

Conversion of soil pH values and exchangeable base cation quantities determined by different methods

by

Teneille Nel

Thesis presented in partial fulfilment of the requirements for the degree of

Master of Agricultural Sciences



Stellenbosch University

Department of Soil Science, Faculty of AgriSciences

Supervisor: Dr Ailsa Hardie

Co-supervisor: Dr Cathy Clarke

December 2021

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third-party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: December 2021

Summary

Soil management recommendations based on chemical test results rely on comparisons with norm values developed using specific analytical methods and soil taxonomic classes. Conversions of soil pH and exchangeable base cation concentration between different methods are required for the standardization of data from scientific literature, soil test reports and soil databases. The combined effects of soil physicochemical properties on the accuracy and parsimony of soil pH and cation concentration conversions are unknown. Rapid soil analyses can be obtained by infrared (IR) spectroscopy, but few studies have evaluated the accuracy of IR spectral-based models to determine soil pH and exchangeable base cations in Western Cape soils. The aim of this study was to develop equations for conversion of soil pH values and exchangeable base cation concentrations between different methods of measurement. An additional aim was to develop IR spectral-based models for determination of soil pH and exchangeable base cation content of Western Cape soils. Soil pH was measured in different matrices (water, 0.1M KCl and 0.01M CaCl₂) of varying soil: solution ratios (1:1, 1:2.5, 1:4 and 1:5). The exchangeable base cation content was measured according to the ammonium acetate (1M, pH = 7), compulsive exchange, ammonium acetate (0.2M, pH = 7), Ambic-I, Mehlich-III, Bray-II and Citric acid (1%) methods. Models converting soil pH and cation content between values obtained by different methods were developed using simple linear regression (SLR). In addition, multivariate linear regression (MLR) equations incorporating soil physicochemical properties as parameters were developed. Spectral measurements of the soil samples were taken with three spectroscopic instruments in the near- to mid-IR range. IR spectral-based models predicting exchangeable base cation content and soil pH were calibrated using partial least squares regression. SLR equations for predicting soil pH (RMSE = 0.12-0.32) were more accurate than the conversion factors. MLR equations constructed by stepwise regression to convert exchangeable base cation content (RMSE = 1.4-1.5 cmol_c kg⁻¹, 0.4 cmol_c kg⁻¹, 16-20 mg kg⁻¹ and 27-39 mg kg⁻¹ for Ca, Mg, Na and K, respectively) were more accurate than SLR models. Exchangeable K content was accurately converted between the most extraction methods. Models converting base cation content may be chosen based on the purpose of application and relative size of the prediction error. IR spectroscopy is recommended to determine exchangeable Ca and Mg content in Western Cape soils (RMSE = 1.2-1.4 cmol_c kg⁻¹ and 0.37 cmol_c kg⁻¹ for Ca and Mg, respectively). Subsetting soil sample sets by similarities in texture and organic carbon content led to more accurate predictions of soil pH and exchangeable Ca content (RMSE decreased by up to 0.19 pH units and 0.79 cmol_c kg⁻¹, respectively). In summary, use of the developed SLR as well as MLR models are recommended to convert soil pH and exchangeable base cation concentrations of Western Cape soils. Models should only be as complex as practical context requires. Moreover, IR spectroscopy may be used for determination of exchangeable Ca and Mg content in Western Cape soils.

Opsomming

Grondbestuurshandleiding wat op chemiese toetsresultate gebaseer is, is afhanklik van vergelyking met kritiese drempelwaardes wat d.m.v. spesifieke ontledingsmetodes en grondtipes ontwikkel is. Omskakeling van grond pH-waardes en uitruilbare basiese kationinhoud tussen verskillende metodes is vir standaardisering van data wat in die wetenskaplike literatuur, grondtoetsverslae en gronddatabasisse voorkom, nodig. Die saamgestelde effekte van grondfisiologiese eienskappe op die akkuraatheid en parsimonie van pH- en kationinhoud-omskakelings, is onbekend. Grondontledings kan vinnig m.b.v. infrarooi-(IR)-spektroskopie bepaal word, maar daar is min studies wat die akkuraatheid van IR modelle vir die bepaling van grond pH-waardes en uitruilbare kationinhoud in Wes-Kaapse gronde geëvalueer het. Die doel van hierdie studie is om vergelykings vir die omskakeling van grond pH-waardes en uitruilbare basiese kationinhoud tussen verskillende metingsmetodes, te ontwikkel. 'n Addisionele doel was om IR spektraal-gebaseerde modelle vir die bepaling van grond pH-waardes en uitruilbare basiese kationinhoud in Wes-Kaapse gronde, te ontwikkel. Grond pH-vlak is in verskillende matrikse (water, 0.1M KCl en 0.01M CaCl₂) met verskeie grond: oplossing verhoudings (1:1, 1:2,5, 1:4 en 1:5) gemeet. Die uitruilbare basiese kationinhoud is volgens die ammoniumasetaat (1M, pH = 7), kompulsiewe uitruil, ammoniumasetaat (0.2M, pH = 7), Ambic-I, Mehlich-III, Bray-II en Sitroensuur (1%) metodes gemeet. Modelle is d.m.v. eenvoudige lineêre regressie ontwikkel vir die omskakeling van die grond pH-vlak of kationinhoud tussen waardes wat deur verskillende metodes verkry word. Meerveranderlike lineêre regressie vergelykings wat verskillende grondfisiologiese eienskappe insluit as veranderlikes, is ook ontwikkel. Spektrale metings van die monsters is met drie verskillende spektroskopiese instrumente in die nabye tot middel-IR omvang geneem. IR spektraal-gebaseerde modelle vir die voorspelling van uitruilbare basiese kationinhoud en grond pH-vlak is d.m.v. gedeeltelike regressie gekalibreer. Eenvoudige lineêre vergelykings vir die bepaling van die grond pH-vlak (RMSE = 0.12-0.32) was akkurater as omskakelingsfaktore. Meerveranderlike modelle vir die omskakeling van uitruilbare basiese katione wat d.m.v. stapsgewyse regressie ontwikkel is (RMSE = 1.4-1.5 cmol_c kg⁻¹, 0.4 cmol_c kg⁻¹, 16-20 mg kg⁻¹ en 27-39 mg kg⁻¹ vir Ca, Mg, Na and K, onderskeidelik), was akkurater as die lineêre regressie modelle. Uitruilbare K inhoud was tussen die meeste ekstrasiemetodes akkuraat omgeskakel. Modelle vir die omskakeling van basiese kationinhoud kan gekies word op grond van die doel van toepassing en relatiewe grootte van die voorspellingsfout. IR spektroskopie word aanbeveel vir die bepaling van uitruilbare Ca- en Mg-inhoud in Wes-Kaapse gronde (RMSE = 1.2-1.4 cmol_c kg⁻¹ en 0.37 cmol_c kg⁻¹ vir Ca and Mg, onderskeidelik). Onderverstelling van grondmonsterstelle volgens soortgelykhede van grond tekstuur en organiese koolstof-inhoud, het gelei tot akkurater bepaling van grond pH-waardes en uitruilbare Ca-inhoud (RMSE het afgeneem tot en met 0.19 pH eenhede en 0.79 cmol_c kg⁻¹ onderskeidelik). Ter afsluiting word gebruik van die ontwikkelde lineêre regressie sowel as multiveranderlike modelle voorgestel vir die omskakeling van grond pH-vlak en uitruilbare basiese katione in Wes-Kaapse gronde. Modelle moet slegs so kompleks wees as wat praktiese konteks vereis. Boonop kan IR spektroskopie gebruik word vir die bepaling van die uitruilbare Ca- en Mg-inhoud in Wes-Kaapse gronde.

Acknowledgements

I wish to express my sincere gratitude and appreciation to the following:

- My Abba, for being my rock.
- My family, for loving me at my best, my worst, and every moment in between – since before I could even hold a pen.
- My supervisor at Stellenbosch University (Dr A. G. Hardie) for organizing the project funding, troubleshooting, inspiring me to grow, and believing in me.
- My co-supervisor at Stellenbosch University (Dr C. Clarke) for proofreading, de-bugging, sound advice, and encouragement.
- The laboratory staff at Elsenburg Analytical Services (Ms. R. Swart and Ms. R. Topley), for providing the soil samples.
- The administrative and technical staff at the Department of Soil Science, Stellenbosch University (Ms J. van der Merwe, Mrs L. De Witt, Mr F. Arnolds, Mr H. Achilles, and Mr V. Siyo), for practical assistance, logistical gymnastics, and moral support.
- The spectroscopy expert at the Department of Viticulture and Oenology, Stellenbosch University (Dr Alexandre-Tudó), for equipment- and software-use training.
- The statistics experts at the Centre for Statistical Consultation, Stellenbosch University (Prof M. Kidd) and at the Department of Genetics, Stellenbosch University (Mr. S. van der Westhuizen), for de-mystifying data analysis.
- My host supervisor at Faculty of Bioscience Engineering, KU Leuven (Dr E. Smolders), for teaching me invaluable lessons.
- The academic staff at the Department of Soil Science, Stellenbosch University (Dr A. Rozanov, Mr. E. L. Lategan, and Dr E. Hoffman), for their input and support, and for making me part of the family.
- My mentor, Roeline van Schalkwyk, for keeping things real.
- My friends, for many laughs and tears, and reminding me about the important things in life.
- The exchange coordinators at the International Office of KU Leuven (Mr. M. Tips and Ms. K. Rock) and at Stellenbosch University (Ms. S. van der Westhuizen and Ms. E. Ryklief), for giving me wings.

List of tables

Table 2-1 Wavenumbers and wavelengths of VIS-NIR-MIR absorptions of major soil components, adapted from Soriano-Disla et al. (2014)	29
Table 3-1 Soil pH measurement methods	34
Table 3-2 Summary of soil physicochemical properties of samples used for comparison of methods of pH determination data (N = 57).....	36
Table 3-3 Conversion equations between common pH measurement methods with R ² and RMSE values (**P < 0.001).....	44
Table 3-4 r-values showing correlation between pH measurement methods and soil properties...	47
Table 3-5 Conversion factors between common pH measurement methods and increase in RMSE ⁱ -values compared to corresponding SLR models	48
Table 4-1 Procedures involved in different cation extraction methods	54
Table 4-2 Summary of soil physicochemical properties of samples used for comparison of cation extraction methods (N = 57).....	56
Table 4-3 Median ratios of nutrient concentrations determined by different methods versus the M-III method ([element]method / [element]M-III) using data subsets for Ca (N = 55), Mg (N = 55), Na (N = 52) and K (N = 52), as well as the full dataseti for Na (N = 57). Compulsive exchange method denoted as BaCl ₂	61
Table 4-4 Criteria for evaluation of suitability of models for conversion of exchangeable base cation concentrations.....	63
Table 4-5 SLR equation for conversion of exchangeable Mg content in units of cmol _c kg ⁻¹ (**P < 0.001).....	66
Table 4-6 SLR equations for conversion of exchangeable Na content in units of mg kg ⁻¹ for soils within normal EC _e range (**P < 0.001).....	66
Table 4-7 SLR equation for conversion of exchangeable Na content in units of mg kg ⁻¹ for saline soils (**P < 0.001).....	66
Table 4-8 SLR equations for conversion of exchangeable K content in units of mg kg ⁻¹ (**P < 0.001).....	66
Table 4-9 MLR models with soil pH as a regressor for conversion between exchangeable Mg content in units of cmol _c kg ⁻¹ (**P < 0.001)	68
Table 4-10 MLR models with soil EC _e as a regressor for conversion between exchangeable Na content in units of mg kg ⁻¹ in soils within normal EC _e range.....	68
Table 4-11 MLR models with soil pH as a regressor for conversion between exchangeable K content in units of mg kg ⁻¹ (**P < 0.001, *P < 0.01, *P < 0.05).....	69
Table 4-12 MLR models selected by stepwise regression for conversion between exchangeable Ca quantification methods with R ² and RMSE (cmol _c kg ⁻¹) values (**P < 0.001, *P < 0.01, *P < 0.05)	70

Table 4-13 MLR models selected by stepwise regression for conversion between exchangeable Mg quantification methods with R^2 and RMSE ($\text{cmol}_c \text{ kg}^{-1}$) values (**P < 0.01, *P < 0.05).....	70
Table 4-14 MLR models selected by stepwise regression for conversion between exchangeable Na quantification methods for soils within normal EC_e range with R^2 and RMSE (mg kg^{-1}) values (**P < 0.01, *P < 0.05).....	70
Table 4-15 MLR models selected by stepwise regression for conversion between exchangeable Na quantification methods for saline soils with R^2 and RMSE (mg kg^{-1}) values (**P < 0.01, *P < 0.05).....	71
Table 4-16 MLR models selected by stepwise regression for conversion between exchangeable K quantification methods with R^2 and RMSE (mg kg^{-1}) values (**P < 0.01, *P < 0.05).	71
Table 4-17 Approximate threshold values of Na content associated with sodic soils ($\text{ESP} > 15$) ..	73
Table 5-1 Specifications and settings of IR spectroscopy instruments	76
Table 5-2 IR wavelength regions correlated to soil chemical properties for overall best-performing models	78
Table 5-3 Figures of merit for best-performing models predicting soil chemical properties from IR spectra by each instrument. Best-performing models for each property overall (regardless of instrument) are shaded	81

List of tables: Appendix A

Table A- 1 Maximum coefficient of variation of repeated pH measures	95
Table A- 2 Difference between AIC-values ⁱ of simple and corresponding multiple linear regression models for soil pH conversion including E _{Ce} variable	95
Table A- 3 Difference between AIC ⁱ - & RMSE ⁱⁱ -values of simple and corresponding multiple linear regression models for soil pH conversion including %clay as variable	96
Table A- 4 Difference between RMSE ⁱ -values of simple and corresponding multiple linear regression models for soil pH conversion including parameters selected by stepwise regression .	96

List of tables: Appendix B

Table B- 1 Difference between RMSE ⁱ -values of simple linear regression models calibrated from datasets including versus excluding outlier samples, for prediction of exchangeable Ca content (cmol _c kg ⁻¹)	99
Table B- 2 Difference between RMSE ⁱ -values of simple linear regression models calibrated from datasets including versus excluding outlier samples, for prediction of exchangeable Mg content (cmol _c kg ⁻¹)	99
Table B- 3 Difference between RMSE ⁱ -values of simple linear regression models calibrated from datasets including versus excluding outlier samples, for prediction of exchangeable Na content (mg kg ⁻¹)	99
Table B- 4 Difference between RMSE ⁱ -values of simple linear regression models calibrated from datasets including versus excluding outlier samples, for prediction of exchangeable K content (mg kg ⁻¹)	99
Table B- 5 Summary of soil physicochemical properties of Ca data subset (N = 55)	100
Table B- 6 Summary of soil physicochemical properties of Mg data subset (N = 55).....	100
Table B- 7 Summary of soil physicochemical properties of Na data subset (N = 52)	100
Table B- 8 Summary of soil physicochemical properties of K data subset (N = 54)	100
Table B- 9 Pearson correlation coefficient (r) values showing correlation between methods for extraction of Ca.....	101
Table B- 10 Pearson correlation coefficient (r) values showing correlation between methods for extraction of Mg	102
Table B- 11 Pearson correlation coefficient (r) values showing correlation between methods for extraction of Na in soils within normal E _{Ce} range	102
Table B- 12 Pearson correlation coefficient (r) values showing correlation between methods for extraction of Na in saline soils	102

Table B- 13 Pearson correlation coefficient (r) values showing correlation between methods for extraction of K.....	102
Table B- 14 Pearson correlation coefficient (r) values showing correlation between soil properties and Ca extracted by different methods.....	103
Table B- 15 Pearson correlation coefficient (r) values showing correlation between soil properties and Mg extracted by different methods.....	103
Table B- 16 Pearson correlation coefficient (r) values showing correlation between soil properties and Na extracted by different methods in soils within normal EC _e range.....	103
Table B- 17 Pearson correlation coefficient (r) values showing correlation between soil properties and Na extracted by different methods in saline soils.....	104
Table B- 18 Pearson correlation coefficient (r) values correlation between soil properties and K extracted by different methods.....	104
Table B- 19 Difference between AIC-values ⁱ of SLR and corresponding MLR models including pH as regressor for conversion of Ca content.....	104
Table B- 20 Difference between AIC-values ⁱ of SLR and corresponding MLR models including pH as regressor for conversion of Mg content.....	105
Table B- 21 Difference between AIC-values ⁱ of SLR and corresponding MLR models including EC _e as regressor for conversion of Na content for soils within normal EC _e range.....	105
Table B- 22 Difference between AIC-values ⁱ of SLR and corresponding MLR models including EC _e as regressor for conversion of Na content for saline soils.....	105
Table B- 23 Difference between AIC-values ⁱ of SLR and corresponding MLR models including pH as regressor for conversion of K content.....	105
Table B- 24 Difference between RMSE-values of SLR and corresponding MLR models including soil pH ⁱ and various combinations of soil properties ⁱⁱ as regressors for conversion of exchangeable Ca content.....	106
Table B- 25 Difference between RMSE-values of SLR and corresponding MLR models including soil pH ⁱ and various combinations of soil properties ⁱⁱ as regressors for conversion of exchangeable Mg content.....	106
Table B- 26 Difference between RMSE-values of SLR and corresponding MLR models including soil EC _e ⁱ and various combinations of soil properties ⁱⁱ as regressors for conversion of exchangeable Na content for soil within normal EC _e range.....	107
Table B- 27 Difference between RMSE-values of SLR and corresponding MLR models including soil EC _e ⁱ and various combinations of soil properties ⁱⁱ as regressors for conversion of exchangeable Na content for saline soils.....	107
Table B- 28 Difference between RMSE-values of SLR and corresponding MLR models including soil pH ⁱ and various combinations of soil properties ⁱⁱ as regressors for conversion of exchangeable K content.....	108

List of tables: Appendix C

Table C- 1 A review of validation statistics of models for prediction of soil chemical properties based on spectral response in IR regions of the electromagnetic spectrum	113
Table C- 2 Figures of merit for models predicting soil chemical properties from NIR spectra generated by the MPA instrument	116
Table C- 3 Figures of merit for models predicting soil chemical properties from NIR spectra generated by the Matrix-F instrument	118
Table C- 4 Figures of merit for models predicting soil chemical properties from MIR spectra generated by the Alpha instrument	120
Table C- 5 Figures of merit for restricted-range models predicting soil chemical properties from MIR spectra generated by the MPA instrument	122
Table C- 6 Figures of merit for restricted-range models predicting soil chemical properties from MIR spectra generated by the Matrix-F instrument.....	123
Table C- 7 Figures of merit for restricted-range models predicting soil chemical properties from MIR spectra generated by the Alpha instrument.....	124
Table C- 8 Performance of models validated with either local or non-local sample subsets as an indication of the effect of soil texture on predictions of exchangeable base cations and soil pH from spectral data	125
Table C- 9 Performance of models validated with either local or non-local sample subsets as an indication of the effect of SOM on predictions of exchangeable base cations and soil pH from spectral data	126

List of figures

Figure 2-1 Scatter plot of pH in CaCl ₂ vs pH in water with the additive calibration curve superimposed (solid line). The broken line shows the 1:1 relationship (Image source: Henderson & Bui, 2002).....	11
Figure 3-1 Distribution soil physicochemical properties of samples used for comparison of methods of pH determination data (N = 57).....	37
Figure 3-2 Distribution of pH measured in different soil: solution ratios in H ₂ O (a), 0.01M CaCl ₂ (b) and 1M KCl (c). Boxes show first and third quartiles, with median line in bold. Whiskers show minimum and maximum values.	40
Figure 3-3 Boxplots showing distribution of soil pH levels in different matrices for a 1:1 soil: solution ratio (a) and boxplots showing the same distributions grouped by texture class (b). Boxes show first and third quartiles, with median line in bold. Whiskers show minimum and maximum values.	41
Figure 3-4 Boxplots showing distribution of soil pH levels in different matrices for soil: solution ratios of 1:2.5 (a), 1:4 (b) and 1:5(c). Boxes show first and third quartiles, with median line in bold. Whiskers show minimum and maximum values.	42
Figure 3-5 Plots showing conversions between pH measured by common methods to illustrate confidence ⁱ and prediction ⁱⁱ intervals (a), as well as effects of textural grouping (b), EC levels (c) and %Clay content (d) on conversions	45
Figure 4-1 Relative amounts of base cations extracted by different methods using data subsets for Ca (a, N = 55), Mg (b, N = 55), Na (c, N = 52) and K (d, N = 52), as well as using the full dataset for Na (e, N = 57). Compulsive exchange method denoted as BaCl ₂ . Boxes show first and third quartiles, with median line in bold. Whiskers show minimum and maximum values (excluding outliers w.r.t. range). Superscripts show means that are not significantly different when compared using an a posteriori Dunn test.	62
Figure 4-2 Confidence ⁱ and prediction ⁱⁱ intervals for models ⁱⁱ converting exchangeable Mg (a), Na (b) and K (c) between values obtained by different extraction methods	65

List of figures: Appendix B

Figure B- 1 Boxplots showing distribution of E _{Ce} (a), %N (b), pH _{KCl} (c), %C (d) and %Clay (e) of soils used for comparison of cation extraction methods. Boxes show first and third quartiles, with median line in bold. Whiskers show minimum and maximum values.	97
Figure B- 2 PCA biplots with 95% confidence ellipse for quantities of Ca (a), Mg (b), Na (c) and K (d) extracted by various methods ⁱ (represented as vectors). PCA 1 and 2 (vertical and horizontal axes) represent new components resulting from conversion of the original explanatory variables (dimensional reduction).....	98

Figure B- 3 Distribution of CV (%) replicate measurements of Ca (a), Mg (b), Na (c) and K (d) extracted by different methods. Boxes show first and third quartiles, with median line in bold. Whiskers show minimum and maximum values. 101

List of figures: Appendix C

Figure C- 1 Infrared radiation wavelength (cm^{-1}) versus absorbance units for scans of the full soil cation dataset with the Alpha (a), Matrix (b) and MPA (c) spectroscopic instruments. Full soil pH dataset was highly similar, differing by replacement of 6 samples only	109
Figure C- 2 Graphs for calibration & validation of infrared spectral-based models to determine of Ca (a & b), Mg (c & d), Na (e & f) and K (g & h) content.....	110
Figure C- 3 Graphs for calibration (a) & validation (b) of infrared spectral-based models to determine of soil pH	111
Figure C- 4 Spectra after pre-processing, used for calibration of models predicting exchangeable Ca (A, 1st der. / SLS), Mg (B, COE), Na (C, SNV) and K (D, 1st der. / MSC) and soil pH (E, $N_{\text{min-max}}$)	112

List of equations

Equation 2-1.....	5
Equation 2-2.....	7
Equation 2-3.....	8
Equation 2-4.....	13
Equation 2-5.....	13
Equation 2-6.....	21
Equation 2-7.....	22
Equation 2-8.....	22
Equation 2-9.....	23
Equation 2-10.....	23
Equation 2-11.....	26
Equation 2-12.....	26
Equation 2-13.....	26
Equation 2-14.....	27

List of acronyms used in this text

1M – ammonium acetate

ANOVA – analysis of variance

ANN – artificial neural networks

AAS – atomic absorption spectroscopy

CCA – canonical correlation analysis

CEC – cation exchange capacity

COE – constant offset elimination

DRIFT – diffuse reflectance infrared Fourier transform

ECEC – effective cation exchange capacity

EC – electrical conductivity

ESP – exchangeable sodium percentage

ICP-AES – inductively coupled plasma atomic emission spectroscopy

ICRAF-ISRIC – International Council for Research in Agroforestry and International Soil Reference and Information Centre

IR – infrared

IQR – inter-quartile range

M-III – Mehlich-III

MIR – mid-infrared

Nmin-max – min-max normalization

MSC – multiplicative scattering correction

MLR – multivariate linear regression

NIR – near-infrared

NIR-HIS – near-infrared hyperspectral imaging

PLSR – partial least squares regression

CV (%) – percentage coefficient of variation

PZSE – point of zero salt effect

PCA – principal component analysis

PCA – principal component regression

Q – quartile

RPD – ratio of performance to deviation

RPIQ – ratio of performance to interquartile distance

RMSE – root mean square error

RMSECV – root mean square error of cross-validation

RMSEP – root mean square error of prediction

SWIR-HIS – short-wave infrared hyperspectral imaging

SLR – simple linear regression

SAR – sodium adsorption ratio

SOC – soil organic carbon

SOM – soil organic matter

SSA – specific surface area

SD – standard deviation

SEP – standard error of prediction

SNV – standard normal variate

SLS – straight line subtraction

TIR – thermal infrared

Vis-NIR – visible- to near-infrared

Table of Contents

1.	Introduction	1
1.1	Background	1
1.2	Problem statement and rationale.....	2
1.3	Aim and objectives	3
2.	Literature review	4
2.1	Measurement of soil pH.....	4
2.1.1	Soil acidity	4
2.1.2	Concept of pH measurement	6
2.1.3	Construction and evaluation of linear regression models.....	7
2.1.4	Factors affecting soil pH measurement.....	8
2.2	Measurement of exchangeable base cations	12
2.2.1	Importance of base cations	12
2.2.2	Chemistry of exchangeable base cations.....	13
2.2.3	Measurement of CEC.....	16
2.2.4	Comparison of methods of extraction of base cations	17
2.3	Infrared spectroscopy	20
2.3.1	Infrared spectroscopy introduction	20
2.3.2	Infrared spectroscopic analysis methods	23
2.3.3	Models for prediction of soil properties based on infrared spectra	25
2.4	Gaps in knowledge	30
3.	Relationships between soil pH measurement methods.....	32
3.1	Introduction.....	32
3.2	Materials and methods	33
3.2.1	Soil sampling and characterization.....	33
3.2.2	Soil pH measurement.....	34
3.2.3	Statistical analyses.....	35
3.3	Results and discussion.....	35
3.3.1	General soil properties	35
3.3.2	Comparison of soil pH measurement methods	38
3.3.3	Conversion between soil pH measurement methods.....	43
3.4	Conclusions	49
4.	Relationships between soil base cation extraction methods	50
4.1	Introduction.....	50
4.2	Materials and methods	52
4.2.1	Soil sampling and characterisation.....	52
4.2.2	Extraction of soil base cations.....	52
4.2.3	Measurement of soil base cations.....	55

4.2.4	Statistical analyses.....	55
4.3	Results and discussion.....	56
4.3.1	General soil properties	56
4.3.2	Comparison of soil base cation extraction methods.....	58
4.3.3	Conversion between soil base cation extraction methods using SLR models	63
4.3.4	Conversion between soil base cation extraction methods using MLR models.....	67
4.3.5	Recommendation of appropriate use of exchangeable base cation conversion models 72	
4.4	Conclusions.....	73
5.	Infrared spectroscopic determination of soil pH and extractable base cations.....	74
5.1	Introduction.....	74
5.2	Materials and methods	75
5.2.1	Infrared spectral measurement method	75
5.2.2	Statistical analysis	76
5.3	Results and discussion.....	77
5.3.1	Correlation of soil properties with spectra	77
5.3.2	Evaluation of performance of general IR models	78
5.3.3	Evaluation of performance of IR models calibrated by similar samples	81
5.4	Conclusions.....	83
6.	Conclusions and recommendations.....	84
6.1	Conclusions.....	84
6.2	Recommendations for future studies.....	85
	Appendix A: Soil pH conversion appendices	95
	Appendix B: Exchangeable base cation conversion appendices.....	97
	Appendix C: IR spectral model appendices	109

1. Introduction

1.1 Background

One of the most important factors determining the chemical and biological properties of a soil is its pH. The soil solution acidity level affects the availability of nutrients and toxic compounds for uptake by plants, as well as the mobility and fate of pollutants in the soil (Brady and Weil, 2017). Not only is pH a soil quality indicator, but also an important parameter in development of pedotransfer functions (Mojid *et al.*, 2019). There are several standard methods of soil pH determination. In the US pH is measured in deionized water using a 1:1 soil: solution ratio, while measurement in a 1:2.5 ratio with 0.1M calcium chloride (CaCl_2) solution is preferred in the UK. In South Africa a 1:2.5 ratio in 1M potassium chloride (KCl) solution is widely used (The Non-Affiliated Soil Analysis Working Committee, 1990) and in Australia a 1:5 ratio in 0.01M CaCl_2 is used (Rayment and Lyons, 2011). The quantity of exchangeable base cations, which are plant-available forms of calcium, magnesium, sodium and potassium (Ca, Mg, Na and K, respectively), are commonly measured for evaluation of soil fertility. Base cations are a source of nutrients that are essential for plant growth and affect physical properties of the soil, such as tilth (Brady and Weil, 2017). Various chemical extractants have been designed to measure concentrations of these cations in soil testing laboratories. The standard ammonium acetate (1M, pH = 7, Soil Survey Laboratory Staff, 1992), Bray-II (Bray & Kurtz, 1945) and Mehlich-III (M-III, Mehlich, 1984) extraction methods are universally accepted, while South African laboratories also use the ammonium acetate (0.2M, pH = 7, The Non-Affiliated Soil Analysis Working Committee, 1990), Ambic-I (Van der Merwe *et al.*, 1984) and Citric acid (1%, The Non-Affiliated Soil Analysis Working Committee, 1990) methods. The compulsive exchange method (Gillman & Sumpter, 1986) is not employed for routine soil testing, however it has featured in scientific studies globally (Dohrmann, 2006; Gillman & Sumpter, 1986; Hendershot & Duquette, 1986; Jaremko & Kalembasa, 2014; Sumner & Miller, 1996).

IR spectroscopy is a fast, cost-effective, and environmentally friendly alternative to conventional “wet” methods of soil analysis, which involve chemical extraction. Numerous soil properties have been predicted by IR spectroscopic models with varying degrees of success (Johnson *et al.*, 2019; Mohamed *et al.*, 2018; Ng *et al.*, 2019). IR spectroscopy does not directly measure soil properties. The IR instrument exposes samples to radiation of discrete wavelengths. The transmittance of light by the sample, is detected and plotted against the incident radiation wavelength, to form an IR spectrum (Clark, 1999; Shepherd and Walsh, 2002). The spectral signature is characteristic for a given material, as it is a result of the unique interaction of molecular constituents of the material with incident radiation (Mohamed *et al.*, 2018). The physical or chemical property of interest is then predicted based on empirical models calibrated from previous input data (Gobrecht *et al.*, 2014). The degree of transmittance of light by a sample is mathematically related to the concentration of certain

chemical components (Gobrecht *et al.*, 2014). Thus, IR spectra may be used for quantification as well as identification of certain analytes (Mohamed *et al.*, 2018).

1.2 Problem statement and rationale

The interpretation of soil acidity and fertility levels from laboratory test results, is vital for consultants to make appropriate agricultural site assessments and recommendations regarding soil management practices. Critical values and ranges for soil properties have been developed from calibrations on soils with unique sets of physicochemical characteristics. Soil pH and base status is thus often dependant on both the chemical composition of the extractant used by a specific method, as well as soil properties. The method chosen by a laboratory for the determination of soil properties depends on various factors. These include client need (e.g. soil type), historical practices, availability of equipment and chemicals, efficiency of procedures and standardization of soil report formats across a region (Sumner and Miller, 1996). Different methods are often developed with the intention of mimicking *in-situ* conditions. For example, extractants may have an ionic strength or pH like the soils which were used to develop that extraction method. Few studies compare soil pH values and exchangeable base cation concentrations obtained by international methods with those obtained by South African methods. This had led to problems for consultants in the evaluation of soil test results obtained from laboratories using foreign methods, especially for continued monitoring of soils over time. Furthermore, most literature reporting conversions of soil chemical properties between different methods calibrated equations with soils from Europe and North America. These soils differ in chemical behaviour compared to (sub)tropical and (semi-)arid zones due to differences in clay mineralogy and other factors (Sharpley, 1989). Thus, there is a need for conversion equations between international soil test methods for soils in Mediterranean climatic zones such as the Western Cape province of South Africa.

The implications of differing degrees of precision and accuracy of these methods are unknown. The effects of soil pH, soil organic carbon (SOC) content, electrical conductivity (EC) and texture on the quantification of extractable base cations have yet to be thoroughly examined. Conversion of soil test results between different methods would not only benefit consultants. Researchers would also be aided in transforming data from global scientific literature to use in local studies. In addition, generalization of data formats serves to streamline the assimilation of soil survey data for development of large-scale mapping systems. Conversion of historic data allows temporal evaluation of soil quality when new methods are employed in ongoing studies. There is also a demand for more rapid and environmentally friendly techniques of measuring soil chemical properties, such as infrared (IR) spectroscopic methods. IR spectroscopy also shows potential for use in portable scanners to enable *in-situ* measurements and mapping of soil chemical properties with remote-sensed data. There are few studies evaluating the performance of models for prediction of soil pH and exchangeable base cations calibrated from soils originating from a winter-rainfall Mediterranean

climatic zone. Infrared spectral models calibrated from different chemical references need to be evaluated in terms of the practical implications of model performance.

1.3 Aim and objectives

The effects of soil physicochemical properties on conversions of soil pH and base cation concentrations between different extraction methods will be investigated for a set of soils from the Western Cape province of South Africa. This will include investigation of soil properties such as electrical conductivity (EC), texture, organic matter content and acidity. The methods of measuring pH will include combinations of commonly used matrices (water, 0.1M KCl and 0.01M CaCl₂) of varying soil: solution ratios (1:1, 1:2.5, 1:4 and 1:5). The quantity of exchangeable base cations will be determined by six recognized extraction methods, namely (i) AA 1.0 M, pH 7 (ii) compulsive exchange (iii) M-III (iv) Citric acid, (v) Ambic-I and (vi) Bray-II. The performance of models for prediction of soil pH and exchangeable cations based on IR spectral data obtained from different sources will be evaluated. IR spectral data of the soils will be obtained using three IR spectrophotometers. Two of these instruments collect NIR spectra and the third operates in the MIR spectral region. The soil pH and exchangeable base cation measurements will be used to calibrate models to predict these properties from IR spectral data. The relevance of the accuracy of models predicting these soil properties is not often discussed in literature; hence this study will assess the implications of model choice for practical application. The overall aim of this study is to relate different methods of soil pH determination and exchangeable cations to one another mathematically by developing conversion equations. The suitability of IR spectroscopy for determination of soil pH and exchangeable base cation content of Western Cape soils will also be evaluated.

2. Literature review

2.1 Measurement of soil pH

2.1.1 Soil acidity

2.1.1 i) *Importance of acidity*

The soil pH controls the availability of both nutrients and toxic compounds for plant uptake. The mobility of chemicals in the environment is determined by their solubility, rate of breakdown and tendency to adsorb to colloids. All these processes are controlled by the soil solution pH; hence the acidity of a soil influences its fertility, as well as the degree of contamination of surface water and groundwater reservoirs (Brady and Weil, 2017). Specific ion toxicity occurs at low pH levels; this is particularly the case for manganese (Mn) and iron (Fe). The availability of most micronutrients increases with decreasing pH, except for molybdenum (Mo) and boron (B). Macronutrient availability is usually optimal within the slightly to moderately alkaline pH range, while availability of micronutrients decreases at high pH (Foth, 1958). Plants vary in their tolerance for extreme pH conditions and concentrations of trace elements. The choice of crop for agricultural production purposes should be determined by the soil pH of the production site to minimize the necessity of expensive measures of soil pH amelioration.

The biological activity in soils is also reliant on the soil pH. Most fungi are adapted to survive over a wide pH range; however, bacteria do not flourish under acidic conditions and earthworms are sensitive to Al toxicity. The reaction of aluminium (Al) in soils is largely dependent on soil pH. Al is highly toxic to most organisms and is released from soil particle mineral structures at low pH. At soil solution pH > 5, hydroxy Al on exchange sites of colloids form an uncharged species which is released from colloid surfaces (McBride, 1994). Soil fauna populations are responsible for decomposition of soil organic matter (SOM) in the soil and regulating rates of mineralization and immobilization of soil nutrients, affecting availability of nutrients for plant uptake. Soil biota such as mycorrhizal fungi and nitrogen-fixing bacteria play a fundamental role in soil nutrient cycling. Other microbes such as pathogenic organisms and plant root parasites are detrimental to plant growth (Foth, 1958).

The soil pH does not influence the maximum potential cation exchange capacity (CEC) of a soil, as this is a fixed property. However, as soil solution alkalinity increases, protons are released from variable charge sites on surfaces and edges of layer silicates and metal oxide minerals. Al also dissociates from clay surfaces and is released from SOM complexes, precipitating as an uncharged hydroxy species. This also raises the soil's CEC, as previously bound hydroxy-Al polymers block exchange sites and even reduces expanding of clays when adsorbed in mineral interlayers (Foth, 1958). Thus, while CEC remains constant, the effective CEC (ECEC) in soils increases with a rise in pH. The ECEC measured at the soil pH is generally lower than the CEC value, due to occupation

of cation exchange sites by acidic cations. The base cation saturation (%BS) of a soil at a specified pH given by:

$$\%BS = \frac{[Ca]+[Mg]+[K]+[Na]}{CEC} \quad \text{Equation 2-1}$$

where Ca, Mg, Na, K and CEC are in units of $\text{cmol}_c \text{ kg}^{-1}$

Calculation of %BS of cation exchange sites using the maximum CEC is often used for soil taxonomy purposes. On the other hand, calculation of base saturation using the ECEC has important implications regarding soil fertility, as this represents the amount of nutrients that can be stored by soil particles at the current soil pH (Brady and Weil, 2017).

