99.

Sun, Y., Yuan, M., Liu, X., Su, M., Wang, L., Zeng, Y., Zang, H., et al. (2021), “A sample selection method specific to unknown test samples for calibration and validation sets based on spectra similarity”, *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, Elsevier B.V., Vol. 258 No. 119870, pp. 1–8.

Thomas, C.L., Hernandez-Allica, J., Dunham, S.J., McGrath, S.P. and Haefele, S.M. (2021), “A comparison of soil texture measurements using mid-infrared spectroscopy (MIRS) and laser diffraction analysis (LDA) in diverse soils”, *Scientific Reports*, Nature Publishing Group UK, Vol. 11 No. 1, pp. 1–12.

Vasu, D., Singh, S.K., Sahu, N., Tiwary, P., Chandran, P., Duraisami, V.P., Ramamurthy, V., et al. (2017), “Assessment of spatial variability of soil properties using geospatial techniques for farm level nutrient management”, *Soil and Tillage Research*, Elsevier B.V., Vol. 169, pp. 25–34.

Vestergaard, R.-J., Vasava, H.B., Aspinall, D., Chen, S., Gillespie, A., Adamchuk, V. and Biswas, A. (2021), “Evaluation of Optimized Preprocessing and Modeling”, *Sensors*, Vol. 21 No. 20, p. 6745.

Viscarra Rossel, R.A., Behrens, T., Ben-dor, E., Brown, D.J., Demattê, J.A.M., Shepherd, K.D., Shi, Z., et al. (2016), “A global spectral library to characterize the world’s soil”, *Earth-Science Reviews*, Vol. 155, pp. 198–230.

Wang, G., Wang, W., Fang, Q., Jiang, H., Xin, Q. and Xue, B. (2018), “The application of discrete wavelet transform with improved partial least-squares method for the estimation of soil properties with visible and near-infrared spectral data”, *Remote Sensing*, Vol. 10 No. 6, available at: <https://doi.org/10.3390/rs10060867>.

Wijewardane, N. k., Ge, Y., Wills, S. and Libohova, Z. (2018), "Predicting Physical and Chemical Properties of US Soils with a Mid-Infrared Reflectance Spectral Library", *Soil Science Society of America Journal*, Vol. 82 No. 3, pp. 722–731.

Ye, H., Huang, W., Huang, S., Huang, Y., Zhang, S., Dong, Y. and Chen, P. (2017), "Effects of different sampling densities on geographically weighted regression kriging for predicting soil organic carbon", *Spatial Statistics*, Elsevier B.V., Vol. 20, pp. 76–91.

Zeng, R., Rossiter, D.G., Yang, F., Li, D.C., Zhao, Y.G. and Zhang, G.L. (2017), "How accurately can soil classes be allocated based on spectrally predicted physio-chemical properties?", *Geoderma*, Elsevier, Vol. 303 No. March, pp. 78–84.

5 Improving fertilizer application for Moroccan rainfed wheat growing regions using crop models and NIR soil spectroscopy

Authors: Tadesse Gashaw Asrat¹, Ruben Sakrabani¹, Ron Corstanje¹, Cathy L. Thomas², Fassil Kebede³, Stephan M. Haefele²

¹Cranfield University, Cranfield, UK; ²Rothamsted Research, Harpenden, UK;

³College of Agriculture and Environmental Science, AgroBioscience Program, Mohammed VI Polytechnic University, Ben Guerir, Morocco.

5.1 Abstract

Effective nutrient management is essential for enhancing crop production, particularly in semi-arid regions like Morocco, where changing climate patterns and erratic rainfall significantly impact yields and production costs. Improved nutrient practices not only increase productivity and profitability but also safeguard soil health and reduce environmental risks. This study introduces an innovative approach to nutrient management by integrating spectral soil analysis with crop response modeling to enable data-driven decision-making. Pot experiments were conducted using Cambisol, Kastanozem, and Vertisol soils from Morocco, along with the 'Achtar' wheat cultivar, to evaluate the effectiveness of quadratic plateau functions and hyperbolic nutrient uptake models.

Additionally, the study assessed the accuracy of a handheld NIR spectrometer in predicting key soil properties for estimating indigenous nutrient supply capacity and refined parameters for the QUEFTS model, which had not previously been applied to wheat in Morocco. Both crop and nutrient uptake models demonstrated strong predictive performance, with R^2 values exceeding 0.7, successfully capturing the relationship between wheat grain yield and total nutrient uptake.

Optimal fertilizer rates and potential yields were estimated for the studied soils, revealing that nutrient uptake was higher in fertilized plots compared to nutrient-omitted plots. This increased uptake may be attributed to enhanced crop growth, allowing for greater soil exploration and improved physiological mechanisms for nutrient acquisition. These findings highlight the importance of evaluating nutrient

use efficiency across various application rates and soil nutrient levels, considering the soil's inherent nutrient supply capacity.

The NIR soil spectroscopy effectively predicted the indigenous supply of nitrogen, phosphorus, and potassium, offering results comparable to conventional soil analysis methods in a cost-effective and timely manner. This demonstrates the potential of NIR spectroscopy to complement crop models and support informed nutrient management decisions, enhancing the precision of fertilizer recommendations across diverse soil types.

Keywords: NIR spectroscopy, crop model, quadratic plateau function, the QUEFTS model, Morocco, "Achtar" wheat variety

5.2 Introduction

Most soils are deficient in at least one or two nutrient elements for intensive arable crops. Further on, continuous crop harvest alongside various biogeochemical processes, deplete essential plant nutrients such as nitrogen, phosphorus, potassium from cultivated lands over time. This deficiency and depletion cause the need for replenishing nutrients to sustain agricultural production effectively. Fertilizers serve as the primary means of introducing nutrients back into the crop production cycle which also supports improving or maintaining the overall soil health condition. Velayutham (2017) reported the global fertilizer consumption trend over five decades and found a significantly increasing use from 9.4 kg ha⁻¹ in 1967-68 to 117 kg ha⁻¹ in 2007-08, and further to 141 kg ha⁻¹ in 2013-14, respectively. This change has played a crucial role in ensuring food security in the face of rapid population growth and has contributed significantly to the advancement of rural economies (John and Babu, 2021). However, achieving sustainable agricultural productivity entails more than just mere application of these fertilizers. It requires tailored nutrient management strategies that consider trade-offs between maximizing yields and economic returns and minimizing the environmental impacts associated with nutrient imbalances, nutrient runoff and water resource pollution. This can be achieved through explicit understanding of soil-plant interactions at the field level and adapting management decisions based specific soil characteristics. Such approach enables the development of

management practices that are both effective and environmentally responsible, enabling to quantify trade-offs and arrive at solutions that are acceptable to all stakeholders (Fixen et al., 2015).

With ongoing climate change, particularly in semi-arid and arid regions like Morocco where a decreasing in rainfall and rising temperatures are anticipated, a tailored and site-specific approach is necessary. This includes implementing effective agronomic practices to enhance sustainability and productivity (Mrabet et al., 2012; Epule et al., 2016; The World Bank Group, 2021). Mamassi et al. (2023) emphasized the variability in agro-ecological conditions for cereal production in Morocco, noting that while 30% of the areas have favourable conditions (rainfall >400 mm/annum), 24% face intermediate conditions (rainfall <400 mm/annum). Significant production constraints result in a large yield gap in wheat, with reported attainable yields averaging at 1.2 t ha^{-1} , half of the potential 2.6 t ha^{-1} achievable under optimal agronomic practices (Silva et al., 2023). Current nutrient application rates mainly involve nitrogen and phosphorus, applied in low to modest quantities, with farmers applying 0–30 kg/ha N, P_2O_5 , and potassium (K_2O) (Pala et al., 2011). These recommendations often rely on generalized, non-informed decision-making processes that fail to consider the complexity of site-specific soil and crop interactions. Coupled with the projected decline in rainfall amount (Epule et al., 2016), this low-input approach highlights the critical importance of site-specific nutrient management strategies that address the suitability and productivity potential of specific regions to meet growing food production demands.

The primary goal of tailored recommendations is to identify factors limiting yield and propose site- and crop-specific strategies that optimize fertilizer use, minimize risks, and enhance profitability (Yost et al., 2000; Hochmuth et al., 2011; Maman et al., 2018; Kiboi et al., 2021). Two core approaches are typically employed: the basic production function and the basic nutrient uptake function (Nijland et al., 2008). The production function aims to maximize crop yield by deriving optimal nutrient application rates through curve-fitting techniques such as Mitscherlich-Bray, linear-plateau, logistic, and quadratic models, based on

experimental yield responses for respective nutrients (Sonar and Babhulkar, 2002; Avelu et al., 2003; Hochmuth et al., 2011; Correndo et al., 2023). These models, while valuable for determining fertilizer rates by incorporating economic factors, often generalize nutrient recommendations without integrating the site-specific soil nutrient dynamics.

However, crop response models have inherent limitations in accurately accounting for the differential nutrient supply from soils. They assume a linear or quadratic relationship between fertilizer inputs and crop yield without adequately considering key soil processes, such as mineralization, nutrient fixation, or residual effects from previous fertilizer applications (Dobermann and Cassman, 2002). Furthermore, these models do not incorporate soil information directly, limiting their scalability for site-specific recommendations. This generalized approach may lead to overestimation of nutrient requirements for marginal yield increases, particularly in soils with moderate to high nutrient reserves, resulting in excessive fertilizer application, increased production costs, and environmental risks (Zaremba and Smoleński, 2000; Avelu et al., 2003; Mahdi, 2008; Lark et al., 2020). Dhakal and Lange (2021) reviewed crop-yield response functions and emphasized the need for models that reflect real soil processes and nutrient availability. While functions like the quadratic response stochastic plateau yield function provide stable predictions with lower optimal nutrient rates (Lyons et al., 2019), they still fail to directly integrate soil nutrient supply variability. This results in models focusing primarily on fertilizer inputs rather than a holistic nutrient management approach, leading to recommendations that overlook nutrient recovery efficiency, soil organic matter contributions, and site-specific fertility conditions. Thus, while these models can capture diminishing marginal returns, their oversimplification of soil-nutrient interactions limits their reliability in fine-tuning nutrient recommendations across diverse agro-ecological conditions.

The second approach focuses on nutrient use efficiency, which plays a pivotal role in balancing soil nutrient supply, fertilizer application, and eventual crop uptake. Schut and Giller (2020) emphasized the importance of optimizing input quantities and timing, aligning them with plant demand and soil nutrient dynamics.

This concept underpins the Quantitative Evaluation of the Fertility of Tropical Soils (QUEFTS) model, which estimates nutrient requirements for target yields by integrating soil nutrient supply and applied inputs (Smaling and Janssen, 1993; Haefele et al., 2003; Fixen et al., 2015; Ichami et al., 2019). QUEFTS predicts crop yields by factoring macronutrients (NPK) interactions, crop physiological nutrient use efficiency (PhE), and region-specific maximum yields (Janssen, 2011). Its versatility has been validated in both temperate and tropical agroecosystems (Sattari et al., 2014). A key advantage of QUEFTS lies in its use of soil information for soil fertility assessment, allowing the derivation of fertilizer recommendations by quantifying the soil's nutrient supply capacity.

However, QUEFTS focuses on soil fertility rather than soil productivity — the latter being influenced not only by nutrient availability but also by other site-specific factors that influence soil productivity potential like soil structure and, water availability within an agroecology. Besides, its reliance on regional or agroecological maximum yields may not capture within-region soil productivity variability (Njoroge et al., 2017). Variations in local soil conditions, management practices, and crop responses often lead to discrepancies between predicted and observed nutrient uptake (Tittonell et al., 2008; Sattari et al., 2014). Additionally, QUEFTS assumes that nutrient uptake differences between fertilized and unfertilized plots are solely due to applied nutrients, overlooking the enhanced root growth and nutrient access in fertilized plots. This can result in underestimating the soil's inherent nutrient supply capacity (Haefele et al., 2003). Recent adjustments using quantile regression for local maximum yields may further introduce inaccuracies when soil variability is insufficiently accounted for (Ravensbergen et al., 2021).

To address the challenges of optimizing nutrient management, this study proposes integrating soil productivity potential into the QUEFTS model by combining crop response models with the assessment of differential nutrient recovery for each nutrient. While crop response models effectively characterize yield responses to nutrient levels based on soil productivity potential, they often lack direct incorporation of soil fertility factors, which are essential for

understanding the underlying mechanisms and relationships between soil properties and crop responses to applied inputs. By integrating these models, the study aims to enhance the ability to assess soil fertility and address productivity variations by understanding the soil's inherent productivity potential. Additionally, QUEFTS uses multiple soil properties to quantify indigenous nutrient supply, and integrating NIR-derived soil data can provide valuable insights for better-informed fertilizer decisions.

NIR soil spectroscopy offers a cost-effective and scalable approach for real-time soil nutrient assessment. By combining NIR-derived soil data with crop models, the nutrient supply capacity of the soil can be more accurately determined, leading to tailored and site-specific fertilizer recommendations (Asrat et al., 2023). This integration creates a dynamic feedback loop, where soil spectroscopy enhances model inputs, and the combined models refine fertilizer strategies, promoting efficient nutrient use and maximizing crop performance. The study's objectives are to: 1) assess the primary limiting nutrients and determine their optimal application rates for rainfed wheat across Morocco's key soil types—Cambisol, Kastanozems, and Vertisol; 2) estimate QUEFTS and Quadratic Plateau (QP) model parameters, incorporating variations in soil productivity to refine fertilizer recommendations for different soil types; and 3) evaluate the effectiveness of NIR soil spectroscopy in predicting soil nutrient supply capacity, thus improving decision-making with high-resolution soil information for wheat production in Morocco's agricultural systems.

5.3 Materials and methods

5.3.1 Pot experimentation

5.3.1.1 General approach, soil types and treatment setup

Pot experiments were necessary due to logistical constraints in conducting large-scale field trials and the need for precise control over environmental variables, particularly soil nutrient availability and moisture levels. However, it is essential to acknowledge that pot experiments may not fully replicate field conditions, particularly in terms of root development, soil moisture dynamics, and

microclimatic variability. Therefore, while the results provide valuable insights, caution is needed when extrapolating findings to field conditions, and follow-up field validation is recommended.

This pot experiment was designed to assess the response of rainfed wheat to different nitrogen (N), phosphorus (P), and potassium (K) application rates across three major soil types found in Morocco: Cambisol, Kastanozems, and Vertisol. The "Full" NPK application rates were established for each soil type based on their specific properties and the current fertilizer recommendations for rainfed wheat in Morocco (Table 5.1). Two additional rates were tested: half of the full rate (designated "0.5") and one-and-a-half times the full rate (designated "1.5"), along with a nutrient omission treatment, resulting in a four-level nutrient response study. Zinc was tested at two levels: omission and 10 kg ha⁻¹. This setup created 11 treatment combinations, each replicated three times per soil type. To enhance precision, the combined NPK treatment was replicated six times (Trt-4 and Trt-12, Table 5.1), bringing the total number of pots to 108.

To minimize border effects, additional pots filled with compost were placed around the experimental block as border plants. This arrangement ensured that any observed outcomes were solely attributed to variations in nutrient levels and soil types under study.

Table 5.1: Soil characteristics of the soil types used for nutrient study using pot experimentation.

Soil type	Cambisols	Kastanozems	Vertisols
pH-H ₂ O	8.3	6.8	8.2
Olsen P (g/kg)	5.8	12.6	20.2
DTPA Zn (g/kg)	1.27	0.97	1.50
Exchangeable K (mmol/kg)	3.0	6.0	6.1
CEC (cmol +/kg)	23.7	59.5	48.9
SOC (g/kg)	12.2	10.0	10.0
Clay %	20.5	51.1	45.7
Textural class (U)	Clay Loam	Heavy Clay	Heavy Clay

Base fertilizer amounts considered for 7 t ha ⁻¹ wheat grain yield			
N (kg/ha)	135	135	125
P ₂ O ₅ (kg/ha)	80	80	90
K ₂ O (kg/ha)	30	30	70
Zn (kg/ha)	10	10	10

Table 5.2: Treatment set up and target response.

Treatment	N-P-K-Zn	Response		Micronutrient	
Trt-1	0-full-full-0	PK	N	-Zn	
Trt-2	½-full-full-0				
Trt-3	1½-full-full-0				
Trt-4	Full-full-full-0	NPK	Common to NPK	-Zn	
Trt-5	Full-0-full-0	NK	P	-Zn	
Trt-6	Full-½-full-0				
Trt-7	Full-1½-full-0				
Trt-8	Full-full-0-0	NP	K		-Zn
Trt-9	Full-full-½-0				
Trt-10	Full-full-1½-0				
Trt-11	Full-full-full full	NPK with Zn		+ Zn	
Trt-12	Full-full-full-0	NPK	Common to NPK	-Zn	

***Full' is the base fertilizer in Table 5.1 and the numbers (½ and 1½) multiplied this amount to get the fertilizer rates.*

5.3.1.2 Experimental unit and design

The experimental design was a split plot design where the soil types were assigned as the sub-factor (main plot) and the treatments were assigned as the main factor (sub plot) (Figure D-1). Straight fertilizer, one for each nutrient, were applied at planting as a basal application and mixed with the soil thoroughly. The

fertilizers used were ammonium nitrate for N, Triple Super Phosphate for P, Muriate of potash for K and Zinc Sulphate for Zn. Seeds were pre-germinated in pellet trays in a standard in-house compost and five seedlings were transplanted at five days after sowing to the respective experimental 3-litre pot (19 cm top diameter, 13 cm base diameter and 15 cm height). The Moroccan spring wheat variety 'Achtar' was used for this study as test crop.

5.3.1.3 Details of the glasshouse conditions and water application

The wheat plants were grown in a controlled glasshouse facility at Rothamsted Research, during January to April 2023. Temperature in the glasshouse was maintained at 20° C Day/15° C night. There was supplemental overhead lighting with sixteen bulbs that provided a 16/8 hr day/night photoperiod throughout the growing period.

The pots were arranged on a raised bench, and each placed on an independent saucer which was used as watering point. Plants were watered to 10% of the soil weight, three times weekly with reverse osmosis (RO) water. The experiment was conducted in an 11 m² GMO/Quarantine GH glasshouse.

Nutrient amounts added to each experimental pot were based on the bulk density of the soil, if the top 20 cm of 1 ha soil weigh 2,200,000 kg, calculated with a bulk density of 1100 kg m⁻³ for the top 20 cm soil layer.

5.3.1.4 Plant growth monitoring, harvesting and data collected

Data pertaining to crop development were gathered, encompassing various stages from planting to physiological maturity. This included recoding the planting data, emergence date (defined as the point when 50% of the pot was covered by the seedlings), anthesis date (the date at which more than 50% of tillers developed anthers), grain filling date, and physiological maturity (when 50% of tillers had reached maturity). Additionally, the number of tillers (defined as those with effective heads setting seed), plant height (measured for five representative tillers), grain yield, moisture content at harvest, spike length, and seed count per spike were documented. Straw productivity was determined based on the harvest

of leaves and stalks, followed by drying at 70°C for 72 hours. Pictorial nutrient response of the crops at grain filling stage is depicted in Figure 5.1.

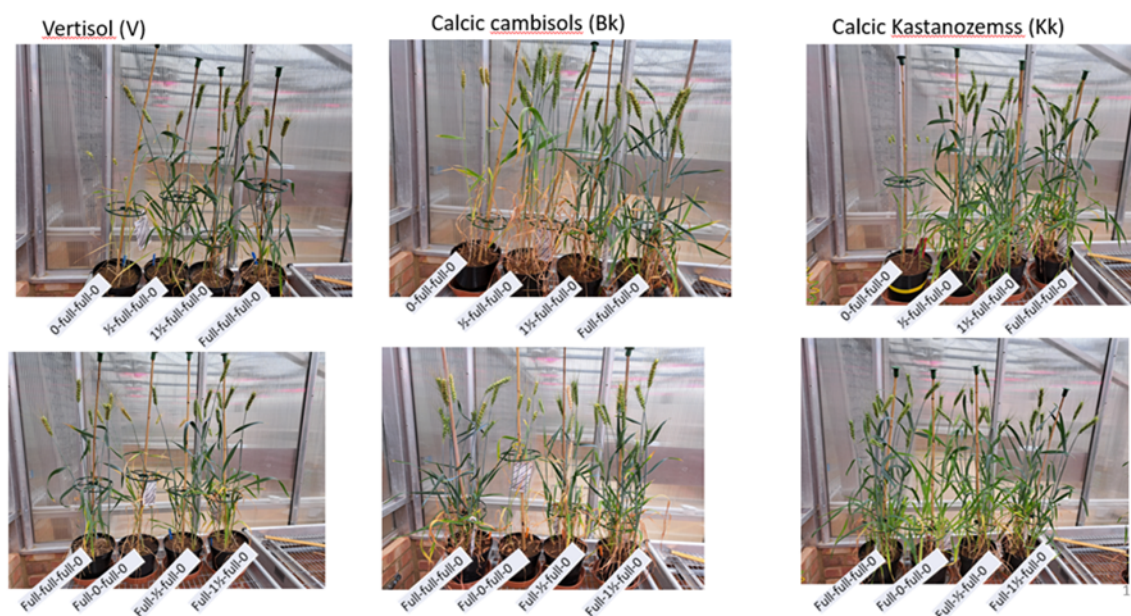


Figure 5.1: Pictorial response of wheat crop for applied nitrogen (upper row) and phosphorus (bottom row) across the soil types.

5.3.1.5 Yield and yield component data management

Based on the moisture content at harvesting of the subsample taken from the plots, standardized grain yield was calculated to the moisture content of 13.5% (Mulvaney and Devkota, 2020) using (5-1).

$$\text{Standardized grain yield} = \frac{(\text{Harvested yield}) * (1 - \text{Moisture content})}{(1 - \text{Standard moisture})} \quad (5-1)$$

Harvest index (HI) was calculated from the ratio of grain yield to total above ground biomass to evaluate the consistency of the productivity of grain yield in the range of the expected harvest index (0.45 – 0.55).

5.3.1.6 Plant sample collection and analysis

Grain and straw (leaf and stalk) were milled after oven drying using an agate ball mixer mill using a grinding time in the range of 2 minutes. Nitrogen was determined using Leco combustion. The grain and straw nutrient densities were determined using portable X-ray florescence (pXRF), at the Dry Spectral Laboratory at Rothamsted Research, as detailed by Acquah et al. (2022) who

found strong correlations and low prediction errors for most of the nutrients in plant samples.

5.3.2 Soil sample analyses

The soils used for the pot study were air-dried, crushed, sieved to less than 2 mm, and further processed (milling and weighing) for laboratory analysis as required. The conventional chemical analysis of the soils was done in three replicates to determine the soil properties with higher precision at the NRM soil lab of UK (<https://cawood.co.uk/nrm/>). The tested soil properties were exchangeable potassium (Exch.K), soil pH-H₂O, soil phosphorus (OlsenP), available zinc (DTPA.Zn) and Cation Exchange Capacity (CEC).

The ammonium acetate extracts (ISO 13536:1995) of exchangeable potassium (Exch.K) and magnesium (Exch.Mg) were determined by atomic emission and atomic absorbance, respectively, and CEC was determined with the ammonium acetate method (Ciesielski and Sterckeman, 1997). Soil pH was measured in a 1:2.5 soil: water suspension (ISO 10390: 2005) one hour after mixing, using a semi-micro sealed pH electrode from Fisher scientific. Sodium bicarbonate extracted soil phosphorus (OlsenP) was determined following the Olsen method (Olsen, 1954) using the phospho-molybdenum blue method on a continuous colorimetric flow analyser. Available zinc (DTPA.Zn) and manganese (DTPA.Mn) were determined by atomic absorption spectrophotometer after complexing with DTPA (Lindsay and Norvell, 1972).

Meanwhile, soil particle size fractions (clay, silt and sand) were determined by laser diffraction on a L-960 particle-size analyser (Horiba scientific Ltd.) in the Dry Spectral Laboratory at Rothamsted Research (Thomas et al., 2021). The dry combustion method was used for determining total carbon (Nelson and Sommers, 1996) and total nitrogen (Bremner, 1996) with a Leco dry combustion analyser. Inorganic carbon was analysed using wet acidification with a Skalar Primacs AIC 100 (Skalar Analytical BV, Breda, Netherlands). Total soil organic carbon (SOC) was then calculated by subtracting inorganic carbon from total soil carbon.

5.3.3 Actual nutrient uptake and recovery efficiency

The amount of nutrient taken up by the crop was quantified from the dried weight of wheat grain and straw for respective nutrients and levels ((5-2)). The equation was used after inspecting the harvest index ranges and values on the soil types and nutrient levels (Sattari et al., 2014).

$$U_{FRi} = Dm_{FRgi} * x_{FRgi} + DM_{FRsi} * x_{FRsi} \quad (5-2)$$

Where U_{FRi} is the total nutrient uptake nutrient i with the applied fertilizer rate (FR_i) of application, Dm_g and DM_s are dry-matter production (kg), x_g and x_s are the mass fraction ($g\ kg^{-1}$) of nutrient in grain and straw, respectively.

Recovery efficiency (RE) is calculated to evaluate how effectively plants utilize the applied nutrient relative to what is available from the soil. It helps quantify the portion of the applied nutrient that is taken up by the crop, thereby providing insights into nutrient use efficiency and potential nutrient losses to the environment. In this study, when a nutrient was not applied, the plant's nutrient uptake was used as a baseline, representing the soil's inherent nutrient supply capacity under limited conditions (5-3). This baseline uptake was subtracted from the total nutrient uptake when fertilizer was applied, allowing the calculation of RE. This approach accounts for nutrient losses, such as leaching, volatilization, or immobilization, and helps identify how much of the applied nutrient is actually absorbed by the crop. By doing so, RE serves as a key indicator for optimizing fertilizer applications, reducing nutrient losses, minimizing environmental risks, and improving the sustainability of agricultural practices (Ravensbergen et al., 2021).

$$RE_{Oi} = \frac{U_{FRi} - U_{Oi}}{FR_i} \quad (5-3)$$

Where RE_{Oi} is the recovery efficiency of the nutrient for applied nutrient in reference to the uptake at the omitted condition, U_{FRi} is the uptake under the fertilized condition, U_{Oi} is the uptake under nutrient omitted condition, FR_i is the amount of nutrient applied.

In our study, we considered the differential nutrient uptake from the soil supply by the crop and adopted the assumption proposed by Dobermann and Cassman (2002) to quantify recovery efficiency RE_i and was computed with (5-4).

$$RE_{Oi} = \frac{U_{FRi} - U_{Oi}}{FR_i} \quad (5-4)$$

Where INS_i is the indigenous nutrient (i) supply capacity of the soil quantified from the QUEFTS approach in section 5.3.4.

5.3.4 The QUEFTS approach to quantify INS and internal nutrient use efficiency

The indigenous nutrient supply capacity of soils for each nutrient was derived using the following equations ((5-5, (5-6, (5-7) (Sattari et al., 2014).

Calculation of the pH correction factor for each nutrient followed was: -

$$f_N = 0.25(pH - 3) \quad (5-5)$$

$$f_P = 1 - 0.5(pH - 6)^2 \quad (5-6)$$

$$f_K = 0.625(3.4 - 0.4pH) \quad (5-7)$$

Whereby f_N is the correction factor for N, f_P the correction factor for P, f_K the correction factor for K, and pH is pH-H₂O.

The estimation of the indigenous nutrient supply (INS) capacity for each nutrient was computed with the following relationships as:

$$S_N = a_N f_N C_{org} ; a_N = 2 \left(2^{\frac{T-9}{9}} \right) \quad (5-8)$$

$$S_P = a_P f_P C_{org} + \beta_P \text{OlsenP} ; a_P = 0.35 \quad (5-9)$$

$$S_K = \frac{a_K f_K K_{Exch}}{\gamma_K + \beta_K C_{org}} ; a_K = 500 ; \gamma_K = 2 ; \beta_K = 0.9 \quad (5-10)$$

Where S_N , S_P and S_K are soil supplies of crop available N, P and K, respectively; α , β and γ are empirical parameters, T is the annual mean temperature, and f is a pH dependent correction factor describing the pH-dependency of soil organic

matter mineralization (Janssen et al., 1990). A mean annual temperature of 18.05 °C was used (The World Bank Group, 2021).

5.3.5 Crop model fitting for parameter estimations

5.3.5.1 The quadratic response plateau model (QRP)

The initial crop model assessed for soil and nutrient response modeling was the quadratic plateau function, designed to capture the crop grain yield response to nutrient application. This model assumes a quadratic relationship between grain yield and nutrient levels up to a maximum potential yield, accounting for diminishing marginal yield responses as nutrient application increases (Alivelu et al., 2003). The response function was fitted using nonlinear least squares regression on experimental data for N, P and K at four nutrient application levels, ((5-11).

$$y_i = \min(a + b * FR_i + c * FR_i^2, jp) + \epsilon_i \quad (5-11)$$

where y_i is the crop yield (kg/ha), a a baseline yield from inherent soil nutrient supply, b is the incremental yield per unit of applied nutrients, c is the quadratic term representing the curvature of the response, jp denotes the yield plateau level, and $\epsilon_i \sim N(0, \sigma_e^2)$ is the random disturbance term that assumed to be independent and identically distributed. Equation (5-12) was used to estimate c .

$$c = -\frac{1}{2} * \frac{b}{jp} \quad (5-12)$$

To estimate the model parameters, initial values were determined by examining grain yield against nutrient levels: a was estimated based on the minimum observed grain yield; b was determined by estimating the slope in the linear increase in grain yield to nutrient levels; c was derived from the curvature of the yield response curve; and jp was identified at a point where the grain yield began to plateau, beyond which further nutrient application did not improve yield.

Each parameter was independently tested for statistical significance at $\alpha = 0.05$. If the parameter was found to be non-significant, model adjustments would be made by either simplifying or re-specifying the model. Once significant

parameters were identified, the model could be used to estimate the fertilizer requirements for achieving specific yield targets.

5.3.5.2 Nutrient uptake parameter estimation

Nutrient uptake by plants exhibits a hyperbolic response to applied fertilizers based on various factors such as nutrient availability, soil properties, climate conditions, and crop characteristics (Moll et al., 1996; Congreves et al., 2021). To determine the crop nutrient demand for each nutrient (NPK), estimates for the maximum uptake from the applied nutrient was estimated with a hyperbolic relation for the three soil types independently with equation (5-13).

$$Total\ Uptake_i = \frac{a_i * Nutrient\ rate_i}{b_i + Nutrient\ rate_i} \quad (5-13)$$

Whereby a_i is a parameter representing the maximum nutrient uptake achievable under optimal conditions, and b_i is a parameter representing the nutrient level where uptake of the nutrient is at half of the maximum uptake.

Next, a nonlinear least squares regression was fitted to the experimental data to estimate a_i and b_i with equation (5-14):

$$fit = nlsLM\left(Total\ Uptake_i \sim \frac{a_i * Nutrient\ rate_i}{b_i + Nutrient\ rate_i} + \epsilon_i, start = lists(a_{si}, b_{si})\right) \quad (5-14)$$

where a_{si} and b_{si} represent the initial model parameter given by inspecting the relationship between $Total\ Uptake_i$ and $Nutrient\ rate_i$. The model parameter estimates were evaluated for their statistical significance at $\alpha < 0.05$.

5.3.6 NIR applicability for indigenous nutrient supply estimation

Asrat et al., (2024) developed a Moroccan NIR soil spectral library tailored for predicting soil characteristics from samples that were air-dried and sieved to 2mm. They established a framework for utilizing this soil spectral library to construct specific predictive model per soil sample location based on the spatial proximity of the new samples to existing soil spectral databases. Their approach involved selecting a subset of calibration samples based on spatial proximity assuming that soil variation exhibit spatially autocorrelation, and hence

performance of soil spectral prediction might vary depending on spatial scale (Savvides et al., 2010; Summerauer et al., 2021). This methodology is comprehensively detailed in their paper.

In this research study, soil samples of Cambisol (7), Kastanozems (41) and Vertisol (17) were chosen from the validation set which possessed both NIR predicted soil properties using the geostatistical approach, as outlined above, and conventional soil analysis. The indigenous nutrient supply capacity for nitrogen, phosphorus and potassium were derived for these samples using both the NIR predicted and conventional soil analysis methods. This was done to assess and validate the agreement between the NIR-predicted soil properties and conventional soil analysis indigenous nutrient supply result discrepancies for nitrogen, phosphorus, and potassium. These assessments were crucial for supporting informed decision-making in conjunction with the crop models developed in this study as outlined in section **Error! Reference source not found..**

5.3.7 Statistical analysis and model performance evaluation

The analysis of crop parameters (grain yield, straw yield, grain nutrient densities for N, P, K, and Zn), and responses to different levels and types of applied nutrients was conducted using a split plot analysis of variance, according to the experimental design Table D.1. When significant effects of factors were observed, mean separation was carried out to assess the effects of individual factor levels, maintaining a significance level of $\alpha = 0.05$. Duncan's Multiple Range Test (HSD test) was used to perform pairwise comparisons of treatment means following the ANOVA. To minimize the risk of Type I errors due to multiple comparisons, this test controls for the family-wise error rate. This is a common practice in statistical analysis when making multiple comparisons, ensuring that the probability of incorrectly rejecting a null hypothesis (i.e., identifying a significant difference where none exists) remains at an acceptable level.

To evaluate the crop response and nutrient uptake models, the coefficient of determination (R^2), and the Root Means Square Error (RMSE) were computed to evaluate the model performance between predicted (y) and observed values (x).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - x_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (5-15)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2} \quad (5-16)$$

5.4 Results

5.4.1 Crop yield response to applied nutrients and yield limiting nutrient per soil type

This investigation utilized a split plot analysis of variance, detailed in Table 5.3, revealing significant effects of nutrient treatments and soil types on the variables studied, mostly at a significant level of a $p < 0.01$. Notably, the ANOVA result indicated that grain potassium concentration was solely influenced by applied nutrients and showed no variation across soil types, despite differing application rates. Additionally, the ANOVA showed that there was no variation in grain potassium concerning the interaction between applied nutrients and soil types. Subsequent analysis involved comparing mean differences in the effects of applied nutrients and rates for each soil type independently.

Based on the results in Table 5.4, it is evident that nitrogen and phosphorus played crucial roles in determining wheat grain productivity across soil types. Nitrogen, particularly, emerged as the most crucial nutrient, with its deficiency significantly reducing grain yield regardless of soil type. The necessity of nitrogen application was underscored by the observation that without nitrogen, grain yields were only about 0.5 t ha^{-1} across all soil types. Similarly, the lower rate of nitrogen resulted in limited yields, further emphasizing the indispensable role of nitrogen in wheat productivity for these soil types in Morocco. The lower nitrogen rates resulted in 2.37 , 3.87 and 1.62 t ha^{-1} grain yields for Cambisol, Kastanozems and Vertisol, respectively. Nitrogen rates showed a similar effect on straw yield productivity as well as most grain concentrations of nutrients (Table D.2; Table D.3; Table D.4; Table D.5). There were some exceptions in nitrogen responses

of grain and straw yields for the highest rate of nitrogen, reducing the production well below 2 t ha⁻¹ on Vertisols.

Phosphorus followed nitrogen in terms of its impact on wheat production, indicating that adequate levels of phosphorus are required to maximize grain yield production on these soil types (Table 5.4). Withholding phosphorus application limited the grain productivity to below 2 t ha⁻¹ and affected straw yield and grain nutrient concentrations like nitrogen in most of the cases (Table D.2; Table D.3; Table D.4; Table D.5).

In contrast, potassium exhibited varying and complex effects on wheat productivity and grain potassium concentration across all soil types, with no clear trend associated with application rates (Table 5.4; Table D.2; Table D.3; Table D.4; Table D.5). While not statistically significant, zinc application tended to increase yield production, particularly for Kastanozems and Vertisol by up to 1 t ha⁻¹ with an application of 10 kg ha⁻¹ zinc. However, the zinc concentration in wheat grain was statistically significant for all soil types on zinc-fertilized plots, suggesting potential benefits for improving grain quality for human consumption, despite not visibly impacting productivity levels (Table 5.4; Table D.2; Table D.3; Table D.4; Table D.5).

Table 5.3 Analysis of variance for the split plot wheat grain and straw yield and grain nutrient concentration to treatments and soil types.

Source	DF	Mean squares and significance					
		GY t-ha	SY t-ha	G_N	G_P	G_K	G_Zn
Soil type	2	34.59***	27.11***	20.97*	3.11**	0.055 ^{ns}	0.00092**
Nutrient	10	10.72***	5.92***	35.05***	0.52***	1.498***	0.00266***
Nutrient * Soil type	20	0.89***	0.60***	8.69***	0.11**	0.3087 ^{ns}	0.00011***
Residuals	69	0.28	0.14	1.50	0.05	0.2112	0.00003

GY t-ha = Grain yield (t ha⁻¹); SY t-ha = Grain yield (t ha⁻¹); G_N = Grain nitrogen concentration (g kg⁻¹); G_P = Grain phosphorus concentration (g kg⁻¹); G_K = Grain potassium concentration (g kg⁻¹); G_Zn = Grain zinc concentration (g kg⁻¹). One asterisk (*), two asterisks (**), and three asterisks (***) denotes significant at p < 0.05, 0.01, and 0.001, respectively and ns is not significant at p < 0.05.

Table 5.4 Effect of nutrient combinations and nutrient levels on grain yield and straw productivity of wheat across soil types.

Cambisol				Kastanozems				Vertisol			
Treatment	GY t-ha	Treatment	SY t-ha	Treatment	GY t-ha	Treatment	SY t-ha	Treatment	GY t-ha	Treatment	SY t-ha
Full_1.5_full_0	3.59 ^a	Full_1.5_full_0	3.09a	Full_full_full full	5.66 ^a	1.5_full_full_0	4.35a	Full_full_full full	2.96a	Full_full_0.5_0	2.36a
Full_full_full_0	3.41 ^a	Full_full_full_0	2.81a	Full_full_0.5_0	5.27 ^a	Full_full_1.5_0	4.27ab	Full_full_0.5_0	2.91a	Full_full_full full	2.34a
Full_full_0_0	3.17 ^a	Full_full_1.5_0	2.66a	Full_1.5_full_0	5.11 ^a	Full_full_0.5_0	4.26ab	Full_1.5_full_0	2.88a	Full_0.5_full_0	2.11a
Full_full_0.5_0	3.14 ^a	Full_full_0.5_0	2.53a	Full_full_1.5_0	4.69 ^a	Full_1.5_full_0	4.11ab	Full_0.5_full_0	2.72ab	Full_full_0_0	2.02a
Full_full_full full	2.98 ^{ab}	Full_full_full full	2.45a	1.5_full_full_0	4.66 ^a	Full_full_full full	4.05ab	Full_full_0_0	2.67ab	Full_1.5_full_0	2.00a
1.5_full_full_0	2.83 ^{abc}	1.5_full_full_0	2.41ab	Full_full_full_0	4.57 ^a	Full_full_full_0	3.78ab	Full_full_1.5_0	2.26ab	Full_full_1.5_0	1.95a
Full_0.5_full_0	2.64 ^{abc}	Full_0.5_full_0	2.18abc	Full_full_0_0	4.25 ^a	Full_full_0_0	3.61ab	Full_full_full_0	2.25ab	Full_full_full_0	1.93a
Full_full_1.5_0	2.39 ^{abc}	0.5_full_full_0	1.89abc	Full_0.5_full_0	4.17 ^a	Full_0.5_full_0	3.57ab	0.5_full_full_0	2.04ab	1.5_full_full_0	1.76a
0.5_full_full_0	2.37 ^{abc}	Full_full_0_0	1.82abc	0.5_full_full_0	3.87ab	0.5_full_full_0	3.13bc	1.5_full_full_0	1.96ab	0.5_full_full_0	1.62a
Full_0_full_0	0.84 ^{bc}	Full_0_full_0	0.75bc	Full_0_full_0	1.87bc	Full_0_full_0	2.31c	Full_0_full_0	1.46bc	Full_0_full_0	1.37ab
0_full_full_0	0.63 ^c	0_full_full_0	0.70c	0_full_full_0	0.31c	0_full_full_0	0.36d	0_full_full_0	0.32c	0_full_full_0	0.40b

GY t-ha = Grain yield (t ha⁻¹); SY t-ha = Straw yield (t ha⁻¹).

5.4.2 Actual uptake from applied and indigenous nutrient supply

Table 5.5 presents nutrients uptake in wheat grain and straw from plots where the respective nutrients (N, P, or K) were omitted, aiming to indicate the nutrients absorbed by the plant from the soil across the different soil types. The results showed that omitting nitrogen and phosphorus reduced the uptake of nitrogen, phosphorus, and potassium, whereas potassium omission had no effect on nitrogen and phosphorus uptake. Interestingly, the lowest potassium uptake was noted in Cambisol when nitrogen and phosphorus were sufficient, suggesting a potential relationship between potassium uptake and soil types. For Cambisol and Kastanozems, the current potassium rate of application for these soils was 30 kg ha⁻¹ (Table 5.2). The inherent nutrient supply (INS) capacity of soils for each nutrient across different soil types was determined following Step 1 of the QUEFTS approach and the result is presented in Table 5.6. Only a portion of the supply of nitrogen and phosphorus were utilized in the omission plots, despite the severe impact on crop performance. This result is consistent with the findings of Poulton et al. (2023), who reported a regression R^2 value of 0.3 between soil N and wheat grain yield. This discrepancy suggested that crops require an optimum nutrient level to fully exploit the available nitrogen and phosphorus, as these nutrients are integral to various physiological plant activities. In this study, the crop's uptake of N, P, and K from soil reserves followed a specific ratio, which should be taken into account when making nutrient management. Across most soils, the uptake proportion for all nutrients ranged between 0.2 and 0.35 relative to the INS. An exception was noted concerning phosphorus uptake in Kastanozems where nearly 87% of the INS of P was taken up by the plant. This suggests that the INS estimated using the QUEFTS approach may have been underestimated. In this analysis, SOC was employed to estimate soil supply. Nonetheless, it is recommended that total phosphorus is considered instead of determining phosphorus soil supply in regions or farmlands where substantial phosphorus applications are common (Sattari et al., 2014; Janssen et al., 2014). Hence our finding corroborates that the use of SOC in soil supply estimation of phosphorus was underestimating the available phosphorus in Kastanozems.

Table 5.5: Uptake of each nutrient by the grain and straw of wheat on the untreated plots by soil types.

Soil type	Treatment	Grain N	Straw N	Grain P	Straw P	Grain K	Straw K
		Kg ha ⁻¹					
Cambisols	0_full_full_0	10.03 (1.75)	3.12 (1.01)	1.28 (0.26)	0.7 (0.05)	1.38 (0.39)	13.09 (2.56)
Cambisols	Full_0_full_0	19.46 (1.89)	9.5 (1.52)	1.36 (0.12)	0.26 (0.02)	1.94 (0.3)	13.58 (0.69)
Cambisols	Full_full_0_0	71.75 (15.53)	17.18 (2.43)	7.48 (1.61)	1.46 (0.27)	10.08 (2.13)	15.58 (0.67)
Kastanozems	0_full_full_0	6.85 (0.78)	2.00 (0.24)	0.66 (0)	0.69 (0.08)	0.76 (0.02)	5.08 (0.27)
Kastanozems	Full_0_full_0	38.28 (1.04)	53.8 (2.77)	5.3 (0.15)	3.2 (0.24)	5.61 (0.3)	42.47 (3.57)
Kastanozems	Full_full_0_0	97.69 (8.19)	37.41 (2.85)	12.06 (1.36)	3.72 (0.4)	14.88 (2.46)	45.79 (2.61)
Vertisols	0_full_full_0	5.6 (0.46)	1.81 (0.29)	0.6 (0.02)	0.25 (0.05)	0.88 (0.05)	4.27 (0.55)
Vertisols	Full_0_full_0	33.87 (3.56)	14.27 (0.95)	2.78 (0.44)	0.93 (0.16)	4.38 (0.28)	25.21 (1.83)
Vertisols	Full_full_0_0	62.2 (1.14)	18.88 (1.06)	6.77 (0.33)	2 (0.15)	9.44 (0.66)	31.32 (4.31)

() = se. The bold values are the nutrient uptake when the nutrient was omitted. The values are elemental N, P, or K.

Table 5.6: Indigenous nutrient supply capacity (INS) using QUEFTS estimation approach with uptake at zero nutrient addition for each soil.

Nutrient	Soil types								
	Cambisols			Kastanozems			Vertisols		
	INS	Total uptake*	Proportion	INS	Total uptake *	Proportion	INS	Total uptake *	Proportion
Nitrogen (kg/ha)	48.9	13.2	0.27	37.8	8.85	0.23	40.2	7.4	0.18
Phosphorus – P ₂ O ₅ (kg/ha)	5.3	1.63	0.31	9.8	8.5	0.87	12.1	3.71	0.31
Potassium – K ₂ O (kg/ha)	69.5	25.7	0.37	167.0	60.7	0.36	166.0	40.8	0.25

aN, with the modified derivation, was 4.015; INS indigenous nutrient supply based on QUEFTS; * Uptake omission plot.

5.4.3 Recovery efficiency estimates for applied nutrients across soil types

Table 5.7, Table 5.8, Table 5.9 present the apparent recovery efficiency, determined using the difference method for the applied rates of nitrogen, phosphorus and potassium, respectively. Current application rates, denoted as 'Full', exhibited higher recovery efficiency compared to rates below or above for nitrogen and phosphorus in Cambisols and Kastanozems, except for phosphorus in Kastanozems. Conversely, the recovery efficiency of nitrogen and phosphorus rates in Vertisol decreased as rates increase yet remained optimal compared to default values of 50 and 10%, for nitrogen and phosphorus, respectively, (Janssen et al., 1990). However, it is reasonable to expect better recovery efficiencies for these nutrients, given the absence of losses through percolation and runoff which are persistent in farmlands (Syers et al., 2008).

Recovery efficiency of applied potassium displayed no discernible trend, complicating the assessment of nutrient uptake on Cambisols and Kastanozems, despite the small rates applied. For Vertisol, the recovery efficiency indicated that the applied nutrient rates were not taken up, despite the estimated optimal rate being 70 kg ha⁻¹. This finding, along with the preceding analysis, suggested that potassium deficiency might not be a limiting factor for wheat productivity in Morocco, considering the soils examined in this study.

Notable discrepancies were identified between the two methods used to estimate recovery efficiency. The crop uptake-based method generally underestimated soil nutrient uptake in treated plots, while the INS method tended to overestimate it, potentially resulting in reduced yields. Both methods showed convergence at higher application rates, suggesting that recovery estimation may follow a parabolic trend when estimated using INS. The actual recovery fraction likely lies between the values derived from nutrient uptake in omitted plots and those estimated using the INS approach.

Table 5.7: Recovery efficiency among applied nitrogen levels across soil types.

Treatment	Soil type	Recovery efficiency (%)	
		Using INS	Using U_{0i}

Full_full_0.5_0	Cambisol	15.16	68.07
Full_full_full_0	Cambisol	45.32	71.78
Full_full_1.5_0	Cambisol	21.49	39.13
Full_full_0.5_0	Kastanozems	73.51	116.32
Full_full_full_0	Kastanozems	75.01	96.42
Full_full_1.5_0	Kastanozems	61.16	75.43
Full_full_0.5_0	Vertisol	34.19	86.58
Full_full_full_0	Vertisol	30.13	56.33
Full_full_1.5_0	Vertisol	14.73	32.20

U_{0i} = uptake of nutrients at nutrient omitted pots

Table 5.8: Recovery efficiency of applied phosphorus rates across soil types.

Treatment	Soil type	Recovery efficiency (%)	
		Using INS	Using U_{0i}
Full_0.5_full_0	Cambisol	9.88	31.17
Full_full_full_0	Cambisol	14.26	24.91
Full_1.5_full_0	Cambisol	12.45	19.55
Full_0.5_full_0	Kastanozems	25.32	32.76
Full_full_full_0	Kastanozems	19.45	23.17
Full_1.5_full_0	Kastanozems	16.81	19.29
Full_0.5_full_0	Vertisol	-19.38	23.31
Full_full_full_0	Vertisol	-11.04	10.31
Full_1.5_full_0	Vertisol	-5.20	9.03

U_{0i} = uptake of nutrients at nutrient omitted pots.

Table 5.9: Recovery efficiency of applied potassium rates across soil types.

Treatment	Soil type	Recovery efficiency (%)	
		Using INS	Using U_{0i}
Full_full_0.5_0	Cambisol	-218.7	133.28
Full_full_full_0	Cambisol	-17.1	158.94
Full_full_1.5_0	Cambisol	-19.1	98.22
Full_full_0.5_0	Kastanozems	-740.9	114.81
Full_full_full_0	Kastanozems	-334.5	93.33
Full_full_1.5_0	Kastanozems	-197.0	88.20
Full_full_0.5_0	Vertisol	-409.7	22.66
Full_full_full_0	Vertisol	-217.2	-1.02
Full_full_1.5_0	Vertisol	-144.1	0.03

U_{0i} = uptake of nutrients at nutrient omitted pots

5.4.4 Total plant dry matter and grain yield with total nutrient uptake

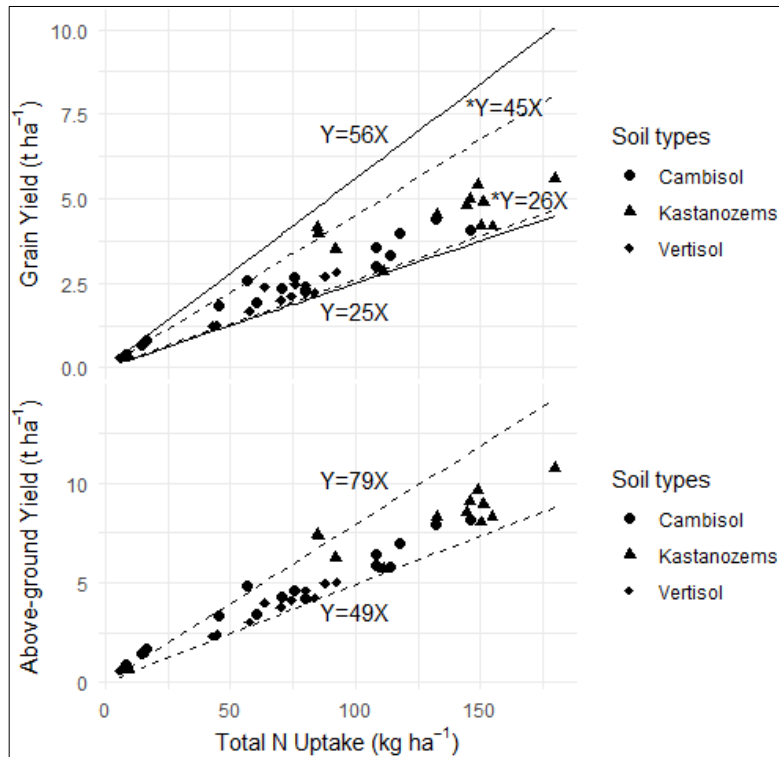
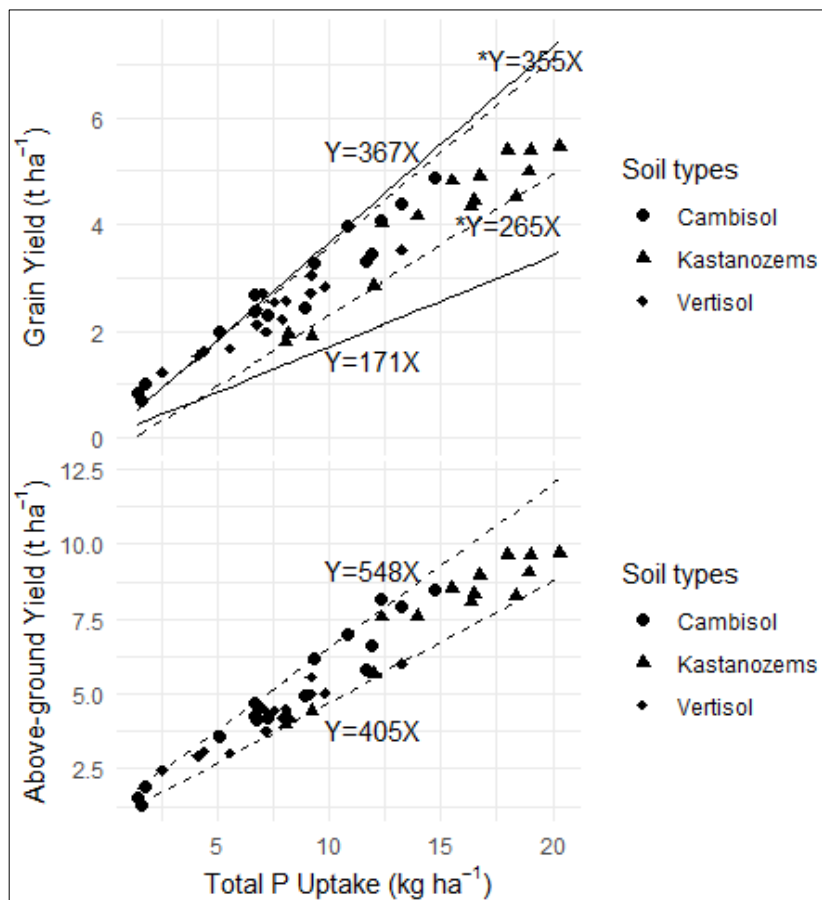
Figure 5.2 illustrates the internal nutrient use efficiencies (NUE) for above-ground biomass and grain yield across different nutrient types (N, P, and K) and soil types. The graph clearly shows how each nutrient is partitioned between biomass

and grain at harvest. Nitrogen (N) and phosphorus (P) are predominantly concentrated in the grain, reflecting their role in reproductive growth and grain filling, while potassium (K) is mainly retained in the biomass, given its critical role in vegetative growth and water regulation.