2.1.1 ii) *Acidification process*

Soil pH is a result of two chemical balances – firstly, the quantity of protons versus hydroxide ions in solution, and secondly, the quantity of acidic versus base cations that are adsorbed to colloid surfaces. The nature of the colloid is an important factor influencing the process of cation exchange, which in turn controls the acidification process (Brady and Weil, 2017). The two cations responsible for acidification are H^+ and Al^{3+} ; other cations do not generate acidity and are thus termed non-acidic or base cations. As pH is the negative logarithm of the activity of protons in solution, it is not immediately obvious why Al is also an acidic cation. A large portion of aluminosilicate minerals is comprised of Al. H^+ ions attack and dissolve the minerals on which they are absorbed, releasing structural Al into solution: $\text{H}^+ + \text{clay} \rightarrow \text{Al}^{3+} + \text{Si}(\text{OH})_4$. When Al^{3+} ions are released into solution, they tend to hydrolyse water. This lowers the soil solution pH to an even greater extent, as three moles of acidity are released per mol of Al (McBride, 1994). Respiration by roots and soil organisms promote acidification include the production of carbon dioxide (CO_2), which dissolves in water to form carbonic acid (H_2CO_3). Mineralization of organic matter forms organic acids which contribute to soil acidity, and subsequent oxidation reactions of mineralized nitrogen and sulfur also release protons (Foth, 1958). Precipitation reacts with atmospheric CO_2 to form carbonic acid, which acidifies soil naturally, while industrial nitrogen- and sulfur-containing gases exacerbates the process by causing ‘acid rain’ in industrially polluted regions. Plant uptake of cations (due to exudation of protons by roots for maintenance of charge). Proton-consuming processes which increase soil solution pH include reduction reactions, weathering of minerals containing base cations, and plant uptake of anions. The last-named has an alkalizing effect due to exudation of bicarbonate ions by roots, reversing the dissociation of H_2CO_3 (Brady & Weil, 2017).

2.1.1 iii) *Pools of acidity*

There are three ‘pools’ contributing to total acidity in soils, namely active, exchangeable, and residual (non-exchangeable) acidity. The reserve (titratable) acidity is the sum of nonexchangeable and exchangeable acidity (McBride, 1994). Active acidity represents the free protons present in soil solution and is easily neutralized by addition of basic species. This directly influences root growth and microbial activity (Brady and Weil, 2017). Exchangeable or salt-replaceable acidity is associated

with H and Al that is easily displaced from the soil exchange complex by addition of cations from an unbuffered salt solution. This includes organically complexed Al, acidic functional groups on humus and Al adsorbed to clay surfaces (McBride, 1994). Soils with high CEC inherently possess higher exchangeable acidity and act as a buffer against an increase in pH with addition of lime. Residual acidity exists as H^+ and Al^{3+} ions which are strongly associated with clays and SOM in a non-exchangeable form. This acidity is only released with a considerable increase in pH, precipitating as amorphous $Al(OH)_3$. The amount of residual acidity may exceed the exchangeable acidity by up to 100 000 times in a high-CEC soil (Brady and Weil, 2017).

2.1.1 iv) *Buffering capacity*

The buffering capacity of the soil is defined as the resistance of soil solution to a change in pH upon addition of H^+ or OH^- ions. The buffering capacities of different soils may be compared by constructing titration curves. These are graphs which plot the change in soil pH with addition of acid or base. A flatter slope (smaller change in pH with acid or base addition) is indicative of a greater soil buffering capacity. Various buffering mechanisms may occur, depending on the soil pH range (Brady and Weil, 2017). At an alkaline pH (6.5-8.3), the addition of protons dissolves carbonate minerals such as calcium carbonate ($CaCO_3$) to form bicarbonate and Ca. This reduces the activity of protons in solution, resulting in resistance to a lowering of pH. In the acid to neutral pH range (5.5-8.3), protons are exchanged for cations which are adsorbed to clay surfaces, thus 'storing' the active acidity (McBride, 1994). It follows that soils with higher clay or SOM content and thus higher CEC, have a greater buffering capacity, as a larger amount of exchangeable and reserve acidity must be neutralized to affect a change in soil pH (Brady and Weil, 2017). In acidic soils of $pH < 5.5$, variable charge functional groups (mineral surface hydroxides and organic functional groups) are protonated to buffer the soil pH (McBride, 1994). The clay is limited in its capacity for storage of charge, after which it dissolves, releasing Al ions. Primary minerals have a lower specific surface area (SSA) covered by charge-accepting chemical groups, and hence a lower buffering capacity than layer silicates (Brady and Weil, 2017).

2.1.2 Concept of pH measurement

pH can be measured in-situ by the colorimetric method. This involves the use of chemical dyes that change colour in response to pH. The colours may be compared to standard colour charts for estimation of pH to a 0.2-0.5-unit degree of accuracy. However, the potentiometric method of pH determination is much more accurate. This technique is based on measurement of the potential difference between two electrodes. The membrane surface of a glass electrode responds to the proton activity of the external solution by generating an electric potential. A reference electrode, with a known potential, is connected to the glass electrode by a salt bridge (junction), which allows ions to pass between the electrode (internal) and external solutions. The potentiometer measures the

potential difference relative to a standard buffer. This system forms an electrochemical cell which may be expressed as follows:

glass electrode | standard buffer solution || salt bridge | reference electrode

The pH is then found by means of the Nernst relationship:

$$\text{pH} = \frac{(\Delta E - \text{constant})}{0.059} \quad \text{Equation 2-2}$$

where ΔE = potential difference in units of volts (McBride, 1994)

The pH value measured in the sediment of a soil suspension made up in water ($\text{pH}_{\text{H}_2\text{O}}$) is often lower than that of the supernatant, especially after separation of soil from water by centrifugation (Tan, 1995, as cited in Al-Busaidi *et al.*, 2005). This phenomenon, known as the suspension effect, is caused by electrodes registering the hydronium activity in solution as well as the double layer surrounding colloidal surfaces, where protons tend to accumulate (Al-Busaidi *et al.*, 2005). For this reason, stirring the soil solution results in slightly lower $\text{pH}_{\text{H}_2\text{O}}$ values than static measurement (McClean, 1982). Use of a salt solution during pH measurement minimizes the suspension effect by flocculation of soil particles (Ellis, 1988 and Al-Busaidi *et al.*, 2005). Electrolytes such as KCl and CaCl_2 are used for this purpose. Soil pH is measured in 0.01M CaCl_2 is considered an accurate representation of soil pH at the typical ionic strength of fertile soils (Schofield & Taylor, 1955). Soil pH may be measured in 1M KCl to test for the presence of free Al, which becomes exchangeable at $\text{pH}_{\text{KCl}} \leq 5.2$. Al is displaced from the exchange complex by the relatively high concentration of K ions. Values of pH_{KCl} lower than $\text{pH}_{\text{H}_2\text{O}}$ is associated with a net negative colloidal charge, while the reverse is an indication of net positive charge and hence a greater anion exchange capacity (AEC) of the soil (Kome *et al.*, 2018).

2.1.3 Construction and evaluation of linear regression models

The conversion between pH-values obtained by two different methods may be mathematically expressed as a function which predicts the unknown pH (dependent variable) from the known pH (explanatory variable). Linear regression involves the fitting of a straight line to a plot of one or more explanatory variables versus the dependent variable. The line's gradient and intercept are such that there is maximum correlation between predictor and response variables and minimum difference between measured and predicted values (Naes *et al.*, 2002). A measure of the fraction of total variation that is accounted for by the model is given by the coefficient of determination (r^2 value, ranging from zero to one) with the remaining variation attributed to random error (Awiti *et al.*, 2008). It is the square of the correlation (r) between the measured and predicted values. The higher the r^2 value, the greater the correlation between the explanatory and response variables (Bellon-Maurel *et al.*, 2010). In the case of test set validation (see Section 2.3.3), the r^2_{cal} and r^2_{val} parameters represent the r^2 -values obtained from calibration and validation, respectively.

The root mean square error (RMSE) is a measure of model prediction accuracy and is based on the squared difference between actual and predicted y-values. Calculated as the square root of the variance of the residuals, RMSE is conveniently expressed in the same units as the response variable (Grace-Martin, 2020). A lower RMSE-value indicates better absolute fit of the observed data to the model (Naes *et al.*, 2002). The upper and lower confidence limits of the standard error may be plotted to produce a confidence interval (associated with the predicted mean) or prediction interval (associated with a single predicted value). These intervals, usually calculated at a 95% confidence level, show whether observed data lie within the uncertainty ranges to illustrate the predictive capability of the model (Clewer and Scarisbrick, 2001).

The Akaike Information Criterion (AIC) compares the relative quality of a collection of models of varying degrees of complexity for the same dataset. AIC evaluates the model's fit while adding a penalty term for the model complexity. Overfitting occurs when a model includes unnecessary explanatory variables, which increases the standard error of regressors and results in poor predictions (Salkind, 2020). AIC is therefore a useful tool in overcoming problems of multicollinearity amongst independent variables. A lower AIC value indicates a better balance of model fit with predictive capability on out-of-sample data. AIC is calculated by

$$\text{AIC} = 2\ln(L) + 2k \qquad \text{Equation 2-3}$$

where:

L = maximum likelihood estimation of fit

k = number of parameters (Burnham & Anderson, 2002)

The use of a regression equation for pH approximation may be replaced by a simpler method of conversion. If the slope does not differ significantly from one, conversion may be achieved by adding a constant. If the intercept does not differ significantly from zero, conversion may be achieved by multiplication of the independent variable by a coefficient.

2.1.4 Factors affecting soil pH measurement

2.1.4 i) *Solution matrix*

In Europe and America, 0.01M CaCl₂ is commonly used for the soil suspension matrix in pH measurement. Calcium ions displace some acidity from cation exchange sites into the bulk solution, resulting in pH_{CaCl₂} readings that may be 0.2-0.5 units lower than pH_{H₂O} for the same sample (Brady and Weil, 2017). Kissel and Miller (2010) found a strong linear relationship between pH_{H₂O} and pH_{CaCl₂} using soil: solution ratios of 1:1. A salt solution of 1M KCl has much higher concentration than 0.01M CaCl₂ and displaces the total exchangeable acidity into the active pool; thus, pH_{KCl} readings

are even lower than $\text{pH}_{\text{CaCl}_2}$. The electrolytes also cause initial reduction of the negatively charged colloidal surface's surrounding double layer potential. This promotes the release of hydronium ions into solution, decreasing the measured pH (Al-Busaidi et al., 2005). The effect of KCl for a specific soil depends on its ion exchange capacity, mineralogy, and soluble salt content. In highly acidic soils, the pH_{KCl} may be higher than the $\text{pH}_{\text{H}_2\text{O}}$ value due to displacement of OH^- ions by Cl^- ions from anion exchange sites of positively charged colloids (Brady and Weil, 2017). Bruce et al. (1989) demonstrated that the conversion between pH values measured in different electrolytes had a slope of unity but an intercept dependent on the solution ionic strength.

For conversion of pH_{KCl} to $\text{pH}_{\text{H}_2\text{O}}$ values, a logarithmic relationship was found to yield reliable results (Galka et al., 2016). Galka et al. (2016) also found higher correlation coefficients for equations including additional parameters such as total Carbon (C), clay content and %BS compared to models using only pH_{KCl} as a predictor. van Lierop (1981) found that near-identical $\text{pH}_{\text{CaCl}_2}$ versus pH_{KCl} values were obtained for a given soil: solution ratio. A relatively precise linear relationship was found between $\text{pH}_{\text{H}_2\text{O}}$ and $\text{pH}_{\text{CaCl}_2}$ (at 1:1 and 1:2 ratios, respectively) and between $\text{pH}_{\text{H}_2\text{O}}$ and pH_{KCl} (identical ratios). The electrostatic attraction between K and negatively charged colloids results in dissimilar rates of diffusion of K and Cl^- ions in a glass-electrode assembly (McBride, 1994). Hence KCl solution produces a junction potential which would have been absent during calibration of the electrode with a standard buffer of different ionic strength. This gives rise to a slightly lower pH reading in a colloidal suspension compared to pure water (Illingworth, 1981; van Lierop. & Mackenzie, 1977).

2.1.4 ii) Soil: solution ratio

Modern soil classification databases require soil pH measurement at a 1:5 soil: solution ratio according to international standards (German Institute for Standardization, 2015). However, European countries typically use a 1:2.5 ratio (Galka et al., 2016) with 1M KCl or 0.01M CaCl_2 , while other ratios (1:1, 1:2 or 1:4) with CaCl_2 have been used in Canada (van Lierop, 1981), the US (Kissel and Miller, 2010) and Australia (Little, 1992). A study by Galka et al. (2016) found near-identical pH values using 1:2.5 versus 1:5 ratios in both water and KCl solution. These results confirmed an earlier study by van Lierop (1981) where 1:1, 1:2 and 1:4 soil: solution ratios resulted in near-identical pH values for a given matrix. Kome et al. (2018) found that soil pH increased with an increase in soil: solution ratio and ascribed this to the dilution effect, which has been suggested to be caused by a decrease in proton dissociation from soil surfaces and less hydrolysis of Al (Thomas, 1982). Linear relationships between $\text{pH}_{\text{H}_2\text{O}}$, $\text{pH}_{\text{CaCl}_2}$ and pH_{KCl} using several combinations of soil: solution ratios (1:1, 1:2.5 and 1:5) were constructed for pH conversion in volcanic soils (Kome et al., 2018).

2.1.4 iii) Soil pH range

Van Lierop (1981) showed that the difference between soil pH values obtained in salt solutions of various ionic strengths increased for soils of a higher pH range. Acidic soils with variable charge characteristics showed the same trend and an explanation was offered by Aitken & Moody (1991) in terms of the point of zero salt effect (PZSE). The PZSE is defined as the pH value at which hydronium titration curves, performed at various ionic strengths, intersect (Parker et al., 1979, as cited in Aitken and Moody, 1991). The PZSE may therefore be understood as the pH at which the ionic strength of soil solution has no effect on its pH. Aitken & Moody (1991) suggested that the greater difference between soil values pH measured in solutions of different ionic strength found in alkaline soils, compared to acidic soils, was a result of alkaline soils being further from their PZSE than acidic soils. Curvilinear relationships were found to describe Aitken and Moody's (1991) data better than linear regression curves. Hence, it was concluded that trendlines are only linear for soils falling within a narrow pH range.

Further research was conducted by Little (1992) with extension of the alkaline soil pH data range. Little (1992) verified the curvature at low pH as recorded by Aitken and Moody (1991) by fitting a cubic polynomial to $\text{pH}_{\text{H}_2\text{O}}$ versus $\text{pH}_{\text{CaCl}_2}$ for soil: solution ratios of 1:5, resulting in a sigmoidal curve which also deviated from linearity at high pH. Little (1992) attributed the non-linearity to buffering mechanisms in the extreme pH regions, resulting in smaller resulting differences between pH measurements in salt solution versus water. An additive model was constructed by Henderson and Bui (2002) using a larger dataset than previous studies. The resulting curve (see Figure 2-1) was sigmoidal over the pH range of 2.5-10.5. The additive model was interpreted in terms of three pH ranges where different chemical behaviour is expected to dominate. The middle pH region represented cation exchange activity and the extreme pH regions represented soil pH buffering mechanisms (carbonate and hydroxy-Al/Fe species above pH 7 and below pH 4.2, respectively). The effect of soil texture on soil pH was more pronounced for highly alkaline soils. This observation is consistent with the pH buffering mechanism theory, as carbonate-based particles fall within the silt- to clay-sized fractions (Henderson & Bui, 2002).

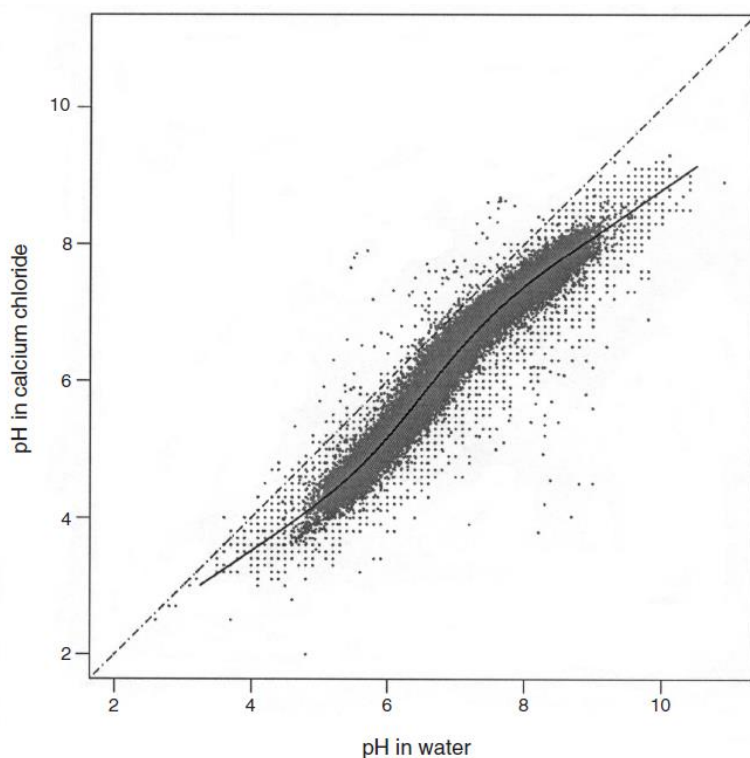


Figure 2-1 Scatter plot of pH in CaCl₂ vs pH in water with the additive calibration curve superimposed (solid line). The broken line shows the 1:1 relationship (Image source: Henderson & Bui, 2002)

2.1.4 iv) Electrical conductivity

The influence of EC of the soil suspension on pH measurement was studied by Kissel and Miller (2010). The pH of a given soil sample suspension was shown to decrease as EC increased according to a non-linear relationship. To determine whether the pH decreased due to an electrode response or due to displacement of hydronium ions from exchange sites by ionic components of the electrolyte, Kissel and Miller (2010) performed the following procedure. The pH_{H₂O} as well as pH_{CaCl₂} of three sample extracts (ratio of 1:1) were measured and converted to hydronium concentration. The soil extracts were then titrated with standard NaOH solution to determine the experimental pH, which was converted to hydronium concentration. The difference between the hydronium concentrations in water and salt solution was calculated for the measured and experimental values. It was found that for alkaline soils, the measured differences between pH values in water versus electrolyte were not the same as the experimental differences. Thus, the differences were only partly explained by cation exchange, and it was concluded that an electrode response must have contributed to the ionic effect. Equations incorporating both EC and pH_{CaCl₂} as regressors improved the accuracy of predictions of soil pH_{H₂O} (Kissel and Miller, 2010).

In a study by Brough *et al.* (2011), a regression of the difference between pH values measured in water and CaCl₂ (Δ pH) versus EC_{H₂O} showed a decrease in Δ pH with an increase in EC. The regression parameter obtained for acidic soils was nearly identical to that in alkaline soils, showing

that the effect of EC is independent of soil pH. pH values obtained in different solution matrices at comparable soil: solution ratios were related by logarithmic functions containing a non-linear EC term. Curvilinear pH prediction models were also constructed using artificial neural networks (ANN). The ANN model performed slightly better than the logarithmic model (Brough *et al.*, 2011). The ANN model was similar in shape to the additive model of Bui *et al.* (2002). Linear functions for converting $\text{pH}_{\text{CaCl}_2}$ to $\text{pH}_{\text{H}_2\text{O}}$ were not significantly improved by use of smoothing spline functions, ANN or random forest (RF) models in a study by Ferguson *et al.* (2014). It was concluded that the advantages of using a simple equation outweighed the precision-related benefits of including factors such as EC, mineralogy, and texture, as generating complex equations requires time and labour for data collection and processing. Kome *et al.* (2018) found that EC values were highly and negatively correlated with $\text{pH}_{\text{H}_2\text{O}}$ and $\text{pH}_{\text{CaCl}_2}$ in non-saline soils, but that multiple linear regression models incorporating EC as a variable did not significantly improve model performance. The authors suggested this to be a result of the relatively small sample set used for model calibrations and argued that if models are not made for similar soils, they may not be valid for individual soil types.

2.2 Measurement of exchangeable base cations

2.2.1 Importance of base cations

Base cations are mainly derived from the weathering of soil minerals. Only a certain proportion of the total soil nutrient concentration is in a labile form – that is, available for assimilation by biological organisms. Plant-available nutrients adsorbed to soil colloid surfaces are in rapid equilibrium with the soil solution (Huang *et al.*, 2012). Root hairs and microbes release H^+ to replace base cations from the soil exchange complex for consequent assimilation of these nutrients by their own adsorptive surfaces. Hence, soil fertility testing involves selective measurement of the labile fraction of the nutrient pool via exchange, dissolution, or complexation with some component of a chemical extractant (Brady and Weil, 2017).

The chief mineral sources of magnesium (Mg) are biotite, dolomite, olivine, serpentine and hornblende minerals (Huang *et al.*, 2012). Mg is a structural component of the chlorophyll molecule and thus essential for the photosynthesis process in plants. Mg is also required for synthesis of oils and proteins, as well as activation of enzymes catalysing plant metabolic reactions. The Ca: Mg ratio influences soil physical properties due to dispersion of soils with high exchangeable Mg content, which is counteracted by the flocculating effects of Ca ions. The Ca: Mg ratio is also an important factor to consider in forage and pasture crops for optimizing nutrition of grazing animals (Brady and Weil, 2017). The main mineral sources of Ca are carbonates, feldspars, and phosphates. Ca tends to dominate the soil exchange complex (Huang *et al.*, 2012). Ca is vital for cell division and elongation processes, membrane permeability and activation of critical enzymes. Calcium-pectate is a structural component of cell wall lamellae in plants. Its role in maintaining integrity of cell walls and membranes

is of importance in protecting the plant against pests, diseases and elemental toxicity (Brady and Weil, 2017).

Potassium (K) ions originate from feldspar and mica minerals. A large proportion of total K is fixed in the interlayers of clay minerals, providing a source of slowly-released nutrients (Huang *et al.*, 2012). K activates enzymes catalysing various cellular reactions in plant tissues, such as starch and protein synthesis, sugar degradation, energy metabolism, photosynthesis, and nitrogen fixation in legumes. K plays a critical role in lowering the water potential of cytoplasmic solution by regulating water uptake by root cells and water loss from stomatal cells. K also improves the general hardiness of plants by increasing their tolerance for drought, frost, pests, and pathogens. The quality of flowers, fruits and vegetables are also improved by proper K nutrition due to enhanced colour and flavour (Brady and Weil, 2017). Shallow saline groundwater, perched water tables, irrigation with water containing high levels of Na, and poor drainage promote accumulation of sodium in soils, especially in arid regions. Soils are classified as sodic when the exchangeable sodium percentage (ESP) exceeds 15, where:

$$\text{ESP (\%)} = \frac{[\text{Na}]}{\text{ECEC}} \times 100 \% \quad \text{Equation 2-4}$$

where Na and ECEC are in units of $\text{cmol}_c \text{ kg}^{-1}$

Soils may also be diagnosed as sodic, when the sodium adsorption ratio (SAR) is greater than 13, where:

$$\text{SAR} = \frac{[\text{Na}]}{\sqrt{[\text{Ca} + \text{Mg}]/2}} \quad \text{Equation 2-5}$$

where Na, Ca and Mg are in units of $\text{cmol}_c \text{ kg}^{-1}$

Sodic soils are highly dispersed, poorly aerated and have low infiltration rates (Huang *et al.*, 2012). These soils are also extremely alkaline and thus show specific nutrient deficiencies and toxicities due to reaction of these elements at high pH. Na contributes to soil salinity which lowers the soil water potential, reducing water uptake by plants. Seedlings are particularly sensitive to salt damage. High levels of Na have specific toxic effects on plants and may compete with other cations for uptake by plant roots (Brady and Weil, 2017).

2.2.2 Chemistry of exchangeable base cations

2.2.2 i) Adsorption and ion exchange

Sorption is a broad term covering different types of interaction of chemical compounds with colloid surfaces, including ion exchange, inner-sphere complexation, and partitioning. Soil clay and humus particles known as micelles possess positive and negative electrostatic charges which attract a swarm of anions and cations present in the soil solution (Foth, 1958). An inner-sphere complex forms when the ion bonds directly with the colloid surface. Alternatively, hydrated cations and anions form an outer-sphere complex with colloids, with water molecules forming a bridge between the ion and

the colloid surface (Brady & Weil, 2017). Hydrated ions are weakly held by electrostatic forces and thus easily replaced by other ions approaching closely from the soil solution. These adsorbed ions are termed “exchangeable”. An exchangeable ion constantly oscillates near the colloid surface and diffuses into solution when replaced by a similarly charged ion (Foth, 1958). The freed ion may be leached or adsorbed by microbial or plant adsorptive surfaces. When a hydrated cation is replaced by another at a negatively charged colloid site, the process is termed cation exchange. Conversely, anion exchange occurs at positively charged exchange sites. All colloids in the soil capable of holding exchangeable ions are termed the exchange complex. The proportion of the exchange complex satisfied by a given cation is termed the “saturation percentage” for that ion. A high percentage saturation of exchange complex by a nutrient results in more rapid displacement of this nutrient and hence, greater availability for plant uptake (Brady & Weil, 2017). The cations that dominate the exchange complex greatly influence soil physicochemical properties, for example the dispersive effects of Mg and Na on soil colloids (Foth, 1958).

2.2.2 ii) Sources of charge on colloids

Soil colloid surfaces possess two types of electrostatic charge, namely permanent (constant) charge and pH-dependent (variable) charge. A permanent negative charge is the result of substitution of one type of cation in the clay crystal structure by another cation with similar size but lower charge. Tetrahedrally or octahedrally-coordinated cations in the layered sheets of phyllosilicates typically undergo this isomorphous substitution. Variable charge sites, on the other hand, are the hydroxyl groups found on the surfaces or broken edges of clay minerals (OH attached to Si, Fe or Al) and organic colloids (OH attached to C). A moderately positive pH-dependent charge may develop due to protonation of hydroxide groups on colloid surfaces at low pH (McBride, 1994). A negative pH-dependent charge may result due to dissociation of protons from acidic hydroxyl groups on colloid surfaces. As the soil solution pH increases, $\text{Al}(\text{OH})_2^+$ on exchange sites also react with hydroxide ions in solution, forming insoluble $\text{Al}(\text{OH})_3$, thus freeing negatively charged cation exchange sites (Foth, 1958).

Certain groups of colloids have characteristic CEC ranges due to structural sources of permanent charge or a known quantity of pH-dependent surface charge. Sesquioxides (non-silicate metal oxides and hydroxides) and the kaolin-serpentine group of layer silicates are prevalent in highly weathered soils. These minerals possess mainly pH-dependent charge and have low CEC (Brady & Weil, 2017). Chlorites are a group of phyllosilicates that balance the negative charge in their interlayers (resulting from isomorphous substitution) with a positively charged metal hydroxy precipitate, thus blocking exchange sites, and lowering the CEC. Montmorillonites, a member of the smectite-saponite group, undergo substitution of trivalent cations by divalent ones, decreasing the net surface charge and favouring cation exchange in the interlayer. Vermiculites undergo isomorphous substitution in several layer sheet-types, leading to a greater negative surface charge

and higher CEC than smectites. Micas undergo the most isomorphous substitution and have such a large net negative charge that K cations are fixed within the interlayers, blocking exchange sites, and lowering the CEC. Illites are formed by the weathering of micas. Some loss of K ions results in slightly higher CEC compared to micas. Sources of non-crystalline minerals such as allophanes and imogolites include recently deposited volcanic ash and acidic, humic soils. The irregularity of these crystal structures causes them to have high specific surface area. A large quantity of variable-charge sites produced by this area results in non-crystalline structures having very high CEC (McBride, 1994).

2.2.2 iii) *Cation exchange principles*

Cation exchange is a reversible chemical reaction that follows a charge equivalence rule. This means that exchange occurs on a charge-for-charge basis; for example, it requires two monovalent cations to replace one divalent cation (Foth, 1958). The ratio law states that for two cations undergoing exchange at equilibrium, the ratio of one cation to another on the colloid will be the same as the ratio of that cation to the other in solution (Brady and Weil, 2017). Although exchange is reversible, it will favour the forward reaction if the released cation is removed from solution (e.g., by strong association with an anion or volatilization). This principle is known as the law of mass action. Selectivity is the preferential adsorption of one cation over another due to chemical properties of the cation and / or colloid exchanger. A cation with a relatively high charge to hydrated radius ratio will be more strongly adsorbed. Weakly held cations oscillate further from the colloid and are more likely to be replaced. The selectivity of clays for cations decreases according to the atomic number in the following order: $\text{Al} > \text{Ba} > \text{Sr} > \text{Ca} > \text{Mg} > \text{Cs} > \text{K} > \text{NH}_4 > \text{K} > \text{Na} > \text{Li}$ (Foth, 1958; Mengel *et al.*, 2001).

The likelihood of a cation being displaced from a colloid also depends on how strongly its complementary cations are adsorbed to the colloid. For example, K is more likely to be displaced if its neighbouring cations are strongly adsorbed Al^{3+} ions than if they were weakly-held Na ions. The weathering of parent material releases K, Mg and Na into solution. Leaching of these cations results in the exchange complex being dominated by Ca, Al, Al hydroxy ions, and H in humid regions, or Ca, Mg and Na in drier regions (McBride, 1994). Micas and vermiculite show greater preference for K, NH_4 and Cs in inter-tetrahedral spaces exposed at weathered crystal edges, while copper (Cu), mercury (Hg) and lead (Pb) have high selective affinities for sites on humus and Fe oxides (Brady and Weil, 2017). The exchanger surface activity coefficient may vary with the mole fractions of exchange sites occupied by cations. Hence, the conditional equilibrium constant (K_c) for an exchange reaction will also vary with composition of the exchanger (Sumner & Miller, 1996).

2.2.3 Measurement of CEC

CEC is a soil chemical property used for classification in soil taxonomy, as well as evaluation of soil fertility and potential environmental impact. It is defined as the total number of exchangeable cationic charges that a soil can adsorb and may be expressed in centimoles of positive charge per kg soil ($\text{cmol}_c \text{ kg}^{-1}$), or milliequivalents per gram (me g^{-1}) (McBride, 1994). CEC may be determined by leaching a soil sample with a concentrated solution of an exchanger cation that completely replaces the exchangeable cations initially held by the soil. This is followed by measurement of either the number of exchanger cations adsorbed to the exchange complex, or the amount of each of the displaced elements in solution. Quantitative analysis of the leachate is performed by techniques such as atomic absorption spectroscopy (AAS), inductively coupled plasma atomic emission spectroscopy (ICP-AES) or colorimetry (Huang *et al.*, 2012). In this way the concentration of exchangeable basic is determined. ICP-AES is the most accurate, efficient, and sensitive technique of those named, however it is also the most expensive. The high stability of elements during the combustion process is a result of the properties of light source plasma (Harris, 2010).

CEC measurements are affected by the chemical components of the extraction solution, the pH of buffered solutions, and the presence of soluble salts or carbonates in soils. Buffered CEC-measurement methods involve maintenance of a certain pH of the extraction solution. This measures the potential CEC. The ECEC is the CEC of the soil as sampled in the field (Brady and Weil, 2017). The pH of acidic soils is increased by extraction with solutions buffered to a higher pH. Variable charge sites on colloid surfaces are then deprotonated, increasing the CEC (Sumner and Miller, 1996). While CEC measurement is affected by a buffered extractant, the base cation concentrations remain unaffected due to differences in methodology. Exchangeable base cations are measured directly in the leachate, while CEC is determined by measuring the quantity of index cations displaced from exchange sites by a second extractant. Only the latter process is affected by deprotonation of variable charge sites in a buffer solution. Organic functional groups on colloid surfaces may also show different cation selectivity with changing pH (Miller *et al.*, 1990). Buffers containing phosphate and acetate anions result in formation of complexes with divalent cations, which also complicates CEC determination (Sumner and Miller, 1996). Measurement of CEC at pH levels below 4.5 should be avoided, due to release of soluble Al. Acidic saturating solutions may overestimate extractable K due to protons being small enough to enter the interlayers of 2:1 clays minerals, displacing non-exchangeable K (Huang *et al.*, 2012). An acidic extractant may also overestimate exchangeable Ca and Mg in arid soils, due to dissolution of Ca- and Mg-rich minerals.

Some CEC measurement methods involve a water / alcohol wash for removing the entrained solution. If the electrolyte is not completely removed, or if the washing solution cause hydrolysis of cations, the measured CEC is underestimated (Sumner and Miller, 1996). Conversely, washing with a solution of low ionic strength may cause dissolution of the clay itself, releasing cations from the mineral structure and resulting in CEC overestimation (Frenkel and Suarez, 1977). Slowly dissolving

salts, such as free gypsum (CaSO_4) or CaCO_3 found in calcareous soils, may also become soluble in the washing solution (Amrhein and Suarez, 1990). The BaCl_2 extractant has been shown to dissolve more limestone and gypsum than other methods, due to precipitation of Ba carbonates and sulphates; Ouhadi & Deiranlou, 2011). This releases additional Ca ions into the extract. Use of an extractant containing K or NH_4 cations may result in an underestimation of CEC, due to fixation of these cations in the interlayer spaces of micaceous mineral structures (Bower, 1950). Underestimation of CEC may also occur when using extractants consisting of Na or NH_4 , due to hydrolysis of these cations which lowers their displacement effectivity (Sumner and Miller, 1996). Polyvalent ions may form hydroxy ions, which results in an incorrect estimation of ECEC if they are assumed to be present as simple ions (Bache, 1976). If extractant solutions contain sulfate (SO_4) or phosphate (PO_4), these anions may form inner-sphere complexes with variable-charge sites on colloid surfaces, leading to overestimation of CEC (Sumner and Miller, 1996).

2.2.4 Comparison of methods of extraction of base cations

2.2.4 i) *Development of extraction methods*

Various attempts have been made to develop CEC measurement methods that correct for the effects of solution pH, the presence of free salts in soils, and an equilibrium electrolyte concentration differing from that of the natural soil solution. The use of an ammonium acetate (NH_4OAc) extractant buffered to a pH of 7, first proposed by Schollenberger and Dreibelbis (1930), leads to the overestimation of CEC in acidic soils (Sumner and Miller, 1996). However, the 1M NH_4OAc (1M AA, pH 7) method as described by Soil Survey Laboratory Staff (1992) remains a standard (Huang *et al.*, 2012). An adaptation of this method, using 0.2M NH_4OAc buffered to a pH of 7 (The Non-Affiliated Soil Analysis Working Committee, 1990), is also used for base cation quantification in the South African province of the Western Cape. Water soluble cations may be determined separately in saline soils (resistance < 460 ohms) and subtracted from the total extractable cations for a measure of the exchangeable cations. Gillman and Sumpter (1986) established the compulsive exchange method, which is designed for use in weathered, saline, and calcareous soils. This method is recommended by the Soil Science Society of America (Sumner and Miller, 1996) as the ionic strength of the equilibrium solution with which CEC is determined is adjusted to match the natural soil solution concentration (ca. 0.006 M) (Sumner and Miller, 1996). However, the procedure is time-consuming and generates toxic waste in the form of hydrated barium chloride ($\text{BaCl}_2 \cdot 2\text{H}_2\text{O}$), thus is not widely used (Ross and Ketterings, 2011). The extractant consists of 0.2M BaCl_2 and 0.2M ammonium chloride (NH_4Cl), with Ba outcompeting other exchangeable cations for adsorption to clays due to high charge to hydrated radius ratio (Mengel *et al.*, 2001).

The Bray-methods were originally developed for extraction of available and reserve phosphates in soil but may be used for multi-elemental extraction. The Bray-II method is commonly used for P-determination in soil analytical laboratories of the Western Cape province of South Africa, as well as

tropical regions such as Malaysia and Columbia (Zapata & Roy, 2004). Bray-II solution, which consists of 0.1M HCl and 0.03 M ammonium fluoride (NH_4F), drives the acidic dissolution of Ca-, Al- and Fe-phosphates by complexation of fluoride with Al and Fe ions and precipitation of Ca with fluoride (Schmidt *et al.*, 2004). The soil extract is acquired more rapidly compared to other methods (Ballard, 1978). Mehlich-III (M-III, Mehlich, 1984) is a common method of measuring exchangeable cations in the US (Huang *et al.*, 2012) and is also used in the Czech Republic for soil P measurement (Csathó & Magyar, 2005). Like Bray-methods, the Mehlich-methods were originally designed to extract P from soils. The M-III extractant was developed with the aim of enabling additional extraction of Cu and reducing corrosivity of the solution. The saturating solution consists of 0.2M acetic acid (CH_3COOH), 0.25 M ammonium nitrate (NH_4NO_3), 0.015 M NH_4F , 0.013 M nitric acid (HNO_3) and 0.001M ethylene diamine tetra-acetic acid (EDTA). The Ambic-I extraction method, based on the International Soil Fertility Evaluation and Improvement (ISFEI) method of Hunter (1974) and modified by Van der Merwe *et al.* (1984), is still used in South African laboratories for P determination, particularly in the province of KwaZulu-Natal. The extractant solution consists of 0.25 M ammonium bicarbonate (NH_4HCO_3), 0.01M $(\text{NH}_4)_2$ EDTA, 0.01M ammonium fluoride (NH_4F) and Superfloc, buffered to a pH of 8.2. The saturating solution solubilizes extracted Al, Fe and Ca species via complexation with EDTA. The Citric acid (1%) extraction procedure was developed by Dyer (1984). Citric acid ($\text{C}_6\text{H}_8\text{O}_7$) solubilizes phosphates of Ca, Al and Fe and simultaneously displaces cations from the soil exchange complex (The Non-Affiliated Soil Analysis Working Committee, 1990). This extraction technique is also currently used in soil analysis laboratories of South Africa for P and exchangeable base cation content determination, particularly in the province of the Western Cape.

2.2.4 ii) Conversion equations

The relationship between AA (1M) and other extraction methods for determination of base status in soils has been studied extensively. Michaelson *et al.* (1987) found nearly one-to-one relationships between M-III and AA-extracted K, as well as Mg, as did Barbagelata and Mallarino (2013). However, extractable K, Ca and Mg determined by the M-III method were higher than that determined by the AA method according to Martins *et al.* (2015). This was especially the case at higher base cation concentrations according to a study by Culman *et al.* (2020). Mamo *et al.* (1996) did not find significant correlation between Na values obtained by M-III and AA extraction methods, however, in a study by Wang *et al.* (2004) the M-III procedure was shown to result in lower Na values compared to the AA method. Gillman *et al.* (1983) and (Horn, 1982) compared extraction using 0.1M BaCl_2 with the AA procedure. These studies found higher quantities of extractable Ca and Mg with the former method. This was ascribed to more efficient displacement of divalent cations, as well as greater dissolution of limestone by the Ba ion compared to the NH_4 ion. Jönsson *et al.* (2002) found that similar amounts of Ca and Mg were displaced when comparing the compulsive exchange method with extraction using an unbuffered salt (NH_4Cl). This disagrees with the theory of superior

displacement of divalent cations by Ba as suggested by Gillman *et al.* (1983) and (Horn, 1982). Michaelson *et al.* (1987) proposed that the nature of the soil affected the regressions for M-III extractable Ca versus AA-Ca. This observation was also offered by Gartley *et al.* (2007). However, Wang *et al.* (2004) found no significant effect of soil pH or texture on K, Ca, Mg and Na values obtained. Likewise, Zhu *et al.* (2016) found no correlation between soil pH, carbonate content, OM or EC and M-III extractable K.