The distribution of points across the graph highlights significant variability in nutrient use efficiency among soil types. Soils with higher productivity potential tend to have points clustered near the upper boundaries of accumulation, indicating greater nutrient uptake efficiency. In contrast, soils with lower productivity potential fall closer to the lower boundaries, reflecting reduced nutrient use efficiency.

This variability underscores the importance of tailoring fertilizer recommendations based on soil productivity potential. A one-size-fits-all approach, where nutrient requirements are estimated using a uniform target yield across diverse soil types, risks overestimating nutrient needs for less productive soils. This could lead to excessive fertilizer application, increasing the risk of nutrient runoff, environmental pollution, and unnecessary input costs. Conversely, it may also underestimate nutrient requirements for high-productivity soils, potentially limiting yield potential.

Overall, Figure 5.2 emphasizes the need for site-specific nutrient management strategies that consider both soil type and nutrient partitioning to enhance fertilizer efficiency, reduce environmental risks, and optimize crop productivity.



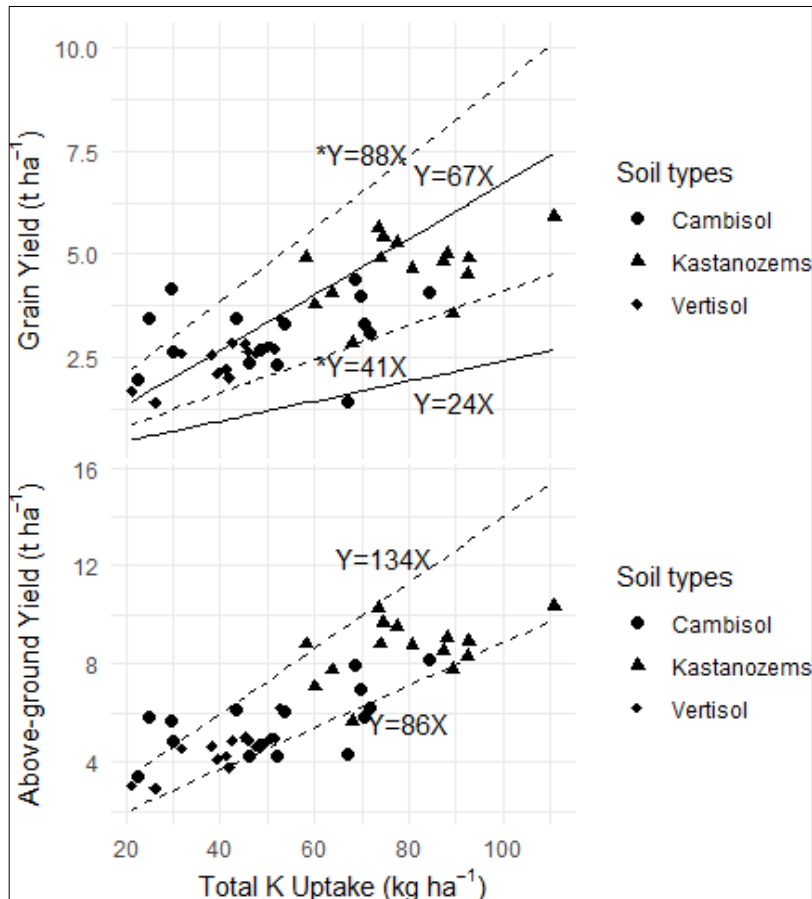


Figure 5.2: Relationships between total plant dry matter or grain yield with total nutrient uptake (N, P or K). The solid lines are from Liu et al. (2006). The dashed lines are the 95 and 5 percentiles from the nutrient study from the three soil types.

5.4.5 Soil-crop-fertilizer model choice for Moroccan rainfed wheat

Following the above analysis of crop response to applied nutrient rates, the subsequent crop modelling for nutrient uptake and yield response were only considered for nitrogen and phosphorus, thereby estimating the relevant model parameters that best explain the optimum level of nutrient use for each soil type.

5.4.5.1 Nutrient uptake models for nitrogen and phosphorus

Crop nutrient uptake, as described in section 5.3.5.2, was modelled in relation to nutrient application, employing a hyperbolic characteristic. The model parameters ("a" and "b") were derived through nonlinear least squares regression using experimental data for each soil type independently. The model fits and parameter estimations for nitrogen and phosphorus are depicted in Figure 5.3 and Figure

5.4, respectively. For soil nitrogen, parameter “b”, representing potential soil contribution to nitrogen uptake (Congreves et al., 2021), was statistically significant at 135 kg N ha⁻¹ for Kastanozems (Table D.6). However, this parameter did not attain statistical significance for Cambisols and Vertisols, likely due to high variability among replicate results, rendering it less informative regarding soil nitrogen's contribution to plant N uptake within the model framework. Conversely, parameter “a”, which describes the linear relationship between nutrient uptake and nitrogen application, was statistically significant with values of 150, 277 and 81 kg N ha⁻¹ for Cambisol, Kastanozems and Vertisol, respectively. The model fit between nitrogen uptake and nitrogen application demonstrated good agreement for Cambisol ($R^2 = 0.69$) and Vertisol ($R^2 = 0.79$), and excellent agreement for Kastanozems ($R^2 = 0.97$). The RMSE for the models were 25.58, 9.64 and 13.92 kg ha⁻¹ nitrogen for Cambisol, Kastanozems and Vertisol, respectively.

Similarly, parameters elucidating the relationship between crop phosphorus uptake and phosphorus applied, as well as, soil phosphorus, were statistically significant, except for parameter “b” of Cambisol and Vertisol (Table D.7). Parameter “a” had values of 18, 17 and 9 kg P ha⁻¹ for Cambisol, Kastanozems and Vertisol, respectively. These findings consistent with soil Olsen P analysis values (Table 5.1) showing a negative correlation, while grain productivity (Table 5.4) exhibited a positive correlation, indicating the potential for phosphorus soil test application using these models for each soil types. The model fit between phosphorus uptake and phosphorus application, as well as soil phosphorus, displayed good agreement for Cambisol ($R^2 = 0.83$) and Kastanozems ($R^2 = 0.78$), and satisfactory agreement for Vertisol ($R^2 = 0.6$). Model performances were affected by variation in replicate readings within each level. With these established relationships, the nutrient uptake from nutrient application could be reliably predicted for an optimum yield targeting for each soil types in rainfed wheat production in Morocco.

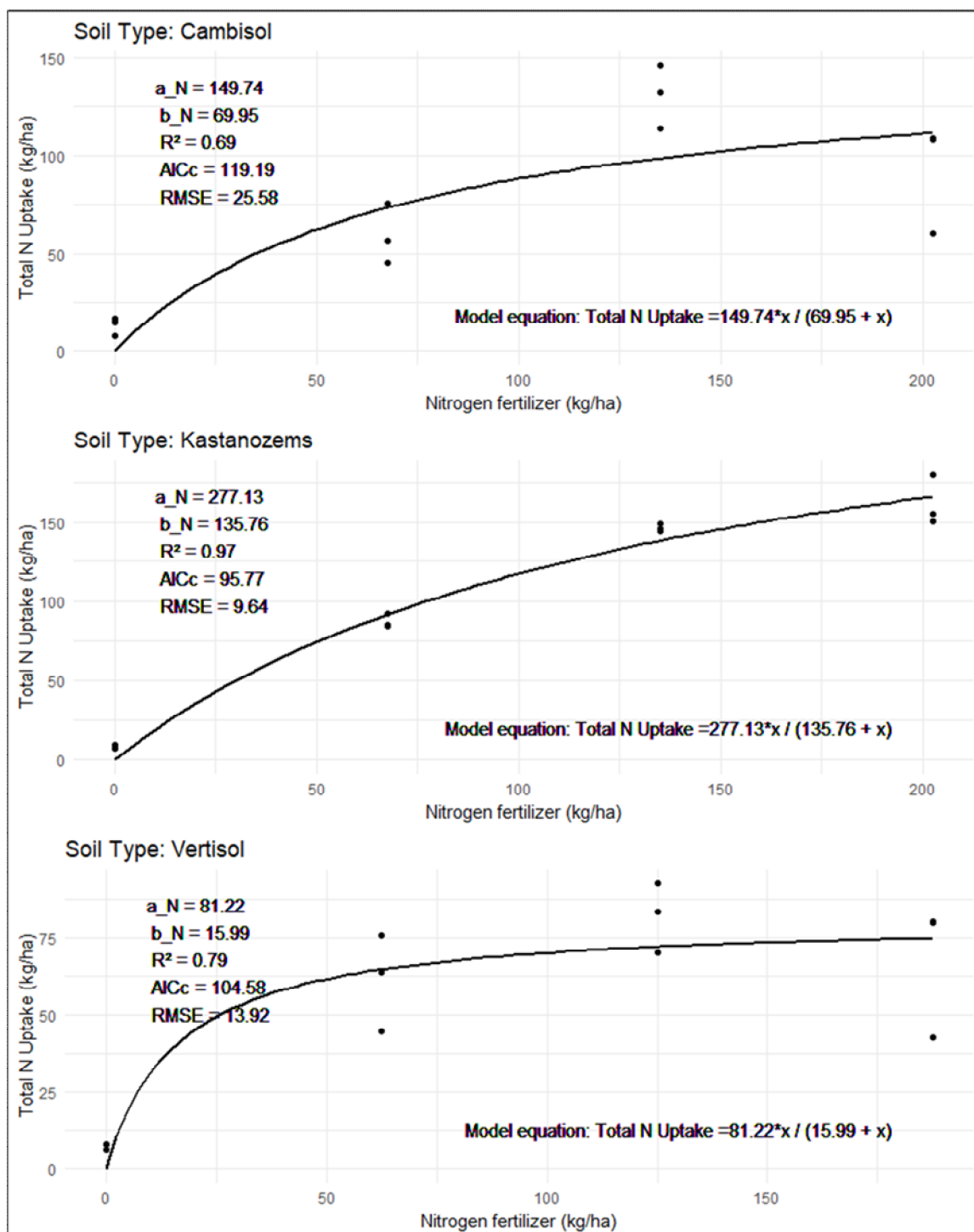


Figure 5.3: A hyperbolic relationship between total nitrogen uptake and application from a nonlinear least squares regression fit. The parameters shown are as defined in section **Error! Reference source not found.**. The statistical shown are R^2 = Coefficient of determination; AICc = corrected version of Akaike information criterion; and RMSE = Root means square error.

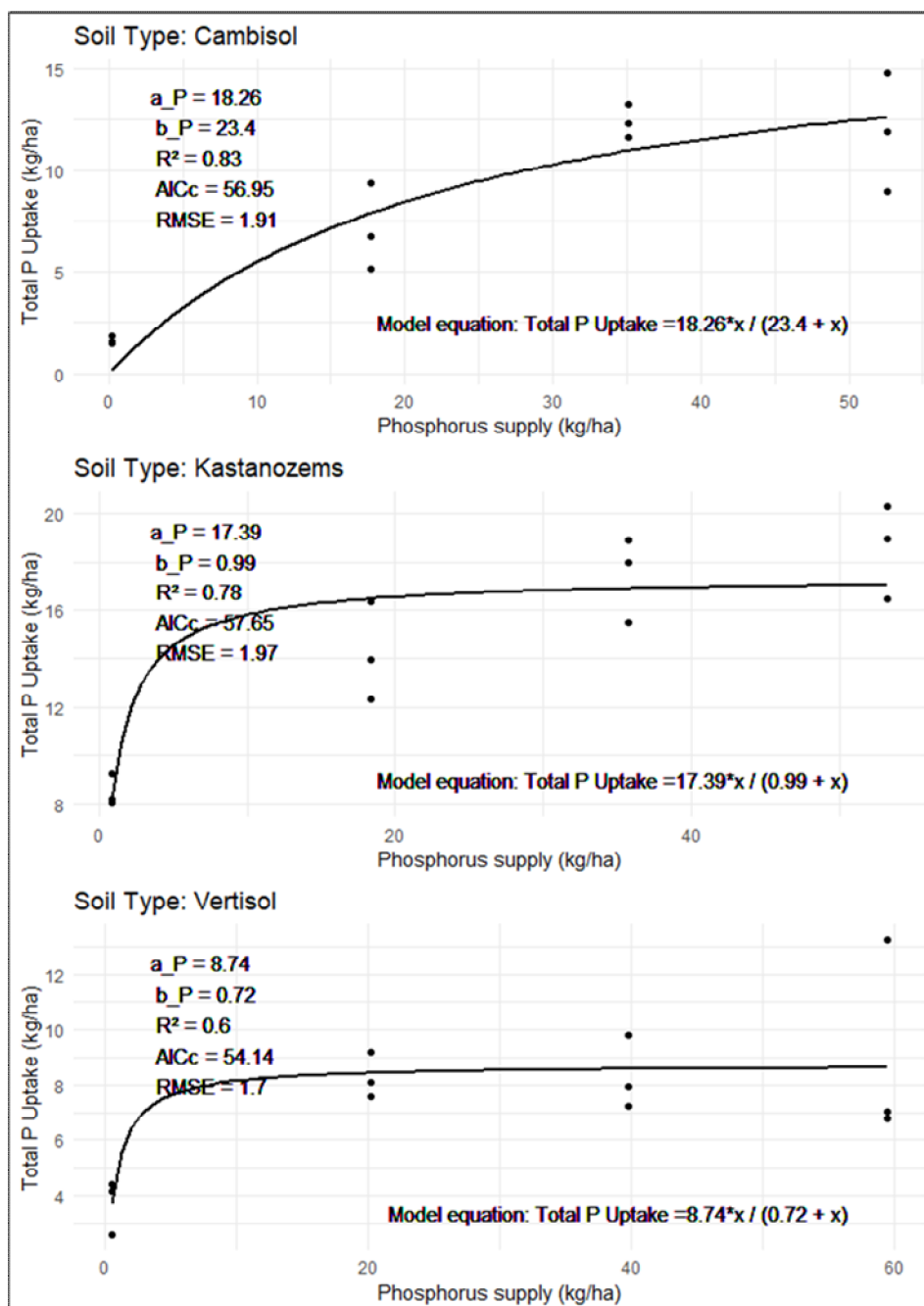


Figure 5.4: A hyperbolic relationship between total phosphorus uptake and application from a nonlinear least squares regression fit. The parameters shown are as defined in section **Error! Reference source not found.**. The statistical shown are R^2 = Coefficient of determination; AICc = corrected version of Akaike information criterion; and RMSE = Root means square error.

5.4.5.2 Crop response models for nitrogen and phosphorus

Model fitting between grain yield and nitrogen and phosphorus application was conducted for all soil types except for phosphorus in Vertisols, utilizing the experimental data and employing a quadratic plateau approach. The model fits and parameter estimates are presented in Figure 5.5 and Figure 5.6 for nitrogen and phosphorus, respectively. The estimated yield for each soil type in nitrogen-omitted plots were 0.31, 0.60 and 0.32 t ha⁻¹ for Cambisol, Kastanozems and Vertisol, respectively, although statistically not significantly different. The linear (“b”) and the quadratic term (“jp”), as outlined in section 5.3.5.1, were statistically significant for all soil types (Table D.8). The model fit between grain yield and nitrogen application exhibited good agreement for Cambisol ($R^2 = 0.69$) and Vertisol ($R^2 = 0.74$), and excellent agreement for Kastanozems ($R^2 = 0.88$). The RMSE of the models were 0.67, 0.61 and 0.43 t ha⁻¹ grain yield for Cambisol, Kastanozems and Vertisol, respectively.

Grain yields for different phosphorus rates were modelled for Cambisol and Katanozems only, as it was not feasible to achieve a best fit for Vertisol. Nonetheless, parameters explaining yield response for phosphorus rates were statistically significant, except parameter “a” for Cambisols (Table D.9). The model fit between grain yield and phosphorus application showed satisfactory agreement for Cambisols ($R^2 = 0.66$) and good agreement for Kastanozems ($R^2 = 0.78$). The RMSE of the models were 0.86 and 0.59 t ha⁻¹ grain yield for Cambisol and Kastanozems, respectively.

Based on these findings and model fits, optimum application rates and expected grain yields were identified for nitrogen for the three soil types, and for phosphorus only for Cambisol and Kastanozem (Table 5.10). The optimum nitrogen rates were determined as 145, 126 and 82 kg N ha⁻¹ for the expected grain yield of 3.33, 4.86 and 2.15 t ha⁻¹ for Cambisol, Kastanozem and Vertisol, respectively. Meanwhile, the optimum phosphorus application rates were 99 and 72 kg P₂O₅ ha⁻¹ for the expected yield of 3.55 and 4.75 t ha⁻¹ for Cambisol and Kastanozem, respectively. This estimation of fertilizer using the model coincide with the soil OlsenP status of the respective soils (Table 5.1).

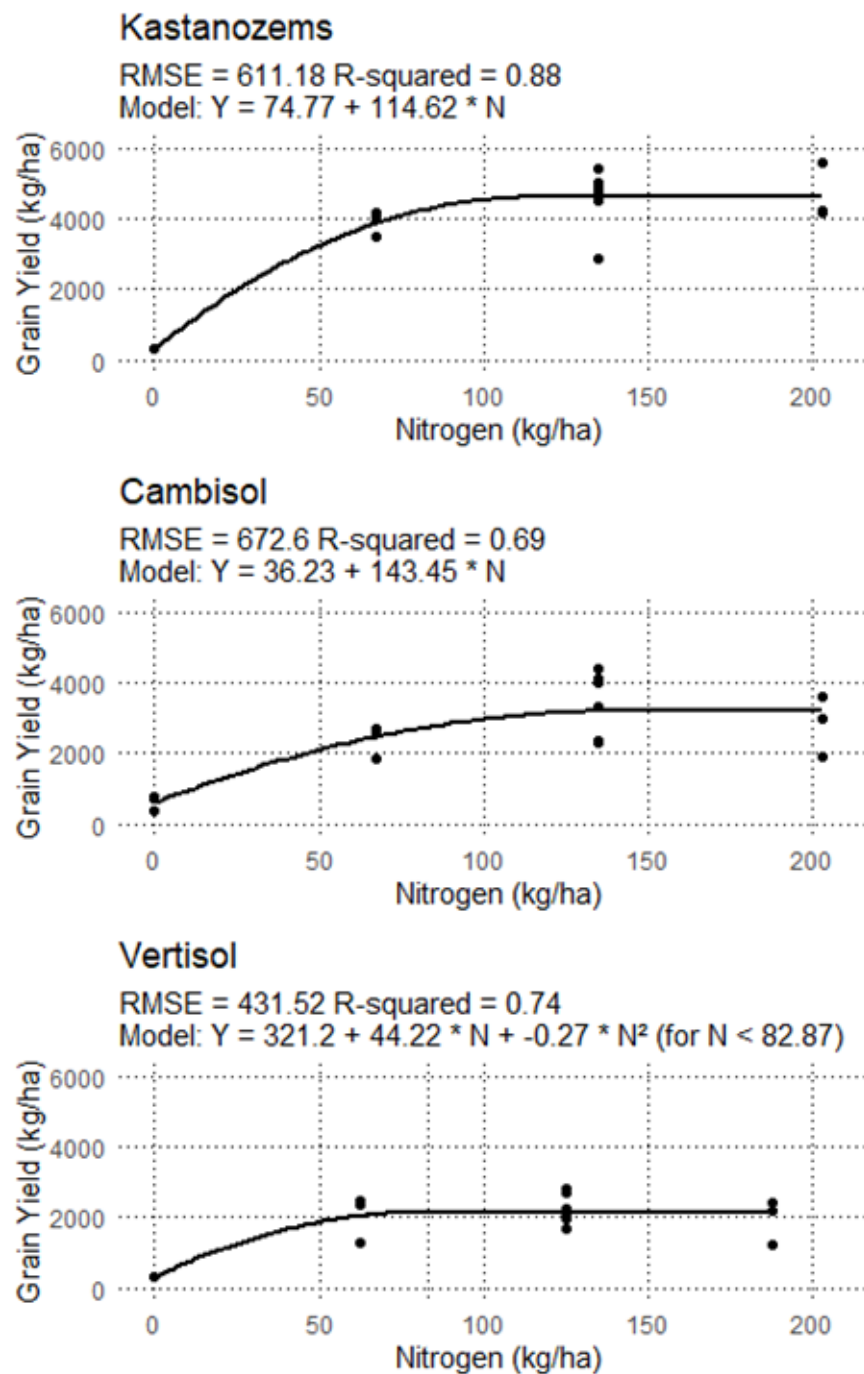


Figure 5.5: A quadratic plateau relationship between grain yield and nitrogen application rates from a nonlinear least squares regression fit. The parameters shown are as defined in section 5.3.5.1. The statistical shown are R^2 = Coefficient of determination; AICc = corrected version of Akaike information criterion; and RMSE = Root means square error.

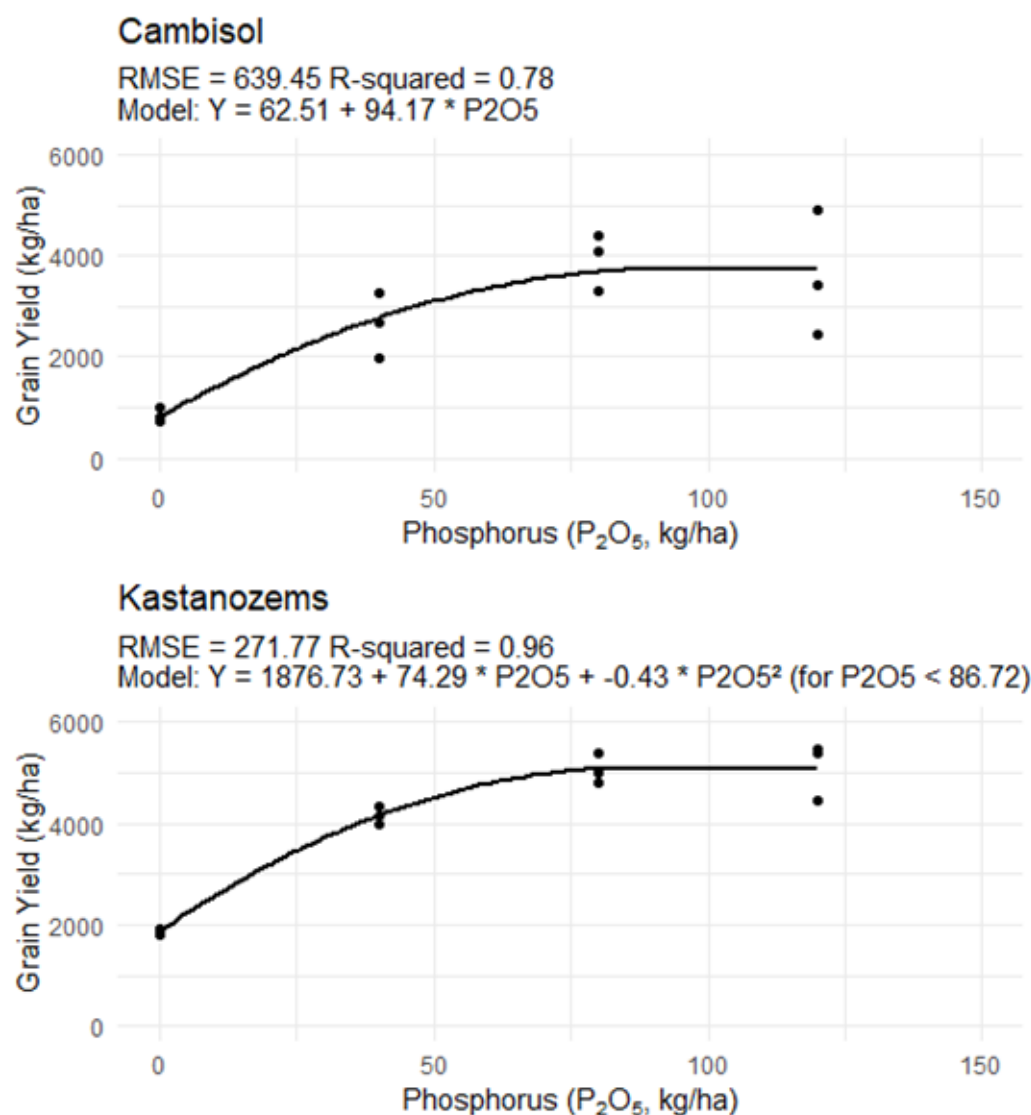


Figure 5.6: A quadratic plateau relationship between grain yield and phosphorus application rates from a nonlinear least squares regression fit. The parameters shown are as defined in section 5.3.5.1. The statistical shown are R^2 = Coefficient of determination; AICc = corrected version of Akaike information criterion; and RMSE = Root means square error.

Table 5.10: Optimum fertilizer rates for predicted yields from response models for Cambisol, Kastanozems and Vertisol.