Studies by Jönsson *et al.* (2002), Ciesielski *et al.* (1997) and Amacher *et al.* (1990) show that BaCl₂ is less effective at displacing K-ions from the exchange complex than ammonium-based extraction agents. This was attributed to the higher concentration of NH₄ (1M) than Ba (0.1M), as well as the similar hydrated ionic radius of the monovalent cations (Jönsson *et al.*, 2002). Several studies (Doll & Lucas, 1973, Ballard and Pritchett, 1975 and Mehlich, 1974 as cited in Ballard, 1978) have shown that quantities of K and Mg extracted by Bray-II solution were quantitatively similar to those extracted by the AA method. Thompson (1995) found high correlation coefficients for the relationship between extractable K obtained by AA, M-III, Ambic-I, citric acid, Bray-II and the ISFEI (precursor to Ambic-I) methods. A study by Simard *et al.* (1989) found that the amount of K released by successive extractions with 5 x 10⁻⁴ M citric acid was highly correlated with the amount of NH₄OAc-extractable K, as well as the coarse clay content of the soil. The latter may be explained by the tendency of certain phyllosilicates to fix K mineral structure interlayers, as discussed in Section 2.2.2. Oxalic (C₂H₂O₄) and citric acids were suggested to reduce the K-fixation capacity of minerals by partially dissolving crystal edges, covering K-fixing sites with calcium oxalate (CaC₂O₄) precipitate, or causing Ca-induced mineral flocculation and slowing diffusion of K to the interlayer (Najafi-Ghiri *et al.*, 2019). Soluble Ca and Mg in calcareous soils may displace the K fixed in clay minerals, thus increasing the extractable K measured.

2.2.4 iii) PCA for complex data

Principal component analysis (PCA) is a method of identifying patterns in complex data by reducing the dimensionality of large datasets without losing important information (Jolliffe & Cadima, 2016). The essence of the data is captured in principal components which account for the most variation in the dataset (Ngo, 2018). The latent variables (orthogonal factors) are found by solving for eigenvalues (Mohamed *et al.*, 2018). Some non-linearity may be overcome by increasing the number of latent variables incorporated into the model (Pasquini, 2018). PCA biplots are useful for examining multidimensional datasets, as they contain information about factors contributing to differences between samples. A PCA biplot is the combination of a PCA plot with a loading plot. The PCA plot occupies the bottom and left axes of the biplot (1st and 2nd principal component scores, respectively). It consists of dots representing individual samples with coordinates corresponding to their principal component scores. Samples may form clusters based on similarity. Outliers may be identified as dots lying outside of a 95% confidence ellipse. These are the samples whose variance cannot be

explained in by the principal components in the same manner as the rest of the dataset. A loading plot shows variables as vectors which show the degree to which each variable contributes to the principal components (x- and y-axes representing the 1st and 2nd principal components, respectively). The magnitude of the vectors represents the amount of variation these variables account for in the data. Vectors point in the positive direction of axes representing the principal component to which they carry positive contribution. Vectors also point towards sample clusters with which the factors they represent are associated (Carter, 2019). Vectors pointing in similar direction are strongly correlated, where strength of correlation decreases from unity to zero for an angle of zero to 90 degrees between vectors.

2.3 Infrared spectroscopy

2.3.1 Infrared spectroscopy introduction

2.3.1 i) *Advantages of IR spectroscopy for soil analysis*

Thus far, only the “wet” methods of soil analysis have been discussed. These involve chemical equilibration, or extraction followed by quantitative analysis of the leachate (Huang *et al.*, 2012). IR spectroscopy has been proven to be a rapid, cost-effective, and non-destructive alternative technique of soil analysis. Spectroscopy is the use of emitted light for obtaining chemical information. The parameter of interest, measured by a reference method, is correlated with the spectra. This relationship is used to generate prediction algorithms (usually by means of regression techniques) so that the parameter may be estimated in new samples using solely spectral information (Aleixandre-tudo *et al.*, 2017). The procedure is environmentally friendly, as no extracting chemicals are needed (Bushong *et al.*, 2015). The only sample preparation steps required for soils are drying and sieving or crushing (Mohamed *et al.*, 2018). A single spectrum contains information about several soil characteristics, enabling simultaneous prediction of these properties (Islam *et al.*, 2003 as cited in Nocita *et al.*, 2015). The reproducible and accurate determination of several soil attributes (CEC, EC, sum of base cations, exchangeable Ca, Mg, K and Al, concentrations of heavy metals, phytoavailable P, soil pH, SOC, nitrate-N content, lime requirement, extractable Fe, moisture content and clay mineralogy) with IR spectroscopy has been recorded by numerous authors (Wan *et al.*, 2020, Chen *et al.*, 2021, Johnson *et al.*, 2019; Mohamed *et al.*, 2018; Ng *et al.*, 2019; Nocita *et al.*, 2015).

IR spectroscopy is not limited to laboratory analyses. *In situ* measurements are possible using portable spectroradiometers (Šestak *et al.*, 2018). Spectral libraries are currently under development for the large-scale soil mapping by means of remote sensing. Airborne sensors covering these regions emit visible- to near-IR (Vis-NIR) as well as thermal IR (TIR) electromagnetic radiation (Viscarra Rossel *et al.*, 2016). Regional, national and continental-scale spectral libraries are already in existence. The ICRAF-ISRIC (International Council for Research in Agroforestry / International

Soil Reference and Information Centre) world spectral library includes data from several continents (Nocita *et al.*, 2015). Multi-spectral remote-sensed data has been combined with laboratory spectra for mapping SOC at regional scales (Peng *et al.*, 2015). NIR hyperspectral imaging (NIR-HIS), also known as short-wave IR-HIS (SWIR-HIS), is a new technology showing potential for agricultural applications (Nocita *et al.*, 2015). Hyperspectral data collected from satellites includes a spatial dimension of the sample surface (Pasquini, 2018).

2.3.1 ii) Principles of infrared spectroscopy

IR spectroscopy is an indirect analytical technique used to predict the concentration of a compound from multivariate spectral data, based on empirical models developed from previous calibration (Gobrecht *et al.*, 2014). Spectra are generated by the interaction of soil with infrared light which has been dispersed into discrete wavelengths, usually by a prism (Shepherd and Walsh, 2002). The spectral signature of a material is a plot of the response of the sample (reflectance or transmittance) versus the wavelength or frequency of incident electromagnetic radiation (Clark, 1999). Spectral signatures of materials differ because of the unique vibrational motion of atoms in their crystal and molecular structural arrangements. Molecules in the sample matter absorb incident light providing them with energy to vibrate. Modes of vibration include motions such as stretching, bending, wagging and scissoring (Workman, 1996). Characteristic vibrational frequencies are associated with certain submolecular groups of atoms, known as group frequencies. Hence, diagnostic spectral bands may be used to confirm the presence of chemical functional groups (Johnston and Aochi, 1996). The response of soil components is usually measured in the near-IR (NIR) to mid-IR (MIR) regions, ranging from 400 – 2500 nm and 2500 – 25 000 nm respectively (Shepherd and Walsh, 2002). The energy of light absorbed by a molecule is equal to the energy difference between the discrete energy levels of different vibrational states of the molecule. The energy difference (AE) is determined by the bond strengths of chemical functional groups in the molecule (Wetzel, 1983). The energy absorbed (Ep) is related to the wavelength of radiation by the equation:

$$E_p = hv = \frac{hc}{\lambda} = AE \quad \text{Equation 2-6}$$

where:

v is the frequency of incident light,

c is velocity of light,

λ is the wavelength, and

h is Plank's constant.

Based on the harmonic oscillator model, the permitted energy states of a molecule are given by:

$$E_v = hv \left(v + \frac{1}{2} \right) \quad \text{Equation 2-7}$$

where: v is the vibrational quantum number ($v = 0, 1, 2, \dots$).

The fundamental vibration (for transition from $v = 0$ to $v = 1$) occurs at a lower frequency for weaker chemical bonds or heavier atoms. Thus, fundamental vibrations for O-H, C-H and N-H dominate the MIR spectral region (Mohamed *et al.*, 2018), allowing detection of primary and secondary minerals, as well as organic compounds. The NIR spectral region is characterized by overtones and combinations of fundamental vibrations of O-H, S-O and C-O. These bonds are present in clay minerals, sulphates, carbonates, water, and carbon dioxide (Viscarra Rossel *et al.*, 2016). MIR spectroscopy is a more accurate and reproducible analysis technique, as it is less affected by particle size, water content, and light penetration (Reeves *et al.*, 2010 as cited in Nocita *et al.*, 2015). However, MIR instruments are not suited to field conditions and are sensitive to soil moisture content differences between samples (Soriano-Disla *et al.*, 2014). Therefore Vis-NIR instruments are more widespread for *in-situ* applications (Reeves *et al.*, 2010 as cited in Nocita *et al.*, 2015). Pirie *et al.*, (2005) and Viscarra Rossel *et al.*, (2006) found better correlation of soil pH and extractable Ca content with spectra in the MIR region compared to the NIR region. This was attributed to higher incidence of spectral bands and greater intensity and specificity of signals in the MIR region. Viscarra Rossel *et al.*, (2006) found that NIR spectroscopy generated more accurate prediction models of extractable K, compared to MIR. Beer's Law is a fundamental principle used in quantitative analysis with IR spectroscopy and is only applicable to analysis of diluted, transparent materials. The spectrometric response is related to the concentration of the analyte in a sample by the linearized form of the Beer-Lambert equation:

$$A(\lambda) = \log \frac{I_0}{I} = Ec\ell \quad \text{Equation 2-8}$$

where:

I is the intensity of transmitted light

I_0 is the intensity of incident light

$A(\lambda)$ is the absorbance at wavelength λ ,

E is the molar extinction coefficient ($L \text{ mol}^{-1} \text{ cm}^{-1}$),

c is the concentration (mol L^{-1}) and

ℓ is the path length (cm).

$E(\lambda)c$ may be replaced by a single variable called the absorption coefficient, which is an indication of the absorption capacity of the sample (Gobrecht *et al.*, 2014). As well as being absorbed or transmitted, photons may be reflected, refracted, or diffracted. In particulate, solid or turbid samples, photons are scattered as they encounter other particles or changing refractive indices of the material (Ciani *et al.*, 2005 as cited in Gobrecht *et al.*, 2014). The extent of scattering depends on physical

and chemical properties of the medium. For this reason, inter-sample variability results in baseline drifts of the ideal absorption spectrum (Gobrecht *et al.*, 2014).

2.3.2 Infrared spectroscopic analysis methods

2.3.2 i) *Common techniques of infrared spectroscopic analysis*

The oldest and simplest method of IR spectroscopy is based on the principle of transmittance (Du and Zhou, 2011). This technique may be used for analyses of solids, liquids, or gases. Soil samples are typically prepared by grinding the soil to a fine powder with potassium bromide (KBr) and pressing it into a translucent disk or “pellet” (Harris, 2010). FTIR spectrometers have improved sensitivity, precision, and rapidity of spectral data acquisition (Johnston and Aochi, 1996). These advances are due to an increased IR frequency range allowed by greater entrance aperture size, interferometry enabled by use of a monochromatic source, greater signal-to-noise (S/N) ratio and simultaneous detection of the entire spectrum (Gregoriou, 2000). Diffuse reflectance IR Fourier transform (DRIFT) spectroscopy is based on the measurement of radiation reflected from internal surfaces at a specified depth in the material. However, this technique is sensitive to variation in soil particle size (Du and Zhou, 2011). The Beer-Lambert Law is applied to DRIFT measurement by replacing the intensity of transmitted radiation $I_T(\lambda)$ with that remitted $I_R(\lambda)$ (Dahm and Dahm, 2001 as cited in Gobrecht *et al.*, 2014). The reflectance (R) is related to the apparent absorbance (A) by the relationship:

$$A = \log \frac{1}{R} \quad \text{Equation 2-9}$$

The reflectance of the sample is related to the concentration of the analyte by the Kubelka-Munk function, which is based on the principle that light penetrating a homogeneous material will either be absorbed or scattered. Measurements at each wavelength for a given sample are multiplied by a constant to compensate for the difference in scatter between equivalent samples; this is known as multiplicative effect assumption. The Kubelka-Munk equation is given by:

$$\frac{K}{S} = \frac{(1-R)^2}{2R} \quad \text{Equation 2-10}$$

where:

K = true absorbance

S = scatter

R = reflected light

The term R may be replaced by I/I_0 , where I and I_0 are the intensities of reflected and incident light, respectively. The Beer-Lambert and Kubelka-Munk models assume that a constant fraction of reflected light is detected, but in reality this may be sample dependent, resulting in an additive scatter effect (Naes *et al.*, 2002). The wavelength of IR light absorbed may be also affected by environmental factors such as light intensity and temperature (Stenberg *et al.*, 2010). Differences in IR instrument manufacture (relative position of illumination and detection systems) and replacement of the radiation source of an instrument may cause baseline drift (Pasquini, 2018). For these reasons, pre-processing transformation of data is usually required to improve correlation of spectra with measured properties and to eliminate irrelevant parts of the spectrum (Lin *et al.*, 2017).

2.3.2 ii) *Pre-processing of spectral data*

Pre-processing is defined as the mathematical manipulation of spectral data prior to analysis (Ma *et al.*, 2019). This should only be done if pre-processing transformation of data results in a significant decrease in prediction error and number of latent variables in calibration (Olivieri, 2015). Multiplicative scattering correction (MSC) and standard normal variate (SNV) are transformations that minimize scattering causing additive / multiplicative effects of suspended particles in solid samples (Pasquini, 2018). MSC executes linear transformation, fitting each spectrum to its mean (Du and Zhou, 2011). Normalization is achieved by dividing each absorbance by a constant to remove systematic variation (Du and Zhou, 2011). Variations of this technique such as minimum-maximum normalization ($N_{\min-\max}$) have been developed (Mayoral *et al.*, 2013). Baseline correction involves joining the points of lowest absorbance on a peak in reproducibly flat parts of the spectrum baseline such that the signal intensities are maximized (Du and Zhou, 2011). Straight line subtraction (SLS) accounts for baseline drift by setting the y-minimum to zero (Mayoral *et al.*, 2013). Resolution is improved by enhancing sharper bands and manipulating gradients. Derivatization decreases the S/N ratio to an increasingly greater extent with higher order derivatives, hence usually only the first and second derivatives are used. Different pre-processing treatments should be tested on the data, individually and in combination, for best results. A smoothing filter may be used to achieve multiple pre-processing objectives simultaneously. The Savitzky–Golay smoothing filter performs baseline correction as well as S/N ratio enhancement (Du and Zhou, 2011; Mayoral *et al.*, 2013). Variable selection prior to modelling may be considered a form of data pre-processing. Outliers may be detected by analysing influence plots and sample leverages and should only be eliminated if reasons for their nature can be verified. This is also useful to determine the effect of sample properties on IR spectra and in aiding model interpretation (Pasquini, 2018).

2.3.3 Models for prediction of soil properties based on infrared spectra

2.3.3 i) *Construction of models for prediction of soil properties from IR spectra*

Since IR spectra of soils have overlapping peaks due to the detection of many components, the Beer-Lambert law cannot be used directly in quantitative analysis. Statistical models describing the relationship between individual soil characteristics and spectral signatures can only be developed by isolation of absorptive bands (Du and Zhou, 2011). This is accomplished by using chemometric techniques of multivariate calibration such as principal component regression (PCR) and partial least squares regression (PLSR) (Mohamed *et al.*, 2018), which achieves the compression of full spectra into a few latent variables by projecting variables to a lower dimension (Vinzi *et al.*, 2010 as cited in Mohamed *et al.*, 2018). PLSR incorporates the principles of PCA, canonical correlation analysis (CCA) and linear regression. Whereas with PCR, spectra are first decomposed into eigenvectors after which regression with the parameter of interest follows, PLSR performs these procedures in a single step. Unlike PCR, the objectives of explaining variation in predictors as well as in predictions are balanced equally. Hence, PLSR results in spectral vectors that are directly related to the soil attribute and is robust in terms of handling noise, missing values and multicollinearity (Viscarra Rossel *et al.*, 2006). PLSR was reported to be the most widely-used modelling method in chemometrics and sensometrics for IR analysis (Rönkkö *et al.*, 2015 as cited in Mohamed *et al.*, 2018). PLSR is computationally faster than some other methods (Stenberg *et al.*, 2010), yields lower prediction error compared to MLR, and enables identification of outliers using diagnostic information in the residuals (Olivieri, 2015). Relationships between soil characteristics are often non-linear, therefore non-parametric data mining calibration techniques such as ANN, regression trees and other supervised learning models, are becoming more widely used (Xu *et al.*, 2016).

2.3.3 ii) *Evaluation of models for prediction of soil properties from IR spectra*

The performance of prediction models and uncertainty of measurements are ideally expressed using parameters that are interpretable and may be compared across studies (Bellon-Maurel and McBratney, 2011). Criteria used to evaluate the accuracy, precision, and predictive capability of models in NIR / MIR studies, include r^2 (coefficient of determination), root mean square error of prediction and cross-validation (RMSEP and RMSECV, respectively), standard error of prediction (SEP), percentage coefficient of variation (CV (%)), ratio of performance to deviation (RPD), ratio of performance to inter-quartile distance (RPIQ), bias and rank. The r^2 value (see Section 2.1.3.) depends on the measurement range and is not the best indicator of model goodness of fit (Bellon-Maurel and McBratney, 2011). The RMSE-value (see Section 2.1.3) is the most commonly-used parameter for evaluating the accuracy of IR analytical methods (Pasquini, 2018). The cross-validation approach involves calibration using a certain portion of the data, after which the model is tested on data which was left out of calibration (Du and Zhou, 2011). This process is repeated with

different portions of the data used for each calibration, until all the samples have been left out at least once (Aleixandre-Tudo *et al.*, 2018). The RMSECV may be computed by the equation:

$$\text{RMSECV} = \sqrt{\sum_{i=1}^N \frac{(\hat{y}_{\text{CV},i} - y_i)^2}{N}} \quad \text{Equation 2-11}$$

where:

$\hat{y}_{\text{CV},i}$ = i^{th} y-value predicted based on calibration equation where sample i was deleted

y_i = i^{th} y-value provided by independent sample set

N = number of samples in the test set (Naes *et al.*, 2002).

In the test set validation procedure, we distinguish between the RMSE of the calibration and validation models by referring to the former as RMSEE (root mean square error of estimation) and the latter as RMSEP. The equation for calculation of the parameter is like that of RMSECV defined above, with $\hat{y}_{\text{CV},i}$ being replaced by \hat{y}_P (predicted y-value) and y_i is obtained from the reference sample set for RMSEE, or out-of-sample data for RMSEP. As opposed to cross-validation, calibration occurs only once for test set validation. SEP is the most commonly used performance indicator in NIR and MIR studies (Bellon-Maurel and McBratney, 2011) which measures uncertainty of predicted values by defining the limits of the confidence interval in which predictions are found (Bellon-Maurel and McBratney, 2011). Since SEP is larger for measured quantities spanning a greater range, three indices have been proposed to standardize the SEP. These are CV (%), RPD and RPIQ. CV (%) indicates the degree of distribution of data around the mean and is given by:

$$\text{CV (\%)} = \frac{SD}{\text{mean}} \quad \text{Equation 2-12}$$

where: SD = standard deviation

RPD is the most popular choice of the given three parameters, however, the relevance of this parameter has been criticized by several authors (Bellon-Maurel *et al.*, 2010; Minasny and McBratney, 2013). The value has been shown to be inversely related to the correlation coefficient and hence it is redundant to report RPD as well as the r^2 value. Nonetheless, this parameter is still widely used in agricultural literature, with a higher RPD indicating a better model fit (Bellon-Maurel and McBratney, 2011). RPD is calculated as:

$$\text{RPD} = \frac{SD}{\text{RMSE}} \quad \text{Equation 2-13}$$

where:

SD = standard deviation of the calibration / validation sample set (for RPD_{cal} / RPD_{val} respectively)

RMSE = RMSECV / RMSEP (for RPD_{cal} / RPD_{val} respectively) (Aleixandre-Tudo *et al.*, 2018)

As the range and SD of reference samples differ between studies, no critical values exist for RPD values of models predicting soil chemical properties from Vis-NIR spectra (Bilgili *et al.*, 2010). Chang *et al.* (2001) suggested that RPD values for these models could be grouped according to the ranges > 2.0 , $1.4\text{--}2.0$, and < 1.4 for indication of excellent, acceptable and poor predictions, respectively. Dunn *et al.* (2002) suggested the slightly different ranges $RPD > 2.0$, $1.6\text{--}2.0$ and < 1.6 . When models are constructed using PLSR from data with log-normal distribution (typical of soil chemical properties), RPIQ may be substituted for RPD to account for skewed distributions. RPIQ is calculated similarly to RPD, by replacing SD with inter-quartile distance (Bellon-Maurel *et al.*, 2010). The RPIQ takes prediction error as well as variation of measured values into account (Johnson *et al.*, 2019). While random error may be associated with imprecision of measurement, bias refers to inaccuracy. Bias is the difference between the average true value and the average predicted value. A positive bias indicates over-estimation of the prediction values, while a negative value represents an under-estimation. Bias is larger when validation is performed on independent samples whose properties differ from the calibration samples. The bias-corrected standard error of prediction (SEP_c) is a measure of the difference between repeated measurements, hence indicative of precision. The square of the SEP value, which is also the RMSEP, is equal to the quadratic sum of the bias and the bias-corrected SEP (SEP_c):

$$SEP^2 = Bias^2 + SEP_c^2 \quad \text{Equation 2-14}$$

Hence, if the bias is low, the RMSEP and SEP_c will be similar (Bellon-Maurel and McBratney, 2011). Increasing the number of samples used for calibration of a prediction model increases the RPD-value, however it may also decrease the r^2 -value (Kamrunnahar *et al.*, 2003). This provides further evidence of the inadequacy of the r^2 -value alone as a model performance indicator. The rank is the number of eigenvectors or predictors used by the model. The combination of regressors must be chosen such that the most parsimonious model is constructed. The disadvantages of overfitting were discussed in Section 2.1.3. However, omitting variables that are not intercorrelated and that could contribute useful information to the model may result in greater bias. Although methods which convert regressors to an uncorrelated set (e.g. PCA) may reduce the rank, it is difficult to interpret the meaning of converted regressors. Analysis of variance (ANOVA) may be used to quantify the significance of inclusion of additional variables as regressors (Caffo, 2015).

2.3.3 iii) Spectral activity related to soil pH and exchangeable base cations

Soil IR spectra usually show prominent absorption bands around 1400, 1900 and 2200 nm (Kamrunnahar *et al.*, 2003; Shepherd & Walsh, 2002). OH groups present in free water are observed at 1400 and 1900 nm, whereas OH groups present in mineral lattices feature at 1400 and 2200 nm (Hunt, 1980, as cited in Kamrunnahar *et al.*, 2003). Wavenumbers of absorption peaks typically correlated to mineral and organic soil components, as well as specific chemical functional groups, are shown in

Table 2-1. Due to co-variation of soil pH with spectrally active soil constituents, soil pH has been accurately predicted by spectral models (Chang et al., 2001; Dunn et al., 2002; Kamrunnahar et al., 2003; Pirie et al., 2005; Shepherd & Walsh, 2002), despite the fact that proton activity itself does not cause a direct spectral response. Soil pH is regulated by the soil's buffering capacity, which largely depends on CEC. Thus, the pH may be a function of clay type and content, as well as SOM content. Soil mineralogy and weathering, especially carbonates, influence soil acidity (Stenberg *et al.*, 2010). Soil acidity is also generated from oxidizable ammonium, organic N, and organic acids (Soriano-Disla *et al.*, 2014). The mechanisms responsible for generation of soil acidity were discussed in detail in Section 2.1.1. Analysis of PLS factor loading weights for prediction of $\text{pH}_{\text{CaCl}_2}$ using MIR spectral data revealed positive contributions due to smectite and organic acids, as well as negative contributions due to quartz and kaolin (Viscarra Rossel et al., 2006). Ma *et al.* (2019) noted that IR spectral wavenumbers contributing to pH prediction models of rice paddy soils were associated with the functional groups OH (phenols, carboxyls and hydroxyls of clay minerals and water), COO^- , and CO_3^{2-} (carboxylates, carbonates, and carboxylic acids).

As with soil pH, exchangeable base cation quantities also do not have direct absorption features in the IR spectrum but are correlated to spectrally active soil components such as clay content and SOM. Ca, Mg and K are major cations in minerals and in interstitial spaces of 2: 1 layer silicate clays, which give strong signals in the vis-NIR-MIR region. Exchangeable Ca is present in heavy clays such as smectites, on the edges of other clay minerals, or in calcite and dolomite minerals. Spectral models calibrated using the MIR region yield slightly better predictions than models calibrated using the Vis-NIR regions according to Soriano-Disla *et al.* (2014). The major sources of Mg in soils are heavy clays, dolomite minerals, and hydrated cations in octahedral structures of micas. Unlike Ca predictions, Mg predictions are slightly improved using the Vis-NIR region (Soriano-Disla *et al.*, 2014). Like Mg, spectral models for exchangeable K predictions performed better when calibrated from Vis-NIR spectra compared to NIR or MIR spectra. This likely due to the absorbance of NIR frequencies by K fixed in the interstitial clay layers of illite minerals and mixed-layer clays. Absorptions of illites versus smectites in the MIR region are also difficult to distinguish. K prediction models are highly site-specific due to nutrient variability in the field, as K is readily leached and applied often as fertilizer. Another factor to consider during sample selection is that different K pools may be extracted by different reference chemical methods (Soriano-Disla *et al.*, 2014). Few authors have compared models calibrated from reference data generated by different chemical extraction methods. Chang et al. (2001) found that M-III extractable cations were predicted more accurately than those extracted by NH_4OAc .

Table 2-1 Wavenumbers and wavelengths of VIS-NIR-MIR absorptions of major soil components, adapted from Soriano-Disla *et al.* (2014)

IR spectral range	Soil component	Wavenumbers (cm ⁻¹)
MIR	Quartz (sand)	1100–1000
	Kaolinite	3690–3620
	Smectite	3620–3630
	Illite	3400–3300
	Carbonates	1430 and 2520
	Iron oxides	600–700
	Iron oxyhydroxides	3100, 900, 800
	Alkyl	2930–2850
	Protein amide	1670 and 1530
	Carboxylic acid	1720
	Water associated	1630
	Carboxylate anion	1600 and 1400
	Aromatic groups	1600–1570
Vis-NIR	Water	7143 and 5263
	Kaolinite	7143 and 4545
	Illite	4545, 4274, and 4090
	Smectite	4545
	Carbonates	4283
	Iron oxides	25 000, 22 222, 20 000, 15 385 and 11 111
	Organic matter	9091, 6250, 5882, 5556, 5000, and 4167–4545

2.3.3 d) *Subsetting of calibration samples*

Heterogeneity of sample sets may complicate prediction of soil properties from IR spectra. Mechanisms of soil acidity generation and cation exchange vary from one soil type to another, therefore covariations upon which indirect calibrations are built may differ according to geographic location (Reeves *et al.*, 1999). In a study by Xu *et al.* (2016), models calibrated by subsets with similar parent materials showed improved SOM predictions compared to global models. Stenberg *et al.* (2010) and Araújo *et al.* (2014) found improvements in accuracy for predictions of SOM and clay content when the spectral library was divided into subsets that were more uniform in texture class and mineralogy, regardless of geographical origin. Stenberg *et al.* (2010) noted that SOC variability also accounted for variation in model prediction accuracy. Araújo *et al.* (2014) split calibration spectra into groups by k-means cluster analysis of their spectral features. Another solution is the nested analysis of covariance of soils, which is the partitioning of the variance of soil properties into scale-specific components. Savvides *et al.* (2010) used this technique to demonstrate that spectral

variations at smaller spatial scales are caused by variations of certain soil properties, while other soil properties cause variations at larger scales. The stronger correlation between CEC and spectral components at field scale as opposed to larger scales reveals the usefulness of IR as a practical analysis technique (Savvides *et al.*, 2010).

2.4 Gaps in knowledge

Different methods of soil pH and base cation content measurement vary in efficiency of extraction of acidic and base cations from the soil exchange complex due to chemical and procedural variations. There is a need for equations for soil pH and cation content conversions calibrated with soils from a Mediterranean climatic zone, which would greatly benefit researchers and soil consultants for continued monitoring of soil quality. Several authors have developed equations for selected soil pH conversions (Aitken & Moody, 1991; Galka *et al.*, 2016; Henderson & Bui, 2002; Kissel & Miller, 2010; Kome *et al.*, 2018; Little, 1992). However, no study has performed comparisons of pH measurement methods using all combinations of the most used matrices (water, 0.01M CaCl₂ and 1M KCl) and soil: solution ratios (1:1, 1:2.5, 1:4 or 1:5). Some literature indicates that the inclusion of soil properties as additional regressors in conversion models may improve soil pH predictions (Brough *et al.*, 2011; Galka *et al.*, 2016). However, the mechanisms underlying the effects of these properties on pH predictions may only be applicable to soil types from which the models were calibrated. The effects of soil properties on soil pH conversions in Western Cape soils have not yet been investigated. Exchangeable base cation quantities in soils may be determined by extraction of samples with 1M AA (pH = 7, Schollenberger & Dreibelbis, 1930), or by the compulsive exchange procedure (Gillman & Sumpter, 1986) which simulates soil pH and ionic strength at field conditions. The convenience of multi-elemental extraction has led to the development of procedures such as the M-III (Mehlich, 1984), Citric acid (1%, Dyer, 1984) and Bray-II (Bray & Kurtz, 1945) methods. Soil testing laboratories in the Western Cape province of South Africa have adopted these, as well as the AA (0.2M, pH = 7) methods. Meanwhile, the Ambic-I method (Van der Merwe *et al.*, 1984) has become a standard multi-elemental extraction technique in the province of KwaZulu-Natal. The comparative effectivities of index cations for base cation extraction need to be compared, as there are discrepancies in scientific literature (Gillman *et al.*, 1983; Horn, 1982; Jönsson *et al.*, 2002). Studies have also found evidence of soil physicochemical properties influencing cation concentration conversions between methods (Gartley *et al.*, 2002; Michaelson *et al.*, 1987; Simard *et al.*, 1989; Wang *et al.*, 2004; Zhu *et al.*, 2016). Whether multivariate models incorporating soil properties as additional regressors are more accurate than simple models for base cation content conversions of soils in a Mediterranean climatic zone needs to be established. Furthermore, the practical relevance of such model improvements for soil quality assessment purposes should be evaluated. IR spectroscopy is a rapid soil analysis method with potential for field applications. As the technique relies on chemometric algorithms to calibrate prediction models from existing data, more research is

needed to gauge how accurately soil properties may be determined by this method in the Western Cape region. The question of viability of models calibrated from NIR- or MIR-generated spectral data for measurement of soil base cation content needs to be addressed. Few studies have calibrated models from different sets of reference data to evaluate how suitable different chemical extraction methods are for calibration of IR prediction models. There is evidence to suggest that co-variations between soil properties and IR spectra depend on the geographic origin and homogeneity of soils (Araújo *et al.*, 2014; Reeves *et al.*, 1999; Stenberg *et al.*, 2010; Xu *et al.*, 2016). There are no studies which have investigated whether subsetting Western Cape calibration soil sample sets based on similarities in soil properties, affects the accuracy of IR prediction models.

3. Relationships between soil pH measurement methods

3.1 Introduction

Soil acidity is a master variable controlling plant nutrient availability, soil biological activity and mobility of toxic compounds (Brady & Weil, 2017). Accurate determination of soil pH is therefore essential for soil quality assessment, however diagnostic norms have been developed based on pH measurements performed according to different methods. This makes interpretation of soil chemical analyses difficult for scientists and consultants when test results are reported in formats that are incompatible with locally developed norms. Soil pH is measured by determination of the potential difference between two electrodes suspended in soil solution, relative to a standard buffer. Salt solutions are commonly used to flocculate soil particles, thereby stabilizing pH readings compared to measurements in soil-water suspensions (Al-Busaidi *et al.*, 2005). pH measurements in 1M KCl is a standard method in South Africa and some European countries, which also allows determination of exchangeable Al displaced by the highly-concentrated salt (Kome *et al.*, 2018). pH measurements are performed in 0.01M CaCl₂ in Canada, the US and Australia to mimic the ionic strength of typical fertile soils (Schofield & Taylor, 1955). Soil pH measurements have been shown to decrease in solutions of higher ionic strength, due to displacement of acidic cations from the soil exchange complex (Brady & Weil, 2017). Many different soil: solution ratios are used globally, the most common being 1:1, 1:2.5, 1:4 and 1:5 (Galka *et al.*, 2016; Kissel & Miller, 2010; van Lierop, 1981; Little, 1992).

Soil pH measurements have been shown to increase with an increase in the soil: solution ratio (Thomas, 1982). Linear (Kissel & Miller, 2010; Kome *et al.*, 2018) and non-linear (Aitken & Moody, 1991; Galka *et al.*, 2016; Henderson & Bui, 2002; Little, 1992) equations have been developed for conversion of soil pH values obtained by different methods. However, all combinations of the most commonly-used matrices and soil: solution ratios have not been compared for a single soil sample set. No comparative studies concerning soil pH measurement methods have been performed for Western Cape soils. Soil pH conversions may be affected by different chemical reactions of various clay minerals and organic functional groups with the solution matrix (McBride, 1994). Inclusion of soil properties which are correlated to soil pH as predictors in pH conversion equations has been shown to improve model prediction accuracy (Brough *et al.*, 2011; Galka *et al.*, 2016). Few studies evaluate whether incorporation of additional regressors improve conversion model accuracy by a significant degree. It has been suggested that the advantages of using more simple equations outweigh the benefits of improved performance of more complex models (Ferguson *et al.*, 2014) and that inclusion of additional regressors may only improve performance of models

tested on soils like the calibration set (Kome *et al.*, 2018). However, it is unknown to what extent the accuracy might be improved for conversions of Western Cape soils and the practical significance thereof. The aim of this study is to develop equations for conversion of soil pH values obtained by the most common pH measurement methods for a set of Western Cape soils. In addition, the improvement of models with incorporation of soil properties as additional regressors will be evaluated.

3.2 Materials and methods

3.2.1 Soil sampling and characterization

Soil samples (N = 57) were obtained from the soil collection of Elsenburg soil analytical laboratory. The sampling strategy was based on a method recommended by statistical expert Prof Martin Kidd (Stellenbosch University) for selection of diverse soils. Soils showing different combinations of high, medium, and low ranges of soil physiochemical properties were chosen to maximize variation in the dataset. The potential CEC (pH = 7) of the soils was estimated by summation of basic cations determined by Citric acid 1% (see Section 4.2.2) and Eksteen titratable acidity (Eksteen, 1969), determined by Elsenburg Analytical Services. The total C and N content, pH, EC, and clay content of each sample was determined by the MSc candidate at the Department of Soil Science, Stellenbosch University. Prior to analysis, soils were dried and passed through a 2 mm sieve to remove coarse material. The clay, sand and silt fractions were determined according to the pipette method (The Non-Affiliated Soil Analysis Working Committee, 1990) with slight changes to the procedure regarding soil preparation.

A limited amount of each soil sample was available for analysis; therefore, avoidance of sample pretreatment (requiring use of a larger initial sample mass) was desirable. A trial was performed to investigate the necessity of SOM removal prior to determination of clay, sand, and silt fractions. The clay, sand and silt fractions of five soil samples (approximately 40 g each) with varied organic matter content were determined according to the pipette method (The Non-Affiliated Soil Analysis Working Committee, 1990). The method was then repeated for the same five soil samples (approximately 20 g each), but without the initial removal of SOM. The clay fractions of the samples (with versus without SOM removal) were compared using a paired 2-sample t-test with RStudio 1.2.5033 (RStudio Team, 2020) software. There was no significant difference between the mean clay content of soil samples determined by the original method, versus that without initial organic material removal ($t_4 = 2.71$, NS). Hence, it was concluded that SOM removal of samples prior to the determination of sand, silt, and clay fractions could be avoided without any significant effect on clay content values obtained.

Total C and N content was determined by dry combustion using an elemental analyzer (Elementar Vario Macro Cube CN). The EC (soil: solution ratio of 1:1) was determined with a Jenway 4510 Conductivity Meter after shaking 10 g of each sample in 10 mL distilled water for 20 min and allowing an hour for equilibration before measurement. These measurements were converted to EC-values that would have been obtained using a saturated paste (EC_e) (Sonmez et al., 2008) in order to classify soils on a basis of salinity according to standard ranges (Brady and Weil, 2017).

3.2.2 Soil pH measurement

Soil pH was measured using a Metrohm 827 pH meter after shaking soil samples for 20 minutes in the desired volume of solution, as shown in Table 3-1. These variations cover the most common standard procedures used in soil science laboratories. This study is the first to compare all methods for the same sample dataset. Soil samples of 10 g were used such that the electrode junction was effectively covered by the supernatant, especially for the more narrow soil: solution ratios. Measurements were performed by suspending the probe in the supernatant after soil solutions had equilibrated for a minimum of one hour. Triplicate measurements were performed on 5 random samples for each pH measurement method. Measurements were performed for both the 1:1 and 1:2.5 soil: solution ratios using the same soil sample by adding solution to the initial sample and repeating the shaking procedure to prepare the wider soil: solution ratio sample. The same method was applied to the measurement of the 1:4 and 1:5 ratios.