Soil type	Nitrogen		Phosphorus	
	Optimum rate (kg ha ⁻¹)	Predicted yield (t ha ⁻¹)	Optimum rate (kg ha ⁻¹)	Predicted yield (t ha ⁻¹)
Cambisol	145	3.33	99	3.55
Kastanozems	126	4.86	72	4.75

Vertisol	82	2.15	-	-
----------	----	------	---	---

5.4.6 NIR soil spectroscopy use towards informed decision making in Morocco

The use of soil properties to determine the indigenous nutrient supply capacity (INS) of soil is crucial for understanding nutrient uptake and its subsequent impact on expected productivity and management decisions. Table 5.11 presents the overall mean estimates, along with standard errors, of INS derived from soil organic carbon (SOC), Olsen extractable phosphorus, exchangeable potassium, and pH-H₂O. These properties were predicted using the NIR soil spectroscopy tools and conventional soil analysis methods, for the three soil types utilized in the pot experimentation and subsequent parameter estimation for wheat grain yield and nutrient uptake.

The mean INS values for all plant nutrients obtained from both methods were largely similar, with minimal differences between soil sample results and a promising agreement between the methods, thereby providing confidence in management decision made when employing the NIR soil spectroscopy. Notably, Kastanozems exhibited the highest estimated INS for plant nutrients, followed by Vertisol and Cambisol, respectively. While the estimated INS for nitrogen was similar for both methods, phosphorus INS was higher based on conventional soil analysis for Cambisol and Vertisol, showing differences of 1.66 and 2.86 kg P₂O₅ ha⁻¹, respectively.

Furthermore, the overall performance and reliability of NIR soil spectroscopy in estimating INS were assessed by evaluating the residuals of the estimated INS per samples between the two methods, which is presented in Figure 5.7. The RMSE of the residual for nitrogen INS between the two methods was 13 kg N ha⁻¹, while it was 6.9 kg P₂O₅ ha⁻¹ for phosphorus, indicating a reasonable precision. However, the RMSE of the residual for potassium INS was 170 kg K₂O ha⁻¹.

Overall, the similarity in estimating the INS between the two methods corroborates the reliability of the NIR soil spectroscopic tools and methods to apply for informed decision-making in these soil types and wider Morocco rainfed wheat growing regions. This could alleviate the scarcity of informed decision

making which might improve or supplement the existing production approach regarding soil specific management of fertilizers in Morocco.

Table 5.11: Overall mean and standard error of indigenous nutrient supply capacity of soils estimated by NIR spectroscopy (p) and the conventional soil analysis (c) for the three soil types.

Soil types	Number of samples	SNc (kg ha ⁻¹)	SNp (kg ha ⁻¹)	SPc (kg ha ⁻¹)	SPp (kg ha ⁻¹)	SKc (kg ha ⁻¹)	SKp (kg ha ⁻¹)
Cambisol	7	39.4 (3.25)	40.0 (3.25)	8.12 (0.83)	6.46 (0.62)	209 (31.3)	247 (36.4)
Kastanozems	41	56.2 (3.25)	54.5 (2.24)	10.5 (1.03)	9.82 (0.61)	254 (23.1)	261 (20.2)
Vertisol	17	46.7 (3.82)	42.9 (3.15)	11.6 (2.78)	8.74 (0.91)	331 (66.9)	318 (51.2)

S= soil supply capacity; NPK = nitrogen, phosphorus and potassium, c = employment of conventional soil analysis method; p = employment of the NIR prediction; () = se.

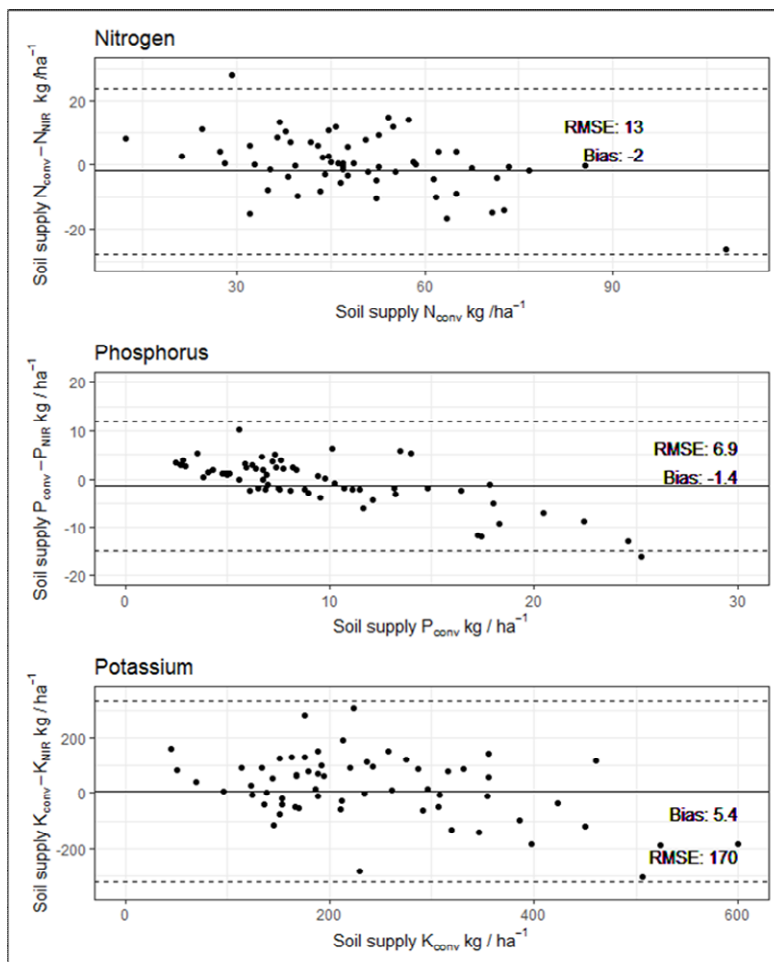


Figure 5.7: Residual in soil supply capacity between conventional method and NIR spectrometer (NeoSpectra), relative to the conventionally determined soil

supply capacity. The dashed horizontal lines present the mean residual with $\pm z^{(\alpha/2)}\sigma$. The horizontal line presents the overall mean residual.

5.5 Discussion

5.5.1 Yield limiting nutrients and nutrient uptake for rainfed wheat growing regions of Morocco

The soils of Morocco generally have high pH values, especially Cambisols and Vertisols, and many topsoil possessed calcic characteristics (Table 5.1; FAO, 2007). Specifically, Vertisols have a high pH with high CEC, and potassium availability was predicted to be low enough to limit wheat productivity (Linguist et al., 2022), therefore favouring higher rates of K for Vertisols than the other two soil types. It was also observed that crop growth was constrained by expanding and contracting characteristics of the Vertisol which might have constrained yields in pots more than it would in field conditions. Due to this, the yield performance and response to applied nutrients might be lower than expected in the case of Vertisols. In respect of yield-limiting nutrients for wheat productivity in Morocco, especially nitrogen and phosphorus had a profound influence across the soil types investigated, limiting wheat yields well below 1.0 t ha⁻¹ irrespective of soil types. And even the low application rate tested gave yield increments up to 2.0 t ha⁻¹, depending on the soil type.

The observed potassium demand and uptake by wheat was high, at a similar level than nitrogen, which indicates that monitoring potassium status in the soil is essential (Dobermann et al., 2003; Sharma et al., 2022). All soils except Cambisols seemed to have large K reserves causing limited and variable K treatment effects on yield and tissue concentrations (Table 5.1; Table D.5). However, it is observed that K depletion might occur in the long run depending on the agricultural system practiced (Rani et al., 2023).

Soil Zinc concentration and availability varied among the three soil types and higher grain zinc concentration were observed in Cambisols compared with the other two soil types, which corroborates the findings reported by Jordan-Meille et al. (2021). Our findings suggest that zinc can increase grain yield production to some extent, and zinc uptake by the crop significantly. Micronutrient depletion in

soils can cause low crop produce quality with respect to micronutrient concentrations but conventional grain production does often not consider micronutrient application as long as there is no clear deficiency (Ryan et al., 1995). However, our results indicate that Zn might be limiting crop growth on some of the Moroccan soils.

5.5.2 Crop models for rainfed wheat growing regions of Morocco

Effective fertilizer management practices are crucial for improving nutrient use efficiency, enhancing productivity, and ensuring favorable economic outcomes for farmers (AHDB, 2017; Liang et al., 2019). The productivity potential of agricultural lands varies significantly due to differences in soil properties, which directly influence crop performance and nutrient availability. Recent studies highlight the importance of examining site-specific nutrient response patterns, recognizing the variability in soil productivity potential within and across farms (Kihara et al., 2016; Haefele and Wopereis, 2005). This variability presents challenges in formulating effective nutrient application strategies that require comprehensive soil data and the development of site-specific relationships. Dobermann and Cassman, (2002) suggested that crop-based estimates of nutrient uptake in nutrient omission trials were preferable to estimates derived from soil testing methods. However, our results indicated that crop-based estimates were significantly lower than actual nutrient uptake when crops were supplied with nutrients, and enhanced growth led to nutrient uptake from the soil beyond the levels observed in omission plots.

To overcome these challenges and improve nutrient management, it is essential to conduct crop response studies alongside thorough soil analysis. Soil characteristics play a dual role in nutrient management by both directly contributing to nutrient uptake and facilitating the availability of applied nutrients. Understanding soil nutrient dynamics enables the calculation of nutrient recovery fractions, which are vital for refining application rates, preventing nutrient imbalances, and minimizing environmental impacts (Haefele et al., 2003; Ezui et al., 2017; Salim and Raza, 2019).

The calculation of nutrient recovery efficiency involves assessing the difference in nutrient uptake between plots receiving nutrients and those without, relative to the amount of nutrient applied (Liang et al., 2019; Rurinda et al., 2020; Ravensbergen et al., 2021). However, an analysis of total nutrient uptake per application rate and soil's inherent nutrient supply revealed that soil also contributes to nutrient uptake, even when nutrients are applied in excess. This may be due to increased nitrogen and phosphorus concentrations in the soil, which can accelerate organic carbon mineralization, thereby releasing additional nutrients not available to crops in the omission plots (Mineau et al., 2013, Rosemond et al., 2015; Griffiths and Johnson, 2018). This finding emphasizes the need to account for the proportional contribution of soil supply in nutrient recovery fraction calculations.

In this study, recovery efficiency was determined by calculating the difference between total nutrient uptake and soil supply capacity, relative to the applied nutrient amount (Dobermann, 2007). This approach addresses the overestimation typically associated with higher nutrient application rates and highlights that nutrient uptake does not increase linearly with nutrient application rates (Moll et al., 1996). A hyperbolic relationship better represents the connection between total nutrient uptake and applied nutrient rates across all soil types. However, the model's fit may not fully capture the observed relationships in every soil type, as the study is pot-based and limited by the number of observation points (nutrient rates considered). This relationship is crucial for understanding the varying nutrient uptake of crops from the soil's inherent supply at different application rates. It is also essential for assessing nutrient surplus, which may lead to environmental contamination if not properly managed.

Other approaches to quantifying nutrient uptake involve a thorough understanding and measurement of agronomic use efficiency with the QUEFTS approach. The internal nutrient use efficiency observed in this study was consistent with findings from other wheat-growing regions (Haefele et al., 2003; Liu et al., 2006), indicating that these parameters for nitrogen (N), phosphorus (P), and potassium (K) could be applicable to wheat production areas in Morocco.

This suggests the potential for extending these insights to optimize nutrient management practices in the region. However, significant variation exists across soil types, with nutrient use efficiency influenced by their inherent productivity potential and nutrient recovery efficiency (Dobermann, 2007).

Integrating nutrient uptake and grain yield models based on specific soil properties allows for tailored fertilizer application recommendations for each soil type and site (Csatho et al., 2009; Fixen et al., 2015). Most notably, Abail et al. (2014) emphasized the reliance on farmers' experience in Morocco and called for the integration of scientific research and technological support to guide decision-making. They further advocated for strengthening the role of agricultural extension services in providing fertilizer-related advice to promote best management practices in Morocco's semi-arid regions. Additionally, increasing fertilizer use in Morocco is crucial to replenish soil nutrients, counteract declining productivity, and address land degradation (Oweis et al., 2011; Mrabet et al., 2012; Epule et al., 2016). This necessitates farmers and advisors to consider the economic and environmental costs of fertilizer use.

5.5.3 Soil spectroscopy for decision support in rainfed wheat growing regions of Morocco

With the rise of Agriculture 4.0, as discussed by Zhai et al. (2020), the main focus is on boosting agricultural productivity through the efficient allocation of resources in response to climate change. Central to this shift is the need for extensive soil and productivity data to enable informed decision-making, striking a balance between economic advantages and environmental concerns. As a result, decisions must be dependable, accessible, and actionable. Fertilizer management, guided by the '4R' principle — right type, right amount, right time, and right placement (Johnston and Bruulsema, 2014) — relies heavily on precise soil information, which is essential for crop modeling and productivity evaluation at both timely intervals and high spatial resolutions.

Given the complex role of soil properties in nutrient uptake and crop response, depending on a single soil property indicator for nutrient management is insufficient. The methods used to determine soil properties and their influence on

different soil systems must be considered (Merrington et al., 2006). To manage this complexity, relational approaches that integrate a variety of soil data, such as soil organic carbon (SOC), pH, clay, and silt content, have been advocated (Janssen et al., 1990; Ravensbergen et al., 2021). However, studying these relationships thoroughly requires soil data that is both readily available and easily accessible.

In this regard, soil spectroscopy presents a promising solution for the cost-effective and widespread quantification of soil properties directly on farms. When properly calibrated, predictive models using soil spectroscopy can accurately estimate multiple soil properties with minimal uncertainty, making it applicable across various agricultural contexts. When combined with robust crop models, these technologies enable stakeholders to make well-informed decisions, addressing the challenges of fertilizer management in the context of climate change.

The study's results demonstrate a strong alignment between the estimated mean indigenous nutrient supply capacity for nitrogen (N), phosphorus (P), and potassium (K), and the indigenous nutrient supply (INS) power in explaining crop uptake and productivity. This indicates the potential of NIR soil spectroscopy as a useful tool for decision support. It is suggested that the experimental results and model insights could be extended to other soil types within the study area, though confirmation from larger datasets is required. Overall, the consistency in INS estimations between conventional methods and NIR soil spectroscopy reinforces the reliability of these tools for informed decision-making, not just for the study's soil types, but also across broader rainfed wheat-growing regions of Morocco, as also noted by Ng et al. (2020) and Asrat et al. (2023). This highlights the potential for informed decision-making to enhance or complement current production strategies tailored to soil-specific fertilizer management in Morocco.

5.6 Conclusions

This study examined the impact of nutrient types and application rates on wheat grain and straw yield, as well as grain density of nitrogen, phosphorus, potassium,

and zinc using major soil types of Morocco and wheat variety. Nitrogen and phosphorus were the most limiting nutrients which limited yield to below 1 t ha⁻¹, seriously restricting wheat productivity. In contrast, potassium did not seem to constrain yields on the soil types considered. Although there was no statistically significant effect on yield of wheat due to the high variability, application of 10 kg ha⁻¹ zinc seemed to improve productivity by up to 1 t ha⁻¹ and resulted in higher concentrations of zinc in wheat grains. Crop response and nutrient uptake model fitting for nitrogen and phosphorus were possible for these soil types with a good model performance to employ them for decision-making. The crop tended to absorb more nutrients from the soil in fertilized plots compared to plots where nutrients were omitted. This suggested that crops have a greater tendency to efficiently absorb nutrients from soils rich in nutrients rather than from poor soils. This finding underscores the need for computing nutrient use efficiency in fertilizer applications. Besides, the NIR soil spectroscopy tools and methods developed for Morocco predicted soil indigenous nutrient supply capacity with a RMSE of 13, 6.9 and 170 kg N, P₂O₅ and K₂O ha⁻¹, respectively. This suggested that there is potential to integrate NIR soil spectroscopy with crop models for better nutrient management for rainfed wheat growing regions of Morocco. However, this study did not validate the precision of the crop and nutrient uptake models and advice generated at farm fields.

5.7 Acknowledgment

Funding: This study was financially supported by OCP Group, as part of the FPO2 “New soil and plant diagnostics tools for better fertilizer recommendations” under collaborative project of Sustainable Agriculture for Africa (SAFA) between Cranfield University, Rothamsted Research and Mohammed VI Polytechnic University.

5.8 Author Contributions

T.G.A., S.M.H, R.S., F.K., R.C. and K.H. conceptualized the study, and acquired and administered project funding. T.G.A. identified soil sampling locations and soil types. H.A. facilitated soil sampling campaign and packaging for sample

export from Morocco to UK. C.T. support in dry spectral analysis, including nutrient density estimation using pXRF. T.G.A. completed the data analyses, including method development, data visualizations and data interpretation. T.G.A wrote the primary draft of the paper with editing and reviewing inputs from other authors.

5.9 Conflict Of Interest

The authors declare that they have no conflict of interest.

5.10 References

Abail, B.Z., Halima, O.I., Boulal, H., Gharous, M. El and Brouj, E. (2014), Fertilizer Use Patterns in the Semi-Arid, Cereal Producing Region of Chaouia, available at:<https://doi.org/10.13140/RG.2.1.2274.0886>.

Acquah, G.E., Hernandez-Allica, J., Thomas, C.L., Dunham, S.J., Towett, E.K., Drake, L.B., Shepherd, K.D., et al. (2022), “Portable X-ray fluorescence (pXRF) calibration for analysis of nutrient concentrations and trace element contaminants in fertilisers”, PLoS ONE, Vol. 17 No. 1 January, pp. 1–20.

AHDB. (2017), “Principles of nutrient management and fertiliser use Nutrient Management Guide (RB209)”.

Alivelu, K., Srivastava, S., Rao, A.S., Singh, K.N., Selvakumari, G. and Raju, N.S. (2003), “Comparison of modified Mitscherlich and response plateau models for calibrating soil test based nitrogen recommendations for rice on typic ustropept”, Communications in Soil Science and Plant Analysis, Vol. 34 No. 17–18, pp. 2633–2643.

Asrat, T.G., Breure, T., Sakrabani, R., Corstanje, R., Hassall, K.L., Hamma, A., Kebede, F., et al. (2024), “A Moroccan soil spectral library use framework for improving soil property prediction : Evaluating a geostatistical approach”, Geoderma, Elsevier B.V., Vol. 452 No. November, p. 117116.

Asrat, T.G., Sakrabani, R., Corstanje, R., Breure, T., Hassall, K.L., Kebede, F. and Haefele, S.M. (2023), “Spectral soil analysis for fertilizer recommendations by coupling with

QUEFTS for maize in East Africa: A sensitivity analysis”, *Geoderma*, Elsevier B.V., Vol. 432 No. February, p. 116397.

Bremner, J.M. (1996), “Nitrogen Total”, *Methods of Soil Analysis: Part 3 Chemical Methods*, pp. 1085–1121.

Ciesielski, H. and Sterckeman, T. (1997), “A comparison between three methods for the determination of cation exchange capacity and exchangeable cations in soils”, *Agronomie*, Vol. 17 No. 1, pp. 9–16.

Congreves, K.A., Otchere, O., Ferland, D., Farzadfar, S., Williams, S. and Arcand, M.M. (2021), “Nitrogen Use Efficiency Definitions of Today and Tomorrow”, *Frontiers in Plant Science*, Vol. 12 No. June, pp. 1–10.

Correndo, A.A., Pearce, A., Bolster, C.H., Spargo, J.T., Osmond, D. and Ciampitti, I.A. (2023), “The soiltestcorr R package: An accessible framework for reproducible correlation analysis of crop yield and soil test data”, *SoftwareX*, Elsevier B.V., Vol. 21, p. 101275.

Csatho, P., Rendas, T.A., Fodor, N. and Nemeth, T. (2009), “Evaluation of Different Fertilizer Recommendation Systems on Various Soils and Crops in Hungary”, *Communications in Soil Science and Plant Analysis*, Vol. 40 No. 11–12, pp. 1689–1711.

Dhakal, C. and Lange, K. (2021), “Crop yield response functions in nutrient application : A review”, *Agronomy Journal*, Vol. 113 No. 6, pp. 5222–5234.

Dobermann, A. (2007), “Nutrient use efficiency – measurement and management”, *Fertilizer Best Management Practices General Principles, Strategy for Their Adoption and Voluntary Initiatives vs Regulations: Papers Presented at the IFA International Workshop on Fertilizer Best Management Practices 7-9 March 2007, Brussels, Belgium*, pp. 1–28.

Dobermann, A. and Cassman, K.G. (2002), “Plant nutrient management for enhanced productivity in intensive grain production systems of the United States and Asia”, *Plant and Soil*, Vol. 247 No. 1, pp. 153–175.

Dobermann, A., Witt, C., Abdulrachman, S., Gines, H.C., Nagarajan, R., Son, T.T., Tan, P.S., et al. (2003), "Estimating indigenous nutrient supplies for site-specific nutrient management in irrigated rice", *Agronomy Journal*, Vol. 95 No. 4, pp. 924–935.

Epule, T.E., Poirier, V., Lafontaine, S., Jemo, M., Dhiba, D., Kechchour, A., Achli, S., et al. (2016), "Recent Trends in the Yield-Nutrient-Water Nexus in Morocco", *Climate Change-Recent Observations*, Vol. IntechOpen, p. 13.

Ezui, K.S., Franke, A.C., Ahiabor, B.D.K., Tetteh, F.M., Sogbedji, J., Janssen, B.H., Mando, A., et al. (2017), "Understanding cassava yield response to soil and fertilizer nutrient supply in West Africa", *Plant and Soil*, Vol. 420 No. 1–2, pp. 331–347.

FAO. (2007), "Digital Soil Map of the World (DSMW)", Rome, Italy, <https://www.fao.org/land-water/land/land-governance/land-resources-planning-toolbox/category/details/en/c/1026564> (Last Access: 27 March 2023), 2007.

Fixen, P., Brentrup, F., Bruulsema, T., Garcia, F., Norton, R. and Zingore, S. (2015), "Nutrient / fertilizer use efficiency : measurement , current situation and trends", *Managing Water and Fertilizer for Sustainable Agricultural Intensification*, Vol. 270 No. January, pp. 1–30.

Griffiths, N.A. and Johnson, L.T. (2018), "Influence of dual nitrogen and phosphorus additions on nutrient uptake and saturation kinetics in a forested headwater stream", *Freshwater Science*, Vol. 37 No. 4, pp. 810–825.

Haefele, S.M. and Wopereis, M.C.S. (2005), "Spatial variability of indigenous supplies for N, P and K and its impact on fertilizer strategies for irrigated rice in West Africa", *Plant and Soil*, Vol. 270 No. 1, pp. 57–72.

Haefele, S.M., Wopereis, M.C.S., Ndiaye, M.K., Barro, S.E. and Ould Isselmou, M. (2003), "Internal nutrient efficiencies, fertilizer recovery rates and indigenous nutrient supply of irrigated lowland rice in Sahelian West Africa", *Field Crops Research*, Vol. 80 No. 1, pp. 19–32.

Hochmuth, G., Hanlon, E. and Overman, A. (2011), "Fertilizer Experimentation, Data Analyses, and Interpretation for Developing Fertilization Recommendations — Examples with Vegetable Crop Research", *Edis*, Vol. 2011 No. 11, available at:<https://doi.org/10.32473/edis-ss548-2011>.

Ichami, S.M., Shepherd, K.D., Sila, A.M., Stoorvogel, J.J. and Hoffland, E. (2019), "Fertilizer response and nitrogen use efficiency in African smallholder maize farms", *Nutrient Cycling in Agroecosystems*, Springer Netherlands, Vol. 113 No. 1, pp. 1–19.

Janssen, B.H. (2011), "Simple models and concepts as tools for the study of sustained soil productivity in long-term experiments. II. Crop nutrient equivalents, balanced supplies of available nutrients, and NPK triangles", *Plant and Soil*, Vol. 339 No. 1, pp. 17–33.

Janssen, B.H., Guiking, F.C.T., van der Eijk, D., Smaling, E.M.A., Wolf, J. and van Reuler, H. (1990), "A system for quantitative evaluation of the fertility of tropical soils (QUEFTS)", *Geoderma*, Vol. 46 No. 4, pp. 299–318.

Janssen, B.H., Sattari, S.Z., Ittersum, M.K. Van, Bouwman, A.F., Smit, A.L. and Janssen, B.H. (2014), "Crop yield response to soil fertility and N , P , K inputs in different environments : Testing and improving the QUEFTS model Field Crops Research Crop yield response to soil fertility and N , P , K inputs in different environments : Testing and improving", *Field Crops Research*, Elsevier B.V., Vol. 157 No. February, pp. 35–46.

John, D.A. and Babu, G.R. (2021), "Lessons From the Aftermaths of Green Revolution on Food System and Health", *Frontiers in Sustainable Food Systems*, Vol. 5 No. February, pp. 1–6.

Johnston, A.M. and Bruulsema, T.W. (2014), "4R nutrient stewardship for improved nutrient use efficiency", *Procedia Engineering*, Elsevier B.V., Vol. 83, pp. 365–370.

Jordan-Meille, L., Holland, J.E., McGrath, S.P., Glendining, M.J., Thomas, C.L. and Haefele, S.M. (2021), "The grain mineral composition of barley, oat and wheat on soils with pH and soil phosphorus gradients", *European Journal of Agronomy*, Elsevier B.V., Vol. 126 No. August 2020, p. 126281.

Kiboi, M.N., Ngetich, F.K., Mucheru-Muna, M.W., Diels, J. and Mugendi, D.N. (2021), "Soil nutrients and crop yield response to conservation-effective management practices in the sub-humid highlands agro-ecologies of Kenya", *Heliyon*, Elsevier Ltd, Vol. 7 No. 6, p. e07156.

Kihara, J., Nziguheba, G., Zingore, S., Coulibaly, A., Esilaba, A., Kabambe, V., Njoroge, S., et al. (2016), "Understanding variability in crop response to fertilizer and amendments in sub-Saharan Africa", *Agriculture, Ecosystems and Environment*, Elsevier B.V., Vol. 229, pp. 1–12.

Lark, R.M., Gillingham, V., Langton, D. and Marchant, B.P. (2020), "Boundary line models for soil nutrient concentrations and wheat yield in national-scale datasets", *European Journal of Soil Science*, Vol. 71 No. 3, pp. 334–351.

Liang, S., Zhang, X., Sun, N., Li, Y., Xu, M. and Wu, L. (2019), "Modeling crop yield and nitrogen use efficiency in wheat and maize production systems under future climate change", *Nutrient Cycling in Agroecosystems*, Springer Netherlands, Vol. 115 No. 1, pp. 117–136.