Table 3-1 Soil pH measurement methods

Matrix	Soil: solution ratios
H ₂ O (distilled)	1:1
	1:2.5
	1:4
	1:5
CaCl ₂ (0.01M)	1:1
	1:2.5
	1:4
	1:5
KCl (1M)	1:1
	1:2.5
	1:4
	1:5

3.2.3 Statistical analyses

Statistical analyses were performed using STATISTICA 13.5.0.17 and RStudio 1.2.5033 (RStudio Team, 2020) software. The minimum, maximum, mean, and SD of general soil properties ($\text{pH}_{\text{KCl}1:2.5}$, EC_e , clay content, %C and %N) were computed. The Pearson coefficients for correlation between general soil properties and soil pH measured by various methods were found. The CV (% , percentage coefficient of variation) of triplicate measurement data was found to determine the repeatability of pH measurement for each method. Kruskal Wallis tests were performed to determine whether pH measurements differed between extract matrices (for constant soil: solution ratio) and whether pH measurements differed between soil: solution ratios for each extract matrix. Post-hoc pairwise Dunn tests with Bonferroni adjustments were done to identify in which matrices pH measurements differed significantly from others.

Simple linear regression (SLR) models were constructed for conversion between the most commonly-used soil pH measurement methods. The r^2 -value, coefficient, and intercept, as well as the significance of inclusion of the predictor (based on two-tailed student's t-statistic) was computed for each model. Confidence and prediction intervals at 95% confidence level were plotted and RMSE was computed for each model. The effect of EC on pH prediction was investigated by construction of multivariate linear regression (MLR) models, including EC_e as an explanatory variable. The AIC-value of each SLR model, as well as that of its corresponding multivariate regression model was computed. The effect of clay content on pH prediction was investigated by construction of MLR models including %clay content as an additional explanatory variable. The AIC- and RMSE-values of each SLR model and its corresponding multivariate regression model were computed.

3.3 Results and discussion

3.3.1 General soil properties

The samples originated from agricultural topsoils in a Mediterranean climatic zone with sandstone, granite, and shale parent materials. Although mineralogical data was not available, the soils were expected to be dominated by low-activity minerals (kaolinite, illite and some Fe oxides). A few highly weathered relic tropical soils would likely have contained some Al (hydr)oxides. The infrared spectra obtained (see Section 2.3) could not be used to perform quantitative mineralogical analyses, as chemometric methods require XRF or other reference data for calibration. The soils embodied a variety of physicochemical properties, as intended by the sampling method. A summary of the distribution of soil properties is shown in Table 3-2 and boxplots illustrating these distributions are shown in Figure 3-1. Loams represented the

dominant texture class. Half of the soils had a low clay content (less than 10%) and the other half ranged between approximately 10-35% clay content, with two outlier soils ranging from approximately 40-50%. The soil CEC values mostly represented those of lower-activity clays as expected to be found in a Mediterranean climate. More than half of the soils had CEC-values less than 10 $\text{cmol}_c \text{ kg}^{-1}$ and the other half fell below 30 $\text{cmol}_c \text{ kg}^{-1}$, with five outlier samples ranging between 30-50 $\text{cmol}_c \text{ kg}^{-1}$. Soils ranged from acid to alkaline (4.0-7.6) were relatively evenly distributed about a median of 6.1, which is a typical pH value of fertile soil. Most of the soils were of moderate salinity levels commonly observed in agricultural soils, however three samples were saline (EC values larger than 4 dS m^{-1}). The samples showed low to moderate C and N contents and two outlier samples had notably greater C and N content. The C content of most soils was within the range reported for dryland soils, where low organic matter accumulation is expected to occur (Lal, 2004).

Table 3-2 Summary of soil physicochemical properties of samples used for comparison of methods of pH determination data (N = 57)

	C	N	Sand	Silt	Clay	pH_{KCl 1:2.5}	ECe (dS m⁻¹)	CEC ($\text{cmol}_c \text{ kg}^{-1}$)
	%							
Min	0.3	0.0	5.4	2.0	0.0	4.0	0.23	2.0
Max	5.9	0.6	94	84	53	7.6	6.5	45
Median	1.0	0.1	52	26	10	6.1	1.1	6.2
Std dev	0.9	0.1	21	18	11	0.9	1.2	12
CV (%)ⁱ	76	82	39	58	81	15	93	100

ⁱpercentage coefficient of variation

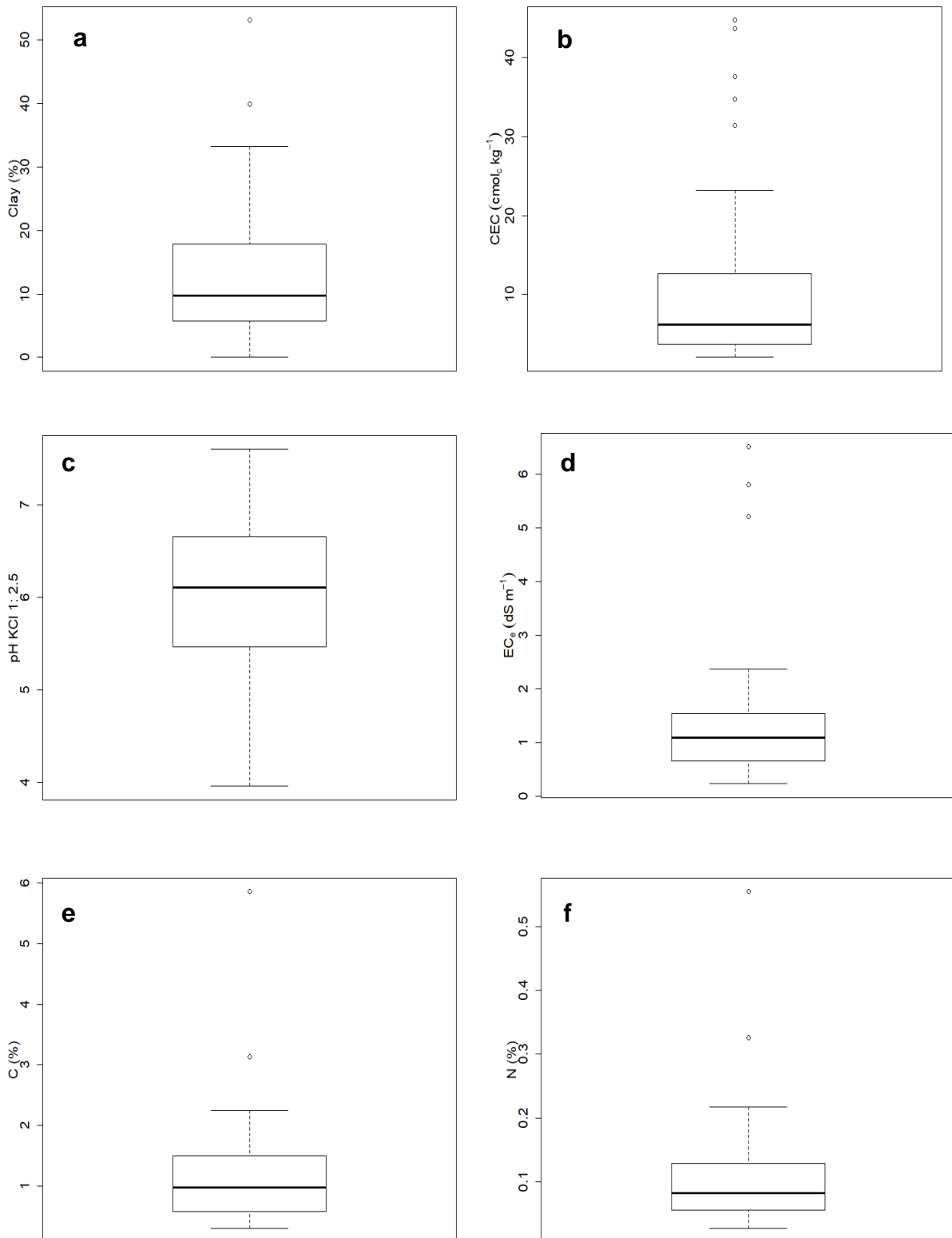


Figure 3-1 Distribution soil physicochemical properties of samples used for comparison of methods of pH determination data (N = 57)

3.3.2 Comparison of soil pH measurement methods

3.3.2 i) *Reproducibility of pH measurement by different methods*

The maximum CV (%) of the replicates of five samples, were considered for evaluation of the reproducibility of each pH measurement method. The maximum CV (%) of pH measurements of sample replicates are shown in Table A- 1. CV (%) < 10 % was taken as an indication of good reproducibility, while CV (%) > 15% was deemed unacceptable. The maximum CV (%) for pH measurements of sample replicates did not exceed 6 % for any of the pH measurement methods This indicates excellent reproducibility of all pH measurement methods; hence all methods are shown to be highly reliable.

3.3.2 ii) *Comparison of soil pH determination methods*

Soil pH measurements using wider soil: solution ratios were practically more convenient to perform as this allowed submersion of the probe in clear supernatant, rather than the slurry created in narrow soil: solution ratios. The 1st, 2nd, and 3rd quartiles of pH measurements using different soil: solution ratios within each matrix are shown in Figure 3-2. Overall, the soil pH is observed to increase slightly with an increase in soil: solution ratio, corroborating the dilution effect described by (Thomas, 1996). However, pH measurements did not differ significantly between various soil: solution ratios in each matrix ($H = 0.39615, 0.7583$ and 4.5518 for pH measured in H_2O , $CaCl_2$ and KCl , respectively; $DF = 2$ and $p > 0.05$ for all).

Boxplots showing the distribution of pH measurements in the complete dataset in each of the different matrices for a soil: solution ratio of 1:1, as well as the same measurements grouped by soil texture class, are shown in Figure 3-3. Boxplots showing the distribution of pH measurements in the complete dataset in each of the different matrices for soil: solution ratios of 1:2.5, 1:4 and 1:5 are shown in Figure 3-4. Textural groupings were not illustrated in the latter figure, as the trend was highly similar as for that observed for measurements at a 1:1 soil: solution ratio. For a given soil: solution ratio, the pH level decreased in the order $pH_{H_2O} > pH_{CaCl_2} > pH_{KCl}$ ($H = 20.202, 24.734, 27.031$ and 20.725 for a soil: solution ratio of 1:1, 1:2.5, 1:4 and 1:5, respectively; $DF = 2$ and $p < 0.05$ for all). This is similar to the findings of Kome *et al.* (2018). Median pH_{CaCl_2} and pH_{KCl} values were significantly lower than the median pH_{H_2O} value as expected due to displacement of acidity from CEC sites by electrolytic cations (Thomas, 1996). The larger difference between pH_{CaCl_2} and pH_{H_2O} compared to the study by Kome *et al.* (2018) may be due to lower buffering capacity of the weathered soils used in this study compared to the volcanic soils of the study by Kome *et al.* (2018). Greater variation was observed for pH measurements in KCl , especially for the more acidic soils (lower two quartiles). This may be explained by displacement of Al ions from the soil exchange complex

by highly concentrated 1 M KCl, which hydrolyzes to generate acidity to a larger extent than would have resulted from the displacement of equimolar quantities of protons. While clays and loams showed a clear decrease in pH measured in KCl compared to CaCl₂, sands did not respond to the same degree as seen in Figure 3-3, possibly due to lower quantities of exchangeable Al having been present on the exchange complex.

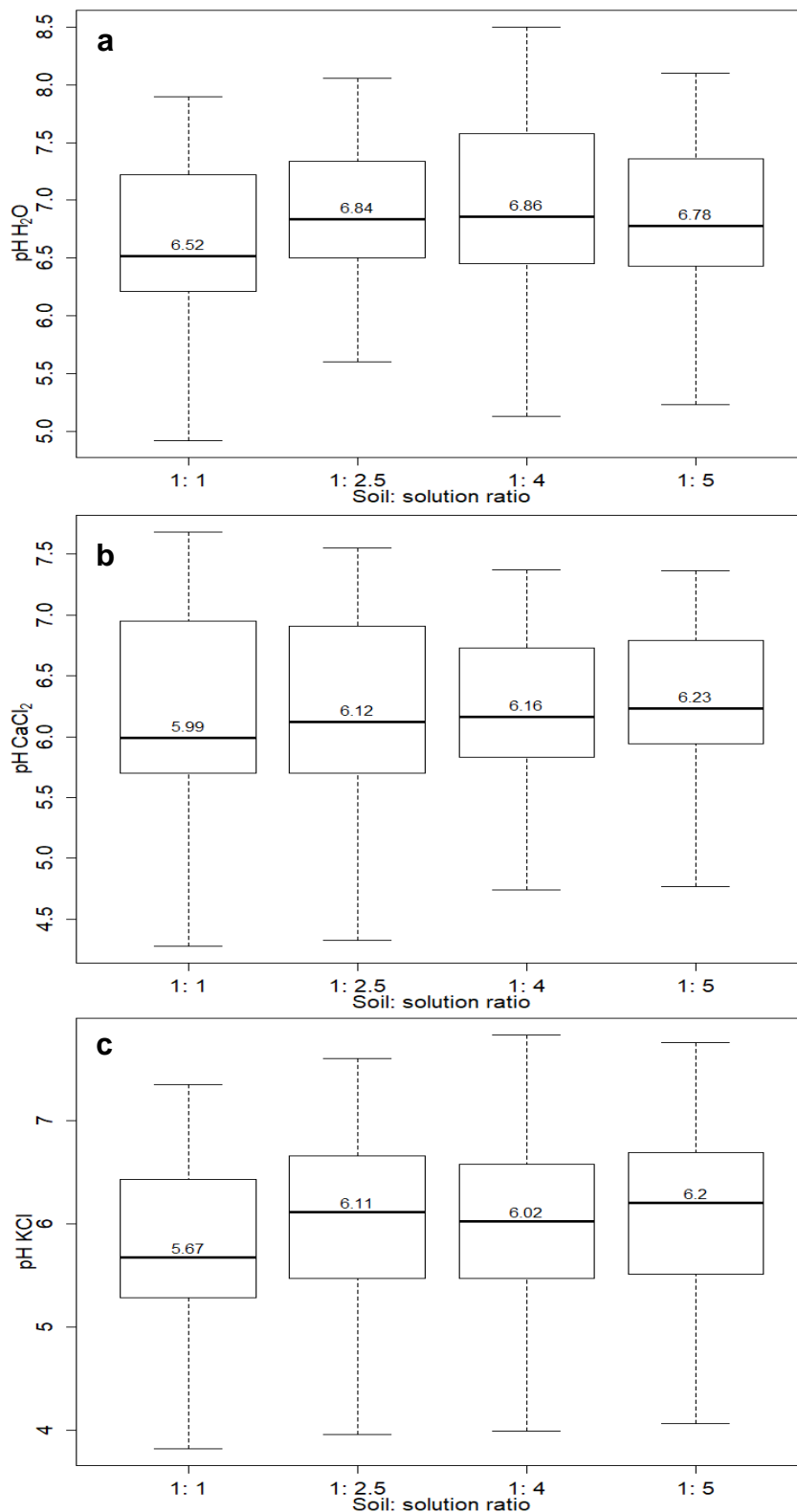


Figure 3-2 Distribution of pH measured in different soil: solution ratios in H₂O (a), 0.01M CaCl₂ (b) and 1M KCl (c). Boxes show first and third quartiles, with median line in bold. Whiskers show minimum and maximum values.

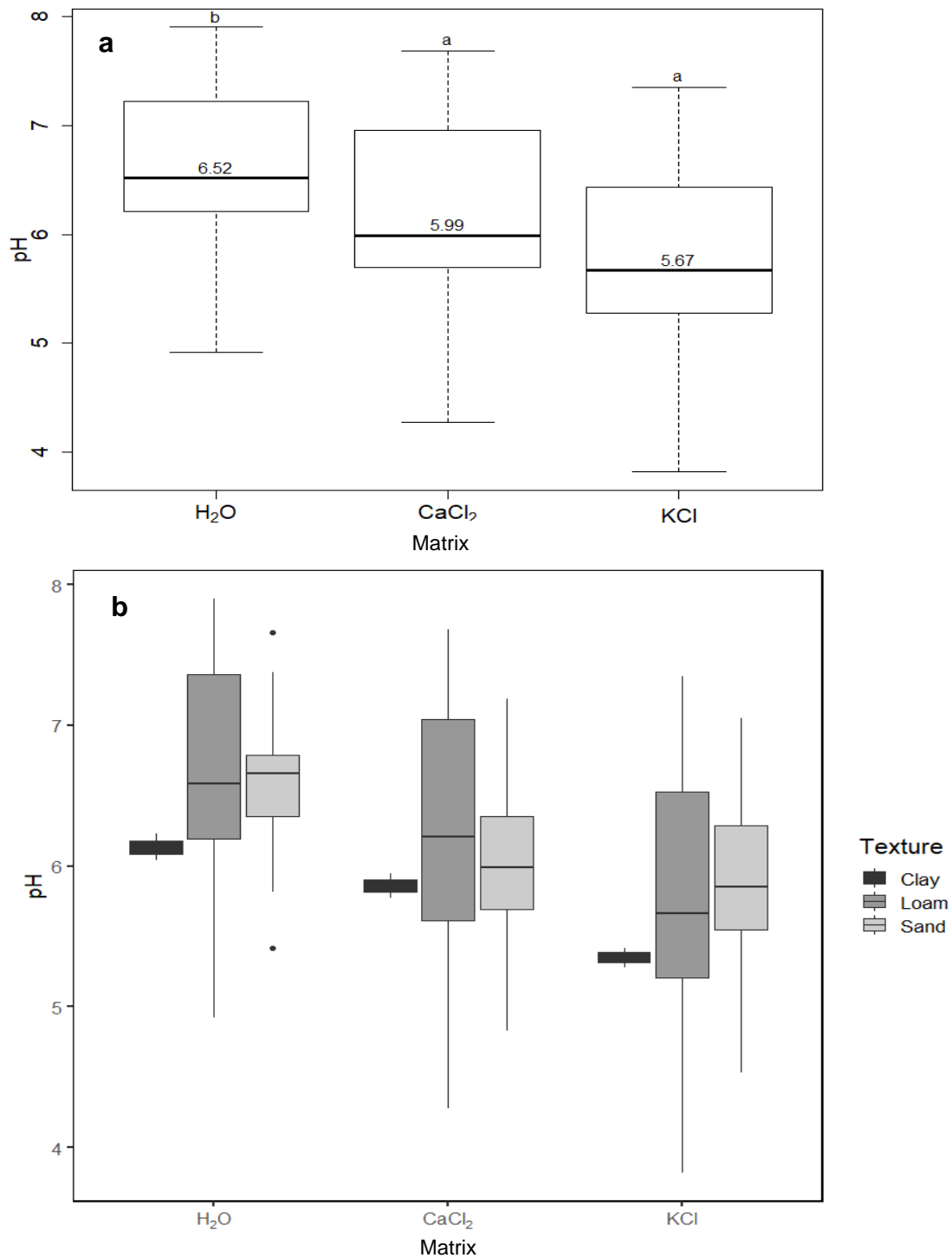


Figure 3-3 Boxplots showing distribution of soil pH levels in different matrices for a 1:1 soil: solution ratio (a) and boxplots showing the same distributions grouped by texture class (b). Boxes show first and third quartiles, with median line in bold. Whiskers show minimum and maximum values.

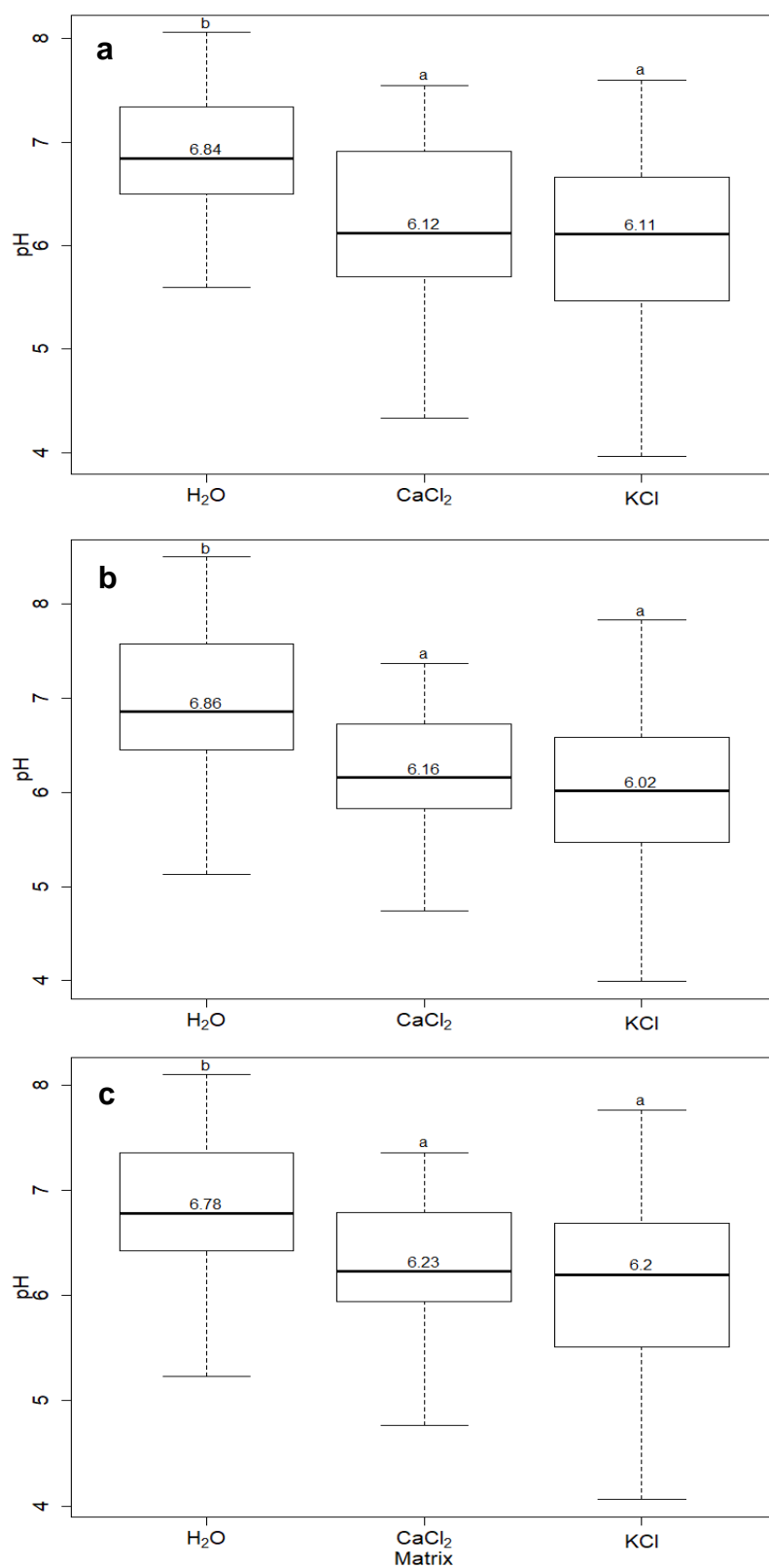


Figure 3-4 Boxplots showing distribution of soil pH levels in different matrices for soil: solution ratios of 1:2.5 (a), 1:4 (b) and 1:5(c). Boxes show first and third quartiles, with median line in bold. Whiskers show minimum and maximum values.

3.3.3 Conversion between soil pH measurement methods

3.3.3 i) SLR models

The coefficient, intercept, r^2 - and RMSE-values of SLR equations for conversion of pH values between the most commonly-used methods are shown in Table 3-3. Preliminary investigation of data showed that linear fits were more appropriate than polynomial functions such as those used by Little (1992). The linearity of conversions confirms results reported in previous studies of van Lierop 1981, Kissel and Miler (2010) and Kome et al. (2018). The linearity also shows that the difference between soil pH values obtained in different electrolytes did not change between soils of different acidity levels, as found by Little (1992). The logarithmic relationships between pH values measured in different matrices as found by Al-Busaidi et al. (2005) and Galka *et al.* (2016) were not observed. This may be due to a greater frequency of samples with extreme pH values used for model calibration in those studies.

The effect of the explanatory variable on the model was highly significant in all cases ($p < 0.001$). The r^2 -values ranged from 0.88 to 0.98, showing that variation was largely explained by the models. The r^2 -values showed that model fits equivalent or higher than those reported by Kome *et al.* (2018). The coefficients of the models were like most of those reported by Kome *et al.* (2018), despite the more physicochemically diverse range of soils used for model calibration in this study. The RMSE-value ranged from 0.12 to 0.32, showing relatively small deviation of actual values from predictions. The RMSEE-values of the models were lower than the RMSEP-values resulting from cross-validation reported by Kome *et al.* (2018), however that study was performed using a more homogeneous set of soils compared to the diverse range of soils represented in this study.

Several pH conversion models are plotted in Figure 3-5 to visualize the uncertainty in predictions and effects of soil properties. The observed data points in Figure 3-5 (a) mostly lie within the prediction and confidence uncertainty ranges, supplying evidence of the capability of the models for predicting individual and mean values, respectively. About 95% of individual error magnitudes lie within $\pm 2 \times$ RMSE of the model output, forming the prediction interval (Kaplan, 2019). The actual error is represented by the vertical distance between the predicted value and the true value at the centre of the interval. In Figure 3-5 (b) we notice that the pH values measured in KCl and CaCl₂ are more similar for sands than for loams, where pH measured in KCl is slightly lower than that measured in CaCl₂. This supports the observations noticed for texture class trends discussed regarding Figure 3-3.

Table 3-3 Conversion equations between common pH measurement methods with R^2 and RMSE values (** $P < 0.001$)

X	Y	Water 1:2.5	KCl 1:2.5	CaCl ₂ 1:2.5	CaCl ₂ 1:4	CaCl ₂ 1:5	Water 1:5	KCl 1:5
Water 1:1	Y = 0.92X + 0.80	Y = 1.1X - 1.42	Y = 0.98X - 0.25	Y = 0.84X + 0.60	Y = 0.81X + 0.90	Y = 0.88X + 0.99	Y = 1.2X - 1.5	
	$R^2 = 0.97^{***}$	$R^2 = 0.93^{***}$	$R^2 = 0.95^{***}$	$R^2 = 0.93^{***}$	$R^2 = 0.92^{***}$	$R^2 = 0.96^{***}$	$R^2 = 0.94^{***}$	
	RMSE = 0.12	RMSE = 0.24	RMSE = 0.17	RMSE = 0.19	RMSE = 0.19	RMSE = 0.14	RMSE = 0.23	
Water 1:2.5		Y = 1.2X - 2.31	Y = 1.0X - 0.93	Y = 0.91X - 0.03	Y = 0.87X + 0.27	Y = 0.95X + 0.29	Y = 1.2X - 2.4	
		$R^2 = 0.94^{***}$	$R^2 = 0.93^{***}$	$R^2 = 0.92^{***}$	$R^2 = 0.92^{***}$	$R^2 = 0.96^{***}$	$R^2 = 0.93^{***}$	
		RMSE = 0.23	RMSE = 0.2	RMSE = 0.19	RMSE = 0.19	RMSE = 0.14	RMSE = 0.24	
KCl 1:2.5			Y = 0.81X + 1.3	Y = 0.72X + 1.8	Y = 0.68X + 2.1	Y = 0.73X + 2.4	Y = 1.0X + 0.04	
			$R^2 = 0.91^{***}$	$R^2 = 0.92^{***}$	$R^2 = 0.90^{***}$	$R^2 = 0.89^{***}$	$R^2 = 0.98^{***}$	
			RMSE = 0.24	RMSE = 0.2	RMSE = 0.21	RMSE = 0.23	RMSE = 0.12	
CaCl ₂ 1:2.5				Y = 0.85X + 0.93	Y = 0.8X + 1.3	Y = 0.86X + 1.5	Y = 1.1X - 0.89	
				$R^2 = 0.93^{***}$	$R^2 = 0.91^{***}$	$R^2 = 0.91^{***}$	$R^2 = 0.91^{***}$	
				RMSE = 0.18	RMSE = 0.2	RMSE = 0.21	RMSE = 0.28	
CaCl ₂ 1:4					Y = 0.91X + 0.63	Y = 0.97X + 0.84	Y = 1.3X - 1.8	
					$R^2 = 0.89^{***}$	$R^2 = 0.89^{***}$	$R^2 = 0.91^{***}$	
					RMSE = 0.22	RMSE = 0.24	RMSE = 0.28	
CaCl ₂ 1:5						Y = 1.0X + 0.52	1.3X - 2.1	
						$R^2 = 0.89^{***}$	$R^2 = 0.88^{***}$	
						RMSE = 0.23	RMSE = 0.32	
Water 1:5							Y = 1.3X - 2.4	
							$R^2 = 0.90^{***}$	
							RMSE = 0.30	

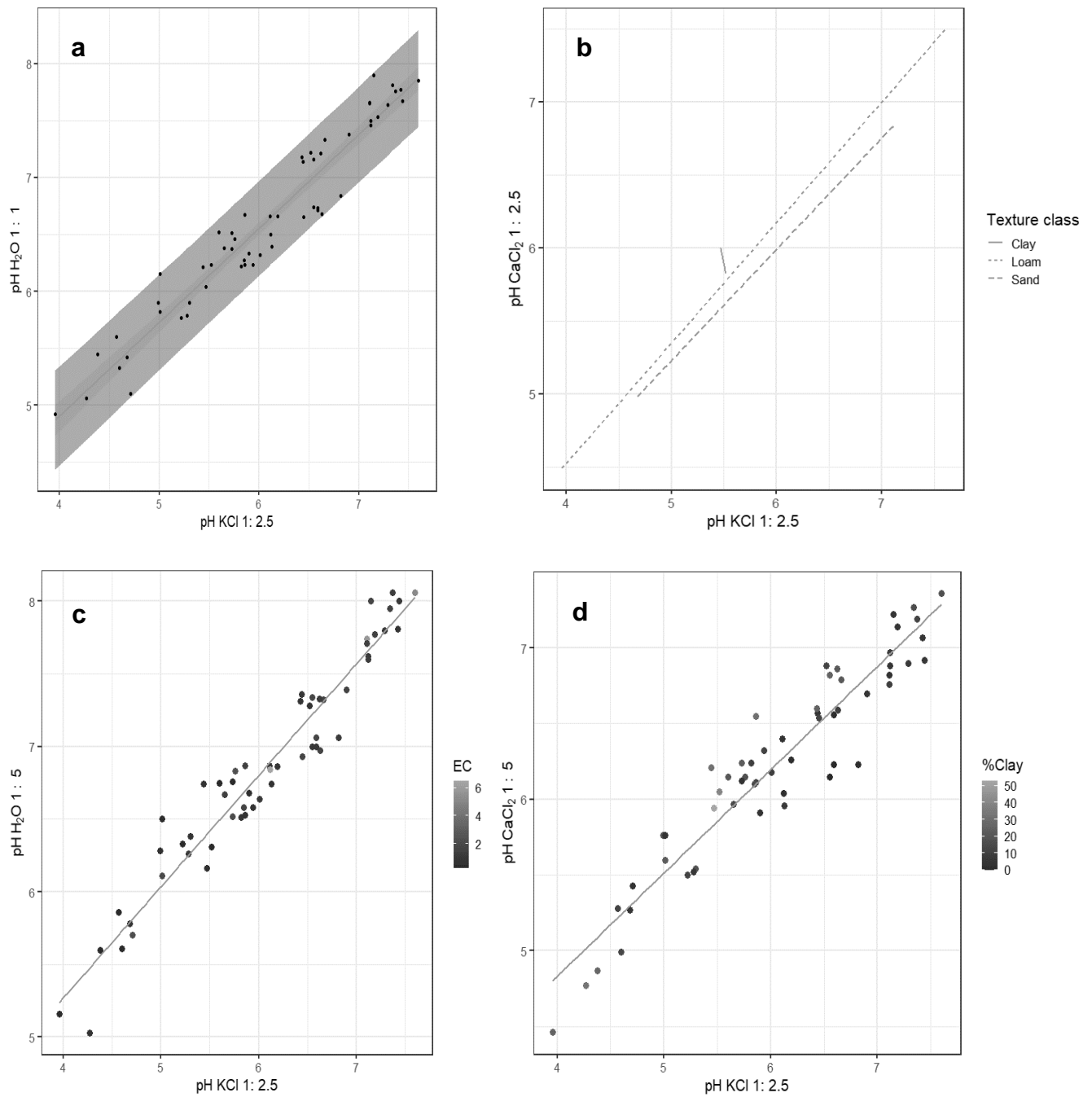


Figure 3-5 Plots showing conversions between pH measured by common methods to illustrate confidenceⁱ and predictionⁱⁱ intervals (a), as well as effects of textural grouping (b), EC levels (c) and %Clay content (d) on conversions

ⁱ Confidence intervals represented by narrow shaded bands

ⁱⁱ Prediction intervals represented by wide shaded bands

3.3.3 ii) MLR models

The Pearson coefficients for the correlation between soil properties (EC_e , clay content, %C and %N) and soil pH measured by various methods are shown in Table 3-4. It is evident that %N and %C were not correlated to soil pH, however EC_e and clay content showed moderate

correlation to soil pH. The positive correlation between EC and soil pH in this study contrasts with findings of Kome *et al.* (2018), where a negative correlation was found between these two parameters for non-saline soils. EC has been shown to influence the relationship between different pH measurement methods due to junction potential effects and a specific electrode response, while soil texture has been suggested to influence pH measurement due to buffering by silt- to clay-sized particles (see Section 2.1.4). However, the degree to which these parameters may improve pH prediction models on a practical level was not discussed. The practical aspects of using pH conversion equations for research and consulting purposes should be considered when deciding on the level of model complexity required for accurate conversions. For this reason, both parsimony and the degree of accuracy was compared between simple and more complex models in this study.

Multiple linear regression (MLR) models including either EC_e or %clay as additional explanatory variables were constructed. The difference between AIC-values of SLR versus corresponding MLR models are shown in Table A- 2 and Table A- 3, where a lower relative AIC-value indicates a better balance of model fit and predictive accuracy. The AIC-values of SLR models were lower than those of MLR models including EC as an additional parameter in most cases. Therefore, most simple conversion models were more parsimonious than models including EC_e as an additional parameter. The AIC-values of more than half of the MLR models including %clay as an explanatory variable were lower than those of the corresponding SLR models. Thus, many MLR models including clay content as a predictor were more parsimonious than SLR models. The MLR models including %clay as a predictor which were not more parsimonious than SLR models, were mostly those converting $pH_{H_2O1:1}$ values. However, even though inclusion of %clay content as a predictor improved the parsimony of pH conversion equations, this only represented model improvement on a theoretical level. To evaluate the practical significance of using MLR models for pH conversion, the difference between RMSE values of MLR versus SLR models was evaluated. While some of the MLR models had lower RMSE-values than corresponding SLR models, the maximum degree of accuracy improvement was only 0.05 pH units. Most laboratories report soil pH accurate to one decimal place, therefore the implications of an error of the second decimal in the pH reading is thus of little significance.

The interactive effect of soil properties was investigated by performing stepwise linear regression in the forward direction, using AIC as a criterion for variable selection. The MLR models for this set of conversions included different combinations of the variables %C, %N and %clay content, however, the combination for each model is not shown for the sake of simplicity. The differences between RMSE-values of SLR versus MLR models are shown in Table A- 4. The RMSE-values of MLR models were only up to 0.06 pH units lower than those

of SLR models. Therefore, incorporating various combinations of soil properties as predictors did not improve model prediction accuracy at a practically useful level. The use of conversion models using only pH as an explanatory variable, is deemed the most effective means of soil pH prediction, compared to models including additional variables.

Table 3-4 Pearson coefficients (r-values) showing correlation between pH measurement methods and soil properties

pH measurement method	%N	%C	%clay	Ec_e
KCl 1:1	0.07	0.11	-0.34	0.39
KCl 1:2.5	0.08	0.12	-0.37	0.36
KCl 1:4	0.07	0.11	-0.35	0.38
KCl 1:5	0.07	0.11	-0.37	0.35
H ₂ O 1:1	0.01	0.01	-0.26	0.33
H ₂ O 1:2.5	0.04	0.04	-0.31	0.35
H ₂ O 1:4	0.02	0.03	-0.30	0.34
H ₂ O 1:5	0.00	0.00	-0.26	0.35
CaCl ₂ 1:1	0.06	0.08	-0.22	0.33
CaCl ₂ 1:2.5	0.05	0.06	-0.21	0.32
CaCl ₂ 1:4	0.06	0.08	-0.26	0.31
CaCl ₂ 1:5	0.02	0.02	-0.23	0.32

3.3.3 iii) Conversion factors

The use of conversion factors (intercept set to zero) is desirable for rapid computation of soil pH values in the consulting industry. The coefficients for conversion of soil pH values obtained by different measurement methods are shown in Table 3-5. The effect of the explanatory variable on the model was highly significant in all cases ($p < 0.001$) and the r^2 -values were unity for all conversions (hence not shown), which is an indication of over-fitting. The differences between RMSE-values of SLR models versus conversion factor equations are also shown in the table. All SLR models showed higher prediction accuracy compared to conversion factor equations. This was especially the case for five conversions with RMSE differences of the first decimal order, which may result in values that are imprecise by an entire pH unit due to rounding. This has important implications for the interpretation of soil analysis reports, as an error in lime requirement calculated for remediating a soil by 0.3-0.5 pH units corresponds to a tonne of pure lime in loamy soil (Miller, 2019) and up to 10 tonnes of lime per ha in clayey soil (Stenberg et al., 2010). Thus, SLR models are preferred over conversion factors when serious errors must be avoided, and it is recommended that conversion factors are used with caution.

Table 3-5 Conversion factors between common pH measurement methods and increase in RMSEⁱ-values compared to corresponding SLR models

X	Y	Water 1:2.5	KCl 1:2.5	CaCl ₂ 1:2.5	CaCl ₂ 1:4	CaCl ₂ 1:5	Water 1:5	KCl 1:5
Water 1:1	Y = 1.0X ΔRMSE = 0.03		Y = 0.92X ΔRMSE = 0.05	Y = 0.94X ΔRMSE = 0	Y = 0.93X ΔRMSE = 0.01	Y = 0.94X ΔRMSE = 0.03	Y = 1.0X ΔRMSE = 0.05	Y = 0.93X ΔRMSE = 0.06
Water 1:2.5			Y = 0.89X ΔRMSE = 0.1	Y = 0.91X ΔRMSE = 0.02	Y = 0.90X ΔRMSE = 0.0	Y = 0.91X ΔRMSE = 0.0	Y = 0.99X ΔRMSE = 0.0	Y = 0.90X ΔRMSE = 0.11
KCl 1:2.5				Y = 1.0X ΔRMSE = 0.06	Y = 1.0X ΔRMSE = 0.14	Y = 1.0X ΔRMSE = 1.7	Y = 1.1X ΔRMSE = 0.2	Y = 1.0X ΔRMSE = 0
CaCl₂ 1:2.5					Y = 1.0X ΔRMSE = 0.03	Y = 1.0X ΔRMSE = 0.06	Y = 1.1X ΔRMSE = 0.07	Y = 0.99X ΔRMSE = 0.02
CaCl₂ 1:4						Y = 1.01X ΔRMSE = 0.01	Y = 1.1X ΔRMSE = 0.01	Y = 1.0X ΔRMSE = 0.07
CaCl₂ 1:5							Y = 1.1X ΔRMSE = 0.01	Y = 0.99X ΔRMSE = 0.07
Water 1:5								Y = 0.9X ΔRMSE = 0.08

ⁱ RMSE_{conversion factor equation} - RMSE_{SLR}

3.4 Conclusions

Highly reproducible pH measurements were obtained in all soil: solution ratios and matrices used in this study. The pH measured in different matrices decreased in the order $\text{pH}_{\text{H}_2\text{O}} > \text{pH}_{\text{CaCl}_2} > \text{pH}_{\text{KCl}}$. Soil pH is observed to increase slightly with an increase in soil: solution ratio because of the dilution effect. Most SLR models for conversion of pH values between different methods showed good model fit and accuracy. Moderate correlation was observed between EC_e , as well as clay content, with soil pH. SLR models were more parsimonious than MLR models including EC_e as a predictor. On the other hand, MLR models including %clay content as an explanatory variable showed greater parsimony than SLR models. MLR models were only slightly more accurate than SLR models, even when combinations of predictors selected by stepwise regression were included in models. Therefore, MLR conversion equations would not be of more practical use in agricultural consultancy. Conversion factor equations for pH prediction are only slightly less accurate than SLR equations in most cases. However, in some instances the lower accuracy could have serious financial implications regarding calculation of lime requirement. It is recommended to use SLR models for soil pH prediction, which are simpler than MLR equations, but more accurate than conversion factors.

4. Relationships between soil base cation extraction methods

4.1 Introduction

Exchangeable base cations (Ca, Mg, Na and K) are adsorbed to soil colloids by electrostatic forces and readily displaced by other ions in the soil solution (Huang *et al.*, 2012). Cationic plant nutrients in soils are taken up by plant roots to perform essential metabolic functions. The cations dominating the soil exchange complex influence soil physicochemical properties such as structural stability and pH buffering capacity. Understanding how exchangeable base cation concentrations affect soil quality is essential for soil management purposes. However, different soil cation content norms have been calibrated using various soil CEC determination methods and soil types. This makes interpretation of soil chemical test results problematic for scientists and soil consultants when cation concentrations are not reported according to local methods, as no studies have examined relationships between all local and international methods for a single dataset. Exchangeable and soluble cations are measured by leaching soil samples with an extractant and subsequently quantifying the cations displaced from the exchange complex by spectroscopic analysis of the leachate (Huang *et al.*, 2012). The 1M NH_4OAc (AA, pH = 7) procedure, first described by Schollenberger and Dreibelbis (1930) is the most widely-known reference method to determine buffered or potential CEC for agricultural soils. An adaptation of this method (0.2M NH_4OAc , pH = 7) is also used in South Africa for quantification of base cations (The Non-Affiliated Soil Analysis Working Committee, 1990).

The AA (1M, pH = 7) method allows taxonomic classification of soils, but tends to result in overestimation of CEC for weathered soils with variable charge properties (Gillman *et al.*, 1983; Jaremko & Kalembasa, 2014). Moreover, the high ionic strength of the extractant's index cation (1M NH_4) also affects CEC measurements. Consequently, several unbuffered methods have been developed to determine the effective CEC (eCEC), which is measured at a pH and ionic strength similar to what would be found *in situ* (Pleysier & Juo, 1980). eCEC methods that operate at low ionic strength require an index cation with high selectivity over soil exchangeable cations. The compulsive exchange (Gillman & Sumpter, 1986) extractant consists of 0.2M barium chloride (BaCl_2) and 0.2M ammonium chloride (NH_4Cl). The Ba ion outcompetes other cations for adsorption to clays, due to a high charge to hydrated radius ratio (Mengel *et al.*, 2001). The method, although recommended by Sumner & Miller (1996) for high reproducibility, is not used in for routine analyses in commercial laboratories as it is time-consuming and generates toxic waste. Several international methods have been developed for simultaneous extraction of P, trace elements and exchangeable base cations,

such as the Mehlich-III (M-III, Mehlich, 1984), Citric acid (1%, Dyer, 1984) and Bray-II (Bray & Kurtz, 1945) methods. These methods have been adopted by South African soil test laboratories for P testing, particularly in the Western Cape province. Although the use of the Citric acid (1%) method has declined in other countries, the M-III method is widely used in the US and the Bray-II method is used in tropical regions such as Malaysia and Columbia (Zapata & Roy, 2004). The Ambic-I method (Van der Merwe *et al.*, 1984) is the preferred multi-elemental extraction method in the province of KwaZulu-Natal. This method is a modification of the International Soil Fertility Evaluation and Improvement (ISFEI) procedure (Hunter, 1974). While the Citric acid (1%) extractant relies on acidic dissolution of phosphates (Dyer, 1984), the Bray-II extraction process is also driven by complexation and precipitation of companion cations by fluoride (Bray & Kurtz, 1945). The M-III and Ambic-I extractants contain EDTA as an additional chelating agent.

The different chemical mechanisms governing operation of these extractants result in different effectivities of the exchangers. Gillman *et al.* (1983) and (Horn, 1982) found greater displacement of divalent cations (Ca and Mg) from the soil exchange complex, as well as dissolution of carbonates by 0.1M BaCl₂ compared to the AA method, although this result is not always obtained. Michaelson *et al.* (1987) and Gartley *et al.* (2007) suggested that the nature of the soil affected regressions for M-III extractable Ca versus AA. However, Wang *et al.* (2004) and Zhu *et al.* (2016) found no correlation between soil physicochemical properties and extractable cation content. Thus, there is a lack of research providing evidence of the effects of soil properties on cation content conversions. There is general agreement that AA-based extractant are more efficient than other methods for K-extraction, due to its high concentration of NH₄ ions, which have a similar hydrated ionic radii to K ions (Amacher *et al.*, 1990; Ciesielski *et al.*, 1997; Jönsson *et al.*, 2002). Bray-II, however, has been shown to extract similar quantities of K and Mg as that obtained by the AA method (Doll & Lucas, 1973, Ballard and Pritchett, 1975 and Mehlich, 1974 as cited in Ballard, 1978). High correlation coefficients were found between extractable K-content of several extraction methods by Thompson (1995). A study by Simard *et al.* (1989) showed high correlation between citric acid-extractable K and coarse clay content of the soil. The importance of conversion accuracy and associated practical implications are not often mentioned in literature. Therefore, a more thorough approach is needed to define acceptable ranges of prediction error. The aim of this study is to develop equations for conversion of soil exchangeable base cation concentrations obtained by the several extraction methods for a set of Western Cape soils and to evaluate the effect of soil properties as regressors for improved model performance. The relevance of model prediction error for potential practical applications will also be discussed.

4.2 Materials and methods

4.2.1 Soil sampling and characterisation

Soil sampling and preparation was performed in an identical manner to that outlined in Section 3.2. The same set of soils were used, except for the replacement of six samples by different ones, as the original sample set did not contain sufficient mass of soils to perform cation extractions. These samples were also obtained from the Elsenburg soil analytical laboratory.

4.2.2 Extraction of soil base cations

The extractable base cation concentrations (Ca, Mg, Na and K) of each soil sample, were determined by seven methods. A summary of these procedures is shown in Table 4-1. This is the first study to compare some of extraction methods used in South Africa, namely Citric acid (1%), Ambic-I and ammonium acetate (0.2M), with international methods. The methods used in this study were the (i) standard ammonium acetate (1M NH_4OAc , pH = 7), (ii) 0.2M ammonium acetate (The Non-Affiliated Soil Analysis Working Committee, 1990), (iii) Mehlich-III (Mehlich, 1984), (iv) Ambic-I (The Non-Affiliated Soil Analysis Working Committee, 1990), (v) citric acid (1%) (The Non-Affiliated Soil Analysis Working Committee, 1990), (vi) compulsive exchange (Soil Survey Laboratory Staff, 1992) and (vi) Bray-II (Bray & Kurtz, 1945) methods. All these methods were performed by the MSc candidate at the department of Soil Science, Stellenbosch University, except for the Citric acid (1%) analysis, which was performed by Elsenburg Analytical Services.

The AA (1M, pH = 7) method was performed for each sample as follows. A 5.0 g soil sample was placed in a Buchner funnel lined with Whatman no. 42 filter paper. The sample was stirred with 25 mL 1M NH_4OAc (pH = 7 \pm 0.1) and leached using a vacuum extractor. After adding an additional 25 mL 1M NH_4OAc and covering the flask opening, the sample stood overnight and was leached the following day. The leachate was made up to the required volume with 1M NH_4OAc for subsequent determination of exchangeable cations. The AA (0.2M, pH = 7) method was performed for each sample as follows. After stirring a 20 g soil sample with 50 mL extractant (pH = 7 \pm 0.1) and allowing it to stand for 1 hour, the sample was leached in a Buchner funnel lined with Whatman no. 42 filter paper, using a vacuum extractor. The leachate was made up to the required volume with 1M NH_4OAc for subsequent determination of cation concentrations. In saline soil samples ($\text{EC} > 4 \text{ dS m}^{-1}$), the corrected (exchangeable) cation content was determined by subtracting the water-soluble cation content from the extractable cation content. The water-soluble cation content was determined by performing an additional extraction on separate samples of the saline soils according to the regular procedure but using distilled water instead of the extractant solution.

The M-III method was performed for each sample as follows. A soil sample with a volume of 2.5 cm³ was shaken with 25 mL extractant and filtered in a Buchner funnel lined with Whatman no. 42 filter paper, using a vacuum extractor. The leachate was made up to the required volume with M-III extractant for subsequent determination of exchangeable cation concentrations. The Ambic-I method was performed for each sample as follows. A 5 g soil sample was shaken by hand with 50 mL extractant, allowed to stand for 20 min and shaken by hand again. The solution was then shaken for 30 min at 180 oscillations per minute and allowed to stand for at least 5 min. The sample was then centrifuged for 5 min and 25 mL of the leachate was transferred to a 50 mL volumetric flask. The samples were centrifuged using a Zonkia SC-3612 machine operating at its maximum speed of 4500 rpm. After adding 2.5 mL concentrated HCl (32 % v / v) and shaking the flask by hand, precipitation of organic material occurred overnight. The solution filtered in a Buchner funnel lined with Whatman no. 42 filter paper using a vacuum extractor. The leachate was made up to the required volume with distilled water for subsequent determination of exchangeable cations.

The compulsive exchange method was performed for each sample as follows. A 2.0 g soil sample was shaken for 2 h with 10 mL 0.2M BaCl₂ / 0.2M NH₄Cl solution and then centrifuged. The supernatant was made up to the required volume with 0.2M BaCl₂ / 0.2M NH₄Cl solution for subsequent determination of exchangeable cations. Soluble salts were removed from saline soil samples (EC > 4 dS m⁻¹) prior to the extraction procedure. This was done by washing the soils, first with 20 mL 70 % ethanol, followed by 20 mL 10 % ethylene glycol (C₂O₂O₆), then discarding the solutions. The Bray-II method was performed for each sample as follows. A 6.0 g soil sample was shaken for 40 s with 45 mL extractant and then passed through a funnel lined with Whatman no. 42 filter paper and allowing filtration by gravity. The extract was retained for subsequent determination of exchangeable cations.

Table 4-1 Procedures involved in different cation extraction methods

Method	Extractant composition	Soil: extractant ratio	Buffer	Extraction procedure	Additional procedure	Reference
Ammonium acetate (1M)	1M NH ₄ OAc	0.1	pH = 7 (NH ₄ OH / CH ₃ COOH)	Leach soil with extractant in leaching tube using vacuum extractor. Add 2 nd portion of extractant, allow to stand overnight & filter using Buchner funnel.		Soil Survey Laboratory Staff, 1992
Ammonium acetate (0.2M)	0.2M NH ₄ OAc	0.04	pH = 7 (NH ₄ OH / CH ₃ COOH)	Stir soil with extractant and allow to stand for 1 h. Filter using Buchner funnel, adding remaining extractant in aliquots over 1 h.		The Non-Affiliated Soil Analysis Working Committee, 1990
Mehlich-III	0.2M CH ₃ COOH 0.25 M NH ₄ NO ₃ 0.015 M NH ₄ F 0.013 M HNO ₃ 0.001M EDTA	0.1		Shake soil with extractant for 5 min and filter using Buchner funnel.		Mehlich, 1984
Ambic-I	0.25 M NH ₄ HCO ₃ 0.01M (NH ₄) ₂ EDTA 0.01M NH ₄ F Superfloc N-100	0.1	pH = 8.2 +- 0.1 (NH ₄ OH)	Shake soil briefly with extractant and allow to stand for 20 min. Shake for 30 min at 180 osc/min and allow to stand for 5 min.	Post-extraction OM removal with HCl	The Non-Affiliated Soil Analysis Working Committee, 1990
Citric acid (1%)	1% citric acid (C ₆ H ₈ O ₇)	0.1		Shake soil with extractant, place in oven at 80 °C and shake every 10 min for 1 h.	Post-extraction OM removal with HCl, HNO ₃ & heating	The Non-Affiliated Soil Analysis Working Committee, 1990
Compulsive exchange	0.2M BaCl ₂ 0.2M NH ₄ Cl	0.2		Shake soil with extractant for 2 h, centrifuge and collect supernatant.	Pre-extraction removal of soluble salts with ethanol & ethylene glycol	Gillman and Sumpter, 1986
Bray-II	0.1M HCl 0.03 M NH ₄ F	0.13		Shake soil with extractant and filter immediately		Bray & Kurtz, 1945

4.2.3 Measurement of soil base cations

Triplicate measurements were performed on six samples for each cation extraction method. The concentrations of extractable cations by the methods performed at Stellenbosch University were determined by atomic absorption spectroscopy (AAS) using the Varian AA240FS instrument. The concentrations of exchangeable cations extracted according to the Citric acid (1%) method by Elsenburg Analytical Services were determined by inductively coupled plasma atomic emission spectroscopy (ICP-AES). The use of distilled water (as opposed to deionized water) for the preparation of extraction solutions was desirable, as distilled water was more readily available. A trial was performed to investigate cation contamination of distilled water versus deionized water. The trial proceeded as follows. The concentrations of base cations (Ca, Mg, Na and K) in five distilled water samples, as well as in five deionized water samples, were determined by atomic absorption spectroscopy (AAS) using the Varian AA240FS instrument. The concentrations of each element were compared (distilled versus deionized water samples) by means of an independent 2-sample t-test using RStudio 1.2.5033 (RStudio Team, 2020) software. The results showed that there was no significant difference between the mean base cation quantities of distilled water samples versus deionized water samples ($t_3 = -0.12, 1.28, -0.38$ and -1.33 for Ca, Mg, Na and K respectively, NS). Hence, it was concluded that distilled water was suitable for the purpose of preparing extraction solutions.

4.2.4 Statistical analyses

Statistical analyses were performed using the latest editions of STATISTICA and RStudio 1.2.5033 (RStudio Team, 2020) software. The minimum, maximum, mean and SD of general soil properties ($\text{pH}_{\text{KCl}:2.5}$, EC_e , clay content, %C and %N) were computed. The Pearson coefficients for correlation between methods of extraction for each of the exchangeable base cations, as well as coefficients for correlation between soil properties (EC_e , clay content, %C and %N) and base cation quantities were found. The CV (%) of replicate measurement data was found to determine the reproducibility of exchangeable cation extraction for each method. PCA biplots were constructed to identify outlier samples in conversions between different methods, as well as to visualize the degree of correlation between different extraction methods. Kruskal Wallis tests were performed to determine whether methods differed in extraction efficiency for each element. Post-hoc pairwise Dunn tests with Bonferroni adjustments were done to identify which methods differed significantly from one another.

SLR models calibrated from datasets excluding outliers were constructed for conversion between cation concentrations determined by different extraction methods. The original

dataset (including outliers) was also used to calibrate SLR models that would enable prediction of Na quantities in saline soils. The r^2 , RMSE, coefficient, and intercept, as well as the significance of inclusion of these values (based on two-tailed student's t-statistic) was computed for each model. Confidence and prediction intervals at 95% confidence level were plotted to visualize the uncertainty in predictions. The effect of soil properties on extractable cation measurement was investigated by construction of MLR models for prediction of Ca, Mg and K content by including soil pH as a predictor. The same procedure was followed for Na content conversions by including EC_e as a predictor. Stepwise regression was used to select optimal combinations of soil properties as predictors for conversion of exchangeable base cation content. The AIC- and RMSE-values of each SLR model, as well as those of corresponding MLR models, were computed.

4.3 Results and discussion

4.3.1 General soil properties

The soils used for comparison of cation extraction methods embodied a large variety of physicochemical properties. A summary of the soil properties is shown in Table 4-2. The distribution of general soil properties was like that of the set used for constructing conversion models for soil pH predictions. The sample set included 3 saline soils ($EC > 4 \text{ dS m}^{-1}$). The large range of clay contents represented by the sample set suited the objective of investigating the effect of soil texture on exchangeable cation conversions. The Eksteen acidity data was not available for the six samples which replaced those of insufficient mass from the pH analysis set, therefore the exact potential CEC distribution was not reported; however, the distribution is also expected to be like that of the set used for constructing conversion models for soil pH predictions.

Table 4-2 Summary of soil physicochemical properties of samples used for comparison of cation extraction methods (N = 57)

	C	N	Sand	Silt	Clay	pH_{KCl 1:2.5}	EC_e (dS m⁻¹)
	%						
Min	0.3	0.0	5.4	2.0	0.0	4.3	0.2
Max	3.2	0.3	97	84	53	7.6	6.5
Median	1.0	0.1	53	26	10	6.2	1.0
Std dev	0.7	0.1	22	18	10	0.9	1.2
CV (%)	68	68	42	58	84	15	116

Preliminary analysis of simple linear conversion equations between extraction methods revealed outlier samples for each cation group. This could be due to certain elements becoming unstable during sample combustion in the AAS analysis procedure to an extent depending on experimental conditions, reducing precision and accuracy of measurement (see Section 2.2.3). Hence it was necessary to select subsets of data from the full dataset for calibration of conversion models. Samples that were outliers in terms of typical distribution of soil attributes in the full cation measurements dataset, are represented by the dots on the boxplots in Figure B- 1. These were included in calibration datasets to increase model robusticity. Shown in Figure B- 2 are PCA biplots of cation quantities extracted by different methods, represented as vectors, with a 95% confidence ellipse. The biplots show the amount of variation in the data explained by different extraction methods, which methods correlate well, and identify samples demonstrating unusual behaviour across conversions (see Section 2.2.4). Based on the PCA biplots, subsets of data were selected to develop conversions for each of the four cations. Samples that did not conform to the conversion trend across extraction methods were investigated. If these soils did not possess any extreme physicochemical properties, their unusual behaviour was attributed to experimental error during extraction by at least one method and were thus eliminated. Less than 10% of samples from the original dataset were removed. To confirm the necessity of sample elimination, SLR models for conversion between cation quantities extracted by different methods were constructed using the full dataset, as well as using subsets after removal of outlier samples. The performance of the Na content conversion models improved with elimination of the outlier samples only when saline samples were also removed.

Hence, two sets of models were developed for predictions of exchangeable Na content. One set of models was calibrated from the full dataset. These models were developed with the aim of maximizing robusticity to improve predictions in highly saline soils. The other set of models was calibrated from a subset excluding outliers and saline samples. These models were developed with the aim of maximizing accuracy, but with predictive capability limited to soils falling of low to moderate salinity. The differences between RMSE-values of models calibrated from the original full dataset versus corresponding subsets are shown in

Table B- 1 to Table B- 4. Positive values indicate higher RMSE of the original SLR models. The RMSE-values of all models converting between the most highly correlated methods decreased after sample elimination, indicating improved prediction accuracy using the subsets. The summaries of the distribution of general soil properties in data subsets used for calibrating conversion models are shown in Table B- 5 to Table B- 8. Inspection of the data will reveal that the distribution of soil properties is highly similar in subsets compared to the original dataset.

4.3.2 Comparison of soil base cation extraction methods

4.3.2 i) *Reproducibility of exchangeable cation quantification by different methods*

The distribution of CV (%) of replicate measurements of exchangeable cation concentrations determined by the AA (1M and 0.2 M), Citric acid (1%), M-III, compulsive exchange, Ambic-I, and Bray-II methods are shown Figure B- 3. The mean and median CV (%) of six replicates were considered for evaluation of the reproducibility of cation concentrations obtained by each method. CV (%) < 10% was taken as an indication of good reproducibility, while CV (%) > 15% was deemed unacceptable. The lower reproducibility of Ca and Mg measurements with the AA (1M) and Bray-II methods may be explained by procedural practicalities. Only a short reaction time is allowed for Bray-II extractant to displace Ca from the soil exchange complex during the standard procedure, while no agitation is involved in the AA (1M) extraction procedure. These factors would interfere with displacement of strongly adsorbed divalent cations. Ca and Mg measurements were more reproducible with the other extraction methods. The validity of Mg concentrations determined with the Bray-II method is further addressed in the next section. Na measurements showed good reproducibility with all methods, but less so with the compulsive exchange method. The reproducibility of K measurements was excellent for all extraction methods. The AA (0.2M) and M-III methods showed the best overall reproducibility of extractable base cation quantification.

4.3.2 ii) *Comparison of different extraction methods*

The median concentrations of nutrients extracted by each method relative to those extracted by the M-III method are shown in Table 4-3 as ratio values. M-III was chosen as the reference method due to its widespread use (Huang *et al.*, 2012), high reproducibility, and practical efficiency. M-III was the most rapid extraction method as only a single step was required for collection of the leachate and no additional phase separation was necessary. Boxplots showing the distribution of base cation concentrations extracted by different methods are

shown in Figure 4-1. The boxplots exclude outlier values in terms of range to allow clear visual comparisons of nutrient concentration ranges between the methods. Outliers in terms of range were identified according to standard protocol (values $< Q1 - IQR \cdot 1.5$ or $> Q3 + 1.5 \cdot IQR$, where Q = quartile and IQR = inter-quartile range). There was no significant difference between Ca concentrations extracted by the different methods ($H = 5.924$, $DF = 6$, $p > 0.05$). The least Ca was extracted by the M-III and Bray-II methods. The extractants used in these methods, designed for phosphorous (P) extraction, contain NH_4F to drive dissolution of Ca-phosphates by precipitation of CaF_2 (Schmidt et al., 2004). By this mechanism, Ca is removed from solution, thus yielding lower quantities. A greater amount of Ca was extracted with Ambic-I compared to M-III despite the alkalinity of the former extractant. This could be due to the higher concentration of EDTA (0.1M) with Ambic-I extractant compared to M-III (0.001M), which solubilizes phosphate anions by complexation of Ca cations (Schmidt et al., 2004).

There was no significant difference between Mg concentrations extracted by the different methods ($H = 9.684$, $DF = 6$, $p > 0.05$). The most Mg was extracted by the AA (0.2M) method and the least by the Bray-II method. As is the case with Ca, precipitation of insoluble MgF_2 may have resulted in the lower Mg concentration yielded with the Bray-II method. Citric acid (1%) extracted equivalent or greater amounts of Mg compared to most other methods. This may be attributed to formation of readily soluble Mg-citrate complexes (Najafi-Ghiri et al., 2019). There was a significant difference between Na concentrations extracted by the different methods ($H = 223.3$, $DF = 6$, $p < 0.05$). Na amounts determined by compulsive exchange differed significantly ($p < 0.05$) from those obtained by all other methods. In addition, Na concentrations determined by Ambic-I and M-III differed significantly ($p < 0.05$) from those determined by the Bray-II and Citric acid (1%) methods. Na concentrations determined by Citric acid (1%) also differed significantly ($p < 0.05$) those obtained by the AA (1M and 0.2M) methods. There was a significant difference between Na concentrations extracted by the different methods for the full dataset including saline soils ($H = 178.71$, $DF = 6$, $p < 0.05$). Similar results were obtained regarding significant differences between methods for saline soils as for the subset excluding these soils.

The most Na was extracted by the compulsive exchange method and the least by the Citric acid (1%) method for soils of normal EC_e range, as well as saline soils. As noted in previous sections, compulsive exchange-extractable Na content was not highly correlated to other methods in this study, neither was it highly reproducible. There was a significant difference between K concentrations extracted by the different methods ($H = 16.94$, $DF = 6$, $p < 0.05$). AA (1M)-extractable K concentrations differed significantly ($p < 0.05$) from those obtained by the Citric acid (1%) method. The most K was extracted by the AA (1M) method and the least by the Citric acid (1%) method. AA (1M) has been shown to effectively displace K due to the

similar size and charge of the NH_4 ion, as well as the high concentration of the index cation (Amacher *et al.*, 1990; Ciesielski *et al.*, 1997; Jönsson *et al.*, 2002).

Table 4-3 Median ratios of nutrient concentrations determined by different methods versus the M-III method ([element]method / [element]M-III) using data subsets for Ca (N = 55), Mg (N = 55), Na (N = 52) and K (N = 52), as well as the full datasetⁱ for Na (N = 57). Compulsive exchange method denoted as BaCl₂

Median	AA (1M)	AA (0.2M)	Citric acid (1%)	BaCl ₂ ⁱⁱ	Ambic-I	Bray-II
Ca	1.4	1.4	1.4	1.3	1.4	1.0
Mg	0.9	1.1	1.0	0.9	1.0	0.8
Na	0.8	1.1	0.3	1.6	1.0	0.5
Na ⁱ	0.8	1.1	0.3	1.7	1.0	0.5
K	1.3	1.1	0.7	1.0	1.0	0.9

ⁱ Full Na dataset

ⁱⁱ Compulsive exchange extraction method

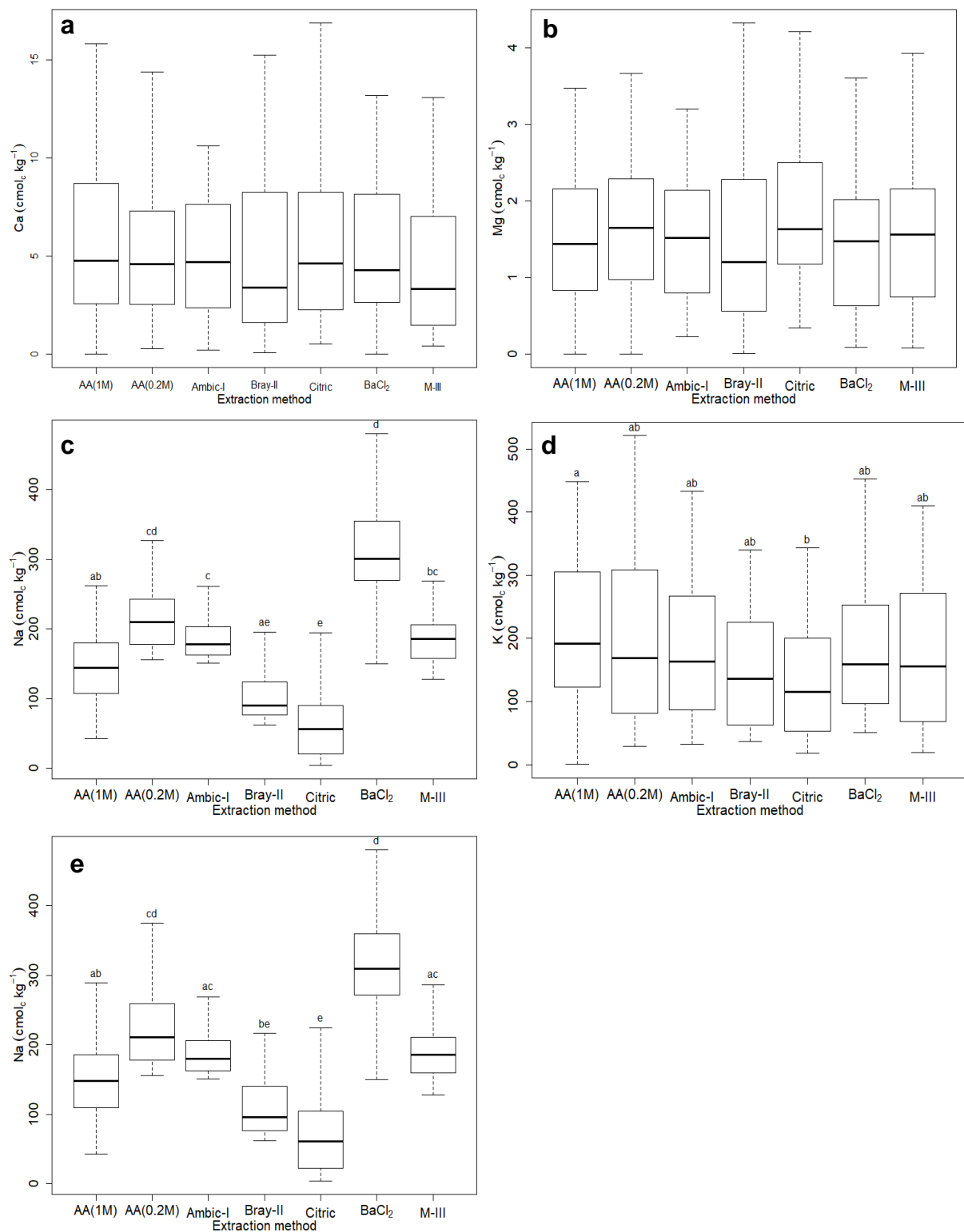


Figure 4-1 Relative amounts of base cations extracted by different methods using data subsets for Ca (a, $N = 55$), Mg (b, $N = 55$), Na (c, $N = 52$) and K (d, $N = 52$), as well as using the full dataset for Na (e, $N = 57$). Compulsive exchange method denoted as BaCl₂. Boxes show first and third quartiles, with median line in bold. Whiskers show minimum and maximum values (excluding outliers w.r.t. range). Superscripts show means that are not significantly different when compared using an a posteriori Dunn test.

4.3.3 Conversion between soil base cation extraction methods using SLR models

Preliminary investigation of data showed that linear regressions were the most appropriate fits for conversion of exchangeable base cation quantities. Since RMSE is reported in the same units as the target property, there are no absolute thresholds in literature. It is recommended that the RMSE should be smaller than 10% of the range of predicted property value (Alexander *et al.*, 2015). Hence, RMSE-values were evaluated based on the size of RMSE relative to the approximate ranges of cation quantities extracted from the soil. An RMSE-value of less than 10% of the associated nutrient content range, was deemed an acceptable degree of error. Conservative ranges (excluding outliers) were chosen to represent most soils used for model calibrations in this study. Model suitability was assessed on a basis of meeting both accuracy and model fit requirements outlined in Table 4-4 below, with elements reported in units commonly used on soil reports and by soil consultants. Models meeting both criteria were deemed suitable for use as conversion equations, while models that did not meet one or both criteria were rejected.

Table 4-4 Criteria for evaluation of suitability of models for conversion of exchangeable base cation concentrations

	Units	Range ⁱ	Max RMSE threshold	Min r^2 threshold
Ca	cmol _c kg ⁻¹	15	1.5	
Mg	cmol _c kg ⁻¹	4	0.4	0.5
Na	mg kg ⁻¹	200	20	
K	mg kg ⁻¹	400	40	

ⁱ Range based on approximate average ranges of nutrient concentrations extracted from the soils used in this study

The coefficient, intercept, r^2 and RMSE values of simple linear regression equations for conversion of exchangeable base cation concentrations obtained by different extraction methods are shown for each of the elements in Table 4-5 to Table 4-8. None of the SLR models for conversion of Ca content had a suitably low RMSE-value according to the criteria concerning accuracy and model fit. Only the model for simple linear conversion of Mg between the AA (1M) and Ambic-I methods was deemed suitable according to the criteria concerning accuracy and model fit. The effect of the explanatory variable on the model was highly significant ($p < 0.001$) and variation was largely explained by the model ($r^2 > 0.80$). Only three models for simple linear conversion of Na in soils of normal EC_e range, showed a suitable level of accuracy and were deemed suitable according to the criteria concerning accuracy and model fit. The effect of the explanatory variable on the models was highly significant ($p < 0.001$) and variation was largely explained by the models ($r^2 > 0.75$). Only the model for simple linear conversion of Na (in saline soils) between the AA (0.2M) and Ambic-I methods was deemed suitable according to the criteria concerning accuracy and model fit. The effect of the

explanatory variable on the model was highly significant ($p < 0.001$) and variation was largely explained by the model ($r^2 > 0.80$). Seven models for simple linear conversion of K were deemed suitable according to the criteria concerning accuracy and model fit. Only Ambic-I extractable K was not convertible between other methods. The effect of the explanatory variable on the models was highly significant ($p < 0.001$) and variation was largely explained by the models ($r^2 > 0.85$).

Examples of confidence and prediction intervals for exchangeable base cation conversions are plotted in Figure 4-2 and may be interpreted as discussed in Section 3.3.3. Outliers are observed for certain conversions, but systematic identification and removal of these samples unique to conversions between specific methods would have resulted in different datasets for over 80 equations. Therefore, the simplest compromise was to remove only samples which were outliers for conversions across all methods, identified with PCA biplots as discussed in Section 4.3.1. The observed values of Mg concentration create a “cone effect” about the prediction line, showing that prediction accuracy decreases for high Mg concentrations. Suitable models for conversion of Ca were not found. It is expected that fewer models were suitable for prediction of Na content in saline soils compared to soils of normal salinity range, as the robusticity was a trade-off for accuracy. The element K was the most easily convertible of all four base cations.