Lindsay, W.L. and Norvell, W.A. (1972), "Development of a DTPA soil test for zinc, iron, manganese and copper.", *Soil Sci.Soc.Am.Proc.*, Vol. 36, pp. 778–783.

Linquist, B.A., Campbell, J.C. and Southard, R.J. (2022), "Assessment of potassium soil balances and availability in high yielding rice systems", *Nutrient Cycling in Agroecosystems*, Springer Netherlands, Vol. 122 No. 3, pp. 255–271.

Liu, M., Yu, Z., Liu, Y. and Konijn, N.T. (2006), "Fertilizer requirements for wheat and maize in China: The QUEFTS approach", *Nutrient Cycling in Agroecosystems*, Vol. 74 No. 3, pp. 245–258.

Lyons, S.E., Tang, Z., Booth, J. and Ketterings, Q.M. (2019), "Nitrogen response models for winter cereals grown for forage", *Journal of Agronomy and Crop Science*.

Mahdi, S.P.M. (2008), "A comparison of three mathematical models of response to applied nitrogen using lettuce", *Journal of Applied Horticulture*, Vol. 10 No. 1, pp. 73–76.

Maman, G., Idriss, S. and Wortmann, C. (2018), "Crop Yield Response to Fertilizer Relative to Soil Properties in Sub-Saharan Africa", *Soil Science Society of America Journal*, Vol. 82 No. 4, pp. 862–870.

Mamassi, A., Balaghi, R., Devkota, K.P., Bouras, H., El-Gharous, M. and Tychon, B. (2023), "Modeling genotype $\times$ environment $\times$ management interactions for a sustainable intensification under rainfed wheat cropping system in Morocco", *Agriculture and Food Security*, BioMed Central, Vol. 12 No. 1, pp. 1–23.

Merrington, G., Fishwick, S., Barracloguh, D., Morris, J., Preedy, N., Boucard, T., Reeve, M., et al. (2006), *The Development and Use of Soil Quality Indicators for Assessing the Role of Soil in Environmental Interactions*, Science Report SC030265, available at: <https://www.soils.org/publications/sssaj/abstracts/73/6/2078>.

Mineau, M.M., Rigsby, C.M., Ely, D.T., Fernandez, I.J., Norton, S.A., Ohno, T., Valett, H.M., et al. (2013), "Chronic catchment nitrogen enrichment and stoichiometric constraints on the bioavailability of dissolved organic matter from leaf leachate", *Freshwater Biology*, Vol. 58, pp. 248–260.

Moll, R.H., Kamprath, E.J. and Jackson, W.A. (1996), "Analysis and Interpretation of Factors Which Contribute to Efficiency of Nitrogen Utilization".

Mrabet, R., Moussadek, R., Fadlaoui, A. and van Ranst, E. (2012), "Conservation agriculture in dry areas of Morocco", *Field Crops Research*, Elsevier B.V., Vol. 132, pp. 84–94.

Mulvaney, M.J. and Devkota, P.J. (2020), *Adjusting Crop Yield to a Standard Moisture Content*, The Institute of Food and Agricultural Sciences (IFAS), University of Florida, United States., Vol. 4, available at: <https://doi.org/10.32473/edis-ag442-2020>.

Nelson, D.W. and Sommers, L.E. (1996), "Total Carbon , Organic Carbon , and Organic Matter", *Methods of Soil Analysis: Part 3 Chemical Methods*, pp. 961–1010.

Ng, W., Husnain, Anggria, L., Siregar, A.F., Hartatik, W., Sulaeman, Y., Jones, E., et al. (2020), "Developing a soil spectral library using a low-cost NIR spectrometer for precision fertilization in Indonesia", *Geoderma Regional*, Elsevier B.V., Vol. 22, p. e00319.

Nijland, G.O., Schouls, J. and Goudriaan, J. (2008), "Integrating the production functions of Liebig , Michaelis – Menten , Mitscherlich and Liebscher into one system dynamics model", *NJAS*, Vol. 55 No. 2, pp. 199–224.

Njoroge, S., Schut, A.G.T., Giller, K.E. and Zingore, S. (2017), "Strong spatial-temporal patterns in maize yield response to nutrient additions in African smallholder farms", *Field Crops Research*, Elsevier, Vol. 214 No. September, pp. 321–330.

Olsen, S.. (1954), "Estimation of available phosphorus in soils by extraction with sodium bicarbonate", *US Department of Agriculture.*, Vol. 939, pp. 1–20.

Pala, M., Oweis, T., Benli, B., Pauw, E. De, Mourid, M. El, Karrou, M., Jamal, M., et al. (2011), *Assessment of Wheat Yield Gap in the Mediterranean : Case Studies from Morocco , Syria , and Turkey, Aleppo, Syria*, available at: <https://www.researchgate.net/publication/267024765%0AAssessment>.

Poulton, P.R., Johnston, A.E. and White, R.P. (2023), "Response of three cereal crops in continuous arable or ley-arable rotations to fertilizer nitrogen and soil nitrogen at Rothamsted's Woburn Ley-arable experiment", *Soil Use and Management*, Vol. 39 No. 2, pp. 771–784.

Rani, K., Datta, A., Jat, H.S., Choudhary, M., Sharma, P.C. and Jat, M.L. (2023), "Assessing the availability of potassium and its quantity-intensity relations under long term conservation agriculture based cereal systems in North-West India", *Soil and Tillage Research*, Elsevier B.V., Vol. 228 No. December 2022, p. 105644.

Ravensbergen, A.P.P., Chamberlin, J., Craufurd, P., Shehu, B.M. and Hijbeek, R. (2021), "Adapting the QUEFTS model to predict attainable yields when training data are characterized by imperfect management", *Field Crops Research*, Elsevier B.V., Vol. 266 No. March, p. 108126.

Rosemond, A.D., Benstead, J.P., Bumpers, P.M., Gulis, V., Kominoski, J.S., Manning, D.W.P., Suberkropp, K., et al. (2015), "Experimental nutrient additions accelerate terrestrial carbon loss from stream ecosystems", *Science*, Vol. 347 No. 6226, pp. 1142–1145.

Rurinda, J., Zingore, S., Jibrin, J.M., Balemi, T., Masuki, K., Andersson, J.A., Pampolino, M.F., et al. (2020), "Science-based decision support for formulating crop fertilizer recommendations in sub-Saharan Africa", *Agricultural Systems*, Elsevier, Vol. 180 No. December 2019, p. 102790.

Ryan, J., Abdel, M., Dafirl, M., Mergouml, M. and Sali, B. (1995), "Response of improved Moroccan corn cultivars to zinc and phosphorus", *Al Awamia*, Vol. 90 No. September, pp. 69–79.

Salim, N. and Raza, A. (2019), "Nutrient use efficiency (NUE) for sustainable wheat production : a review", *Journal of Plant Nutrition*, Taylor & Francis, Vol. 43 No. 2, pp. 297–315.

Sattari, S.Z., van Ittersum, M.K., Bouwman, A.F., Smit, A.L. and Janssen, B.H. (2014), "Crop yield response to soil fertility and N, P, K inputs in different environments: Testing and improving the QUEFTS model", *Field Crops Research*.

Savvides, A., Corstanje, R., Baxter, S.J., Rawlins, B.G. and Lark, R.M. (2010), "The relationship between diffuse spectral reflectance of the soil and its cation exchange capacity is scale-dependent", *Geoderma*, Elsevier B.V., Vol. 154 No. 3–4, pp. 353–358.

Schut, A.G.T. and Giller, K.E. (2020), "Soil-based, field-specific fertilizer recommendations are a pipe-dream", *Geoderma*, Elsevier, Vol. 380 No. August, p. 114680.

- Sharma, S., Kaur, G., Singh, P., Alamri, S., Kumar, R. and Siddiqui, M.H. (2022), “Nitrogen and potassium application effects on productivity, profitability and nutrient use efficiency of irrigated wheat (*Triticum aestivum* L.)”, *PLoS ONE*, Vol. 17 No. 5 May, pp. 1–29.
- Silva, J.V., Jaleta, M., Tesfaye, K., Abeyo, B., Devkota, M., Frija, A., Habarurema, I., et al. (2023), “Pathways to wheat self-sufficiency in Africa”, *Global Food Security*, Vol. 37 No. 100684, pp. 1–12.
- Smaling, E.M.A. and Janssen, B.H. (1993), “Calibration of QUEFTS, a model predicting nutrient uptake and yields from chemical soil fertility indices”, *Geoderma*, Vol. 59 No. 1–4, pp. 21–44.
- Sonar, K.R. and Babhulkar, V.P. (2002), “Application of Mitscherlich-Bray equation for fertilizer use in wheat”, *Communications in Soil Science and Plant Analysis*, Vol. 33 No. 15–18, pp. 3241–3249.
- Summerauer, L., Baumann, P., Ramirez-Lopez, L., Barthel, M., Bauters, M., Bukombe, B., Reichenbach, M., et al. (2021), “The central African soil spectral library: A new soil infrared repository and a geographical prediction analysis”, *Soil*, Vol. 7 No. 2, pp. 693–715.
- Syers, J.K., Johnston, A. and Curtin, D. (2008), *Efficiency of Soil and Fertilizer Phosphorus Use*, *FAO Fertilizer and Plant Nutrition Bulletin*, Vol. 18.
- The World Bank Group. (2021), *Climate Risk Profile: Morocco*, p. 31.
- Thomas, C.L., Hernandez-Allica, J., Dunham, S.J., McGrath, S.P. and Haefele, S.M. (2021), “A comparison of soil texture measurements using mid-infrared spectroscopy (MIRS) and laser diffraction analysis (LDA) in diverse soils”, *Scientific Reports*, Nature Publishing Group UK, Vol. 11 No. 1, pp. 1–12.
- Tittonell, P., Vanlauwe, B., Corbeels, M. and Giller, K.E. (2008), “Yield gaps, nutrient use efficiencies and response to fertilisers by maize across heterogeneous smallholder farms of western Kenya”, *Plant and Soil*, Vol. 313 No. 1–2, pp. 19–37.

Velayutham, M. (2017), "Soil nutrients and crop Response-Curve types in farmers' fields: Key to balanced fertilizer use and sustainable soil fertility management", *Electronic International Fertilizer Correspondent (e-lfc)*, No. 49, pp. 29–35.

Yost, R.S., Tamimi, Y.N., Silva, J.A., Hue, N. V. and Evensen, C.I. (2000), "How Fertilizer Recommendations Are Made", *Plant Nutrient Management in Hawaii's Soils, Approaches for Tropical and Subtropical Agriculture*. Honolulu, HI: College of Tropical Agriculture and Human Resources, University of Hawaii at Manoa, pp. 87–90.

Zaremba, L.S. and Smoleński, W.H. (2000), "Optimal portfolio choice under a liability constraint", *Annals of Operations Research*, Vol. 97 No. 1–4, pp. 131–141.

Zhai, Z., Martínez, J.F., Beltran, V. and Martínez, N.L. (2020), "Decision support systems for agriculture 4.0: Survey and challenges", *Computers and Electronics in Agriculture*, Elsevier, Vol. 170 No. January, p. 105256.

6 Implications of use of soil spectroscopy and its contribution towards informed decisions in agriculture

The overall aim of the study was to enhance fertilizer decision-making by leveraging soil spectroscopic tools and methods, with a focus on exploring the potential of cost-effective, portable NIR instruments like the Neospectra compared with the MIR bench-top Bruker OPTIK Tensor 27 and the vis-NIR portable FieldSpec-4. However, the findings did not yet provide sufficient support to confidently advance with such instruments for application, as their effectiveness is contingent upon the accuracy of soil property predictions essential for informed decision-making. Additionally, it is important to note that the instruments alone are not sufficient; the collected spectra must be processed using calibrated models derived from a spectral library. Further on, crop models incorporating multiple soil parameters need to be developed or recalibrated for the specific region of interest to fully leverage the advancements in soil spectroscopy for supporting agricultural production. Therefore, the subsequent discussion centres on the progresses achieved by this research and explore potential avenues for future research or validation strategies in the field, with inputs from users and stakeholders.

6.1 Can affordable spectrometers provide prediction accuracies adequate for meaningful fertilizer recommendations?

The evolving landscape of soil spectroscopy is accompanied by the development of diverse spectrometers for different applications and conditions, each of them presenting varied costs and operational flexibility (eg. Table 3.1). These soil scanning instruments, operating within distinct electromagnetic spectrum ranges such as vis-NIR, NIR, or MIR, provide a variety of options, ranging from bench-top setups to portable or handheld devices. These instruments exhibit variations in technical specifications, including light source type, detector, measurement principles, energy intensity, and resolutions, all of which influence their performance across soil properties, soil types, and scanning conditions (2.2.2). While MIR spectroscopy provides distinct absorbance peaks, its complexity, cost,

and infrastructure requirements pose challenges, especially for use in developing countries (Reeves, 2010). Conversely, vis-NIR and further inclusion of only a specific NIR range to reduce cost and weight of scanning instruments, such as the Noespectra excluding wave ranges between 500 – 1350 nm, provide robust methods with minimal soil preparation requirements and offer promising cost-effective alternatives, particularly for in-field applications (Sharififar, 2019). Further comparisons on the predictive accuracy and instrument-specific trade-offs were carried out within this study to understand relationships among the prediction performance and informed decision making.

The comparisons made in this study have exploited advanced statistical tools, such as non-parametric bootstrapping by residuals error propagation (3.3.3; B.1.3; B.1.4), to evaluate the instruments and associated spectra quality while predicting soil properties. Further, the output predictions were coupled with the QUEFTS model to understand their value for informed decision making. All QUEFTS input soil properties, except Olsen P were adequately predicted from spectra obtained from all spectrometers, as indicated by CCC values ranging from 0.77 to 0.96 and RPIQ values between 1.7 to 4.7 (Figure 3.1). Soil pH, which is useful for nutrient supply capacity estimation, was predicted with higher precision and consistency, when compared to other soil properties considered, by the vis-NIR, NIR and the MIR spectrometers. Predictions from the NIR spectrometer resulted in the lowest RPIQ values for SOC, exchangeable K and Total N (2.1, 1.7 and 2.2, respectively) and the values were in the range of model stability and reproducibility (Zelikman and Carmina, 2013). Longer confidence intervals were observed for Olsen P predictions by all soil spectrometers, revealing higher uncertainties in predicting Olsen P, and data points fall far from the 1:1 line as Olsen P values increased. Notably, the deviation between the predicted and measured Olsen P increased above 15 mg P kg⁻¹, irrespective of the spectrometer used. Overall, there was a tendency to underestimate higher values of Olsen P (>15 mg P kg⁻¹) and exchangeable K (>25 cmol kg⁻¹ soil) by all soil spectrometers (Hu et al., 2013). The MIR spectroscopy performed slightly better than the vis-NIR or NIR spectroscopy with regards to prediction accuracy and consistency as was also reported by Barra et al. (2021). However, in general

the NIR equipment tested is the most cost-effective method, reasonably accurate, and with a large capacity for quick measurements (Li et al., 2022). Overall, the results indicated that the NIR spectroscopy predicted the QUEFTS-relevant soil properties with good accuracy, confirming that a site-specific fertilizer recommendation could be derived efficiently and much faster if compared to conventional methods.

Given that, the NIR spectrometer gave a fertilizer estimation deviation range, with an 95% confidence interval, between -8 to 8 kg N ha⁻¹ and -25.4 to 19.5 P₂O₅ ha⁻¹ when compared to the conventional soil analysis methods, which is an acceptable precision range for nitrogen but a considerable deviation for phosphorus (Figure 3.6; Figure 5.7). The MIR spectrometer was the most efficient when compared to the vis-NIR and NIR spectrometers for use in fertilizer recommendation, however its practicality for developing countries is limited by the current instrumentation complexity and cost. And the vis-NIR range spectrometer tested, though portable, was still costly with a minimum or no improvement in prediction performance when compared to the NIR spectrometer. This result ascertained that the cheaper, flexible to use and handheld NIR can contribute a lot in access to soil information for informed decision to improve crop productivity through better fertilizer management, which ultimately improves the livelihood of farmers. Overall, the study advanced the applicability of soil spectroscopy for site-specific decision support which has been proposed to develop categorical nutrient recommendations (Ng et al., 2020).

However, there were challenges for the use of soil spectroscopy predictions related to imprecisions and reliability issues associated with the soil spectral library used to perform this comparison. This is due to the prediction dependency on global soil spectral libraries, whereby local predictions would be more precise and reliable (Lobsey et al., 2017; Padarian et al., 2019; Luce et al., 2022). The results of this study associated these challenges with representativeness of the values in the training set for a specific soil property. This can be caused by a skewed distribution of the conventional soil analysis results whereby the attribute ranges were not always well represented with an adequate sample number in the

training set (Brodský et al., 2013). However, other factors, such as scale dependency of the soil spectra can also contribute to the performance of soil prediction using soil spectroscopy (Savvides et al., 2010; Grunwald et al., 2018).

To summarize, the spectroscopic methods tested were efficient and effective for N fertilizer recommendations, and unsatisfactory in precision, often overestimating the P fertilizer rate on soils with a good P supply. The applicability of soil spectroscopy needs further investigation and consideration of larger and better-balanced calibration sets and/or improving the predictive model approach. The confidence intervals indicated that the uncertainty could be better than a blanket or uninformed decision from environmental, productivity and input cost perspectives. Specifically, the NIR spectrometer would be applicable, being low-cost and providing ease of application as compared with the expensive MIR and vis-NIR spectrometers and resulting in recommendations of nitrogen and phosphorus fertilizer comparable to more expensive methods to generate soil information.

6.2 Soil spectral library spatial dependency can be useful to subset calibration sample selection for precise soil property prediction

Calibration of spectral sensors remains imperative for accurate soil property prediction, typically achieved through conventional soil analysis, and local models perform better than regional or higher scale spectral libraries. However, this heightened precision entails increased costs, complexity, and resource demands, hindering widespread adoption. While complete elimination of local calibration samples is unlikely, it is important to improve the sampling of soil spectral libraries, which requires a thoughtful approach to make better use of this valuable resource. To address this issue, this study endeavoured to develop a spatially optimized soil spectral library tailored for specific farmland soil property prediction, leveraging the Moroccan soil spectral library (MSSL) database, obtained through following a stratified spatially balanced coverage sampling design. The soil sampling locations for the spectral library were based on considering factors like soil type elevation, slope percentage, rainfall, and mean

temperature which influence soil formation and underlying environmental resources (C.1.2; C.1.3). Then, various criteria for calibration sample selection were explored, including spatial autocorrelation of spectra, memory-based learner (MBL), and environmental covariates (4.3.5; 4.3.6; 4.3.7; 4.3.8). Additionally, the soil spectra were captured by two scanning instruments, covering the NIR and MIR ranges, for prediction of several agricultural important soil properties.

The study in Chapter 5 synthesised and utilized the spatial autocorrelation analysis of the spectra PC score vector to optimize calibration sample selection within the MSSL for a specific farmland. This spatial autocorrelation analysis quantified the distance over which pairs of spatial soil spectra signatures were correlated (Webster and Oliver, 2007; Haining et al., 2010). This quantitative description aided in determining the suitable distance for calibration sample entry locations within the MSSL, facilitating the development of fine-tuned models to predict soil properties of the new samples by computing the experimental variograms by the method of moments, considering a maximum distance of 120 km and a lag distance of 10 km (Pebesma and Wesseling, 1998; 4.3.1). Thus, the semi-variogram of the PC vectors delineated the spatial variability and autocorrelation extent, indicating the geostatistical approach's reliability for calibration sample selection (Figure 5.5). Differences observed when employing varying cutoff distances resulted in two notable outcomes: firstly, a shorter cutoff distance of 120 km demonstrated a better fit for spherical variograms, while secondly, a longer cutoff of 200 km revealed emerging variogram characteristics for both NIR and MIR spectra. Additionally, the increased spatial autocorrelation of the PC vectors maybe influenced by factors that vary, thus impacting soil spectra as a multiresolution dataset and influenced by various soil chromophores displaying different spatial scales (Lark and Webster, 1999; Song et al., 2021). Moreover, the study highlighted variations in spatial autocorrelation between the MIR (range = 117 km) and NIR regions (range = 67 km), indicating that longer wavelengths tend to exhibit higher spatial variability (Qiu et al., 2013). This might indicate that a proper understanding of the spectra generated by each spectrometer and wavelength regions, and the spatial variability of the soil

resources is fundamental and ought to be considered for better use of this approach.

Primarily, the spatial selection approach led to enhanced prediction performance for soil property-spectral region combinations that exhibited either poor or satisfactory predictive results when using the entire MSSL, indicating greater reliability (Table 4.2). Notably, there was a substantial improvement of 48.1% and 12.1% in prediction agreement for Olsen phosphorus and Mehlich3 phosphorus by the MIR spectra, respectively. In addition, improvements of 25.4% for CEC, 16.7% for pH-H₂O, 8.1% for total nitrogen and 8.0% for exchangeable magnesium were observed for NIR spectra. The new approach addressed the challenges of computationally effectiveness, decreased noise and model inadequacy by selecting fewer samples representative and determined well-distributed attributes of the soil property for the calibration sets from abundant data entries to build a robust and reliable model without losing the prediction accuracy (Sáiz-Abajo et al., 2005).

Thereby the study brought a new approach and addressed the challenge to determine the best calibration samples from a large and diverse SSL to predict soil properties for any unknown sample which only possess geographic coordinates and soil spectra information. Ultimately, this method facilitated predicting soil properties at the farm level in Morocco, enhancing soil prediction accuracy and employment of soil spectroscopy in general. The spatial sample selection and MBL approaches outperform employing the whole MSSL entry and improve prediction accuracy by fine-tuning calibration sample entries to address local variations. This approach is effective and well assessed for the use of the MSSL, though the knowledge and algorithms obtained from the study could be transferred or applied in wider soil spectroscopy applicability and similar studies. It is suggested that further research on spatial dependency analysis of the specific responsible wavelength regions for a soil property and its spatial variation in any geographic extent is recommended.

6.3 Influence of soil-crop-fertilizer models to quantify indigenous nutrient supply and hence encourage use of soil spectroscopic tools and methods

Crop response models are essential for evaluating crop productivity potential based on applied nutrient rates and combinations, while also considering economic and environmentally sustainable input use within a specific agro-environmental context (Dobermann and Cassman, 2002). However, these models typically do not account for the actual soil nutrient status or the residual nutrients carried over from previous growing seasons. This limitation necessitates the integration of nutrient use efficiency (NUE) relationships to balance soil nutrient supply, fertilizer application, and crop uptake. Addressing this gap is crucial for improving the precision of nutrient management strategies. To achieve these goals, the study in Chapter 5 assessed a dual modeling approach. A quadratic plateau function was utilized to model yield productivity, identifying lower optimal fertilizer rates required to achieve maximum yield while providing more stable and reliable predictions per soil productivity potential (Lyons et al., 2019; Dhakal and Lange, 2021; 5.3.5.1). Additionally, a hyperbolic function was employed to capture the relationship between nutrient use efficiency and varying application rates, demonstrating how incremental nutrient additions affect efficiency (Moll et al., 1996; 5.4.3). By integrating these models, the study showcased in linking the soil's indigenous nutrient supply capacity with crop uptake functions, providing a comprehensive framework for balancing soil supply, fertilizer use, and crop nutrient uptake. This approach enhances the predictive accuracy of crop models, offering a more holistic solution to nutrient management challenges.

The crop response and nutrient uptake models were fitted using a nonlinear least squares regression and demonstrated strong performance in predicting the relationship between wheat grain yield and total nutrient uptake, with R^2 values exceeding 0.7. Utilizing these models, optimal fertilizer rates and potential yields were predicted for the studied soils. The crop tended to absorb more nutrients from the soil in fertilized plots compared to plots where nutrients were omitted. This was concluded from low nutrient uptake in nutrient omitted plots and

exceeding nutrient uptake in nutrient receiving plots beyond the applied nutrient rates (Table 5.4; Table 5.5; Table D.3; Table D.4). This suggested that crops have a greater tendency to efficiently absorb nutrients from soils rich in nutrients than from poor soils. This finding underscores the need in computing nutrient use efficiency in fertilizer application. Besides, the study recognized that the maximum yield QUEFTS estimated, to be used to determine the nutrient demand through on-farm omission trials, might have overestimated or underestimated the specific soil productivity potential (Liang et al., 2019; Rurinda et al., 2020; Ravensbergen et al., 2021).

Thus, this study brought up two crop modelling approaches to complement each other to understand how each level of nutrient could influence crop productivity and total nutrient uptake as a function of soil supply and nutrient application across three distinct soil types of Morocco, representing more than 50% of the rainfed farmland. By doing so, it was possible to address the challenges in setting the attainable yield for a specific region that considers the soil productivity potential using a nutrient level function and integrates the crop nutrient uptake per soil productivity potential to calculate the required nutrient levels (Njoroge et al., 2017). This ensures that the estimated nutrient application for a target yield per soil productivity will employ soil nutrient supply capacity and soil productivity into the nutrient management advice which subsequently utilizes the soil spectroscopic tools and methods proposed in Chapter 3 and Chapter 4. The agreement for the estimated INS between the NIR and the conventional soil analysis methods (Figure 3.6; Figure 3.7; Figure 3.8) showed consistency of the NIR soil spectroscopy approach to be reliable and useful for informed decision-making, not only within the study's soil types but also across broader rainfed wheat-growing regions in Morocco. Besides, this study developed new approaches to optimize nutrient applications based on spectral soil analysis combined with crop models. Further verification of the results, though, are required to improve and finetune the crop modelling approaches and applicability of the NIR estimated soil properties and soil functions (INS) to fully deploy the method for use for the rainfed wheat growing regions of Morocco.