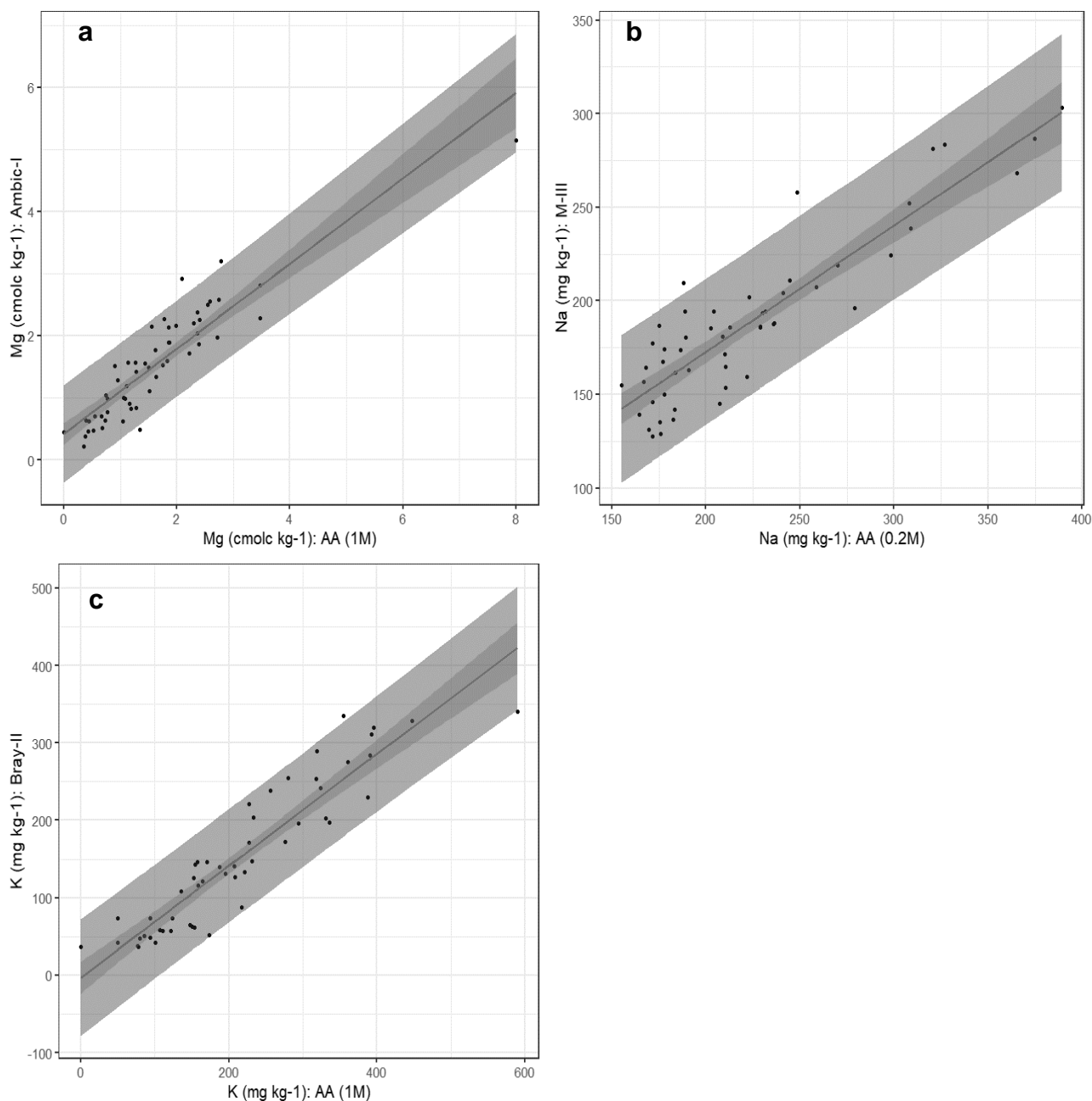


Figure 4-2 Confidenceⁱ and predictionⁱⁱ intervals for modelsⁱⁱ converting exchangeable Mg (a), Na (b) and K (c) between values obtained by different extraction methods

ⁱ Confidence interval represented by narrow shaded bands

ⁱⁱ Prediction interval represented by wide shaded bands

Table 4-5 SLR equation for conversion of exchangeable Mg content in units of $\text{cmol}_c \text{kg}^{-1}$ (** $P < 0.001$)

X	Y	Ambic-I
AA (1M)		$Y = 0.69X + 0.42$ $R^2 = 0.83^{***}$ RMSE = 0.4

Table 4-6 SLR equations for conversion of exchangeable Na content in units of mg kg^{-1} for soils within normal ECe range (** $P < 0.001$)

X	Y	M-III	Ambic-I
AA (0.2M)		$Y = 0.68Y + 38$ $R^2 = 0.81^{***}$ RMSE = 19	$Y = 0.6Y + 57$ $R^2 = 0.79^{***}$ RMSE = 17
Citric acid (1%)			$Y = 0.54Y + 150$ $R^2 = 0.76^{***}$ RMSE = 19

Table 4-7 SLR equation for conversion of exchangeable Na content in units of mg kg^{-1} for saline soils (** $P < 0.001$)

X	Y	Ambic-I
AA (0.2M)		$Y = 0.68X + 38$ $R^2 = 0.96^{***}$ RMSE = 19

Table 4-8 SLR equations for conversion of exchangeable K content in units of mg kg^{-1} (** $P < 0.001$)

X	Y	Citric acid (1%)	M-III	BaCl ₂ ⁱ	Bray-II
AA (1M)					$Y = 0.72X - 2.9$ $R^2 = 0.86^{***}$ RMSE = 35
AA (0.2M)		$Y = 0.72X - 3.3$ $R^2 = 0.91^{***}$ RMSE = 30	$Y = 0.81X + 18$ $R^2 = 0.89^{***}$ RMSE = 38	$Y = 0.69X + 40$ $R^2 = 0.87^{***}$ RMSE = 35	$Y = 0.68X + 21$ $R^2 = 0.91^{***}$ RMSE = 28
Citric acid (1%)					$Y = 0.87X + 34$ $R^2 = 0.86^{***}$ RMSE = 35
M-III					$Y = 0.77X + 18$ $R^2 = 0.87^{***}$ RMSE = 33

ⁱ Compulsive exchange extraction method

4.3.4 Conversion between soil base cation extraction methods using MLR models

Equations incorporating one soil property as a regressor for conversions between quantities of extractable base cations are shown in Table 4-9 to 4-11. Equations incorporating different combinations of soil properties as regressors for conversions between quantities of extractable base cations are shown in Table 4-12 to Table 4-16. Only the models with better performance (AIC and RMSE lower than corresponding SLR models and $R^2 > 0.5$) are tabulated. Although the model p-value was highly significant ($p < 0.001$) for all tabulated conversion equations, the significance of each individual predictor was not always significant. The significance level of the greatest p-value (which could be that of the model, or one of the predictors) is indicated in the tables. The model p-value may depend on the interaction of predictors; thus, inclusion of predictors should not be based strictly on their individual p-values. The Pearson coefficients showing the relationship between soil properties (EC_e , clay content, $pH_{KCl:2.5}$, %C and %N) and cation quantities extracted by various methods are shown in Table B- 14 to Table B- 18 for the elements Ca, Mg, Na and K. Only soil pH was correlated to the quantities of Ca, Mg and K extracted, and only EC_e was correlated to the quantities of Na extracted. These results support the theory that the nature of soils affect the conversions between exchangeable cation content as found by Michaelson *et al.* (1987). MLR equations for conversion of exchangeable Ca, Mg and K content were constructed by including pH as an explanatory variable. MLR models including EC_e as a predictor were constructed for conversions of Na quantities extractable from saline soils as well as soils within normal EC_e range.

4.3.4 i) *Analysis of MLR models incorporating one additional variable for exchangeable base cation conversion*

The differences between AIC-values of SLR versus corresponding MLR models for conversion of exchangeable base cation content are shown in Table B- 19 to Table B- 23. Most AIC-values of MLR models for conversion of Ca and Mg content were lower than those of SLR models. This indicates that most of the MLR models were more parsimonious than corresponding SLR models for Ca and Mg conversions. None of the MLR models for conversion of Ca content showed a suitably low RMSE-value. This shows that Ca dynamics on soil CEC sites is affected by a complex host of factors such as chemical properties of the extractant. The MLR model for conversion of Mg content performed better than the corresponding SLR model and variation was largely explained by the model ($r^2 > 0.80$). No other Mg conversion models were deemed suitable according to the criteria concerning accuracy and model fit. This outcome is like that of SLR conversion models. Just over two-

thirds of the MLR models for conversion of K content showed a lower AIC-value of MLR compared to SLR models. Therefore, models including soil pH as a regressor are not always more effective than models using only exchangeable cation content as a predictor. Nine models for multiple linear conversion of K showed better performance than SLR models as well as a suitable level of accuracy and degree of model fit. Like the SLR models, Ambic-I extractable K was not convertible between other methods. Variation was largely explained by the models ($r^2 > 0.80$).

A little under a half of the MLR models for Na prediction were more parsimonious than SLR models. Hence, the SLR models are not always more effective for Na conversions in soils within normal EC_e range. However, for saline soils, the AIC-values of MLR models were lower than those of SLR models in all except three cases. Therefore, inclusion of EC_e improves effectivity of the Na conversion models of saline soils. Only two models for multiple linear conversion of Na content in normal-range EC_e soils were deemed suitable according to the criteria concerning accuracy and model fit. Both models were for conversion of AA (0.2M) extractable Na. Variation was largely explained by the models ($r^2 > 0.80$). None of the MLR models for conversion of Na content in saline soils showed improved performance compared to SLR models. Like the SLR models, robusticity of MLR models for prediction of Na content in saline soils was traded off for accuracy of predictions in soils with a more typical EC_e -range. The incorporation of one soil property as a predictor for conversion of nutrient content improved model performance, but not consistently, or for conversion between the same sets of methods. Therefore, including a single soil property that is correlated to cation content as a regressor will not necessarily improve predictions of the cation content determined by other methods.

*Table 4-9 MLR models with soil pH as a regressor for conversion between exchangeable Mg content in units of $cmol_c\ kg^{-1}$ (** $P < 0.001$)*

X	Y	Ambic-I
		$Y = 0.71X - 0.13pH + 1.2$
AA (1M)		$R^2 = 0.84^{***}$
		RMSE = 0.4

Table 4-10 MLR models with soil EC_e as a regressor for conversion between exchangeable Na content in units of $mg\ kg^{-1}$ in soils within normal EC_e range

X	Y	Citric acid (1%)	Ambic-I
		$Y = 0.91X + 16EC - 150$	$Y = 0.56X + 8.2EC + 57$
AA (0.2M)		$R^2 = 0.84$	$R^2 = 0.80$
		RMSE = 24	RMSE = 17

Table 4-11 MLR models with soil pH as a regressor for conversion between exchangeable K content in units of mg kg⁻¹ (****P < 0.001, **P < 0.01, *P < 0.05)

X	Y	Citric acid (1%)	M-III	BaCl ₂ ⁱ	Bray-II
AA (1M)		Y = 0.71X + 17pH - 120 R ² = 0.85* RMSE = 38			Y = 0.67X + 21pH - 120 R ² = 0.9*** RMSE = 30
AA (0.2M)		Y = 0.69X + 7.5pH - 46 R ² = 0.91 RMSE = 30	Y = 0.78X + 11pH - 41 R ² = 0.89 RMSE = 37		Y = 0.64X + 14pH - 56 R ² = 0.93** RMSE = 25
Citric acid (1%)				0.94X - 12pH + 117 R ² = 0.84 RMSE = 39	Y = 0.83X + 11pH - 29 R ² = 0.87 RMSE = 34
M-III					Y = 0.73X + 10pH - 37 R ² = 0.88 RMSE = 32
BaCl ₂ ⁱ					Y = 0.78X + 25pH - 140 R ² = 0.85*** RMSE = 36

ⁱCompulsive exchange extraction method

4.3.4 ii) Analysis of MLR models constructed by stepwise regression

The effect of the interaction of parameters was investigated by performing stepwise linear regression in the forward direction, using AIC as a criterion for variable selection. The MLR models for this set of conversions include different combinations of the variables %C, %N, pH KCl and %clay content. The AIC of all models were lower than AIC-values of corresponding SLR models, as expected due to the method of variable selection by stepwise regression. The difference between RMSE-values of SLR versus MLR models are shown in Table B-24 to Table B-28. The comparison is shown for both MLR models constructed by inclusion of the single most correlated soil property, as well as MLR models selected by stepwise regression. Three models including optimal combinations of soil properties as regressors for conversion of exchangeable Ca content were deemed suitable according to the criteria concerning accuracy and model fit. All three models were for conversion of Ca determined by the compulsive exchange method. Variation was largely explained by the models ($r^2 > 0.75$). Two models including optimal combinations of soil properties as regressors for conversion of exchangeable Mg content were deemed suitable according to the criteria concerning accuracy and model fit. Both models were for conversion of Ambic-I extractable Mg. Variation was largely explained by the models ($r^2 > 0.80$).

Six models including optimal combinations of soil properties as regressors for conversion of exchangeable Na content in soils within normal EC_e range were deemed suitable according to the criteria concerning accuracy and model fit. Variation was largely explained by the models ($r^2 > 0.75$). Three models including optimal combinations of soil properties as regressors for conversion of exchangeable Na content in saline soils were deemed suitable according to the criteria concerning accuracy and model fit. Variation was largely explained by the models ($r^2 > 0.95$). Na quantities determined by the AA (0.2M) method were more easily

convertible. Like the SLR models, robusticity of models for prediction of Na content in saline soils was a trade-off for accuracy. Ten models including optimal combinations of soil properties as regressors for conversion of exchangeable K content were deemed suitable according to the criteria concerning accuracy and model fit. Variation was largely explained by the models ($r^2 > 0.85$). Overall, models including a combination of regressors resulting in the optimal balance between model complexity and accuracy consistently performed better than SLR models for conversion of exchangeable cation content between different methods. Hence, should data be available for incorporation of soil physicochemical properties into prediction models, it becomes possible to convert nutrient content with a greater level of accuracy.

Table 4-12 MLR models selected by stepwise regression for conversion between exchangeable Ca quantification methods with R^2 and RMSE ($\text{cmol}_c \text{ kg}^{-1}$) values (** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$)

X	Y	BaCl ₂ ⁱ
AA (1M)		Y = 0.34X + 19%N - 0.64EC + 1.4pH - 6.3 R ² = 0.8*** RMSE = 1.4
AA (0.2M)		Y = 0.36X + 0.57%C + 18%N - 0.57EC + 1.9pH - 9.7 R ² = 0.8 RMSE = 1.5
M-III		Y = 0.28X + 20%N - 0.54EC + 1.4pH - 5.9 R ² = 0.8** RMSE = 1.5

ⁱ Compulsive exchange extraction method

Table 4-13 MLR models selected by stepwise regression for conversion between exchangeable Mg quantification methods with R^2 and RMSE ($\text{cmol}_c \text{ kg}^{-1}$) values (** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$)

X	Y	Ambic-I
AA (1M)		Y = 0.71X - 0.13pH + 1.2 R ² = 0.84* RMSE = 0.4
BaCl ₂ ⁱ		Y = 0.8X + 0.18EC + 0.01%clay R ² = 0.82* RMSE = 0.4

ⁱ Compulsive exchange extraction method

Table 4-14 MLR models selected by stepwise regression for conversion between exchangeable Na quantification methods for soils within normal EC_e range with R^2 and RMSE (mg kg^{-1}) values (** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$)

X	Y	Citric acid (1%)	M-III	Ambic-I
AA (0.2M)		Y = 1.06X + 40%C - 390 %N + 4.7pH - 200 R ² = 0.91 RMSE = 18	Y = 0.66X + 13%C + 25 R ² = 0.85*** RMSE = 17	Y = 0.61X + 7.3pH + 10 R ² = 0.82** RMSE = 16
Citric acid (1%)			Y = 0.52X - 14%C + 270%N + 11EC + 0.54%clay + 120 R ² = 0.81 RMSE = 19	Y = 0.55X - 22%C + 190%N + 4.6pH + 130 R ² = 0.82 RMSE = 16
M-III				Y = 0.82X - 12.41%C + 6.2pH + 13 R ² = 0.78* RMSE = 18

Table 4-15 MLR models selected by stepwise regression for conversion between exchangeable Na quantification methods for saline soils with R^2 and RMSE (mg kg^{-1}) values (** $P < 0.01$, *** $P < 0.001$, * $P < 0.05$)

X	Y	M-III	Ambic-I
		$Y = 0.68X + 12\%C + 24$	$Y = 0.68X + 8.5\text{pH} - 12$
AA (0.2M)		$R^2 = 0.96^{**}$	$R^2 = 0.97^{**}$
		RMSE = 20	RMSE = 17
Citric acid (1%)		$Y = 0.57X - 15\%C + 260\%N + 14\text{EC} + 0.66\%\text{clay} + 120$	
		$R^2 = 0.96^*$	
		RMSE = 19	

Table 4-16 MLR models selected by stepwise regression for conversion between exchangeable K quantification methods with R^2 and RMSE (mg kg^{-1}) values (** $P < 0.01$, *** $P < 0.001$, * $P < 0.05$)

X	Y	Citric acid (1%)	BaCl ₂ ⁱ	Bray-II
		$Y = 0.65X + 1.2\%C + 270\%N - 8.4\text{EC} + 18$	$Y = 0.7X + 210\%$	$Y = 0.68X + 17\text{pH} - 0.82\%\text{cl}$
AA (1M)		$\text{pH} - 1.2\%\text{clay} - 120$	$N + 3.4$	$\text{ay} - 90$
		$R^2 = 0.9$	$R^2 = 0.84$	$R^2 = 0.9$
		RMSE = 31	RMSE = 39	RMSE = 29
AA (0.2M)		$Y = 0.7X + 15\%C - 0.72\%\text{clay} - 7.6$	$Y = 0.65X + 180\%$	$Y = 0.66X - 9.1\%C + 4.5\text{EC}$
		$R^2 = 0.92$	$N + 29$	$+ 11\text{pH} - 41$
		RMSE = 27	$R^2 = 0.88$	$R^2 = 0.94$
			RMSE = 34	RMSE = 24
Citric acid (1%)			$Y = 0.94X - 12\text{pH} + 120$	$Y = 0.87X - 21\%C + 11\text{pH} - 11$
			$R^2 = 0.84$	$R^2 = 0.89$
			RMSE = 39	RMSE = 31
M-III				$Y = 0.78X + 6.8\text{EC} + 7.3$
				$R^2 = 0.88$
				RMSE = 32
BaCl₂ⁱ				$Y = 0.8X - 13\%C + 25\text{pH} - 130$
				$R^2 = 0.86$
				RMSE = 35

ⁱ Compulsive exchange extraction method

4.3.5 Recommendation of appropriate use of exchangeable base cation conversion models

In sections 4.3.3 and 4.3.4, three types of models were given: SLR, MLR incorporating one additional variable, and MLR incorporating any number of additional variables. Model error must be weighed against model complexity when deciding which form of conversion equation to use in practice. If equations are to be used with the end goal of providing amelioration recommendations, the costs of fertilizer and amendment products associated with quantities of nutrients comparable to the average error of prediction should be calculated. In addition, the profit gained (yield or quality improvements due to correction of soil exchangeable nutrient status) should be calculated based on historic production norms. Thus, conversions of this soil-extractable Ca might be less routine than those of Na, Mg and K. This is because Ca deficiencies in the crop are usually confirmed by leaf analyses (Department of Soil Science, 2017) and lime requirement is primarily determined by soil pH.

Whether SLR or MLR equations are to be used for base cation conversions also depends on how far the readings are from the outer ranges of the norm. A critical threshold value could trigger requirement for remediation of nutrient lack or surplus. For example, the norm ranges for Ca and Mg are 4-12 and 0.4-4 cmolc kg⁻¹ respectively (Bemlab, 2016). A greater RMSE-value poses greater risk of triggering a critical threshold value. Whether SLR or MLR equations are to be used for Na conversions also depends on how far the readings are from the critical ESP or SAR threshold value which could trigger requirement for remediation of sodicity. The critical values for soils of various texture classes (ranges dependent on clay mineralogy) are shown in Table 4-17. Accuracy of conversions of Na content are more important in light-textured soils for calculation of gypsum requirement. The effect of the degree of prediction error will be relatively larger in soils with lower CEC, due to the CEC-value's place as denominator in the ESP calculation (see Equation 2-4).

As K is a plant macronutrient, it is usually applied in large quantities. Potassium-based fertilizers are known to be costly, and some crops have a capacity for luxury uptake of this nutrient, thus the optimal fertilizer application rate will depend on whether the gain in yield offers satisfactory return on investment. As perennials have great rooting depth, the fertilizer may be applied up to a depth of 60 cm (van Schoor *et al.*, 2000), tripling the costs. Considering the improvement of exchangeable cation content prediction accuracy with MLR models compared to SLR models, it may be beneficial to use more complex models for conversion of exchangeable base cation concentrations between different extraction methods.

Table 4-17 Approximate threshold values of Na content associated with sodic soils ($ESP > 15$)

Soil texture	Typical CEC ($\text{cmol}_c \text{ kg}^{-1}$) ⁱ	Na critical value (mg kg^{-1})
Sands	3-5	100-170
Loams	10-15	340-520
Silt loams	15-25	520-860
Clays / clay loams	20-50	700-1700

ⁱ CEC values adapted from (Culman *et al.*, 2021); ranges reported depend on clay mineralogy

4.4 Conclusions

The AA (0.2M) and M-III methods showed the best overall reproducibility of extractable base cation quantification. Extractable K is more consistently reproducible to quantify than other base cations. Extractants designed for phosphate dissolution via different chemical mechanisms yielded different Ca and Mg concentrations due to formation of precipitates and soluble complexes. Bray-II extractable Ca and Mg content cannot be interconverted between other extraction methods reliably. K concentrations are highly inter-convertible unless determined by the Ambic-I method. While models calibrated with saline soils for conversion of Na content are more robust, predictions of Na content in soils of typical EC_e -ranges are more accurate. The M-III method is recommended for extraction of base cations due to high reproducibility, rapidity of measurement, convertibility of results between other extraction methods and simultaneous extraction multiple soil nutrients. Moderate correlation was observed between EC_e and extractable Na content, as well as between $pH_{KCl1:2.5}$ and extractable Ca, Mg and K content. MLR models including various soil physicochemical properties as predictors were more accurate than those of most corresponding SLR models for conversions of exchangeable cation content. Models with the desired balance between accuracy and complexity should be chosen with consideration of availability of data, costs associated with soil amelioration products and the proximity of base status relative to critical threshold values.

5. Infrared spectroscopic determination of soil pH and extractable base cations

5.1 Introduction

Infrared (IR) spectroscopy has been shown to be a rapid, cost-effective technique of non-destructive soil analysis. *In situ* measurements by means of portable and spectroscopic instruments and remote-sensed hyperspectral imaging shows potential for large-scale soil mapping. Soil reference data is correlated with spectra collected by emission of infrared radiation onto the soil sample. Chemometric algorithms are used to create prediction models such that the parameters of interest may be predicted simultaneously in new samples using only spectral information (Aleixandre-tudo *et al.*, 2017; Gobrecht *et al.*, 2014). The measurement of several soil attributes (CEC, EC, concentrations of soil nutrients and toxic metals, soil pH, SOC, moisture content and clay mineralogy) by IR spectroscopy, has been reported by numerous authors (Wan *et al.*, 2020, Chen *et al.*, 2021, Johnson *et al.*, 2019; Mohamed *et al.*, 2018; Ng *et al.*, 2019; Nocita *et al.*, 2015). Spectral signatures of materials differ due to characteristic frequencies of energy absorbed by certain chemical functional groups (Johnston and Aochi, 1996). Most soil properties are better detected in the mid-IR (MIR) electromagnetic region compared to the near-IR (NIR) region, resulting in more accurate analyses, however NIR instruments more robust and suitable for field conditions (Nocita *et al.*, 2015; Soriano-Disla *et al.*, 2014; Viscarra Rossel *et al.*, 2016). However, there are not enough studies testing the viability of NIR and MIR spectral-based models to predict soil pH and exchangeable base cation content accurately for Western Cape soils.

With the advent of diffuse reflectance IR Fourier transform (DRIFT) spectroscopy, spectral data may be acquired more rapidly and with greater precision (Johnston & Aochi, 1996). Partial least squares regression (PLSR) is one of the most used chemometric techniques for IR spectral modelling (Rönkkö *et al.*, 2015 as cited in Mohamed *et al.*, 2018). PLSR is based on the concept of dimensionality reduction, decomposing spectra into eigenvectors and regressing these principal components with the parameter of interest. The technique is a robust approach to handling multicollinear variables (Mohamed *et al.*, 2018; Viscarra Rossel *et al.*, 2006). As soil spectra only contain information about soil organic components and mineralogy, soil properties such as pH and exchangeable cation content are not measured directly. However, due to co-variation of these properties with spectrally active soil constituents such as SOC and clay content, these properties have been successfully predicted by IR models (Chang *et al.*, 2001; Dunn *et al.*, 2002; Kamrunnahar *et al.*, 2003; Pirie *et al.*, 2005; Shepherd & Walsh, 2002). Few studies have evaluated the suitability of models calibrated from reference data generated by several chemical extraction methods, especially South

African methods. Some properties such as soil K content are highly site-specific due to variability of field conditions, high leaching rates and seasonal fertilizer applications. Mechanisms of cation exchange may vary between soil types, therefore co-variations upon which calibrations are based may differ from one region to another (Reeves *et al.*, 1999). Xu *et al.* (2016), Stenberg *et al.* (2010) and Araújo *et al.* (2014) reported improved performance of models calibrated by subsets of data that were uniform in mineralogy, texture class and SOC content. There is a lack of infrared studies providing evidence of improved accuracy of IR models calibrated and validated by samples with similar physicochemical properties. The aim of this study is to evaluate the performance of IR models calibrated from different sources of reference and spectral data to predict soil pH and exchangeable base cation content of Western Cape soils. The suitability of prediction accuracy for use in soil quality assessment will also be discussed.

5.2 Materials and methods

5.2.1 Infrared spectral measurement method

The methods of soil pH measurement and determination of exchangeable base cations are detailed in Sections 3.2 and 4.2. The full datasets of soil pH and exchangeable base cation content were used to calibrate models for prediction of these properties from IR spectra. IR spectral measurements were taken with three instruments namely the Alpha, Multi-Purpose Analyzer (MPA-II) and Matrix-F spectrometers (all produced by Bruker Optics, Ettlingen, Germany) to assess the suitability of each instrument for non-destructible analysis of solid soil samples. The Alpha instrument is a compact, robust device with relatively low operational cost. In external reflectance mode, non-contact measurements of the sample may be taken at close range. The MPA instrument uses an integrating sphere with a sample rotator enabling averaging of signal data from different parts of the sample, thereby improving accuracy and precision in the analysis of heterogeneous samples. The Matrix-F instrument allows contactless measurements of solid materials by means of a sensor head. Four tungsten NIR light sources (12 V, 20 W) with a measuring distance of 17 cm create an illumination area (10 to 12 cm diameter) larger than that of a conventional diffuse reflection probe, reducing the effect of variable particle sizes. Monofibers transmit the light back to the Matrix-F detector, which is more cost-effective than the fiber bundles used in conventional reflectance probes. The sensor head is robust due to accessories that protect the instrument from heat, dust, and vibration. This makes the Matrix-F instrument suitable for on-line measurements in commercial settings. Unmilled soil samples with a mass of 10 g were used for spectral data collection. A background scan was performed before each set of IR measurements (64 scans

per measurement) to zero the instrument. A blank surface was used as the background for the MPA-II and Matrix-F instruments, while a gold mirror was used to zero the Alpha instrument. The instrumental specifications and settings are given in Table 5-1.

Table 5-1 Specifications and settings of IR spectroscopy instruments

Instrument	Region	Mode	Range (cm ⁻¹)	Range (nm)	Sample holder	Rs ⁱ	Reps ⁱⁱ
MPA-II	NIR	DRIFT	12 500 - 3600	800 - 2778	Rotating integrating sphere	16	3
Matrix-F	NIR	DRIFT	12 000 - 4000	833 - 2500	Stationary integrating sphere	16	5
Alpha	MIR	External reflectance	7500 - 350	1333 - 28 571	Stationary glass beaker	4	3

ⁱ Spectral resolution

ⁱⁱ No. of spectral scans per soil sample (each spectrum correlated to replicate reference values)

5.2.2 Statistical analysis

OPUS / QUANT2 (OPUS 7.2 from Microsoft, Bruker Optics, Ettlingen, Germany) software was used for multivariate calibration based on a PLSR algorithm. The software includes built-in functions for selecting an optimal combination of spectral pre-processing methods and predictive components resulting in models with lowest prediction error. The pre-processing options available with this software included first and second derivative (1st and 2nd der.), constant offset elimination (COE), straight line subtraction (SLS), vector normalization (SNV), min-max normalization (Nmin-max), MSC and internal standard. The models were calibrated by correlating replicates of each reference value (N = 57) with individual replicate scans, as indicated in Table 5-1. This procedure, described by Alexandre-Tudo *et al.* (2019), challenges models by incorporating spectral variability that may be observed for a single soil due to scattering of infrared radiation by heterogeneous samples (see Section 2.3.1). The resulting datasets used for model construction with spectra generated by the MPA-II and Alpha instruments, totalled N = 171, while that using the Matrix-F instrument totalled N = 285. Fewer replicate measurements were taken with the MPA-II and Alpha instruments compared to the Matrix-F instrument, due to greater spectral reproducibility observed with preliminary results. The Alpha instrument operates in the MIR range, which has been shown to produce more reproducible results compared to NIR spectroscopy (Reeves *et al.*, 2010 as cited in Nocita *et al.*, 2015), as discussed in Section 2.3.1. The rotating sphere used as a sample holder with the MPA-II instrument may have improved reproducibility of spectral measurements, as explained in Section 5.2.1. Outliers were automatically detected according to the Mahalanobis distance. After manual exclusion of these outliers, separate models were constructed. If these

models showed improved performance compared to corresponding models calibrated with the full dataset, their figures of merit and range of predictive capability were also reported.

Different test set validation options available with OPUS software (using a subset of samples for calibration and the remaining samples for validation) were used for evaluation of each model. These included both automatic and structured selection methods. Automatic selection, using the Kennard-Stone algorithm, was applied to each model using a calibration sample subset comprising 64 % of the dataset, leaving 32 % of the samples as a test set. This was repeated using equal proportions (50 %) for calibration and validation. The Kennard-Stone algorithm allocates samples according to the chosen proportions of the subsets such that the subsets show similar distributions of the property of interest (Ramirez-Lopez & Stevens, 2020). Structured selection was applied using a validation: calibration sample number ratio of 1:2 (samples selected in consecutive order), with proportions allocated to subsets equivalent to the 32% / 64% automatic selection scenario. Structured selection was also applied using a validation: calibration sample number ratio of 1:1, with proportions allocated to subsets equivalent to the 50% / 50% automatic selection scenario.

The best-performing model (based on lowest RMSEP reported by different test set validation methods) was selected for prediction of each chemical property of interest. This was done for prediction of properties determined according to different chemical methods (i.e., different reference datasets) using each of the three IR spectroscopy instruments (i.e., different spectral datasets). Wavelength regions used as predictive components in calibration, as well as model performance indicators generated by the software were tabulated. The RPIQ-value of each model (see Section 2.3.3) was calculated using RStudio 1.2.5033 (RStudio Team, 2020) software and tabulated in addition to the automatically-generated figures of merit. The soil reference datasets of the best-performing models were split into subsets based on homogeneity of %clay, C and N content. Validation methods were designed to evaluate the change in prediction accuracy of models calibrated from samples either like or dissimilar to those of the test set. The change in RMSEP-values of these models compared to the original models were tabulated.

5.3 Results and discussion

5.3.1 Correlation of soil properties with spectra

Overlaid spectra of the full sample sets used for calibration of models predicting soil pH and exchangeable cations, including replicate scans, are shown in Figure C- 1. Prominent peaks were observed at approximately 800, 1400, 1600, 1900, 2200, 3000, 3600 and 10 000 nm. These results are consistent with soil spectra obtained by other authors (Chang *et al.*, 2001;

Dunn *et al.*, 2002; Van Groenigen *et al.*, 2003; Johnson *et al.*, 2019; Kamrunnahar *et al.*, 2003; Pirie *et al.*, 2005; Viscarra Rossel *et al.*, 2006). The peaks correspond to absorption of IR radiation by molecular components present in soil constituents such as quartz, iron oxyhydroxides, SOM components, carbonates, and clay minerals such as kaolinite, illite and smectite (see Table 2-1). The pre-processed spectra used for calibration of the best-performing models for prediction of each soil chemical property of interest are shown in Figure C- 2. These spectra, resulting from the mathematical manipulations discussed in Section 2.3.2, were divided into wavelength regions to be used as regression components for calibration of the prediction models. Wavelength regions selected by the optimization process (described in Section 5.2.2) for calibration of the overall best-performing models for prediction of each soil chemical property of interest are shown in Table 5-2.

The wavelengths regions that showed good correlation with soil chemical properties are those associated with clay minerals and SOM, as expected (see Section 2.3.3). Examples of calibration and corresponding validation graphs of the best models for prediction of exchangeable cations and soil pH are shown in Figure C- 3 and Figure C- 4. A strong correlation between spectral components and the predicted soil property was observed for most of these models ($r^2 > 0.7$). It is clear from these graphs that there were “gaps” in the reference data for certain ranges of soil chemical properties, especially in Na and Mg prediction models. This may be because most of the soil samples had low salinity levels.

Table 5-2 IR wavelength regions correlated to soil chemical properties for overall best-performing models

Soil chemical property	Reference methodⁱ	Spectroscopy instrument	Wavelength regions (cm⁻¹)
Ca	BaCl ₂	MPA	10715.1-8038.2 6264-3594.8
Mg	AA (0.1M)	Matrix-F	7197.3-5592.8 4798.2-3995.9 9828-8038.2
Na	M-III	MPA	7158.8-6256.3 5376.8-4482
K	M-III	MPA	9828-6256.3 5376.8-4482
pH	CaCl ₂ (1:4)	MPA	11602.2-9820.2 6264-4482

ⁱ BaCl₂ refers to compulsive exchange method

5.3.2 Evaluation of performance of general IR models

Figures of merit for models constructed for each combination of soil chemical property, chemical reference method and spectral measurement instrument are shown in

Table C- 2 to Table C- 4. Predictions of exchangeable K content showed a slight tendency to be positively biased, while Ca, Mg and Na predictions were predominantly negatively biased. Predictions of soil pH were predominantly positively biased. Hence, exchangeable Ca, Mg and Na content tended to be underestimated, and K content as well as soil pH tended to be overestimated. Most of the soil pH, Ca and Mg content prediction models were calibrated by structured sample selection methods, which could well have resulted in biased training sample sets compared to those that would have been created by automatic selection with the Kennard-Stone algorithm.

For each of the three spectroscopic instruments, the two best-performing models for prediction of Ca were calibrated by reference data from the Ambic-I and compulsive exchange extraction methods and had RMSEP-values that were within the acceptable range recommended in Section 4.3.3. This shows good correlation between both Ambic-I and compulsive exchange-extractable Ca with infrared-detectable soil physicochemical properties. For each of the three spectroscopic instruments, only one of the models for prediction of Mg, calibrated by reference data from the AA (1M) extraction method and spectral data generated by the Matrix-F instrument, had an RMSEP within the acceptable range recommended in Section 4.3.3. None of the infrared models for prediction of Na, K and pH had an RMSEP-value within the acceptable ranges recommended in Sections 3.3.3 and 4.3.3. Hence, predictions of Na, K, and pH from IR spectra are not deemed suitably accurate for practical use for the soils used in this study. Despite the lower accuracy of prediction of these properties, several interesting trends were noted. pH measurements in CaCl_2 matrices led to better calibrations and predictions of soil pH by infrared spectral models compared to pH measurements in water and KCl matrices. This may be due to the concentration of the CaCl_2 solution (0.01M) more accurately representing the ionic strength of soils at typical field conditions. Most of the best-performing IR models for prediction of Na and K-content by the different instruments were calibrated by reference data generated by the M-III extraction method. This suggests better correlation of M-III extractable Na and K with IR spectra compared to other extraction methods. These findings should be confirmed by means of statistical comparison tests, which are only possible with replicate reference datasets for each combination of extraction method and spectral measurement method.

Figures of merit for the best-performing IR-based models for prediction of each of the soil properties of interest by each instrument are shown in Table C- 2 to Table C- 4. The best-performing models overall in this study (shown by the shaded values in Table 5-3) were all based on NIR spectral data generated by the MPA and Matrix-F instruments. This shows promising potential application in portable NIR instruments (see Section 2.3.3) for Western Cape soils. It should be noted that the MIR-based spectral models had lower rank values (rank

< 10) compared to the NIR-based models (rank < 20 in most cases). The meaning of converted regressors after the reduction of dimensionality involved in PCA-based regression methods is not easily interpretable (Caffo, 2015). The lower rank of MIR-based models could indicate that fewer MIR-detectable soil properties were correlated to soil pH and exchangeable base cation content compared to NIR-detectable soil properties. Infrared models calibrated by a combination of Vis-NIR-MIR spectra should also be tested for determination of soil pH and exchangeable base cations in Western Cape soils.

The r^2_{val} -values of the best-performing Na prediction models were higher than any found in the previous studies listed in Table C- 1. The r^2_{val} -values of the best-performing models for prediction of Ca, Mg, K, and pH were higher than the mean r^2_{val} -values found in literature (see Section 2.3.3). Therefore, the validated models showed better fit in general than models reported in previous studies. Strong correlation was observed between spectral components and predicted Ca, Mg and Na content for the best-performing models ($r^2_{\text{cal}} > 0.7$). Spectral components were moderately correlated with predicted K content and soil pH ($r^2_{\text{cal}} > 0.5$). The RPIQ-value is a more suitable model performance indicator than the RPD-value for predictions of soil chemical properties (see Section 2.3.3). The RPIQ-values of the best-performing models for prediction of soil pH and exchangeable cation content were higher than the corresponding RPD values. This shows that the true predictive capability of models as reported by RPD-values in previous studies may be under-estimated. The RPD- and RPIQ-values of the best-performing models for prediction of Na content were higher than any reported in the previous studies listed in Table C- 1. The RPIQ-values of the best-performing Ca, Mg, K, and pH prediction models were higher than any reported in previous studies, and the RPD-values were equal or higher than the mean RPD-values found in literature. This shows that in general, the models were of good quality compared to previous studies.

The prediction accuracy of the best-performing models was compared with models in previous studies (listed in Table C- 1) after converting the RMSEP-values reported in literature to the units used in this study. The RMSE-values for Mg and K predictions in this study were lower than any found in literature, and those for Ca, Mg and pH predictions were lower than the mean RMSE-values found in literature. The RMSE-value for Na predictions was not lower than the mean value in previous studies. Thus, predictions of Ca, Mg, K, and pH were more accurate than most of the previous studies listed in Table C- 1. Models for predictions of Na content did not perform as well as in previous studies, possibly due to an uneven distribution of EC_e -values of soils used in this study. Models showing better predictive performance when calibrated from samples of a limited cation concentration range (i.e., after exclusion of outlier calibration values) are shown in Table C- 5 to Table C- 7. Most of these were models predicting Ca and Mg content from reference spectra generated by the MPA-II and Alpha instruments.

This suggests that predictions of Ca and Mg were less accurate for high concentrations of these elements, as observed in Section 4.2.3. None of the models calibrated from samples of a limited pH range out-performed models calibrated from the full pH range.

Table 5-3 Figures of merit for best-performing models predicting soil chemical properties from IR spectra by each instrument. Best-performing models for each property overall (regardless of instrument) are shaded

Property	Method ⁱ	Instrument	R ² cal	RMSEE ⁱⁱ	Rank	R ² val	RMSEP ⁱⁱ	RPD	RPIQ
Ca	Ambic-I	Alpha	0.66	2.1	2	0.73	1.4	2.1	3.7
	Ambic-I	Matrix	0.79	1.6	5	0.79	1.2	2.7	4.2
	BaCl ₂	MPA	0.82	1.5	8	0.84	1.2	3.0	4.4
K	M-III	Alpha	0.50	97	3	0.71	72	1.9	2.7
	Ambic-I	Matrix	0.66	77	6	0.67	66	1.8	2.9
	M-III	MPA	0.68	80	9	0.80	60	2.2	3.2
Mg	BaCl ₂	Alpha	0.51	0.9	2	0.57	0.57	1.7	2.2
	AA (1 M)	Matrix	0.42	1.2	5	0.73	0.37	2.0	3.0
	Ambic-I	MPA	0.35	1.0	4	0.53	0.50	1.6	2.3
Na	BaCl ₂	Alpha	0.58	65	7	0.51	63	1.4	1.7
	M-III	Matrix	0.78	49	16	0.76	48	2.0	1.3
	M-III	MPA	0.88	38	18	0.88	34	2.9	1.8
pH	CaCl ₂ 1: 5	Alpha	0.88	0.2	8	0.73	0.35	1.9	3.1
	CaCl ₂ 1: 4	Matrix	0.53	0.5	4	0.71	0.37	1.8	2.7
	CaCl ₂ 1: 4	MPA	0.51	0.5	5	0.79	0.31	2.2	3.2

ⁱ BaCl₂ refers to compulsive exchange method

ⁱⁱ Units of RMSE in $\text{cmol}_c \text{ kg}^{-1}$ for exchangeable Ca / Mg, mg kg^{-1} for exchangeable Na / K and pH units for pH

5.3.3 Evaluation of performance of IR models calibrated by similar samples

5.3.3 i) Investigative model construction

The effect of soil texture and SOM on model performance was investigated by comparing the accuracy of models validated with global data versus either local or “non-local” data. Validation procedures were designed such that calibration sets consisted of samples with similar properties, while validation sets consisted of samples with properties either similar (local) or dissimilar (non-local) to the calibration subset. These models are henceforth referred to as investigative models, while the original models reported in this study are referred to as global models. The calibration and validation samples were grouped according to similarities in soil texture class and SOM content, and separate prediction models were subsequently constructed. This procedure was based on the principle that differences in the performance of investigative versus global prediction models provide evidence that soil texture class or SOM content differences affect IR absorption patterns in soils.

Reference datasets were divided into groups representing low, moderate, and high %clay content based on the 33rd and 67th quantile values. Four different validation procedures were followed, each testing the predictive capability of models calibrated from pre-selected %clay content-categories. Three of these procedures simulated predictions of non-local samples and one procedure simulated local predictions. The first method involved a 2:1 calibration: validation sample ratio with calibration from low to moderate clay content-samples and validation with high clay content-samples. The second method tested the opposite scenario, with calibration from high clay content-samples and validation with low to moderate clay content-samples. The third method excluded low clay content-samples from calibration as well as validation sets entirely (local predictions). The last method involved calibration from samples classified as clays and loams, but validation was performed using a test set consisting only of sands.

A similar investigation was conducted regarding the effect of SOM content on model performance. Reference datasets were divided into groups representing low, moderate, and high %C content, as well as %N content. Four validation methods were followed, but in this case two methods tested the performance of local predictions and two cases tested predictions of non-local subsets. The first procedure involved a 2:1 calibration: validation sample ratio with calibration from high %C content-samples and validation with low to moderate %C content-samples. The second procedure excluded low %C content-samples from calibration as well as validation sets entirely (local predictions). The same procedure was followed for the last two validation methods but using %N content, instead of %C content, as a subset-divider.

5.3.3 ii) *Performance of local models constructed from reference data subsets*

The change in RMSEP-values of investigative models, compared to the corresponding global models, are shown in Table C- 8 and Table C- 9, where negative values correspond to a lower RMSEP (greater degree of accuracy) of the investigative models compared to the global models. The RMSEP-values of models predicting exchangeable Mg, Na and K content of non-local samples with respect to soil texture and SOM were not always higher than those of the global models. The RMSEP-values of models predicting exchangeable Mg, Na and K content of content of local samples were not always lower than those of the global models. This suggests that soil texture and SOM did not have a great effect on the predictive capability of spectral-based models to determine of exchangeable Mg, Na and K content of the soils used in this study. The RMSEP-values of models predicting exchangeable Ca content and soil pH of local soils (with respect to clay, %C and %N content) were lower than those of global

models. The RMSEP-values of models predicting exchangeable Ca content and soil pH of non-local soils (with respect to clay, %C and %N content) were higher than those of global models.

Therefore, taking the effects of soil texture and SOM content into consideration improved the accuracy of spectral-based models to determine of exchangeable Ca content and soil pH for soils used in this study. This confirms the effects of soil texture and SOM variability on IR model prediction accuracy found by other authors (Araújo *et al.*, 2014; Stenberg *et al.*, 2010). However, IR spectroscopic models for prediction of soil pH, even from local test sets, was not deemed suitably accurate for practical use (see Section 3.2.2). These results should be confirmed by statistical comparison tests which are only possible with replicate datasets of the reference data, as noted in Section 5.3.2. The performance potential of IR prediction models calibrated from data subsets with homogeneous texture class would be better gauged using a database inclusive of more clay samples. Datasets should ideally be chosen such that the geographic location of individual samples are known. This would enable field scale testing of calibrations based on soil sampling locations, or the use of techniques such as cluster analysis for subset selection.

5.4 Conclusions

NIR models were more accurate than MIR models for prediction of soil pH and exchangeable base cation content for the soils used in this study. Infrared spectral models for determination of Ca content calibrated by reference data generated by the Ambic-I and compulsive exchange extraction methods showed better accuracy compared to models calibrated by other reference methods. The higher RPIQ-values relative to RPD-values of IR spectral-based models to determine of soil pH and exchangeable cation content suggest that the true performance of models as reported by RPD-values in soil science literature may be under-estimated. The level of accuracy of IR-based models developed in this study indicate that IR spectroscopy can replace chemical analytic methods for determination of exchangeable Ca and Mg in Western Cape soils. Exchangeable Na and K content, as well as soil pH values, were not accurately predicted by IR-based models in this study. Predictions of exchangeable Ca and soil pH may be further improved by calibrating models with soils of similar texture or SOM content to those of the test set.

6. Conclusions and recommendations

6.1 Conclusions

Interpretation of soil analytical results by farmers, consultants and researchers is usually done from a local frame of reference. Methods of soil chemical analyses vary across regions; thus, the aim of this study was to develop equations for conversion between local and international methods of soil pH and exchangeable base cation content measurement methods. A second aim was to calibrate models for prediction of soil pH and exchangeable base cation content in Western Cape soils, from infrared spectral measurements. Infrared spectroscopy has gained popularity in the soil science community due to its potential as a rapid, cost-effective, and environmentally friendly technique of soil analysis for use in commercial laboratories, *in-situ* measurements, and remote-sensing. The objectives of this study were accomplished by executing research activities in three units, namely soil pH conversions, exchangeable base cation content conversions and IR spectral-based modelling to predict these chemical properties. Soil pH was measured in several solutions (water, 0.1M KCl and 0.01M CaCl₂) using soil: solution ratios of 1:1, 1:2.5, 1:4 and 1:5. The exchangeable base cation content was measured according to standard chemical methods including the standard ammonium acetate (1M, pH = 7), compulsive exchange, ammonium acetate (0.2M, pH = 7), Ambic-I, Mehlich-III, Bray-II and Citric acid (1%) methods. Simple linear regression was used to develop models for conversion of these soil chemical properties between results that would have been obtained by different methods. Multivariate linear regression was investigated as a means of improving the accuracy of predictions by incorporating other soil physicochemical properties as regressors. Three spectroscopic instruments were used to collect spectral data from the soil samples in the NIR to MIR electromagnetic wavelength regions. Partial least squares regression was used to calibrate models for prediction of exchangeable base cation content and soil pH from spectral components. Several combinations of sample selection, spectral data pre-treatment and validation techniques were tested for optimal model performance. Novel approaches to completing the objectives of this project included the following:

- Development of equations for conversion between numerous pH and base cation content measurement methods using a single diverse soil sample set
- Development of more accurate models for predictions of Na content in soils of normal EC_e-range, compared to more robust models for Na content prediction in saline soils
- Incorporation of optimal combinations of soil physicochemical properties as regressors in cation content and soil pH conversion equations using stepwise regression
- Consideration of practical implications using more complex as opposed to simpler models for conversion of soil chemical properties

- Comparison of the accuracy of global versus local models (based on homogeneity of clay and SOM content in calibration and validation subsets) for prediction of chemical properties of Western Cape soils

The most important findings of this study may be summarized as follows:

- Simple linear equations are recommended for soil pH prediction, as they perform better than conversion factors and are more efficient than multivariate models
- K content is convertible between a greater number of extraction methods compared to other exchangeable base cations
- Bray-II extractable Ca and Mg is not interconvertible between other methods
- M-III is recommended for extraction of exchangeable base cations due to high reproducibility and practical efficiency
- The complexity of base cation conversion models chosen for practical use should depend on the purpose of application
- NIR models outperformed MIR models for prediction of soil pH and exchangeable base cation content in this study
- Infrared spectroscopy may replace chemical analytic methods for determination of exchangeable Ca and Mg in Western Cape soils
- Soil pH and exchangeable Ca content may be predicted more accurately by models calibrated from soils with textural and SOM properties like those of the test samples

6.2 Recommendations for future studies

Further studies of IR spectral-based models to predict soil pH and exchangeable base cation content in Western Cape soils should focus on the following aspects:

- Calibrations should be repeated for several soil datasets to enable statistical comparisons of predictive performance of models calibrated from different reference and spectral data
- Performance of IR spectroscopic models calibrated with data from soils of similar geographic origin should be tested at field scale
- The performance of models calibrated from various combinations of Vis-NIR-MIR ranges should be evaluated

References

- Aghamir, F., Hamidi, S.M., Tehranchi, M.M. & Mirzaiee, R. 2019. Combined Application of Imaging Methods for Estimating Soil Physicochemical Properties. *Eurasian Soil Science*. 52(8):926–934.
- Aitken, R.L. & Moody, P.W. 1991. Interrelations between Soil pH Measurements in Various Electrolytes and Soil Solution pH in Acidic Soils. *Soil Chemistry and Mineralogy*. 29:483–491.
- Al-Busaidi, A., Cookson, P. & Yamamoto, T. 2005. Methods of pH determination in calcareous soils: use of electrolytes and suspension effect. *Australian Journal of Soil Research*. 43:541–545.
- Aleixandre-tudo, J.L., Buica, A., Nieuwoudt, H. & Aleixandre, J.L. 2017. Spectrophotometric Analysis of Phenolic Compounds in Grapes and Wines. *Journal of Agricultural and Food Chemistry*.
- Aleixandre-Tudo, J.L., Nieuwoudt, H., Olivieri, A., Aleixandre, J.L. & du Toit, W. 2018. Phenolic profiling of grapes, fermenting samples and wines using UV-Visible spectroscopy with chemometrics. *Food Control*. 85:11–22.
- Aleixandre-Tudo, J.L., Nieuwoudt, H. & du Toit, W. 2019. Towards on-line monitoring of phenolic content in red wine grapes: A feasibility study. *Food Chemistry*. 270:322–331.
- Alexander, D.L.J., Tropsha, A., Winkler, D.A., Flagship, D.P., Flagship, M. & Sciences, P. 2015. HHS Public Access. 55(7):1316–1322.
- Amacher, M.C., Henderson, R.E., Breithaupt, M.D., Seale, C.L. & LaBauve, J.M. 1990. Unbuffered and Buffered Salt Methods for Exchangeable Cations and Effective Cation Exchange Capacity. *Soil Sci. Soc. Am. J.* 54:1036–1042.
- Amrhein, C. & Suarez, D.L. 1990. rocedure for Determining Sodium-Calcium Selectivity in Calcareous and Gypsiferous Soils. *Soil Science Society of America Journal*. 54(4):999–1007.
- Araújo, S.R., Wetterlind, J., Demattê, J.A.M. & Stenberg, B. 2014. Improving the prediction performance of a large tropical vis-NIR spectroscopic soil library from Brazil by clustering into smaller subsets or use of data mining calibration techniques. *European Journal of Soil Science*. 65(5):718–729.
- Awiti, A.O., Walsh, M.G., Shepherd, K.D. & Kinyamario, J. 2008. Soil condition classification using infrared spectroscopy: A proposition for assessment of soil condition along a tropical forest-cropland chronosequence. 143:73–84.
- Bache, B.W. 1976. The measurement of cation exchange capacity of soils. *Journal of the Science of Food and Agriculture*. 27(3):273–280.
- Ballard, R. 1978. Use of the Bray soil test in forestry. Determination of cation status. *N.Z. J. For. Sci.* 8(3):332–343.
- Barbagelata, P.A. & Mallarino, A. 2013. Field Correlation of Potassium Soil Test Methods Based on Dried and Field-Moist Soil Samples for Corn and Soybean. *Soil Sci. Soc. Am. J.* 77:318–327.
- Bellon-Maurel, V. & 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*. 43(7):1398–1410.
- Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.M. & McBratney, A. 2010.

Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *Trends in Analytical Chemistry*. 29(9):1073–1081.

Bemlab. 2016. *Soil Analysis Norms*.

Bilgili, A.V., Es, H.M. Van, Akbas, F., Durak, A. & Hively, W.D. 2010. Visible-near infrared reflectance spectroscopy for assessment of soil properties in a semi-arid area of Turkey. *Journal of Arid Environments*. 74(2):229–238.

Bower, C.A. 1950. Availability of ammonium fixed in a difficulty exchangeable from by soils of semiarid regions. *Soil Sci Soc Am Proc*. 14:119–122.

Brady, N.C. & Weil, R. 2017. *The Nature and Properties of Soils*. 15th ed. Essex: Pearson Education Ltd.

Bray, R.H. & Kurtz, L.T. 1945. Determination of total, organic and available forms of phosphorus in soils. *Soil Sci*. 59:39–45.

Brough, D.M., Jacquier, D., McBratney, A.B. & Minasny, B. 2011. Models relating soil pH measurements in water and calcium chloride that incorporate electrolyte concentration. *European Journal of Soil Science*. 2:728–732.

Burnham, K.P. & Anderson, D.R. 2002. *Model selection and multimodel inference : a practical information-theoretic approach*. 2nd ed. New York: Springer-Verlag.

Bushong, J.T., Norman, R.J. & Slaton, N.A. 2015. Near-Infrared Reflectance Spectroscopy as a Method for Determining Organic Carbon Concentrations in Soil. *Communications in Soil Science and Plant Analysis*. 46(14):1791–1801.

Caffo, B. 2015. *Regression Models for Data Science in R*. Johns Hopkins Bloomberg School of Public Health. [Online], Available: <https://everythingcomputerscience.com/books/regmods.pdf>.

Carter, D.J. 2019. *Biplots and Interpretation*. [Online], Available: <https://bookdown.org/danieljcarter/socepi/biplots-and-interpretation.html> [2020, August 22].

Chang, C., Laird, D. & Mausbach, M.J. 2001. Near-Infrared Reflectance Spectroscopy – Principal Components Regression Analyses of Soil Properties Near-Infrared Reflectance Spectroscopy – Principal Components.

Chodak, M., Khanna, P., Horvath, B. & Beese, F. 2004. Near infrared spectroscopy for determination of total and exchangeable cations in geologically heterogeneous forest soils. *Journal of Near Infrared Spectroscopy*. 12(5):315–324.

Ciesielski, H., Sterckeman, T., Santerne, M. & Willery, J. 1997. A comparison between three methods for the determination of cation exchange capacity and exchangeable cations in soils. *Agronomie*. 17(1):9–16.

Clark, R.N. 1999. Chapter 1: Spectroscopy of Rocks and Minerals, and Principles of Spectroscopy. In A.N. Rencz (ed.). New York: John Wiley and Sons *Manual of Remote Sensing, Volume 3, Remote Sensing for the Earth Sciences*. 3–58.

Clewer, A.G. & Scarisbrick, D.H. 2001. *Practical Statistics and Experimental Crop Design for Plant and Crop Science*. Chichester: John Wiley & Sons, Ltd.

Csathó, P. & Magyar, M. 2005. Correlation Between Soil P and Wheat Shoot P Contents in a Network of Hungarian Long - Term Field Trials. *Communications in Soil Science and Plant Analysis*. 36:275-293.

- Culman, S., Mann, M. & Brown, C. 2021. *Calculating Cation Exchange Capacity, Base Saturation, and Calcium Saturation*. [Online], Available: <https://ohioline.osu.edu/factsheet/anr-81> [2020, September 02].
- Culman, S.W., Mann, M., Sharma, S., Saeed, M.T., Fulford, A.M., Lindsey, L.E., Brooker, A., Dayton, E., et al. 2020. Calibration of Mehlich-3 with Bray P1 and Ammonium Acetate in the Tri-State Region of Ohio, Indiana and Michigan. *Communications in Soil Science and Plant Analysis*. 51(1):86–97.
- Dangal, S., Sanderman, J., Wills, S. & Ramirez-Lopez, L. 2019. Accurate and Precise Prediction of Soil Properties from a Large Mid-Infrared Spectral Library. *Soil Systems*. 3(1):11.
- Department of Soil Science. 2017. *Plant Nutrition and Fertilisation for BSc Agric*. Stellenbosch University: Faculty of AgriSciences.
- Dohrmann, R. 2006. Cation exchange capacity methodology I: An efficient model for the detection of incorrect cation exchange capacity and exchangeable cation results. *Applied Clay Science*. 34(1–4):31–37.
- Du, C. & Zhou, J. 2011. Application of Infrared Photoacoustic Spectroscopy in Soil Analysis. *Applied Spectroscopy Reviews*. 46(5):405-422.
- Dunn, B.W., Beecher, H.G., Batten, G.D. & Ciavarella, S. 2002. The potential of near-infrared reflectance spectroscopy for soil analysis - a case study from the Riverine Plain of south-eastern Australia. *Australian Journal of Experimental Agriculture*. 42:607–614.
- Dyer, B. 1984. On the analytical determinations of probably available “mineral” plant-food in soil. *J. Chem. Soc.* 65:115.
- Ellis, B.G. 1988. *Soil fertility, soil pH and its management*. New York: John Wiley & Sons, Inc.
- Ferguson, R., Libohova, Z., Wills, S., P., O.N., Nesser, R., Thompson, J.A., West, L.T. & Hempel, J.W. 2014. Converting pH 1:1 H₂O and 1:2 CaCl₂ to 1:5 H₂O to contribute to a harmonized global soil database. *Geoderma*. 213:544–550.
- Foth, H.D. 1958. *Fundamentals of Soil Science*. 8th ed. New York: John Wiley & Sons, Inc.
- Frenkel, H. & Suarez, D.L. 1977. Hydrolysis and decomposition of calcium montmorillonite. *Soil Sci. Soc. Am. J.* 41:887–891.
- Galka, B., Kabala, C., Łabuńska, D., Mańczyńska, P. & Muszyńska, E. 2016. Conversion of Soil pH 1:2.5 KCl and 1:2.5 H₂O to 1:5 H₂O: Conclusions for Soil Management, Environmental Monitoring, and International Soil Databases. *Pol. J. Environ. Stud.* 25(2):647–653.
- Gartley, K.L., Sims, J.T., Olsen, C.T. & Chu, P. 2002. Comparison of soil test extractants used in Mid-Atlantic United States. *Communications in Soil Science and Plant Analysis*. 33(5–6):873–895.
- Gartley, K.L., Sims, J.T., Olsen, C.T. & Chu, P. 2007. Comparison of soil test extractants used in mid-Atlantic United States. *Communications in Soil Science and Plant Analysis*. 33:873–895.
- Gillman, G.P. & Sumpter, E.A. 1986. Modification to the compulsive exchange method for measuring exchange characteristics of soils. *Australian Journal of Soil Research*. 24(1):61–66.
- Gillman, G.P., Bruce, R.C., Davey, B.G., Kimble, J.M., Searle, P.L. & Skjemstad, J.O. 1983. A Comparison of Methods Used for Determination of Cation Exchange Capacity. *Commun. Soil Sci. Plant Anal.* 14(11):1005–1014.

- Gobrecht, A., Roger, J.M. & Bellon-Maurel, V. 2014. Major Issues of Diffuse Reflectance NIR Spectroscopy in the Specific Context of Soil Carbon Content Estimation. A Review. *Advances in Agronomy*. 123:145–175.
- Grace-Martin, K. 2020. *Assessing the Fit of Regression Models*. [Online], Available: <https://www.theanalysisfactor.com/assessing-the-fit-of-regression-models/>.
- Gregoriou, V.G. 2000. Fourier Transform infrared spectroscopy of polymers. In C.D. Craver & C.E. Carraher (eds.). Elsevier B.V. *Applied Polymer Science: 21st Century*. 723–724.
- Van Groenigen, J.W., Mutters, C.S., Horwath, W.R. & Van Kessel, C. 2003. NIR and DRIFT-MIR spectrometry of soils for predicting soil and crop parameters in a flooded field. *Plant and Soil*. 250(1):155–165.
- Haghi, R.K., Pérez-Fernández, E. & Robertson, A.H.J. 2021. Prediction of various soil properties for a national spatial dataset of Scottish soils based on four different chemometric approaches: A comparison of near infrared and mid-infrared spectroscopy. *Geoderma*. 396.
- Harris, D.C. 2010. *Quantitative Chemical Analysis*. 8th ed. New York: W. H. Freeman and Company.
- Hendershot, W.H. & Duquette, M. 1986. A Simple Barium Chloride Method for Determining Cation Exchange Capacity and Exchangeable Cations. *Soil Science Society of America Journal*. 50(3):605–608.
- Henderson, B.L. & Bui, E.N. 2002. An improved calibration curve between soil pH measured in water and CaCl₂. *Aust. J. Soil Res.* 40:1399–1405.
- Horn, D.P. 1982. Cation Exchange Capacity Measurements. *Commun. Soil Sci. Plant Anal.* 13(10):851–862.
- Huang, P.M., Li, Y. & Sumner, M.E. Eds. 2012. 13 - Soil Fertility Evaluation. In 2nd ed. CRC Press *Handbook of Soil Sciences - Resource Management and Environmental Impacts*.
- Hunter, A.H. 1974. *Tentative ISFEI extraction procedure*. Raleigh: N.C. State University.
- Illingworth, J. 1981. A common source of error in pH measurement. *Biochemical Journal*. 195(1):259–262.
- Janik, L.J., Merry, R.H. & J.O., S. 1998. Can mid infrared diffuse reflectance analysis replace soil extractions? *Australian Journal of Experimental Agriculture*. 38:681–696.
- Jaremko, D. & Kalembasa, D. 2014. A comparison of methods for the determination of cation exchange capacity of soils. *Ecological Chemistry and Engineering S.* 21(3):487–498.
- Johnson, J.M., Vandamme, E., Senthilkumar, K., Sila, A., Shepherd, K.D. & Saito, K. 2019. Near-infrared, mid-infrared or combined diffuse reflectance spectroscopy for assessing soil fertility in rice fields in sub-Saharan Africa. *Geoderma*. 354.
- Johnston, C.T. & Aochi, Y.O. 1996. Chapter 10 - Fourier Transform Infrared and Raman Spectroscopy. In Madison: Soil Science Society of America, Inc. *Methods of soil analysis Part 3 – Chemical Methods*. 269–321.
- Jolliffe, I.T. & Cadima, J. 2016. Principal component analysis: a review and recent developments. *Phil. Trans. R. Soc. A*. 374.
- Jönsson, U., Rosengren, U., Nihlgård, B. & Thelin, G. 2002. A comparative study of two methods for determination of pH, exchangeable base cations, and aluminum. *Communications in Soil Science and Plant Analysis*. 33(19–20):3809–3824.

- Kamrunnahar, I., McBratney, A. & Balwant, S. 2003. Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared reflectance spectroscopy. *Australian Journal of Soil Research*. 41(6):1101–1114.
- Kaplan, D. 2019. *Stats for Data Science*. [Online], Available: <https://dtkaplan.github.io/SDS-book/mean-square-error.html>.
- Kissel, D.E. & Miller, R.O. 2010. Comparison of Soil pH Methods on Soils of North America. *Soil Sci. Soc. Am. J.* 74(1):310–316.
- Kome, G.K., Enang, R.K., Yerima, B.P.K. & Lontsi, M.G.R. 2018. Models relating soil pH measurements in H₂O, KCl and CaCl₂ for volcanic ash soils of Cameroon. *Geoderma Regional*. 14:1–8.
- Lal, R. 2004. Carbon sequestration in dryland ecosystems. *Environmental Management*. 33(4):528–544.
- van Lierop., W. & Mackenzie, A.F. 1977. Soil pH measurement and its application to organic soils. *Can. J. Soil Sci.* 57:55–64.
- van Lierop, W. 1981. Conversion of organic soil pH values measured in water, 0.01 M CaCl₂ or 1N KCl. *Can. J. Soil. Sci.* 61:577–579.
- Lin, Z.D., Wang, Y.B., Wang, R.J., Wang, L.S., Lu, C.P., Zhang, Z.Y., Song, L.T. & Liu, Y. 2017. Improvements of the Vis-NIRS Model in the Prediction of Soil Organic Matter Content Using Spectral Pretreatments, Sample Selection, and Wavelength Optimization. *Journal of Applied Spectroscopy*. 84(3):529–534.
- Little, I.P. 1992. The relationship between soil pH measurements in calcium chloride and water suspensions. *Australian Journal of Soil Research*. 30(5):587–592.
- Lu, P., Wang, L., Niu, Z., Li, L. & Zhang, W. 2013. Prediction of soil properties using laboratory VIS-NIR spectroscopy and Hyperion imagery. *Journal of Geochemical Exploration*. 132:26–33.
- Ma, F., Du, C.W., Zhou, J.M. & Shen, Y.Z. 2019. Investigation of soil properties using different techniques of mid-infrared spectroscopy. *European Journal of Soil Science*. 70(1):96–106.
- Mamo, T., Richter, C. & Heiligtag, B. 1996. Comparison of extractants for the determination of available phosphorus, potassium, calcium, magnesium and sodium in some Ethiopian and German soils. *Communications in Soil Science and Plant Analysis*. 27(9):2197–2212.
- Marín-González, O., Kuang, B., Quraishi, M.Z., Muñoz-García, M.Á. & Mouazen, A.M. 2013. On-line measurement of soil properties without direct spectral response in near infrared spectral range. *Soil and Tillage Research*. 132:21–29.
- Martins, P.O., Slaton, N.A., Roberts, T.L. & Norman, R.J. 2015. Comparison of Field-Moist and Oven-Dry Soil on Mehlich-3 and Ammonium Acetate Extractable Soil Nutrient Concentrations. *Soil Science Society of America Journal*. 79:1792–1803.
- Mayoral, M., Galvez-Sola, L., Morales, J., Marhuenda-Egea, F.C. & Barber, J.X. 2013. Estimation of parameters in sewage sludge by near-infrared reflectance spectroscopy (NIRS) using several regression tools. *Talanta*. 110:81–88.
- McBride, M.B. 1994. *Environmental Chemistry of Soils*. New York: Oxford University Press.
- McClean, E.O. 1982. *Methods of Soil Analysis. Part 2. Chemical and Microbiological Properties*. Madison: American Society of Agronomy Inc.

- Mehlich, A. 1984. Mehlich 3 soil test extractant: A modification of Mehlich 2 extractant. *Communications in Soil Science and Plant Analysis*. 15(12):1409–1416.
- Mengel, K., Kirkby, E.A., Kosegarten, H. & Appel, T. 2001. The Soil as a Plant Nutrient Medium. In Dordrecht: Kluwer Academic Publishers *Principles of Plant Nutrition*. 15–110.
- Van der Merwe, A.J., Johnson, J.C. & Ras, L.S.K. 1984. An $\text{NH}_4\text{HCO}_3\text{-NH}_4\text{F-(NH}_4\text{)}$, EDTA method for the determination of extractable P, K, Ca, Mg, Cu, Fe, Mn and Zn in soils. In Pretoria.: Soil and Irrigation Research Institute *Soil and Irrigation Research Institute information bulletin B2/2*.
- Michaelson, G.J., Ping, C.L. & Mitchell, G.A. 1987. Correlation of Mehlich 3, Bray 1, and ammonium acetate extractable P, K, Ca, and Mg for Alaska agricultural soils. *Communications in Soil Science and Plant Analysis*. 18(9):1003–1015.
- Miller, L. 2019. *Lime and liming – managing soil health*. [Online], Available: <https://grdc.com.au/resources-and-publications/grdc-update-papers/tab-content/grdc-update-papers/2019/02/lime-and-liming-managing-soil-health>.
- Miller, W.P., Newman, K.D. & Frenkel, H. 1990. Flocculation Concentration and Sodium/Calcium Exchange of Kaolinitic Soil Clays. *Soil Science Society of America Journal*. 54:346–351.
- Minasny, B. & McBratney, A.B. 2013. Why you don ' t need to use RPD. *Pedometron*. (33):14–15.
- Mohamed, E.S., Saleh, A.M., Belal, A.B. & Gad, A.A. 2018. Application of near-infrared reflectance for quantitative assessment of soil properties. *The Egyptian Journal of Remote Sensing and Space Sciences*. 21:1–14.
- Mojid, M.A., Hossain, A.B.M.Z., Wyseure, G.C.L. & Ashraf, M.A. 2019. Pedo-transfer functions with multiple linear regressions to predict solute-transport parameters. *Eurasian Journal of Soil Science*. 8(3):196–207.
- Naes, T., Isaksson, T., Fearn, T. & Davies, T. 2002. *A user-friendly guide to Multivariate Calibration and Classification*. Chichester: NIR Publications.
- Najafi-Ghiri, M., Niazi, M., Khodabakhshi, M., Boostani, H.R. & Owliaie, H.R. 2019. Mechanisms of potassium release from calcareous soils to different salt, organic acid and inorganic acid solutions. *Soil Research*. 57:301–309.
- Ng, W., Minasny, B., Montazerolghaem, M., Padarian, J., Ferguson, R., Bailey, S. & 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*. (352):251–267.
- Ngo, L. 2018. *Principal component analysis explained simply*. [Online], Available: <https://blog.bioturing.com/2018/06/14/principal-component-analysis-explained-simply/> [2020, August 22].
- Nocita, M., Stevens, A., van Wesemael, B., Aitkenhead, M., Bachmann, M., Barthès, B., Dor, E. Ben, Brown, D.J., et al. 2015. Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring. *Advances in Agronomy*. 132:139–159.
- Olivieri, A.C. 2015. Practical guidelines for reporting results in single- and multi-component analytical calibration: A tutorial. *Analytica Chimica Acta*. 868:10–22.
- Pasquini, C. 2018. Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. *Analytica Chimica Acta*. 1026:8–36.
- Peng, Y., Xiong, X., Adhikari, K., Knadel, M. & Grunwald, S. 2015. Modeling Soil Organic

- Carbon at Regional Scale by Combining Multi-Spectral Images with Laboratory Spectra. *PLoS ONE*. 10(11).
- Pirie, A., Singh, B. & Islam, K. 2005. Spectroscopic Techniques To Predict Several Soil Properties. *Australian Journal of Soil Research*. 43:713–721.
- Pleysier, J.L. & Juo, A.S. 1980. A single-extraction method using silver-thiourea for measuring exchangeable cations and effective CEC in soils with variable charges. *Soil Science*. 129(4):205–211.
- Ramirez-Lopez, L. & Stevens, A. 2020. *kenStone: Kennard-Stone algorithm for calibration sampling*. [Online], Available: [https://rdrr.io/cran/prospectr/man/kenStone.html#:~:text=The Kennard–Stone algorithm allows,from the list of points](https://rdrr.io/cran/prospectr/man/kenStone.html#:~:text=The+Kennard+Stone+algorithm+allows,from+the+list+of+points.). [2020, September 28].
- Rayment, G.E. & Lyons, D.J. 2011. *Soil Chemical Methods - Australasia*. Collingwood: CSIRO Publishing.
- Reeves, J.B., McCarty, G.W. & Meisinger, J.J. 1999. Near infrared reflectance spectroscopy for the analysis of agricultural soils. *Journal of Near Infrared Spectroscopy*. 7(3):179–193.
- Ross, D.S. & Ketterings, Q. 2011. Chapter 9. Recommended Methods for Determining Soil Cation Exchange Capacity. In 3rd ed. The Northeast Coordinating Committee for Soil Testing *Recommended Soil Testing Procedures for the Northeastern United States*.
- RStudio Team. 2019. [Online], Available: <http://www.rstudio.com/>.
- Salkind, N.J. 2020. Overfitting. In SAGE Publications *Encyclopedia of Research Design*.
- Sanderman, J., Savage, K. & Dangal, S.R.S. 2020. Mid-infrared spectroscopy for prediction of soil health indicators in the United States. *Soil Science Society of America Journal*. 84(1):251–261.
- Savvides, A., Corstanje, R., Baxter, S.J., Rawlins, B.G. & Lark, R.M. 2010. The relationship between diffuse spectral reflectance of the soil and its cation exchange capacity is scale-dependent. *Geoderma*. 154(3–4):353–358.
- Schmidt, C.J.J., Adriaanse, F.G. & du Preez, C.C. 2004. Relationships between Ambic 1 and Bray 1 extractable phosphorus in some South African arable soils. *South African Journal of Plant and Soil*. 21(2):71–79.
- Schofield, R.K. & Taylor, A.W. 1955. The Measurement of Soil pH. *Soil Sci Soc Am Proc*. 19.
- Schollenberger, C.J. & Dreibelbis, F.R. 1930. Analytical methods in base-exchange investigations in soils. *Soil Sci*. 50:13–24.
- van Schoor, L., Conradie, K. & Raath, P. 2000. Riglyne vir die interpretasie van grondontledingsverslae vir wingerd. *Wineland*. 96–99.
- Šestak, I., Mesić, M., Zgorelec, Ž., Perčin, A. & Stupnišek, I. 2018. Visible and near infrared reflectance spectroscopy for field-scale assessment of Stagnosols properties. *Plant, Soil and Environment*. 64(6):276–282.
- Seybold, C.A., Ferguson, R., Wysocki, D., Bailey, S., Anderson, J., Nester, B., Schoeneberger, P., Wills, S., et al. 2019. Application of Mid-Infrared Spectroscopy in Soil Survey. *Soil Science Society of America Journal*. 83(6):1746–1759.
- Shepherd, K.D. & Walsh, M.G. 2002. Development of reflectance spectral libraries for characterization of soil properties. *Soil Science Society of America Journal*. 66(3):988–998.

- Sila, A.M., Shepherd, K.D. & Pokhariyal, G.P. 2016. Evaluating the utility of mid-infrared spectral subspaces for predicting soil properties. *Chemometrics and Intelligent Laboratory Systems*. 153:92–105.
- Simard, R.R., De Kimpe, C.R. & Zizka, J. 1989. The kinetics of nonexchangeable potassium and magnesium release from Quebec soils. *Can. J. Soil Sci.* 69:663–675.
- Soil Survey Laboratory Staff. 1992. Soil Surv. Invest. Repts. 42. In Washington, DC: USDA-SCS *Soil Survey Laboratory Reference Manual*.
- Sonmez, S., Buyuktas, D., Okturen, F. & Citak, S. 2008. Assessment of different soil to water ratios (1:1, 1:2.5, 1:5) in soil salinity studies. *Geoderma*. 144(1–2):361–369.
- Soriano-Disla, J.M., Janik, L.J., Viscarra Rossel, R.A., MacDonald, L.M. & 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*. 49(2):139–186.
- Stenberg, B., Rossel, R.A.V. & Mouazen, A.M. 2010. Visible and near infrared spectroscopy in soil science. *Advances in Agronomy*. 107(10):1–44.
- Sumner, M.E. & Miller, W.P. 1996. Chapter 40: Cation Exchange Capacity and Exchange Coefficients. In D.L. Sparks, A.L. Page, P.A. Helmke, R.H. Loeppert, P.N. Soltanpour, M.A. Tabatabai, C.T. Johnston, M.E. Sumner, J.M. Bartels, & J.M. Bigham (eds.). Madison: Soil Science Society of America, Inc. *Methods of soil analysis Part 3– Chemical Methods*. 1201–1229.
- The Non-Affiliated Soil Analysis Working Committee. 1990. *Handbook of Standard Soil Testing Methods for Advisory Purposes*. Pretoria: Soil Science Society of South Africa.
- Thomas, G.W. 1982. *Methods of Soil Analysis, Part 2. Chemical and Microbiological Properties – Agronomy Monograph no. 9*. 2nd ed. Madison: ASA-SSSA.
- Thomas, G.W. 1996. Soil pH and soil acidity. In 3rd ed. D.L. Sparks, A.L. Page, P.A. Helmke, R.H. Loeppert, P.N. Soltanpour, M.A. Tabatabai, C.T. Johnston, & M.E. Sumner (eds.). Madison: American Society of Agronomy Inc. *Methods of Soil Analysis: Part 3– Chemical and Microbiological Properties*. 475–490.
- Thompson, G.R. 1995. A comparison of methods used for the extraction of K in soils of the Western Cape. *South African Journal of Plant and Soil*. 12(1):20–26.
- Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J. & Skjemstad, J.O. 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma*. 131(1–2):59–75.
- Viscarra Rossel, R.A., Behrens, T., Ben-Dor, E., Brown, D.J., Demattê, J.A.M., Shepherd, K.D., Shi, Z., Stenberg, B., et al. 2016. A global spectral library to characterize the world's soil. *Earth-Science Reviews*. (155):198–230.
- Wang, J.J., Harrell, D.L., Henderson, R.E. & Bell, P.F. 2004. Comparison of Soil-Test Extractants for Phosphorus, Potassium, Calcium, Magnesium, Sodium, Zinc, Copper, Manganese, and Iron in Louisiana Soils. *Communications in Soil Science and Plant Analysis*. 35:145–160.
- Wetzel, D.L. 1983. Near-Infrared Reflectance Analysis. *Analytical Chemistry*. 55(12):1165–1176.
- Workman, J.J. 1996. Interpretive spectroscopy for near infrared. *Appl. Spec. Rev.* 31:251–320.
- Xu, S., Shi, X., Wang, M. & Zhao, Y. 2016. Effects of subsetting by parent materials on

prediction of soil organic matter content in a hilly area using vis-NIR spectroscopy. *PLoS ONE*. 11(3):1–17.

Zapata, F. & Roy, R.N. Eds. 2004. Soil testing for phosphate rock application. In Rome: FAO Land and Water Development Division *Use of Phosphate Rocks for Sustainable Agriculture*. [Online], Available: <http://www.fao.org/3/y5053e/y5053e00.htm#Contents>.

Zhu, Q., Ozores-Hampton, M. & Li, Y. 2016. Comparison of Mehlich-3 and Ammonium Bicarbonate-DTPA for the Extraction of Phosphorus and Potassium in Calcareous Soils from Florida. *Communications in Soil Science and Plant Analysis*. 47(20):2315–2324.