6.4 References

- Barra, I., Haefele, S.M., Sakrabani, R. and Kebede, F. (2021), "Soil spectroscopy with the use of chemometrics, machine learning and pre-processing techniques in soil diagnosis: Recent advances—A review", *TrAC - Trends in Analytical Chemistry*, Elsevier Ltd, Vol. 135 No. 116166, pp. 1–13.
- Brodský, L., Vašát, R., Klement, A., Zádorová, T. and Jakšík, O. (2013), "Uncertainty propagation in VNIR reflectance spectroscopy soil organic carbon mapping", *Geoderma*, Vol. 199, pp. 54–63.
- Dhakal, C. and Lange, K. (2021), "Crop yield response functions in nutrient application : A review", *Agronomy Journal*, Vol. 113 No. 6, pp. 5222–5234.
- Dobermann, A. and Cassman, K.G. (2002), "Plant nutrient management for enhanced productivity in intensive grain production systems of the United States and Asia", *Plant and Soil*, Vol. 247 No. 1, pp. 153–175.
- Grunwald, S., Yu, C. and Xiong, X. (2018), "Transferability and Scalability of Soil Total Carbon Prediction Models in Florida, USA", *Pedosphere*, Soil Science Society of China, Vol. 28 No. 6, pp. 856–872.
- Haining, R.P., Kerry, R. and Oliver, M.A. (2010), "Geography , Spatial Data Analysis , and Geostatistics : An Overview", *Geographical Analysis*, Vol. 42 No. 1, pp. 7–31.
- Hu, G., Sudduth, K.A., Myers, D.B., Agriscience, C. and Nathan, M. (2013), "Factors Affecting Soil Phosphorus and Potassium Estimation by Reflectance Spectroscopy", *American Society of Agricultural and Biological Engineers*, Vol. In 2013 Ka No. April 2016.
- Lark, R.M. and Webster, R. (1999), "Analysis and elucidation of soil variation using wavelets", *European Journal of Soil Science*, Vol. 50 No. 2, pp. 185–206.
- Li, S., Rossel, R.A.V. and Webster, R. (2022), "The cost-effectiveness of reflectance spectroscopy for estimating soil organic carbon", *European Journal of Soil Science*, Vol. 73 No. 1, pp. 1–16.
- Liang, S., Zhang, X., Sun, N., Li, Y., Xu, M. and Wu, L. (2019), "Modeling crop yield and nitrogen use efficiency in wheat and maize production systems under future climate change", *Nutrient Cycling in Agroecosystems*, Springer Netherlands, Vol. 115 No. 1, pp. 117–136.
- Lobsey, C.R., Viscarra Rossel, R.A., Roudier, P. and Hedley, C.B. (2017), "Rs-Local Data-Mines Information From Spectral Libraries To Improve Local Calibrations", *European Journal of Soil Science*, Vol. 68 No. 6, pp. 840–852.
- Luce, S.M., Ziadi, N. and Viscarra Rossel, R.A. (2022), "GLOBAL-LOCAL : A new approach for local predictions of soil organic carbon content using large soil spectral libraries", *Geoderma*, Elsevier B.V., Vol. 425, p. 116048.
- Lyons, S.E., Tang, Z., Booth, J. and Ketterings, Q.M. (2019), "Nitrogen response models for winter cereals grown for forage", *Journal of Agronomy and Crop Science*.
- Moll, R.H., Kamprath, E.J. and Jackson, W.A. (1996), "Analysis and Interpretation of Factors Which Contribute to Efficiency of Nitrogen Utilization".
- Ng, W., Husnain, Anggria, L., Siregar, A.F., Hartatik, W., Sulaeman, Y., Jones, E., *et al.* (2020), "Developing a soil spectral library using a low-cost NIR spectrometer for precision

- fertilization in Indonesia”, *Geoderma Regional*, Elsevier B.V., Vol. 22, p. e00319.
- Njoroge, S., Schut, A.G.T., Giller, K.E. and Zingore, S. (2017), “Strong spatial-temporal patterns in maize yield response to nutrient additions in African smallholder farms”, *Field Crops Research*, Elsevier, Vol. 214 No. September, pp. 321–330.
- Padarian, J., Minasny, B. and McBratney, A.B. (2019), “Transfer learning to localise a continental soil vis-NIR calibration model”, *Geoderma*, Elsevier, Vol. 340 No. January, pp. 279–288.
- Pebesma, E.J. and Wesseling, C.G. (1998), “Gstat: A program for geostatistical modelling, prediction and simulation”, *Computers and Geosciences*, Vol. 24 No. 1, pp. 17–31.
- Qiu, B., Zeng, C. and Chen, C. (2013), “Comparative spatio-spectral heterogeneity analysis using multispectral and hyperspectral airborne images”, *Geo-Spatial Information Science*, Vol. 16 No. 2, pp. 83–90.
- Ravensbergen, A.P.P., Chamberlin, J., Craufurd, P., Shehu, B.M. and Hijbeek, R. (2021), “Adapting the QUEFTS model to predict attainable yields when training data are characterized by imperfect management”, *Field Crops Research*, Elsevier B.V., Vol. 266 No. March, p. 108126.
- Reeves, J.B. (2010), “Near- versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: Where are we and what needs to be done?”, *Geoderma*, Elsevier B.V., Vol. 158 No. 1–2, pp. 3–14.
- Rurinda, J., Zingore, S., Jibrin, J.M., Balemi, T., Masuki, K., Andersson, J.A., Pampolino, M.F., *et al.* (2020), “Science-based decision support for formulating crop fertilizer recommendations in sub-Saharan Africa”, *Agricultural Systems*, Elsevier, Vol. 180 No. December 2019, p. 102790.
- Sáiz-Abajo, M.J., Mevik, B.H., Segtnan, V.H. and Næs, T. (2005), “Ensemble methods and data augmentation by noise addition applied to the analysis of spectroscopic data”, *Analytica Chimica Acta*, Vol. 533 No. 2, pp. 147–159.
- Savvides, A., Corstanje, R., Baxter, S.J., Rawlins, B.G. and Lark, R.M. (2010), “The relationship between diffuse spectral reflectance of the soil and its cation exchange capacity is scale-dependent”, *Geoderma*, Elsevier B.V., Vol. 154 No. 3–4, pp. 353–358.
- Sharififar, A. (2019), “Evaluating a low- - cost portable NIR spectrometer for the prediction of soil organic and total carbon using different calibration models”, *Soil Use and Management*, Vol. 35 No. 4, pp. 607–616.
- Song, Y., Shen, Z., Wu, P. and Rossel, R.A.V. (2021), “Wavelet geographically weighted regression for spectroscopic modelling of soil properties”, *Scientific Reports*, Nature Publishing Group UK, No. 0123456789, pp. 1–11.
- Webster, R. and Oliver, M.A. (2007), *Geostatistics for Environmental Scientists: Second Edition*, *Geostatistics for Environmental Scientists: Second Edition*, available at: <https://doi.org/10.1002/9780470517277>.
- Zelikman, E. and Carmina, E. (2013), “The spectral response characteristics of the soils and their possible estimation by using partial least square regression (PLSR) analysis”, *International Journal of Geomatics and Geosciences*, Vol. 3 No. 3, pp. 438–453.

7 Conclusion and future research directions

7.1 Overall Conclusion

This research has significantly advanced our understanding of the integration of soil spectroscopy with fertilizer decision-making, particularly in the context of Morocco's rainfed wheat-growing regions. By evaluating various spectroscopic techniques — such as vis-NIR, NIR, and MIR—this study has highlighted their effectiveness, efficiency, and applicability in predicting key soil properties necessary for informed fertilizer management. The integration of spectral data with crop nutrient uptake models, such as QUEFTS, has shown promise in offering site-specific fertilizer recommendations that could lead to improved environmental sustainability and agricultural productivity.

Objective 1 aimed to evaluate handheld spectral tools and methods for in-field soil analysis. The study found that while portable vis-NIR and MIR spectrometers provided high-quality predictions, handheld NIR spectrometers offered a more accessible and cost-effective solution, especially for low-income regions. This work contributes to understanding how affordable spectrometers can be effectively utilized in agricultural decision-making.

Objective 2 focused on optimizing spectral libraries for the prediction of soil variables at different scales. The findings showed that large spectral libraries could benefit from sub-setting methods to refine calibration samples, enhancing prediction accuracy. These methods were especially useful for improving predictions for specific soil properties like OlsenP and Exch.K, underscoring the importance of spatially and contextually informed sampling strategies.

Objective 3 involved testing spectral methods for rapid and economical identification of crop-available nutrients. The study demonstrated that NIR soil spectroscopy could predict soil nutrient supply capacity, offering a potential tool for efficiently managing nutrient applications. This aligns with the broader goal of improving nutrient use efficiency in fertilizer applications, especially for nitrogen and phosphorus, which are critical for wheat productivity in Morocco.

Objective 4 examined the development and application of spectral-based models for fertilizer recommendations. The research confirmed that integrating spectral data with crop models like QUEFTS could generate more precise fertilizer recommendations, thus optimizing nutrient management practices and potentially reducing environmental impacts through site-specific model parameter estimation.

7.2 Research Impact and Future Directions

This study makes a meaningful contribution to the growing body of research on soil spectroscopy in agriculture. It underscores the potential for integrating low-cost, portable spectroscopic tools with crop models to improve nutrient management in resource-limited settings. The research also highlights gaps in current knowledge, particularly regarding the integration of soil spectroscopy with crop models for precise, site-specific recommendations. Future studies could further explore the accuracy of nutrient estimates generated by these techniques through crop performance data in the field, and evaluate their applicability across different regions and soil types.

Moreover, as soil spectral data offer a wealth of information, further research should focus on optimizing spectral libraries and refining calibration methods to improve prediction accuracy. Investigating the spatial and temporal variability of soil properties and their spectral signatures at finer scales will be crucial for enhancing the precision of future soil management recommendations. Additionally, exploring the potential of combining various spectroscopic techniques (e.g., vis-NIR and MIR) with emerging machine learning models could provide even more robust and scalable solutions for precision agriculture.

Ultimately, this work sets the stage for future research that not only refines the tools and methodologies for soil property prediction but also evaluates the real-world impact of these innovations on agricultural productivity and environmental sustainability.

Appendices

Appendix A : Ethical approval letter



14 March 2020

Dear Mr Asrat ,

Reference: CURES/9715/2020

Title: Dry Spectral Diagnostic Tools and Methods for Precise Fertilizer Application

Thank you for your application to the Cranfield University Research Ethics System (CURES).

We are pleased to inform you your CURES application, reference CURES/9715/2020 has been reviewed. You may now proceed with the research activities you have sought approval for.

If you have any queries, please contact CURES Support.

We wish you every success with your project.

Regards,

CURES Team

Appendix B : Spectral soil analysis for fertilizer recommendations by coupling with QUEFTS for maize in East Africa: a sensitivity analysis

B.1 Materials and methods

B.1.1 Removing outliers

There were four outliers (#4) detected which were far from the rest of the dataset (#346) in the PC1 and PC2 score plots. Hence, we removed these outliers and used #346 samples for this study (Figure B-1).

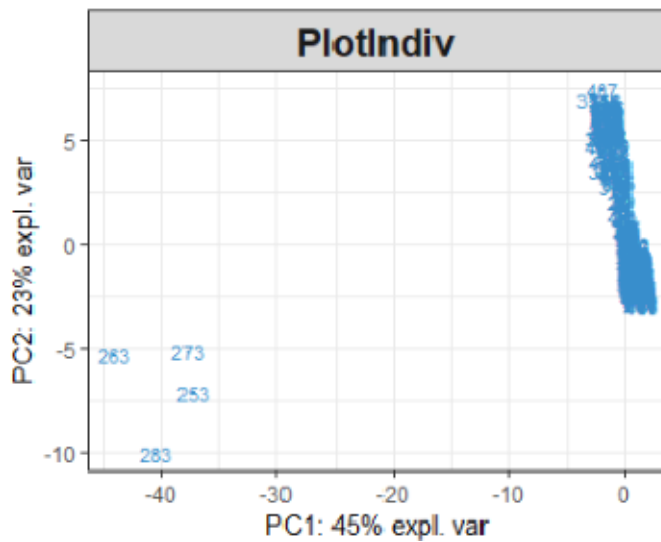


Figure B-1: Outlier detection from PC score plots of PC1 and PC2. The numbers are the sample codes.

B.1.2 Sub-setting of the dataset into model calibration and independent test sets

The dataset was split into model calibration and independent test sets by three sampling methods to get a better representation of the variation contained in the spectra space from the respective instruments. The sampling methods compared were 1) Kennard-Stone and 2) conditioned Latin Hypercubic Sampling (cLHS) which used the PCA scores and 3) the multivariate soil property using Euclidean distance in the attribute space of the considered soil properties. In all the sampling methods, we used the ratio of 75 to 25% split of calibration and

independent test set, respectively. The representation of the calibration and independent test sets were evaluated graphically of the distribution of scatter points in the PC1 vs. PC2 score plot which contributed for the superior variation among the spectra components (Figure B-2). cLHS gave a better representation of the calibration and independent test sets in the first two PC score plot across the instruments which we considered the best sample splitting method to evaluate the prediction performance of the instruments for the QUEFTS input soil properties (pH, Exch.K, SOC, TotalN, OlsenP) and fertilizer recommendation estimation.



Figure B-2: Model calibration and independent test sets sample distribution at the first two PC score space using cLHS sampling.

B.1.3 Derivation of the leverage and residuals for the calibration set

The leverage (h) for the calibration set values was calculated following:

$$h = t_p(T'T)^{-1}t_p' \quad (\text{B.1})$$

Where T were the scores of the factors in the calibration set and t_p the scores of the factors for a prediction sample (in this case also the calibration set). The residuals of the calibration set were then computed as:

$$\epsilon_j = \frac{y_{cj} - \hat{y}_{cj}}{\sqrt{1 - h_j}} - \mu_\epsilon \quad for \quad j = 1, \dots, n. \quad (B.2)$$

Where y_{cj} was the fitted response from the PLS model for the calibration set for position j within and each residual is subtracted with the overall mean μ_ϵ .

B.1.4 Derivation of standard deviation from PLSR calibration to QUEFTS fertilizer recommendation model for N and K

The standard deviations of the soil nitrogen supply (SN_j) is given by:

$$\sigma_{SN_j} = \sqrt{\left(\frac{\sigma_{\widehat{SOC}}^2}{\widehat{SOC}^2}\right)} SN_j \quad (B.3)$$

where $\sigma_{\widehat{SOC}}$ is the standard deviation of the estimated soil organic carbon content from the PLS regression.

The standard deviation for the first term in Equation 4 in the main manuscript ($\zeta = fK_j500\widehat{K_{exch}}$) is given by:

$$\sigma_\zeta = \sqrt{\left(\frac{\sigma_{\widehat{K_{exch}}}^2}{\widehat{K_{exch}}^2}\right)} \zeta \quad (B.4)$$

and thus, the standard deviation for the soil potassium supply by:

$$\sigma_{SK_j} = \sqrt{\left(\frac{\sigma_\zeta^2}{\zeta^2} + \frac{\sigma_{\widehat{SOC}}^2}{\widehat{SOC}^2}\right)} SK_j \quad (B.5)$$

B.2 : Results

B.2.1 Descriptive statistics of the calibration and independent test sets across the sampling methods and sample origins

The highest skewedness was observed for OlsenP (3.27) followed by exchangeable K (2.22) for the whole dataset (data not shown) when the sample set was split based on the spectra score space using the cLHS sampling (Table B.1). Consequently, the mean and standard deviation of OlsenP was much lower for the calibration set than for the independent test set. Such statistical distribution variation among the calibration and test set could affect the prediction performance for OlsenP. However, the mean of the soil attributes for all properties, except OlsenP, were close to each other or equal for the calibration and test sets based on sample sub-setting using the spectra score space. Besides, the distribution of all other soil variables in the calibration and independent test sets were close to each other.

Table B.1: Descriptive statistics of the soil property attributes across countries, model calibration and independent test set subset by cLHS.

Soil property (unit)	Origin	Sample set	N	Min	Median	Mean	Max	sd
Exchangeable K (cmol kg ⁻¹ soil)	Ethiopia	Calibration set	10	1.02	2.28	3.12	7.77	2.18
		Independent test set	5	0.57	2.56	2.62	4.79	1.87
	Kenya	Calibration set	191	1.28	3.27	3.81	11.50	2.08
		Independent test set	64	1.23	3.03	3.42	10.79	1.96
	Tanzania	Calibration set	59	9.11	15.03	16.17	31.82	5.01
		Independent test set	17	11.15	16.68	19.81	45.83	9.87
OlsenP (mg kg ⁻¹ soil)	Ethiopia	Calibration set	10	2.44	3.88	4.05	6.56	1.32
		Independent test set	5	3.48	5.17	5.02	6.24	1.23
	Kenya	Calibration set	191	1.36	7.80	10.85	68.46	9.87
		Independent test set	64	1.50	5.66	8.84	35.62	7.84
	Tanzania	Calibration set	59	0.42	3.54	9.28	89.96	17.12
		Independent test set	17	1.70	4.34	10.02	50.96	13.90
pH	Ethiopia	Calibration set	10	4.91	5.39	5.45	6.27	0.44

		Independent test set	5	4.68	5.45	5.70	7.21	1.12
	Kenya	Calibration set	191	4.48	5.35	5.42	7.35	0.55
		Independent test set	64	4.37	5.23	5.30	6.53	0.46
	Tanzania	Calibration set	59	6.79	8.19	8.12	8.89	0.48
		Independent test set	17	7.02	8.09	8.09	8.76	0.40
SOC (%)	Ethiopia	Calibration set	10	1.27	1.51	1.58	2.17	0.29
		Independent test set	5	1.46	1.63	1.72	2.17	0.33
	Kenya	Calibration set	191	0.50	1.37	1.38	2.36	0.44
		Independent test set	64	0.53	1.46	1.45	2.51	0.49
	Tanzania	Calibration set	59	0.61	1.41	1.44	2.45	0.34
		Independent test set	17	1.16	1.43	1.49	1.86	0.21
Total N (%)	Ethiopia	Calibration set	10	0.11	0.13	0.14	0.20	0.03
		Independent test set	5	0.10	0.14	0.14	0.20	0.04
	Kenya	Calibration set	191	0.04	0.11	0.11	0.19	0.03
		Independent test set	64	0.05	0.12	0.11	0.18	0.03
	Tanzania	Calibration set	59	0.08	0.10	0.10	0.17	0.02
		Independent test set	17	0.07	0.11	0.11	0.15	0.02

B.2.2 Loadings of the first two components

There were distinct variations and contributions of few wavelength regions for the first two PC components for each soil property across the respective soil scanning instruments. There were somehow similarities of the graphs between pH and Exch.K for a specific spectrum recording instrument (Figure B-3 and Figure B-4). And such similarities also were observed between OlsenP, SOC and TotalN (Figure B-5, Figure B-6 and Figure B-7). Hence, the soil P determination using soil spectroscopy could be partly related to the soil organic matter regions in the soil spectra analysis.



Figure B-3: Loadings of the first two PCs for pH across the instruments.



Figure B-4: Loadings of the first two PCs for exchangeable potassium across the instruments.



Figure B-5: Loadings of the first two PCs for OlsenP across instruments.



Figure B-6: Loadings of the first two PCs for SOC across instruments.



Figure B-7: Loadings of the first two components for Total N across instruments.

Appendix C : A Moroccan soil spectral library use-framework for improving soil property diagnostics: Evaluating a geostatistical approach

C.1 Materials and methods

C.1.1 Area coverage of dominant FAO soil types

Table C.1: Area coverage of dominant FAO soil types from 2017/2018 cropping season crop cover map.

Soil type	Cropped area coverage (km ²)	Percent
Calcic Kastanozems	13729	24.9
Calcic Cambisols	8788	15.9
Chromic Luvisols	8067	14.6
Vertisols	5765	10.5
Calcic Xerosols	5666	10.3
Eutric Planosols	3098	5.6
Luvic Phaeozems	2941	5.3
Eutric Gleysols	1973	3.6
Calcic Fluvisols	1718	3.1
Others	3397	6.2

C.1.2 Stratified balanced coverage sampling (StrBCS) design

We used equal inclusion of probabilities for all points whereby π estimate of the mean ((C.1) is equal to the sample mean Grafström and Tillé (2013).

$$\hat{V}(\hat{z}) = \frac{n}{n-p} \frac{p}{p+1} \sum_{k \in S} (1 - \pi_k) \left(\frac{e_k}{\pi_k} - \bar{e}_k \right)^2 \quad (\text{C.1})$$

with p the number of balancing variables, $\pi_k = n/N$, $k = 1, \dots, N$, e_k the regression model residual of unit k (Equation 2), and $\bar{e}_k$ is the local mean of the residuals of this unit, computed by (C.2.

$$e_k = z_k - x_k^T \hat{b} \quad (\text{C.2})$$

with x_k a vector of length p with the balancing variables for unit k , and $\hat{b}$ the estimated population regression coefficients, given by (C.3): -

$$\hat{b} = \left(\sum_{k \in S} c_k \frac{x_k x_k^T}{\pi_k \pi_k} \right)^{-1} \sum_{k \in S} \sum_{k \in S} c_k \frac{x_k z_k}{\pi_k \pi_k} \quad (\text{C.3})$$

Where c_k a weight for unit k , one of which is $c_k = (1 - \pi_k)$

$$\bar{e}_k = \frac{\sum_{k=1}^{p+1} (1 - \pi_k) \frac{e_k}{\pi_k}}{\sum_{k=1}^{p+1} 1 - \pi_k} \quad (\text{C.4})$$

C.1.3 Soil sampling environmental covariates

Spatial covariates considered for selecting representative soil sampling locations for the above-mentioned objectives included: -

1) Five years mean (2015-2019) annual rainfall with a 5km spatial resolution: - Yearly rainfall was downloaded from Climate Hazards Centre InfraRed Precipitation with Station data (CHIRPS) at https://data.chc.ucsb.edu/products/CHIRPS-2.0/africa_dekad/tifs/ for period of 2015 to 2019 for the Africa continent. The resolution of the dataset was a 5km-by-5km grid. The mean value of the five years annual rainfall was used as a covariate in the sampling point selection criteria.

2) FAO soil unit/dominant soil types with 1:5,000,000 scale: - The digital soil map of the world dataset was downloaded to have a prior knowledge on possible soil types in cropped lands of Morocco. The dataset interactive to GIS environment was selected from <http://www.fao.org/geonetwork/srv/en/metadata.show?id=14116> .

3) Digital Elevation Model (DEM)-30m: - One hundred fourteen tiles of 30m resolution DEM were download for Morocco from <https://earthexplorer.usgs.gov/> . The 90m resolution DEM was downloaded for those tiles with no data/holes/ for their 30m DEM (e.g. for 31_08 tile) and resampled/downscaled to 30m DEM using Arc GIS Pro software. The tiles were merged using the 'mosaic to new raster' function at the raster processing tool.

4) Slope percentage and Elevation derived from 30m DEM: - After transforming the coordinate system from geographic WGS 1984 to projected UTM zone 29N of the clipped raster using the Morocco boundary shapefile, slope and elevation characteristics were generated from the DEM using the spatial analyst tool by using the respective functions.

5) Five years annual minimum and maximum temperature with 50 km spatial resolution: - this dataset was downloaded as '.NC' file from <https://downloads.psl.noaa.gov> website. The file was arranged in a way that it holds multidimensional matrix of data (location, time, and daily temperatures vz., maximum and minimum temperatures). The maximum and minimum temperatures were extracted using the 'ncdf4' and 'fields' packages in R. The average temperature was then calculated from the maximum and minimum daily temperatures.

6) Geographic spread. The sampling locations will be selected based on their presence in space and optimized by the geographic spread.

C.1.4 Soil scanning instrument description

A Bruker Tensor 27 bench-top instrument (Tensor II) was used to collect the spectral data in the MIR region (Bruker Tensor 27; Bruker Optik GmbH, Ettlingen, Germany) using a nitrogen purged integrating sphere to cool the MCT (mercury cadmium telluride) mid-band detector. The high throughput screening accessory (HTS-XT), which scans 95 samples in one plate, was used with a spectral resolution of 4 cm⁻¹ and scan time of 32 s per sample.

NeoSpectra Micro (Neospec) is a hand-held Near InfraRed (NIR) spectral sensor manufactured by Si-ware, containing a single-chip Michelson interferometer with monolithic opto-electro-mechanical (OEM) structure (Si-ware, 2017). It has a wavelength resolution of 16 nm and a wavelength accuracy of ± 1.5 nm. The sample scanning was performed with the settings set to boxcar, 3N zero padding and a 10 second integration resulting in 1024 data points per spectrum using linear interpolation. Three replicate measurements per sample were recorded by moving the scanning spot over the soil sample in a petri dish. The spectral

signatures were collected from airdried and sieved samples (2 mm stainless steel sieve).

C.1.5 Selection of spectra pre-processing methods

Compute dissimilarities and select moving-window based on nearest-neighbor RMSE.

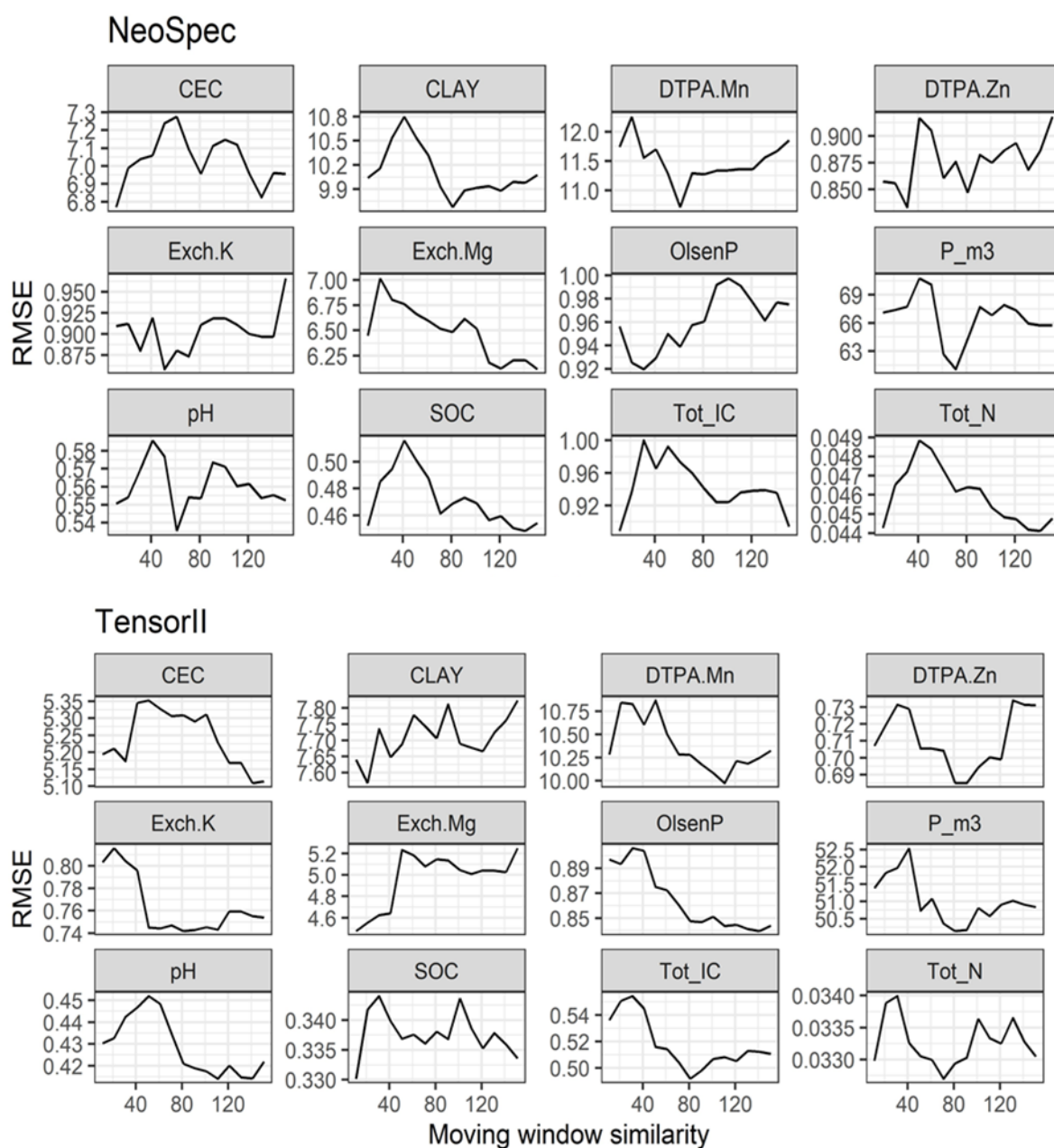


Figure C-1: Moving window size with the RMSE of calibration across soil properties and soil scanning instruments (Neospec -> Neospectra; Tensor-II -> Bruker Tensor 27).

C.1.6 Covariate cluster analysis

Table C.2: Cluster size of training and independent test sets based on nearest neighbours clustering from environmental covariates.

Subsets	Cluster groups					Total
	1	2	3	4	5	
Train set	119	48	156	91	110	119
Test set	18	8	22	12	15	18

C.2 Results

C.2.1 Soil texture distribution in the textural triangle

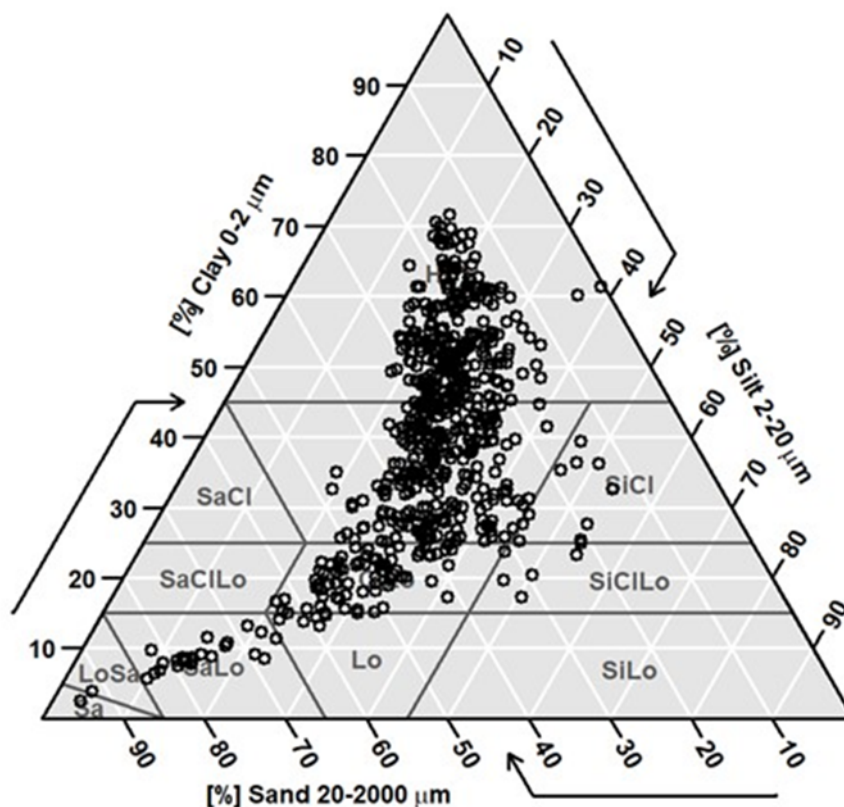


Figure C-2: Soil textural triangle of the MSSL according to ISSS classification.

C.2.2 Descriptive statistics of the calibration and independent test sets across the sampling methods and sample origins

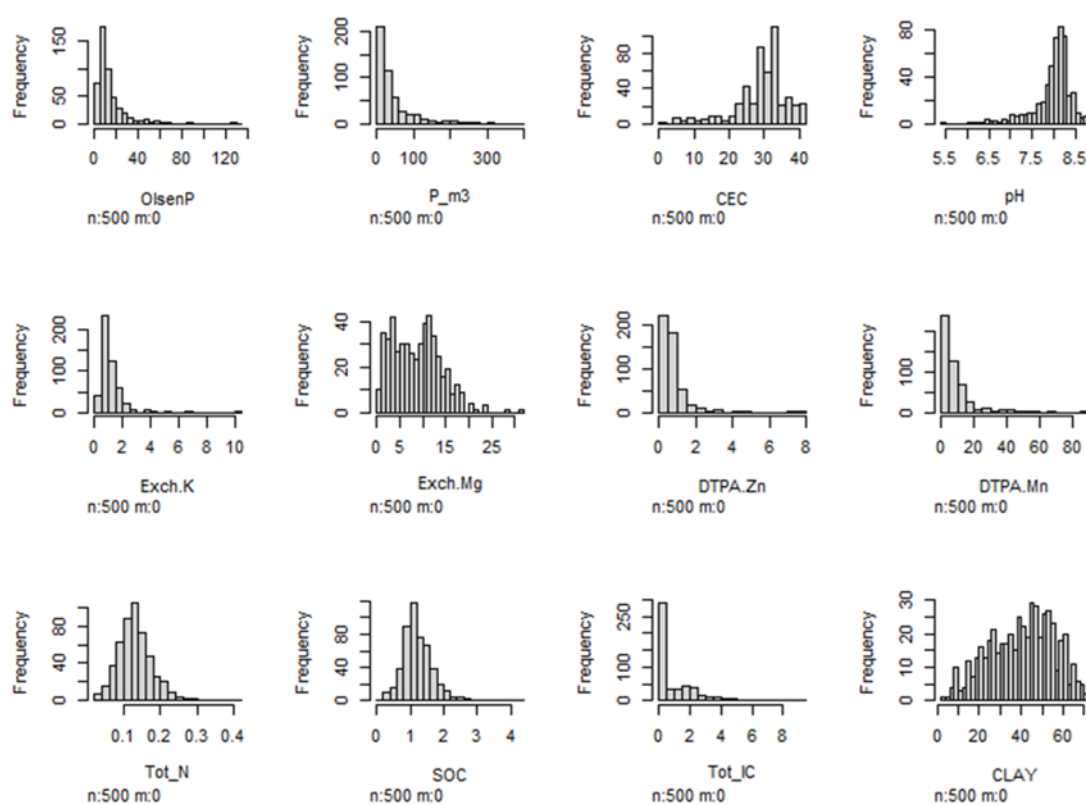


Figure C-3: Statistical distribution of soil properties in training sets (SSL).

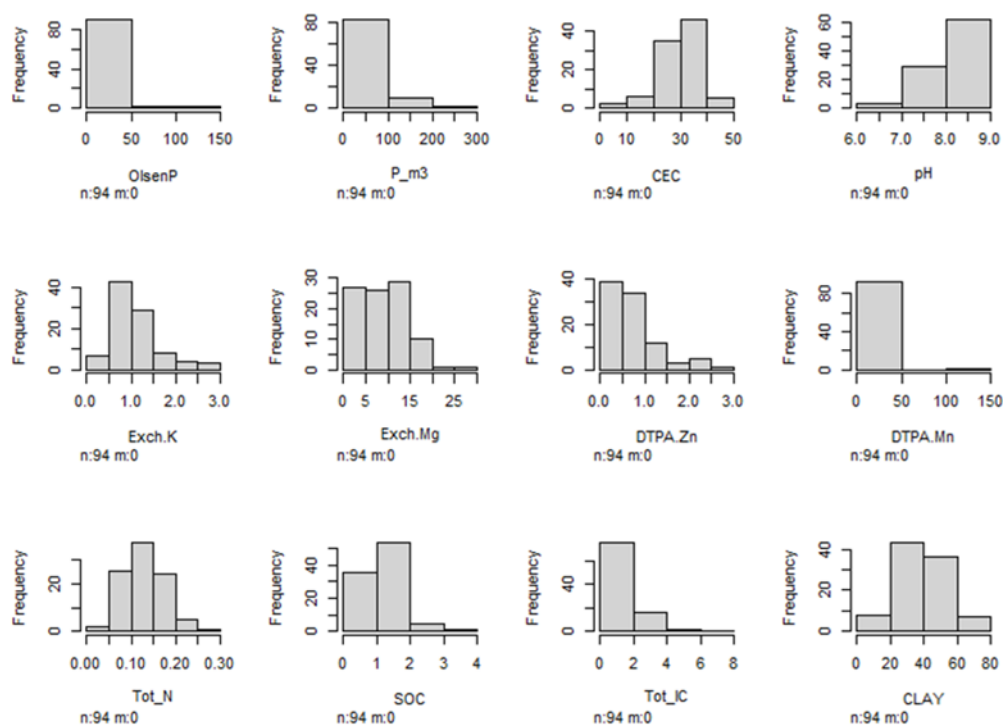


Figure C-4: Statistical distribution of soil properties for independent test sets.

C.2.3 PC scores explaining most of the variation in the spectra space

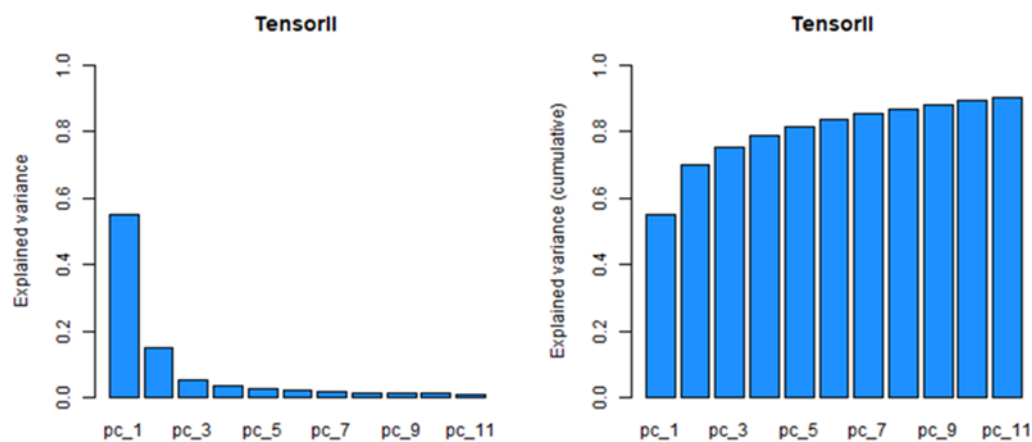


Figure C-5: The proportion of variance explained by first eleven PC's and their cumulative variance explained (99%) for Tensor-II (Bruker Tensor 27).

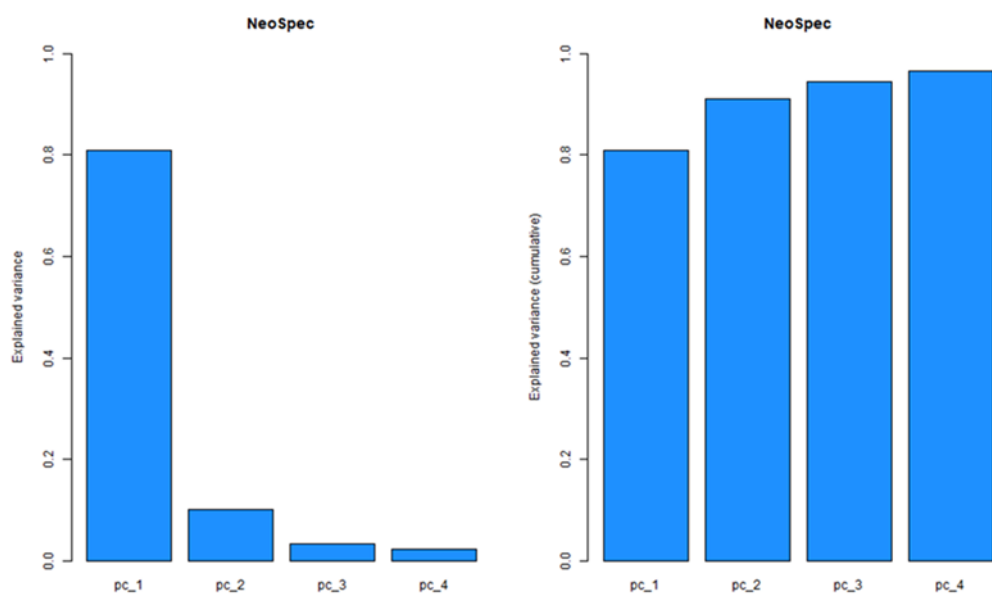


Figure C-6: The proportion of variance explained by first four PC's and their cumulative variance explained (99%) for NeoSpec (Neospectra).

C.2.4 Prediction performance from NIR reflectance with SG + 1D

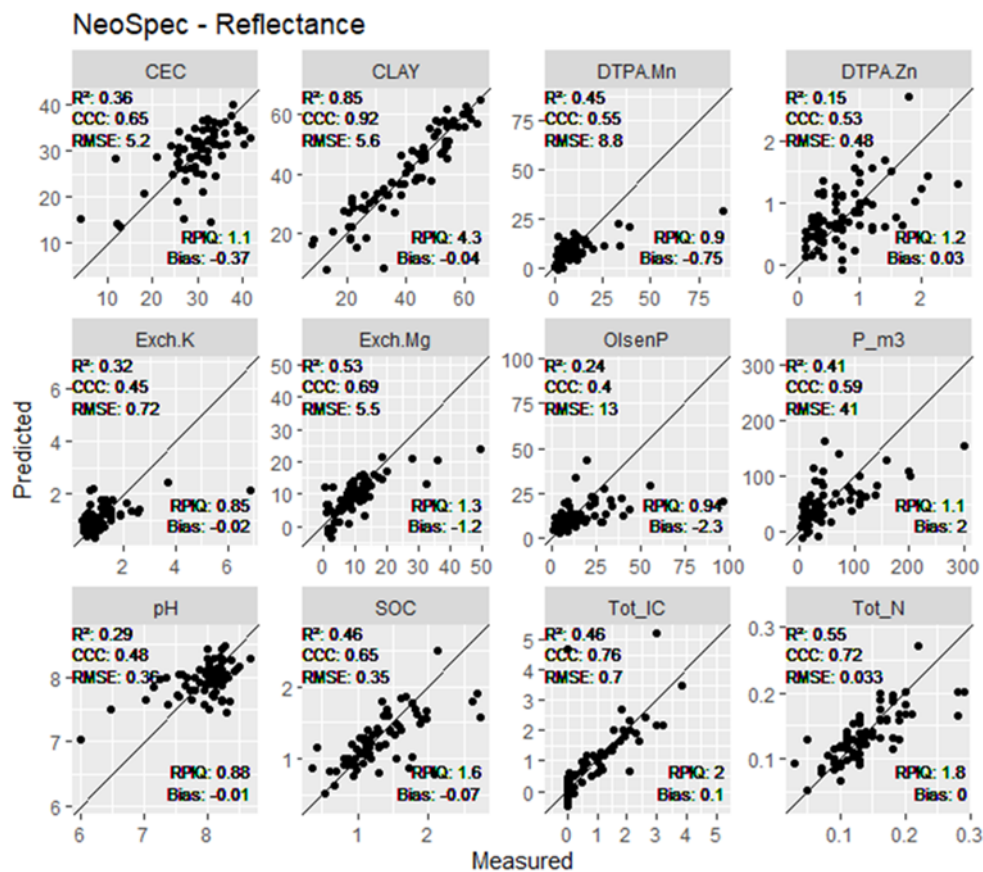
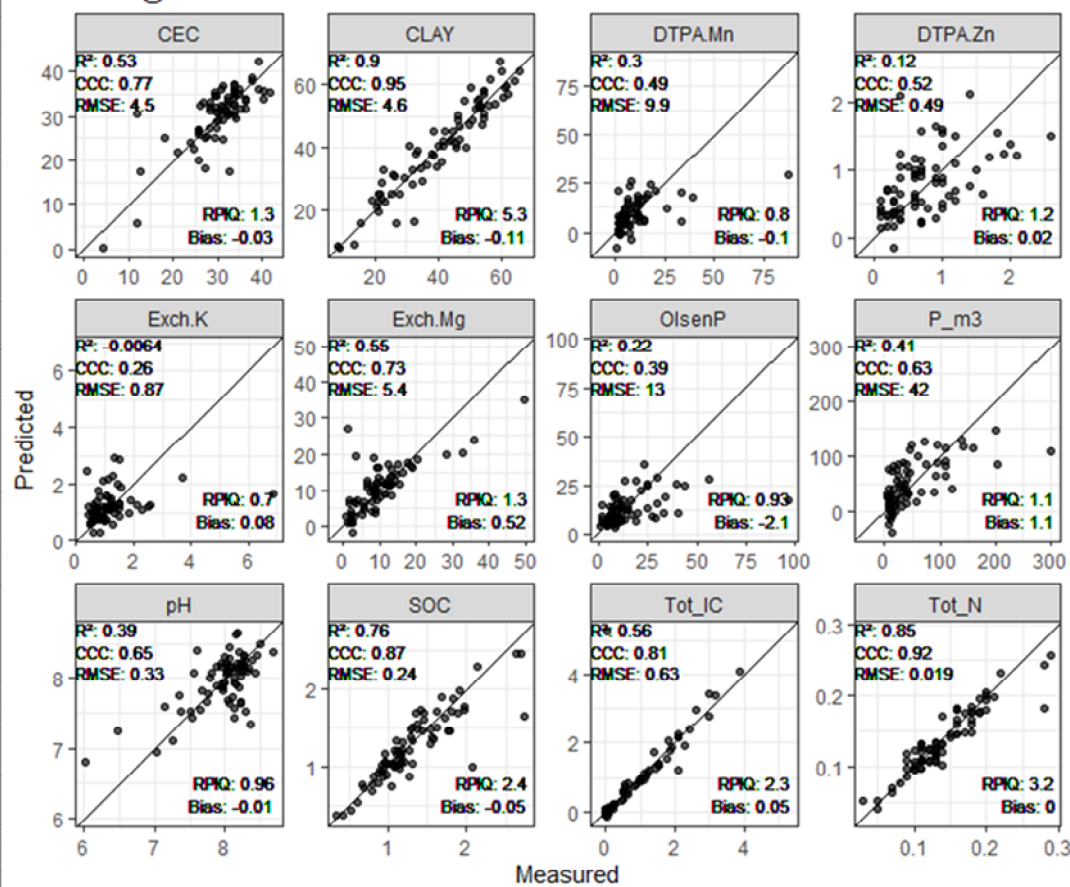


Figure C-7: Prediction performance for the soil properties (Neospec -> Neospectra) using a reflectance spectrum using the entire MSSL. The statistics shown are the R^2 , concordance correlation coefficient (CCC), the root mean square error (RMSE), Bias and the ratio of performance to interquartile range (RPIQ).

C.2.5 Prediction performance for spatial, MBL and cluster based calibration sample entry selection

Tensorll

cutoff @120 km



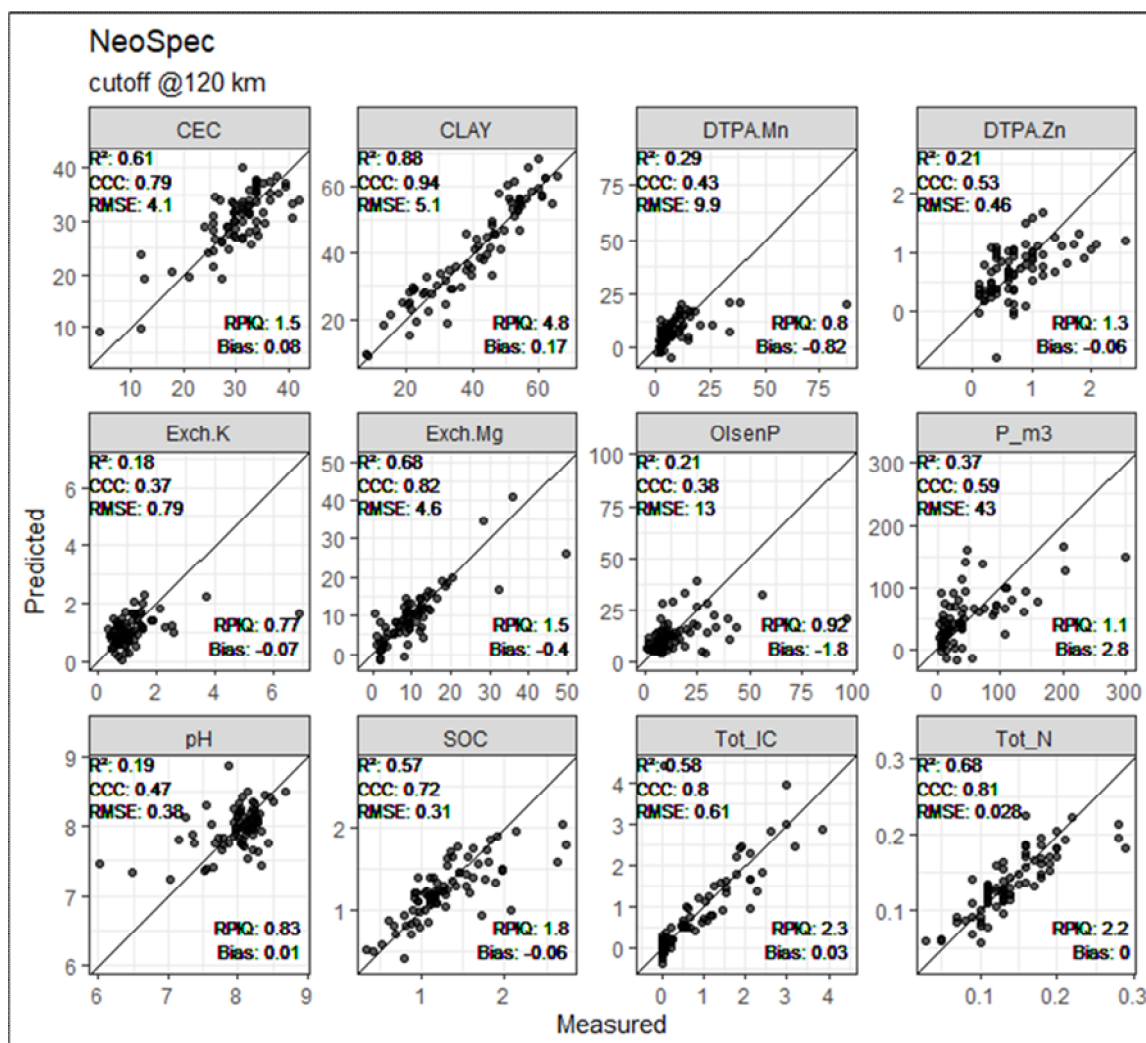
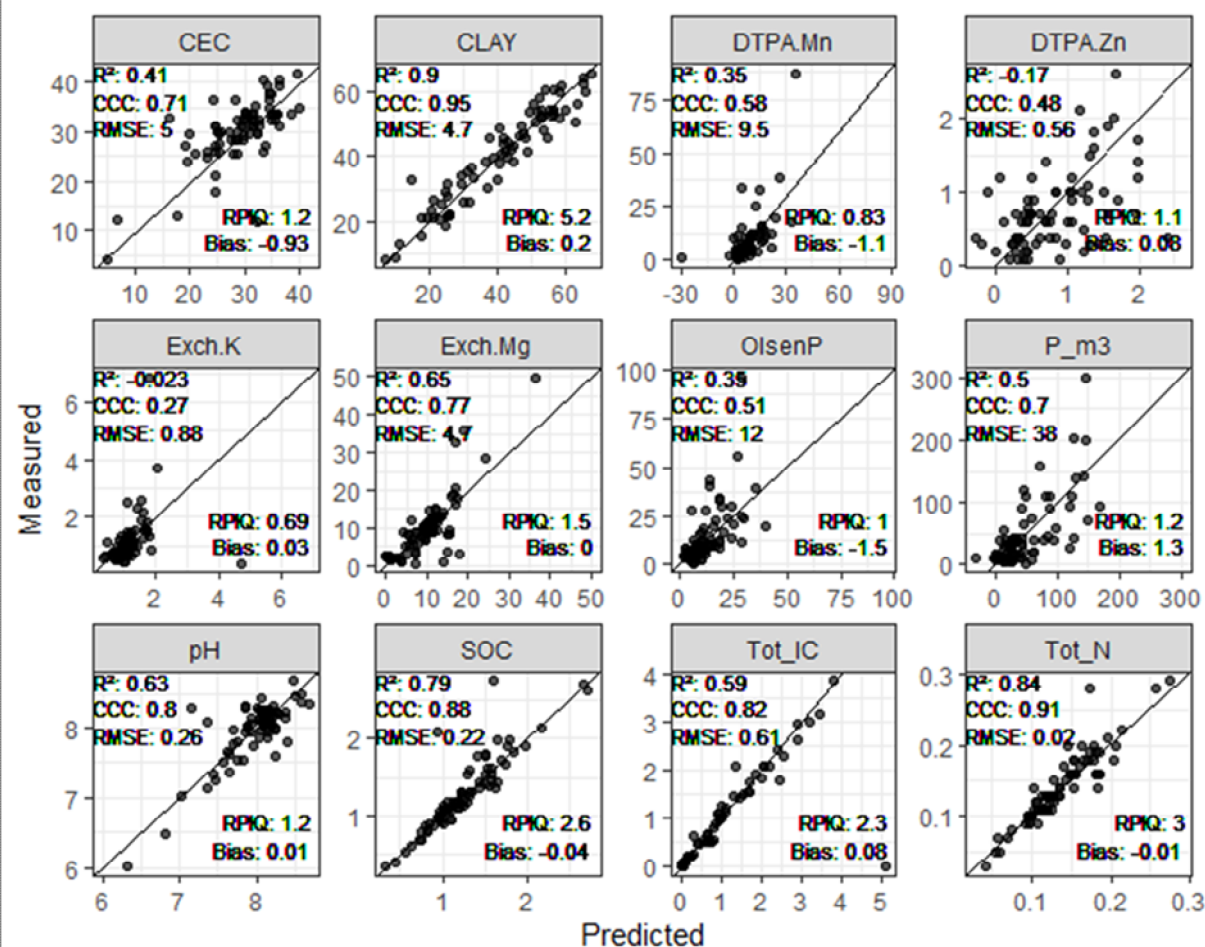


Figure C-8: Prediction performance for the soil properties and soil scanning instruments (Neospec -> Neospectra; TensorII -> Bruker Tensor 27) from calibration sample entry selection using the spatial non-stationarity analysis of the spectra PC vector. The statistics shown are the R^2 , concordance correlation coefficient (CCC), the root mean square error (RMSE), Bias and the ratio of performance to interquartile range (RPIQ).

TensorI

MBL



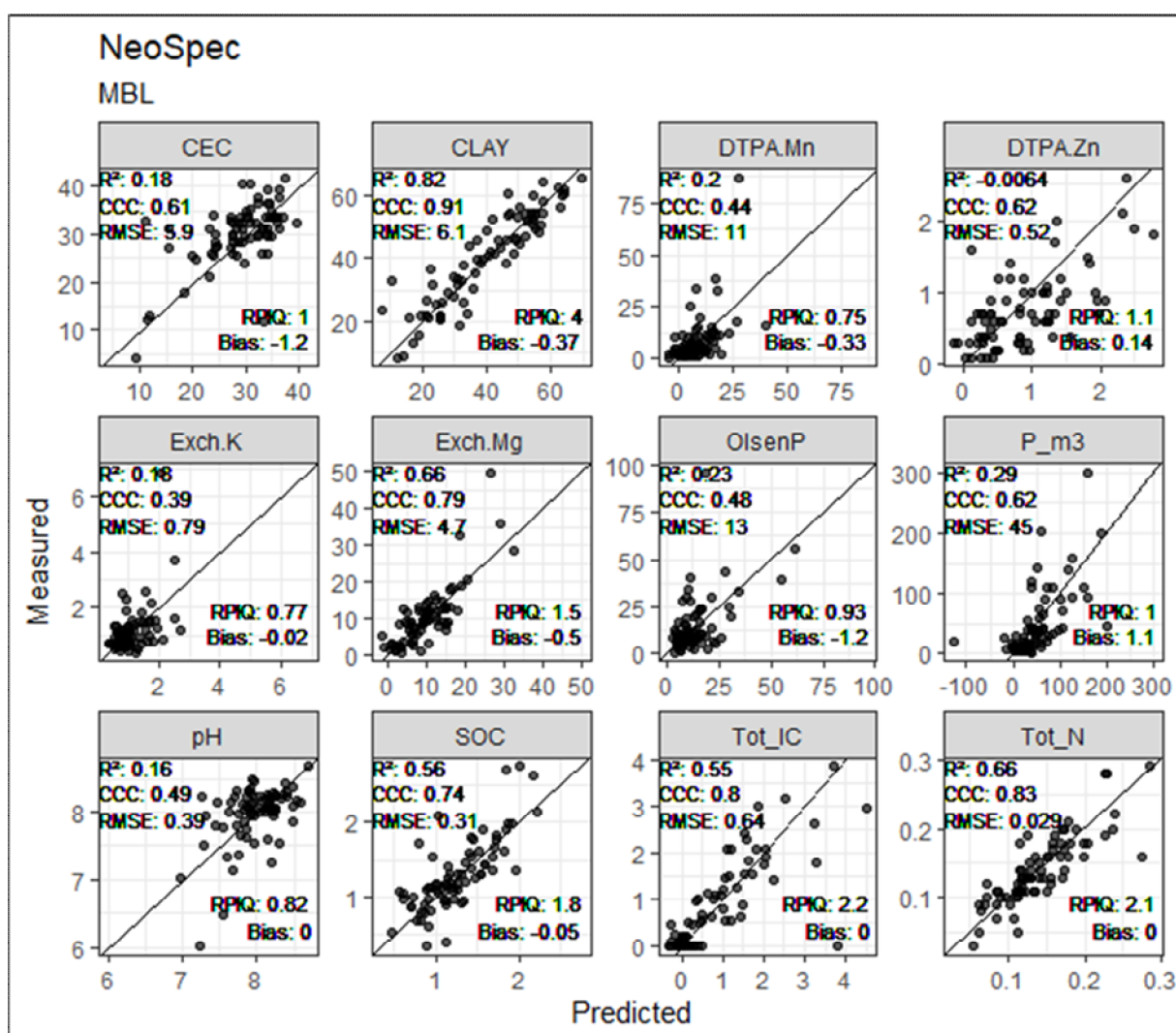
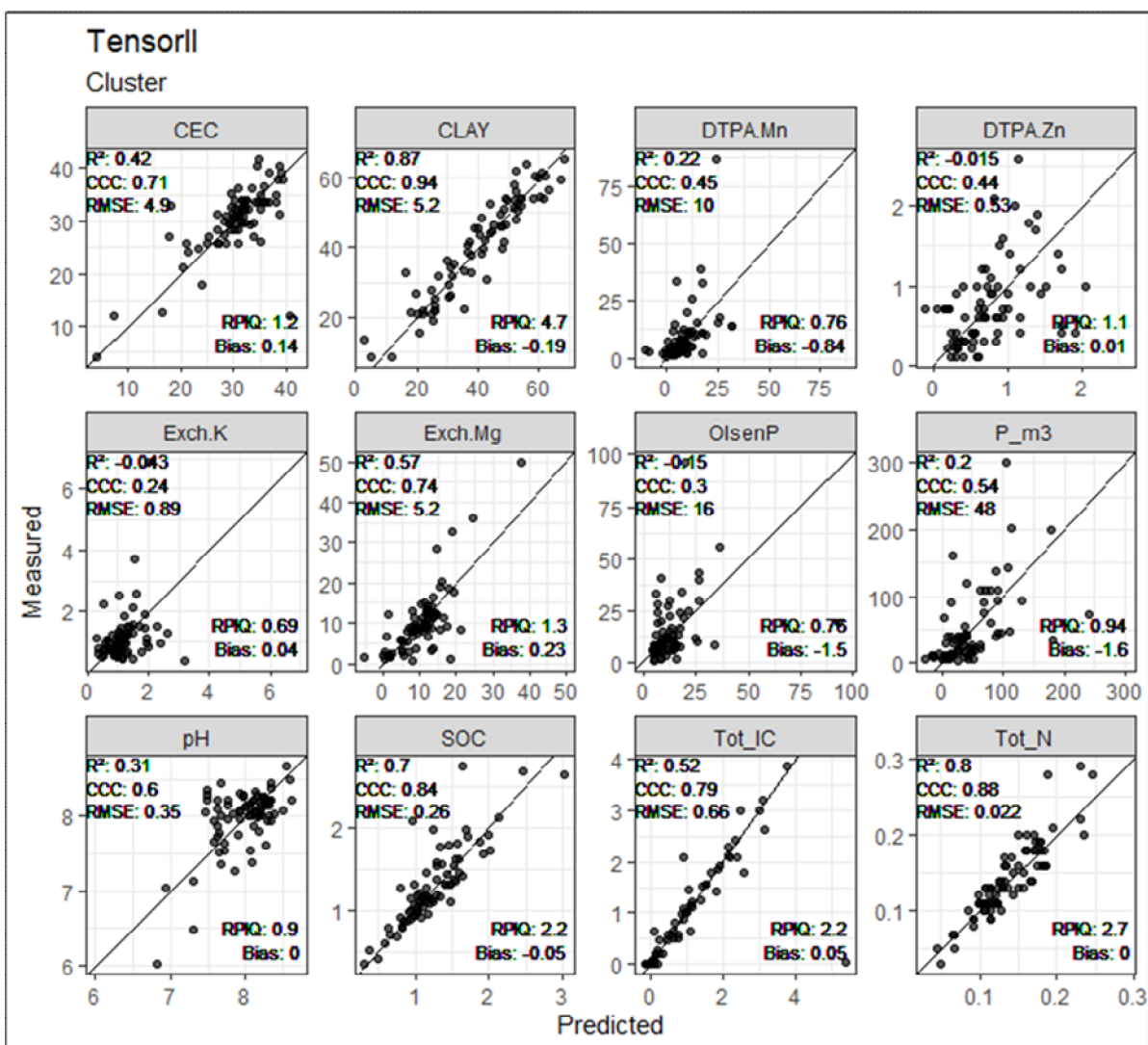


Figure C-9: Prediction performance for the soil properties and soil scanning instruments (Neospec -> Neospectra; TensorII -> Bruker Tensor 27) using spectra similarity sample entry selection method. The statistics shown are the R^2 , concordance correlation coefficient (CCC), the root mean square error (RMSE), Bias and the ratio of performance to interquartile range (RPIQ).



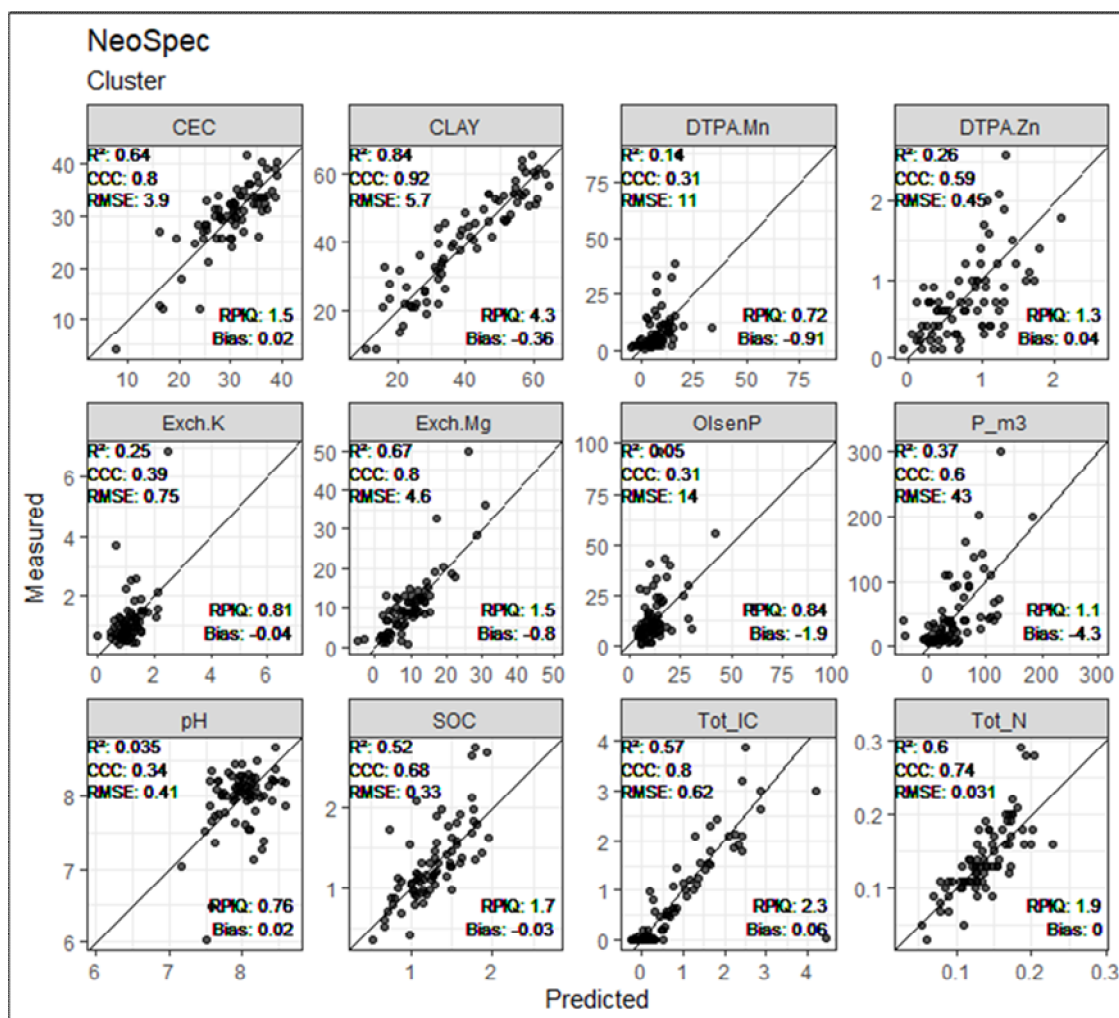


Figure C-10: Prediction performance for the soil properties and soil scanning instruments (Neospec -> Neospectra; TensorII -> Bruker Tensor 27) from calibration sample selection using the cluster grouping of environmental covariates. The statistics shown are the R^2 , concordance correlation coefficient (CCC), the root mean square error (RMSE), Bias and the ratio of performance to interquartile range (RPIQ).

Appendix D : Improving fertilizer application using crop models and NIR soil spectroscopy for Moroccan rainfed wheat growing regions

D.1 Materials and Methods

	Wplot 1						Wplot 2						Wplot 3						Legend	
Block 1	8	9	4	7	12	5	10	11	2	4	5	9	1	12	7	6	3	4	Cambisol	
	11	6	3	1	10	2	12	1	7	8	3	6	2	10	5	11	8	9	Kasta	
Block 2	11	1	9	3	5	7	8	6	7	12	9	2	3	11	2	1	10	9	Vertisol	
	12	4	8	10	6	2	11	4	3	1	10	5	5	7	8	12	6	4		
Block 3	9	1	10	5	3	7	11	7	3	6	10	5	6	10	3	5	4	11		
	6	12	8	2	4	11	2	12	1	9	8	4	9	1	8	7	12	2		

Figure D-1: Split-plot experimental layout for pot nutrient study across Blocks and soil types.

Table D.1: Analysis of variance for nutrient study across soil types and treatments for the split plot design.

Analysis of variance	
Source of variation	d.f.
Block stratum	2
Block.W_ Plots stratum	
Soil type	2
Residual	4
Block.W_ Plots.Pot stratum	
Nutrient	10
Soil type * Nutrient	20
Residual	67
Total	107

D.2 Results

D.2.1 Nutrient concentration response of wheat grain and straw nitrogen, phosphorus, potassium and zinc content for nutrient levels across three soil types

Table D.2: Duncan multiple range mean separation for wheat grain concentration of zinc.

Cambisol		Kastanozems		Vertisol	
Treatment	Grain Zn (g kg ⁻¹)	Treatment	Grain Zn (g kg ⁻¹)	Treatment	Grain Zn (g kg ⁻¹)
Full_full_full full	0.083^a	Full_full_full full	0.077^a	Full_full_full full	0.075^a
Full_0_full_0	0.052 ^b	0_full_full_0	0.046 ^b	1.5_full_full_0	0.023 ^b
0_full_full_0	0.034 ^{bc}	Full_0_full_0	0.034 ^{bc}	Full_0_full_0	0.022 ^b
Full_full_1.5_0	0.032 ^c	1.5_full_full_0	0.029 ^{bc}	Full_full_1.5_0	0.021 ^b
Full_0.5_full_0	0.031 ^c	Full_0.5_full_0	0.020 ^c	0_full_full_0	0.021 ^b
Full_full_0.5_0	0.028 ^c	Full_full_0_0	0.019 ^c	Full_full_0.5_0	0.021 ^b
Full_full_full_0	0.028 ^c	Full_full_full_0	0.019 ^c	Full_0.5_full_0	0.020 ^b
1.5_full_full_0	0.025 ^c	0.5_full_full_0	0.018 ^c	0.5_full_full_0	0.019 ^b
Full_full_0_0	0.025 ^c	Full_full_1.5_0	0.017 ^c	Full_full_full_0	0.019 ^b
Full_1.5_full_0	0.025 ^c	Full_1.5_full_0	0.016 ^c	Full_1.5_full_0	0.018 ^b
0.5_full_full_0	0.022 ^c	Full_full_0.5_0	0.015 ^c	Full_full_0_0	0.018 ^b

Table D.3: Duncan multiple range mean separation for wheat grain concentration of nitrogen.

Cambisol		Kastanozems		Vertisol	
Treatment	Grain N (g kg ⁻¹)	Treatment	Grain N (g kg ⁻¹)	Treatment	Grain N (g kg ⁻¹)
Full_full_1.5_0	29.73 ^a	Full_1.5_full_0	27.57 ^a	Full_full_1.5_0	28.63 ^a
Full_full_full full	28.23 ^a	1.5_full_full_0	26.80 ^{ab}	Full_full_full_0	28.42 ^a
Full_1.5_full_0	28.07 ^a	Full_full_full_0	26.62 ^{ab}	Full_0.5_full_0	28.10 ^a
Full_full_full_0	28.06 ^a	Full_full_0_0	26.60 ^{ab}	Full_full_0.5_0	27.60 ^a
Full_full_0.5_0	27.30 ^{ab}	Full_full_1.5_0	25.93 ^{ab}	1.5_full_full_0	27.60 ^a
Full_0.5_full_0	26.80 ^{ab}	Full_full_0.5_0	25.43 ^{ab}	0.5_full_full_0	27.30 ^a
Full_0_full_0	26.70 ^{ab}	Full_full_full full	25.10 ^{ab}	Full_1.5_full_0	27.07 ^a
1.5_full_full_0	26.33^{ab}	0_full_full_0	25.07^{ab}	Full_full_0_0	26.93 ^a
Full_full_0_0	26.07 ^{ab}	Full_0.5_full_0	23.93 ^{ab}	Full_full_full full	26.90 ^a
0.5_full_full_0	23.43^b	Full_0_full_0	23.63 ^{ab}	Full_0_full_0	26.67 ^a
0_full_full_0	18.70^c	0.5_full_full_0	22.37^b	0_full_full_0	20.07^b

Table D.4: Duncan multiple range mean separation for wheat grain concentration of phosphorus.

Cambisol		Kastanozems		Vertisol	
Treatment	Grain P (g kg ⁻¹)	Treatment	Grain P (g kg ⁻¹)	Treatment	Grain P (g kg ⁻¹)
Full_1.5_full_0	2.97 ^a	Full_full_full_0	3.34 ^a	Full_full_1.5_0	2.97 ^a

Full_full_1.5_0	2.95 ^a	Full_0.5_full_0	3.28 ^a	Full_1.5_full_0	2.95 ^{ab}
Full_full_full full	2.88 ^a	Full_full_0_0	3.28 ^a	Full_full_0_0	2.92 ^{abc}
1.5_full_full_0	2.77 ^{ab}	Full_0_full_0	3.27 ^a	Full_full_full_0	2.90 ^{abc}
Full_full_0_0	2.72 ^{ab}	Full_full_full full	3.25 ^a	0.5_full_full_0	2.87 ^{abc}
Full_full_0.5_0	2.72 ^{ab}	0.5_full_full_0	3.25 ^a	Full_0.5_full_0	2.87 ^{abc}
Full_full_full_0	2.68 ^{ab}	1.5_full_full_0	3.24 ^a	Full_full_0.5_0	2.80 ^{abc}
0.5_full_full_0	2.56 ^{ab}	Full_full_0.5_0	3.22 ^a	Full_full_full full	2.72 ^{abc}
Full_0.5_full_0	2.53 ^{ab}	Full_1.5_full_0	3.22 ^a	1.5_full_full_0	2.62 ^{abc}
0_full_full_0	2.35 ^{bc}	Full_full_1.5_0	3.17 ^a	0_full_full_0	2.17 ^{bc}
Full_0_full_0	1.88 ^c	0_full_full_0	2.45 ^b	Full_0_full_0	2.17 ^c

Table D.5: Duncan multiple range mean separation for wheat grain concentration of potassium.

Cambisol		Kastanozems		Vertisol	
Treatment	Grain K (g kg ⁻¹)	Treatment	Grain K (g kg ⁻¹)	Treatment	Grain K (g kg ⁻¹)
Full_full_1.5_0	4.59 ^a	Full_full_full_0	4.29 ^a	0.5_full_full_0	4.14 ^a
Full_full_full full	4.21 ^a	Full_full_full full	4.28 ^a	Full_full_0_0	4.07 ^a
Full_1.5_full_0	4.16 ^{ab}	Full_full_0_0	4.03 ^a	1.5_full_full_0	4.06 ^a
Full_full_0.5_0	4.15 ^{ab}	Full_full_0.5_0	4.00 ^a	Full_0.5_full_0	3.95 ^a
0.5_full_full_0	4.12 ^{ab}	1.5_full_full_0	3.93 ^a	Full_1.5_full_0	3.91 ^a
1.5_full_full_0	3.93 ^{abc}	Full_0.5_full_0	3.85 ^a	Full_full_0.5_0	3.88 ^a
Full_full_full_0	3.75 ^{abc}	Full_1.5_full_0	3.83 ^a	Full_full_full_0	3.79 ^a
Full_0.5_full_0	3.73 ^{abc}	Full_full_1.5_0	3.69 ^a	Full_full_full full	3.78 ^a
Full_full_0_0	3.66 ^{abc}	0.5_full_full_0	3.66 ^a	Full_full_1.5_0	3.77 ^a
Full_0_full_0	2.63 ^{bc}	Full_0_full_0	3.46 ^a	Full_0_full_0	3.48 ^a
0_full_full_0	2.47 ^c	0_full_full_0	2.80 ^a	0_full_full_0	3.17 ^a

D.2.2 Model summary for QUEFTS

Table D.6: Model summary - Formula: Total_N_Uptake ~ QUEFTS model N (n_fert, a_N, b_N).

Cambisol				
Parameters	Estimate	Std. Error	t value	Pr(> t)
a_N	149.74	54.58	2.744	0.0207 *
b_N	69.95	73.08	0.957	0.3610
Residual standard error: 28.02 on 10 degrees of freedom				
Number of iterations to convergence: 10				
Achieved convergence tolerance: 1.49e-08				
Kastanozems				
Parameters	Estimate	Std. Error	t value	Pr(> t)
a_N	277.12	37.58	7.375	2.38e-05 ***
b_N	135.76	37.88	3.584	0.00498 **
Residual standard error: 10.56 on 10 degrees of freedom				
Number of iterations to convergence: 5				
Achieved convergence tolerance: 1.49e-08				
Vertisol				
Parameters	Estimate	Std. Error	t value	Pr(> t)
a_N	81.22	15.57	5.215	0.000393 ***

b_N	15.99	22.63	0.707	0.495859
Residual standard error: 15.24 on 10 degrees of freedom				
Number of iterations to convergence: 5				
Achieved convergence tolerance: 1.49e-08				

Table D.7: Model summary - Formula: Total_P_Uptake ~ QUEFTS model P (P_supply, a_P, b_P).

Cambisol				
Parameters	Estimate	Std. Error	t value	Pr(> t)
a_P	18.259	5.019	3.638	0.00455 **
b_P	23.401	16.112	1.452	0.17703
Residual standard error: 2.095 on 10 degrees of freedom				
Number of iterations to convergence: 11				
Achieved convergence tolerance: 1.501e-06				
Kastanozems				
Parameters	Estimate	Std. Error	t value	Pr(> t)
a_P	17.3920	0.8076	21.535	1.04e-09 ***
b_P	0.9918	0.3134	3.165	0.0101 *
Residual standard error: 2.157 on 10 degrees of freedom				
Number of iterations to convergence: 5				
Achieved convergence tolerance: 4.06e-06				
Vertisol				
Parameters	Estimate	Std. Error	t value	Pr(> t)
a_P	8.7370	0.6663	13.114	1.26e-07 ***
b_P	0.7237	0.3899	1.856	0.0931 .
Residual standard error: 15.24 on 10 degrees of freedom				
Number of iterations to convergence: 5				
Achieved convergence tolerance: 1.338e-06				

D.2.3 Model summary for Quadratic plateau response

Table D.8: Model summary - Formula: GY_ha ~ qp(n_fert, a, b, jp).

Kastanozems				
Parameters	Estimate	Std. Error	t value	Pr(> t)
a	314.41	394.52	0.797	0.440970
b	74.77	17.22	4.342	0.000958 ***
jp	114.62	24.45	4.687	0.000526 ***
Residual standard error: 683.3 on 12 degrees of freedom				
Number of iterations to convergence: 6				
Achieved convergence tolerance: 1.49e-08				
Cambisol				
Parameters	Estimate	Std. Error	t value	Pr(> t)
a	599.61	433.56	1.383	0.1919
b	36.23	15.60	2.322	0.0387 *
jp	143.45	58.75	2.442	0.0311 *
Residual standard error: 752 on 12 degrees of freedom				
Number of iterations to convergence: 14				
Achieved convergence tolerance: 1.49e-08				
Vertisol				

Parameters	Estimate	Std. Error	t value	Pr(> t)
a	321.20	278.55	1.153	0.2713
b	44.22	20.62	2.144	0.0532 .
jp	82.87	38.71	2.141	0.0535 .
Residual standard error: 482.5 on 12 degrees of freedom				
Number of iterations to convergence: 8				
Achieved convergence tolerance: 1.49e-08				

Table D.9: Model summary - Formula: GY_ha ~ qp(p2o5_fert, a, b, jp).

Kastanozems				
Parameters	Estimate	Std. Error	t value	Pr(> t)
a	1875.57	383.82	4.887	0.000374 ***
b	79.02	26.25	3.010	0.010866 *
jp	72.76	21.97	3.312	0.006200 **
Residual standard error: 664.8 on 12 degrees of freedom				
Number of iterations to convergence: 10				
Achieved convergence tolerance: 1.49e-08				
Cambisol				
Parameters	Estimate	Std. Error	t value	Pr(> t)
a	858.34	466.82	1.839	0.0908 .
b	54.29	22.82	2.380	0.0348 *
jp	99.02	43.48	2.277	0.0419 *
Residual standard error: 820.7 on 12 degrees of freedom				
Number of iterations to convergence: 6				
Achieved convergence tolerance: 1.49e-08				