Appendix A: Soil pH conversion appendices

Table A- 1 Maximum coefficient of variation of repeated pH measures

Matrix	Max CV (%)
KCl 1:1	2.1
KCl 1:2.5	1.8
KCl 1:4	2.8
KCl 1:5	1.1
H ₂ O 1:1	0.5
H ₂ O 1:2.5	2.1
H ₂ O 1:4	0.9
H ₂ O 1:5	0.9
CaCl ₂ 1:1	1.7
CaCl ₂ 1:2.5	2.9
CaCl ₂ 1:4	5.8
CaCl ₂ 1:5	1.1

Table A- 2 Difference between AIC-valuesⁱ of simple and corresponding multiple linear regression models for soil pH conversion including ECe variable

X	Y	Water 1:2.5	KCl 1:2.5	CaCl ₂ 1:2.5	CaCl ₂ 1:4	CaCl ₂ 1:5	Water 1:5	KCl 1:5
Water 1:1		1.1	0.40	2.0	1.9	2.0	1.4	1.2
Water 1:2.5			1.3	1.7	1.5	1.9	2.0	1.8
KCl 1:2.5				1.5	0.70	1.7	2.0	1.7
CaCl ₂ 1:2.5					2.0	1.8	0.80	0.50
CaCl ₂ 1:4						1.4	0.20	-0.35
CaCl ₂ 1:5							1.0	0.70
Water 1:5								1.6

ⁱ AIC MLR - AIC SLR

Table A- 3 Difference between AICⁱ- & RMSEⁱⁱ-values of simple and corresponding multiple linear regression models for soil pH conversion including %clay as variable

X	Y	Water 1:2.5	KCl 1:2.5	CaCl ₂ 1:2.5	CaCl ₂ 1:4	CaCl ₂ 1:5	Water 1:5	KCl 1:5
Water 1:1	ΔAIC	-6.5	-13	0.01	1.9	1.8	1.9	-14
	$\Delta RMSE$	-0.01	-0.030	0.00	0.00	0.00	0.00	-0.03
Water 1:2.5	ΔAIC		-2.4	-6.4	0.59	-2.3	-1.5	-2.1
	$\Delta RMSE$		-0.01	0.00	0.00	-0.01	0.00	-0.01
KCl 1:2.5	ΔAIC			-13	-5.4	-8.4	-2.6	2.0
	$\Delta RMSE$			-0.030	-0.01	-0.02	-0.01	0.00
CaCl₂ 1:2.5	ΔAIC				-0.49	1.6	-0.38	-19
	$\Delta RMSE$				0.00	0.00	0.00	-0.05
CaCl₂ 1:4	ΔAIC					1.9	1.8	-8.6
	$\Delta RMSE$					0.00	0.00	-0.02
CaCl₂ 1:5	ΔAIC						0.78	-11
	$\Delta RMSE$						0.00	-0.03
Water 1:5	ΔAIC							-7.2
	$\Delta RMSE$							-0.02

ⁱ AIC MLR - AIC SLR, denoted by ΔAIC ⁱⁱ RMSE MLR - RMSE SLR, denoted by $\Delta RMSE$ Table A- 4 Difference between RMSEⁱ-values of simple and corresponding multiple linear regression models for soil pH conversion including parameters selected by stepwise regression

X	Y	Water 1:2.5	KCl 1:2.5	CaCl ₂ 1:2.5	CaCl ₂ 1:4	CaCl ₂ 1:5	Water 1:5	KCl 1:5
Water 1:1		-0.01	-0.06	-0.01	-0.01	0.00	0.00	-0.05
Water 1:2.5			-0.04	0.00	0.00	-0.01	0.00	-0.03
KCl 1:2.5				-0.03	-0.01	-0.04	-0.03	0.00
CaCl₂ 1:2.5					0.00	0.00	-0.01	-0.05
CaCl₂ 1:4						0.00	-0.01	-0.02
CaCl₂ 1:5							0.00	-0.05
Water 1:5								-0.05

ⁱ RMSE MLR - RMSE SLR

Appendix B: Exchangeable base cation conversion appendices

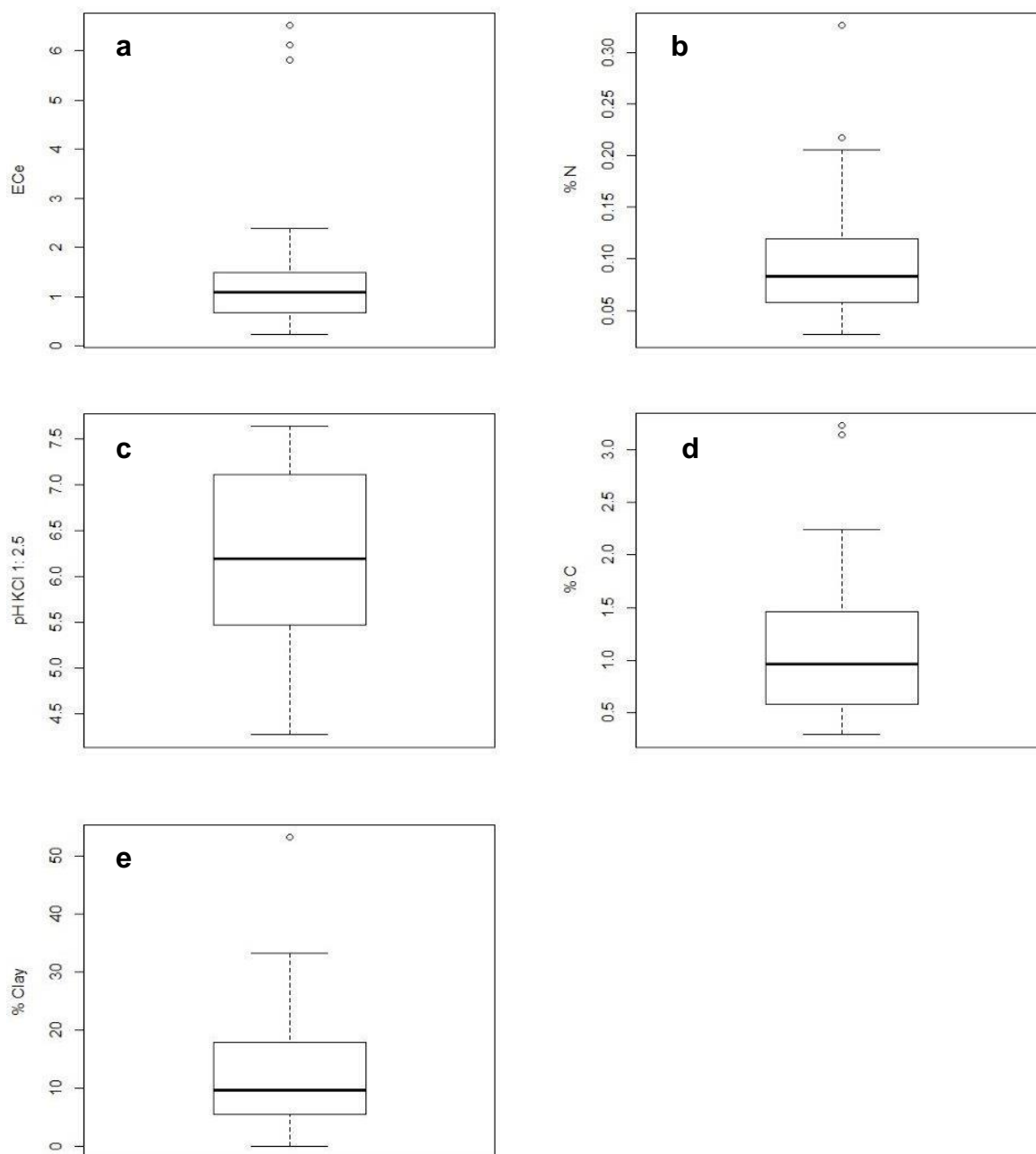


Figure B- 1 Boxplots showing distribution of ECe (a), %N (b), pH_{KCl} (c), %C (d) and %Clay (e) of soils used for comparison of cation extraction methods. Boxes show first and third quartiles, with median line in bold. Whiskers show minimum and maximum values.

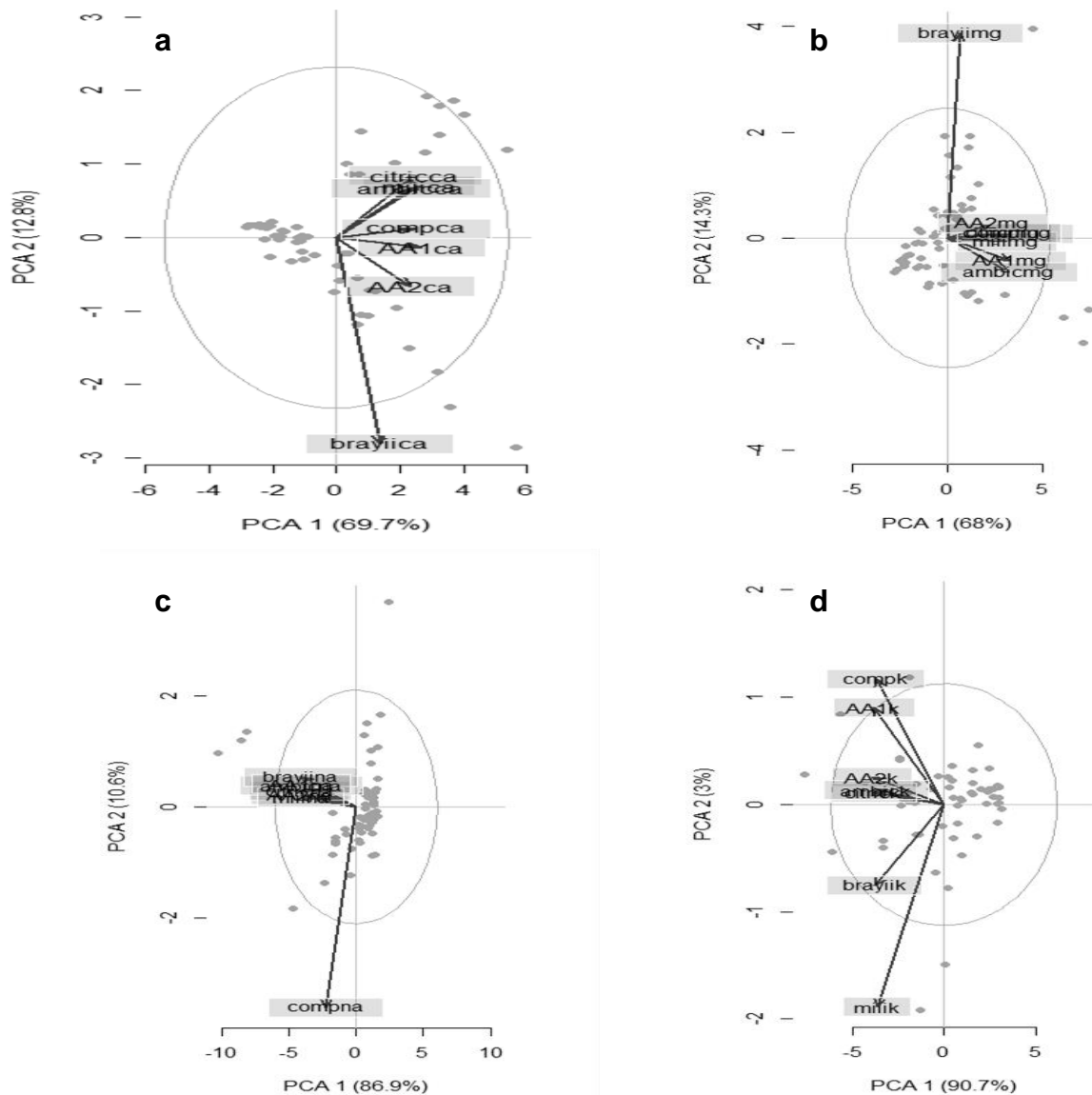


Figure B- 2 PCA biplots with 95% confidence ellipse for quantities of Ca (a), Mg (b), Na (c) and K (d) extracted by various methodsⁱ (represented as vectors). PCA 1 and 2 (vertical and horizontal axes) represent new components resulting from conversion of the original explanatory variables (dimensional reduction)

ⁱAA1, AA2, citric, comp, ambic and miii represent the ammonium acetate (1M and 0.02 M), citric acid (1%), compulsive exchange, Ambic-I and M-III methods respectively

Table B- 1 Difference between RMSEⁱ-values of simple linear regression models calibrated from datasets including versus excluding outlier samples, for prediction of exchangeable Ca content (cmol_c kg⁻¹)

X	Y	AA (0.2M)	Citric acid (1%)	M-III	BaCl ₂ ⁱⁱ	Ambic-I
AA (1M)		0.41	-0.10	0.03	0.11	0.28
AA (0.2M)			0.08	0.44	-0.01	0.52
Citric acid (1%)				0.11	0.04	0.33
M-III					0.22	0.18
BaCl ₂ ⁱⁱ						0.79

ⁱ RMSE (original model) - RMSE (model excluding outliers)

ⁱⁱ Compulsive exchange extraction method

Table B- 2 Difference between RMSEⁱ-values of simple linear regression models calibrated from datasets including versus excluding outlier samples, for prediction of exchangeable Mg content (cmol_c kg⁻¹)

X	Y	AA (0.2M)	Citric acid (1%)	M-III	BaCl ₂ ⁱⁱ	Ambic-I
AA (1M)		0.10	0.04	0.01	0.09	0.04
AA (0.2M)			-0.01	0.06	0.19	0.05
Citric acid (1%)				0.13	0.04	0.15
M-III					0.17	0.04
BaCl ₂ ⁱⁱ						0.01

ⁱ RMSE (original model) - RMSE (model excluding outliers)

ⁱⁱ Compulsive exchange extraction method

Table B- 3 Difference between RMSEⁱ-values of simple linear regression models calibrated from datasets including versus excluding outlier samples, for prediction of exchangeable Na content (mg kg⁻¹)

X	Y	AA (0.2M)	Citric acid (1%)	M-III	BaCl ₂ ⁱⁱ	Ambic-I	Bray-II
AA (1M)		4.9	2.1	2.3	23	5.5	6.5
AA (0.2M)			3.6	2.4	29	1.3	3.2
Citric acid (1%)				0.35	24	6.9	7.4
M-III					25	5.6	7.6
BaCl ₂ ⁱⁱ						56	71
Ambic-I							4.6

ⁱ RMSE (original model) - RMSE (model excluding outliers)

ⁱⁱ Compulsive exchange extraction method

Table B- 4 Difference between RMSEⁱ-values of simple linear regression models calibrated from datasets including versus excluding outlier samples, for prediction of exchangeable K content (mg kg⁻¹)

X	Y	AA (0.2M)	Citric acid (1%)	M-III	BaCl ₂ ⁱⁱ	Ambic-I	Bray-II
AA (1M)		1.0	2.1	19	5.5	3.3	15
AA (0.2M)			0.6	18	11	1.3	15
Citric acid (1%)				17	9.0	-0.69	10
M-III					19	8.8	17
BaCl ₂ ⁱⁱ						4.0	14
Ambic-I							12

ⁱ RMSE (original model) - RMSE (model excluding outliers)

ⁱⁱ Compulsive exchange extraction method

Table B- 5 Summary of soil physicochemical properties of Ca data subset (N = 55)

	C	N	Sand	Silt	Clay	pH_{KCl 1:2.5}	ECe (dS m⁻¹)
	%						
Min	0.3	0.0	5.4	2.0	0.0	4.3	0.2
Max	3.1	0.3	94	84	53	7.6	6.5
Median	0.3	0.0	5.4	2.0	9.8	6.1	1.0
Std dev	0.6	0.1	22	19	10	0.9	1.3
CV (%)	56	59	40	58	82	15	97

Table B- 6 Summary of soil physicochemical properties of Mg data subset (N = 55)

	C	N	Sand	Silt	Clay	pH_{KCl 1:2.5}	ECe (dS m⁻¹)
	%						
Min	0.3	0.0	5.0	5.4	1.2	4.3	0.2
Max	3.2	0.3	46	97	84	7.6	6.5
Median	1.0	0.1	12	53	27	6.1	1.0
Std dev	0.7	0.1	5.6	21	18	0.9	1.3
CV (%)	61	60	47	38	58	15	98

Table B- 7 Summary of soil physicochemical properties of Na data subset (N = 52)

	C	N	Sand	Silt	Clay	pH_{KCl 1:2.5}	ECe (dS m⁻¹)
	%						
Min	0.0	0.3	12	1.2	0.0	4.3	0.2
Max	0.3	3.2	97	84	53	7.6	2.2
Median	0.1	1.0	53	26	9.7	6.1	1.0
Std dev	0.1	0.7	21	18	10	0.9	0.5
CV (%)	60	61	37	60	86	15	51

Table B- 8 Summary of soil physicochemical properties of K data subset (N = 54)

	C	N	Sand	Silt	Clay	pH_{KCl 1:2.5}	ECe (dS m⁻¹)
	%						
Min	0.3	0.0	12	1.2	0.0	4.3	0.2
Max	3.2	0.3	97	84	53	7.6	6.5
Median	0.9	0.1	53	27	9.7	6.2	1.1
Std dev	0.7	0.1	22	19	11	0.9	1.3
CV (%)	61	61	38	60	86	15	98

Figure B- 3 Distribution of CV (%) replicate measurements of Ca (a), Mg (b), Na (c) and K (d) extracted by different methods. Boxes show first and third quartiles, with median line in bold. Whiskers show minimum and maximum values.

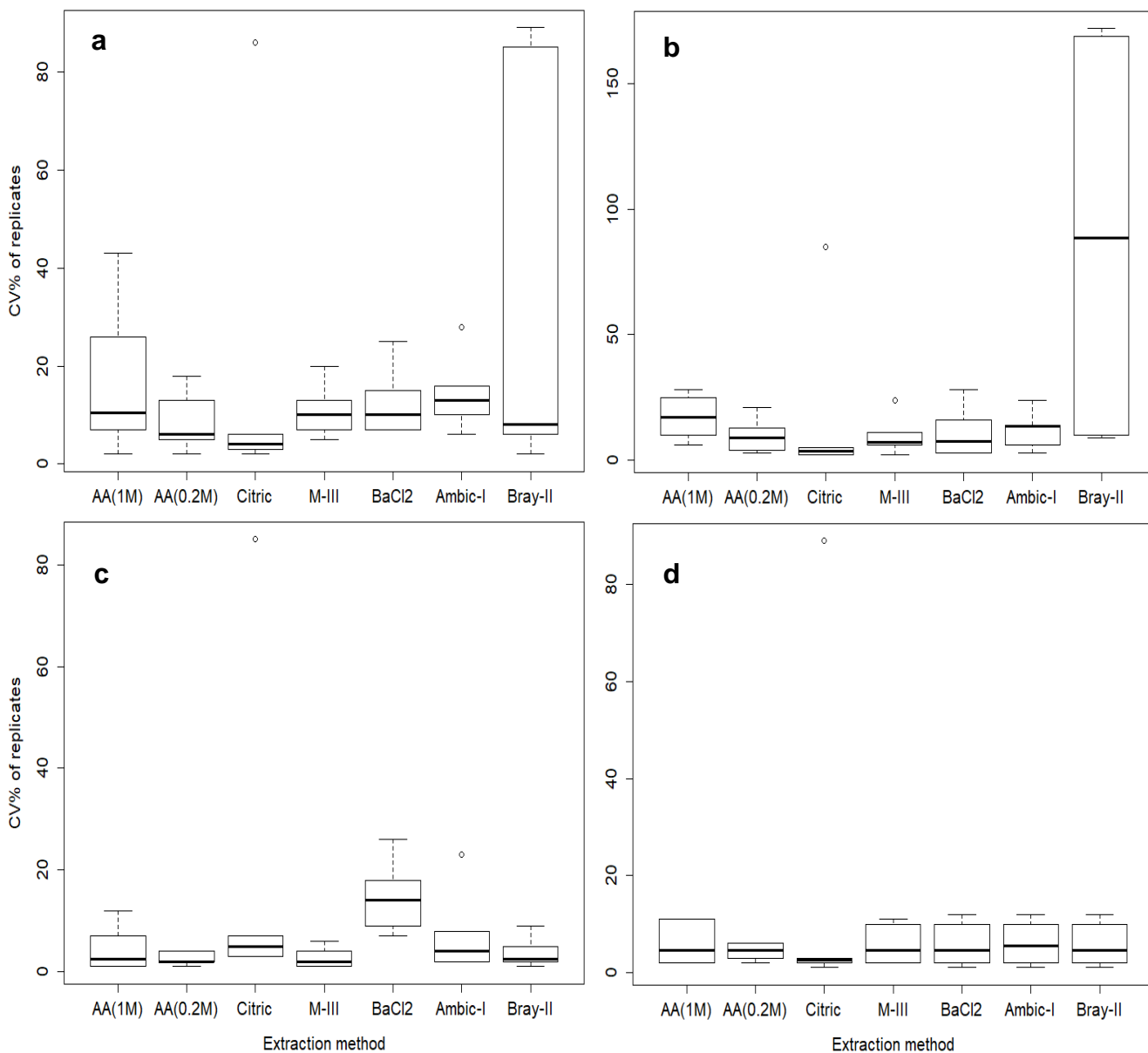


Table B- 9 Pearson correlation coefficient (r) values showing correlation between methods for extraction of Ca

	AA (0.2M)	Citric acid (1%)	M-III	BaCl₂ⁱ	Ambic-I	Bray-II
AA (1M)	0.81	0.76	0.81	0.80	0.75	0.55
AA (0.2M)	1.0	0.59	0.72	0.68	0.60	0.62
Citric acid (1%)	0.59	1.0	0.88	0.69	0.65	0.31
M-III	0.72	0.88	1.0	0.80	0.73	0.38
BaCl₂ⁱ	0.68	0.69	0.80	1.0	0.84	0.42
Ambic-I	0.60	0.65	0.73	0.84	1.0	0.39

ⁱ Compulsive exchange extraction method

Table B- 10 Pearson correlation coefficient (*r*) values showing correlation between methods for extraction of Mg

	AA (0.2M)	Citric acid (1%)	M-III	BaCl₂ⁱ	Ambic-I	Bray-II
AA (1M)	0.31	0.86	0.84	0.82	0.91	0.22
AA (0.2M)	1.0	0.16	0.22	0.66	0.41	0.34
Citric acid (1%)	0.16	1.0	0.87	0.64	0.70	0.27
M-III	0.22	0.87	1.0	0.61	0.73	0.31
BaCl₂ⁱ	0.66	0.64	0.61	1.0	0.87	0.38
Ambic-I	0.41	0.70	0.73	0.87	1.0	0.19

ⁱ Compulsive exchange extraction method

Table B- 11 Pearson correlation coefficient (*r*) values showing correlation between methods for extraction of Na in soils within normal ECe range

	AA (0.2M)	Citric acid (1%)	M-III	BaCl₂ⁱ	Ambic-I	Bray-II
AA (1M)	0.76	0.71	0.64	0.50	0.74	0.70
AA (0.2M)	1.0	0.91	0.90	0.67	0.89	0.87
Citric acid (1%)	0.91	1.0	0.85	0.55	0.87	0.88
M-III	0.90	0.85	1.0	0.60	0.85	0.80
BaCl₂ⁱ	0.67	0.55	0.60	1.0	0.55	0.58
Ambic-I	0.89	0.87	0.85	0.55	1.0	0.85

ⁱ Compulsive exchange extraction method

Table B- 12 Pearson correlation coefficient (*r*) values showing correlation between methods for extraction of Na in saline soils

	AA (0.2M)	Citric acid (1%)	M-III	BaCl₂ⁱ	Ambic-I	Bray-II
AA (1M)	0.95	0.95	0.93	0.45	0.95	0.95
AA (0.2M)	1.0	0.98	0.98	0.50	0.98	0.98
Citric acid (1%)	0.98	1.0	0.97	0.48	0.97	0.97
M-III	0.98	0.97	1.0	0.50	0.97	0.96
BaCl₂ⁱ	0.50	0.48	0.50	1.0	0.46	0.43
Ambic-I	0.98	0.97	0.97	0.46	1.0	0.97

ⁱ Compulsive exchange extraction method

Table B- 13 Pearson correlation coefficient (*r*) values showing correlation between methods for extraction of K

	AA (0.2M)	Citric acid (1%)	M-III	BaCl₂ⁱ	Ambic-I	Bray-II
AA (1M)	0.94	0.91	0.90	0.91	0.88	0.92
AA (0.2M)	1.0	0.95	0.94	0.93	0.91	0.96
Citric acid (1%)	0.95	1.0	0.93	0.91	0.88	0.93
M-III	0.94	0.93	1.0	0.89	0.86	0.93
BaCl₂ⁱ	0.93	0.91	0.89	1.0	0.79	0.89
Ambic-I	0.91	0.88	0.86	0.79	1.0	0.88

ⁱ Compulsive exchange extraction method

Table B- 14 Pearson correlation coefficient (*r*) values showing correlation between soil properties and Ca extracted by different methods

	%N	%C	%clay	Ec _e	pH _{KCl 1:2.5}
AA (0.1M)	0.21	0.20	-0.10	0.17	0.69
AA (0.2M)	0.09	0.07	0.06	0.04	0.52
Citric acid (1%)	0.17	0.29	-0.21	0.17	0.65
M-III	0.20	0.23	-0.12	0.10	0.69
BaCl₂ⁱ	0.45	0.44	-0.09	-0.06	0.64
Ambic-I	0.29	0.29	-0.15	-0.06	0.62

ⁱ Compulsive exchange extraction method

Table B- 15 Pearson correlation coefficient (*r*) values showing correlation between soil properties and Mg extracted by different methods

	%N	%C	%clay	Ec _e	pH _{KCl 1:2.5}
AA (0.1M)	0.14	-0.01	0.33	0.30	0.22
AA (0.2M)	0.25	0.00	0.32	-0.29	-0.02
Citric acid (1%)	0.12	0.13	0.03	0.41	0.54
M-III	0.14	0.07	0.21	0.36	0.32
BaCl₂ⁱ	0.29	0.02	0.37	-0.03	0.14
Ambic-I	0.18	-0.04	0.39	0.20	0.08

ⁱ Compulsive exchange extraction method

Table B- 16 Pearson correlation coefficient (*r*) values showing correlation between soil properties and Na extracted by different methods in soils within normal EC_e range

	%N	%C	%clay	Ec _e	pH _{KCl 1:2.5}
AA (0.1M)	0.11	-0.01	-0.08	0.47	-0.03
AA (0.2M)	0.29	0.08	0.09	0.47	-0.11
Citric acid (1%)	0.26	0.25	0.00	0.53	0.03
M-III	0.43	0.28	0.13	0.50	-0.03
BaCl₂ⁱ	0.24	0.01	-0.01	0.38	-0.06
Ambic-I	0.21	0.07	0.04	0.50	0.08
Bray-II	0.24	0.13	0.07	0.49	0.04

ⁱ Compulsive exchange extraction method

Table B- 17 Pearson correlation coefficient (*r*) values showing correlation between soil properties and Na extracted by different methods in saline soils

	%N	%C	%clay	Ec _e	pH _{KCl 1:2.5}
AA (0.1M)	-0.14	-0.16	-0.06	0.90	0.15
AA (0.2M)	-0.07	-0.12	0.01	0.90	0.13
Citric acid (1%)	-0.07	-0.05	-0.03	0.91	0.16
M-III	0.00	-0.04	0.03	0.90	0.14
BaCl₂ⁱ	0.19	0.01	0.04	0.34	-0.10
Ambic-I	-0.10	-0.12	0.00	0.90	0.20
Bray-II	-0.12	-0.12	-0.01	0.92	0.19

ⁱ Compulsive exchange extraction method

Table B- 18 Pearson correlation coefficient (*r*) values correlation between soil properties and K extracted by different methods

	%N	%C	%clay	Ec _e	pH _{KCl 1:2.5}
AA (0.1M)	0.39	0.18	0.04	-0.05	0.34
AA (0.2M)	0.40	0.22	-0.05	-0.11	0.41
Citric acid (1%)	0.47	0.31	-0.13	-0.09	0.45
M-III	0.38	0.20	-0.07	-0.09	0.46
BaCl₂ⁱ	0.46	0.25	-0.03	-0.10	0.32
Ambic-I	0.32	0.14	0.01	-0.08	0.42
Bray-II	0.31	0.15	-0.12	0.01	0.50

ⁱ Compulsive exchange extraction method

Table B- 19 Difference between AIC-valuesⁱ of SLR and corresponding MLR models including pH as regressor for conversion of Ca content

X	Y	AA (0.2M)	Citric acid (1%)	M-III	BaCl ₂ ⁱⁱ	Ambic-I
AA (1M)		1.5	-1.8	-3.4	-0.27	-0.46
AA (0.2M)			-13	-16	-11	-10
Citric acid (1%)				-4.4	-5.1	-4.9
M-III					-0.29	-1.1
BaCl₂ⁱⁱ						-0.07

ⁱ AIC MLR - AIC SLR

ⁱⁱ Compulsive exchange extraction method

Table B- 20 Difference between AIC-valuesⁱ of SLR and corresponding MLR models including pH as regressor for conversion of Mg content

X	Y	AA (0.2M)	Citric acid (1%)	M-III	BaCl ₂ ⁱⁱ	Ambic-I
AA (1M)		1.4	-35	-1.7	1.7	-3.4
AA (0.2M)			-17	-4.6	-0.36	1.5
Citric acid (1%)				-5.2	-3.8	-13
M-III					1.6	-1.4
BaCl ₂ ⁱⁱ						1.6

ⁱ AIC MLR - AIC SLR

ⁱⁱ Compulsive exchange extraction method

Table B- 21 Difference between AIC-valuesⁱ of SLR and corresponding MLR models including ECe as regressor for conversion of Na content for soils within normal ECe range

X	Y	AA (0.2M)	Citric acid (1%)	M-III	BaCl ₂ ⁱⁱ	Ambic-I	Bray-II
AA (1M)		0.01	-3.5	-2.8	0.06	-1.6	-1.4
AA (0.2M)			-2.4	-0.33	1.5	-0.46	0.15
Citric acid (1%)				1.4	1.2	1.5	1.9
M-III					1.3	0.52	0.52
BaCl ₂ ⁱⁱ						-6.1	-4.9
Ambic-I							1.1

ⁱ AIC MLR - AIC SLR

ⁱⁱ Compulsive exchange extraction method

Table B- 22 Difference between AIC-valuesⁱ of SLR and corresponding MLR models including ECe as regressor for conversion of Na content for saline soils

X	Y	AA (0.2M)	Citric acid (1%)	M-III	BaCl ₂ ⁱⁱ	Ambic-I	Bray-II
AA (1M)		-4.4	-8.4	-6.8	0.37	-4.6	-13
AA (0.2M)			-3.4	-0.31	-3.6	0.26	-9.6
Citric acid (1%)				0.68	-2.3	-0.38	-6.9
M-III					-3.3	-2.5	-12
BaCl ₂ ⁱⁱ						-88	-100
Ambic-I							-11

ⁱ AIC MLR - AIC SLR

ⁱⁱ Compulsive exchange extraction method

Table B- 23 Difference between AIC-valuesⁱ of SLR and corresponding MLR models including pH as regressor for conversion of K content

X	Y	AA (0.2M)	Citric acid (1%)	M-III	BaCl ₂ ⁱⁱ	Ambic-I	Bray-II
AA (1M)		-2.1	-4.9	-5.6	2.0	-2.2	-15
AA (0.2M)			-0.3	-0.9	0.2	1.0	-8
Citric acid (1%)				1.2	-1	1.7	-1.7
M-III					-0.3	1.7	-1.2
BaCl ₂ ⁱⁱ						-2.8	-14
Ambic-I							-3

ⁱ AIC MLR - AIC SLR

ⁱⁱ Compulsive exchange extraction method

Table B- 24 Difference between RMSE-values of SLR and corresponding MLR models including soil pHⁱ and various combinations of soil propertiesⁱⁱ as regressors for conversion of exchangeable Ca content

X	Y		AA (0.2M)	Citric acid (1%)	M-III	BaCl ₂ ⁱⁱⁱ	Ambic-I
AA (1M)		$\Delta RMSE_{MLR}$	-0.01	-0.22	-0.01	-0.04	-0.04
		$\Delta RMSE_{STEP}$	-0.06	-0.85	1.1	-0.56	-0.25
AA (0.2M)		$\Delta RMSE_{MLR}$		-1.1	-0.01	-0.26	-0.25
		$\Delta RMSE_{STEP}$		-1.8	1.1	-0.92	-0.50
Citric acid (1%)		$\Delta RMSE_{MLR}$			-0.01	-0.15	-0.13
		$\Delta RMSE_{STEP}$			0.38	-0.74	-0.39
M-III		$\Delta RMSE_{MLR}$				-0.04	-0.06
		$\Delta RMSE_{STEP}$				-0.51	-0.23
BaCl ₂ ⁱⁱⁱ		$\Delta RMSE_{MLR}$					-0.03
		$\Delta RMSE_{STEP}$					-0.03

ⁱ RMSE MLR - RMSE SLR of models including pH_{KCl 1:2.5}, denoted by $\Delta RMSE_{MLR}$

ⁱⁱ RMSE MLR - RMSE SLR of models including different combinations of soil properties as regressors, denoted by $\Delta RMSE_{STEP}$

ⁱⁱⁱ Compulsive exchange extraction method

Table B- 25 Difference between RMSE-values of SLR and corresponding MLR models including soil pHⁱ and various combinations of soil propertiesⁱⁱ as regressors for conversion of exchangeable Mg content

X	Y		AA (0.2M)	Citric acid (1%)	M-III	BaCl ₂ ⁱⁱⁱ	Ambic-I
AA (1M)		$\Delta RMSE_{MLR}$	-0.00	-0.26	-0.02	-0.09	-0.02
		$\Delta RMSE_{STEP}$	-0.13	-0.34	-0.02	-0.26	-0.02
AA (0.2M)		$\Delta RMSE_{MLR}$		-0.29	-0.07	0.14	-0.00
		$\Delta RMSE_{STEP}$		-0.46	-0.23	0.14	-0.11
Citric acid (1%)		$\Delta RMSE_{MLR}$			-0.04	0.17	-0.08
		$\Delta RMSE_{STEP}$			-0.06	0.08	-0.12
M-III		$\Delta RMSE_{MLR}$			-0.04	0.20	-0.02
		$\Delta RMSE_{STEP}$			-0.06	0.20	-0.04
BaCl ₂ ⁱⁱⁱ		$\Delta RMSE_{MLR}$					-0.00
		$\Delta RMSE_{STEP}$					-0.06

ⁱ RMSE MLR - RMSE SLR of models including pH_{KCl 1:2.5}, denoted by $\Delta RMSE_{MLR}$

ⁱⁱ RMSE MLR - RMSE SLR of models including different combinations of soil properties as regressors, denoted by $\Delta RMSE_{STEP}$

ⁱⁱⁱ Compulsive exchange extraction method

Table B- 26 Difference between RMSE-values of SLR and corresponding MLR models including soil EC_e^i and various combinations of soil propertiesⁱⁱ as regressors for conversion of exchangeable Na content for soil within normal EC_e range

X	Y		AA (0.2M)	Citric acid (1%)	M-III	BaCl ₂ ⁱⁱⁱ	Ambic-I	Bray-II
AA (1M)		$\Delta RMSE_{MLR}$	-0.71	-2.2	-1.5	-1.1	-0.88	-0.99
		$\Delta RMSE_{STEP}$	-4.7	-4.8	-6.3	-5.0	-1.7	-2.2
AA (0.2M)		$\Delta RMSE_{MLR}$		-1.1	-	-0.28	-0.41	-0.38
		$\Delta RMSE_{STEP}$		-7.4	0.42	-0.00	-1.3	-0.82
Citric acid (1%)		$\Delta RMSE_{MLR}$			-	-0.48	-0.09	-0.03
		$\Delta RMSE_{STEP}$			0.14	-0.00	-2.5	-0.00
M-III		$\Delta RMSE_{MLR}$			-	-0.41	-0.28	-0.37
		$\Delta RMSE_{STEP}$			-3.8	-3.7	-2.01	-0.00
BaCl ₂ ⁱⁱⁱ		$\Delta RMSE_{MLR}$				-	-2.4	-2.3
		$\Delta RMSE_{STEP}$					-3.1	-3.3
Ambic-I		$\Delta RMSE_{MLR}$					-	-0.2
		$\Delta RMSE_{STEP}$						-0.00

ⁱ RMSE MLR - RMSE SLR of models including EC_e , denoted by $\Delta RMSE_{MLR}$

ⁱⁱ RMSE MLR - RMSE SLR of models including different combinations of soil properties as regressors, denoted by $\Delta RMSE_{STEP}$

ⁱⁱⁱ Compulsive exchange extraction method

Table B- 27 Difference between RMSE-values of SLR and corresponding MLR models including soil EC_e^i and various combinations of soil propertiesⁱⁱ as regressors for conversion of exchangeable Na content for saline soils

X	Y		AA (0.2M)	Citric acid (1%)	M-III	BaCl ₂ ⁱⁱⁱ	Ambic-I	Bray-II
AA (1M)		$\Delta RMSE_{MLR}$	-2.3	-4.0	-2.6	-1.1	-1.7	-4.7
		$\Delta RMSE_{STEP}$	-6.0	-6.5	-7.5	-5.4	-3.4	-6.9
AA (0.2M)		$\Delta RMSE_{MLR}$		-1.3	-	-3.6	-0.28	-2.4
		$\Delta RMSE_{STEP}$		-4.1	0.43	-6.0	-1.7	-2.9
Citric acid (1%)		$\Delta RMSE_{MLR}$			-	-2.8	-0.53	-2.1
		$\Delta RMSE_{STEP}$			0.26	-7.6	-2.7	-4.2
M-III		$\Delta RMSE_{MLR}$			-	-3.4	-0.99	-4.0
		$\Delta RMSE_{STEP}$			-3.7	-3.4	-2.8	-5.7
BaCl ₂ ⁱⁱⁱ		$\Delta RMSE_{MLR}$				-	-48	-61
		$\Delta RMSE_{STEP}$					-1.0	-66
Ambic-I		$\Delta RMSE_{MLR}$					-	-2.9
		$\Delta RMSE_{STEP}$						-3.6

ⁱ RMSE MLR - RMSE SLR of models including EC_e , denoted by $\Delta RMSE_{MLR}$

ⁱⁱ RMSE MLR - RMSE SLR of models including different combinations of soil properties as regressors, denoted by $\Delta RMSE_{STEP}$

ⁱⁱⁱ Compulsive exchange extraction method

Table B- 28 Difference between RMSE-values of SLR and corresponding MLR models including soil pHⁱ and various combinations of soil propertiesⁱⁱ as regressors for conversion of exchangeable K content

X	Y	AA (0.2M)	Citric acid (1%)	M- III	BaCl ₂ ⁱⁱⁱ	Ambic- I	Bray- II
AA (1M)	$\Delta RMSE_{MLR}$	-1.7	-2.6	-3.4	-0.01	-2.1	-5.3
	$\Delta RMSE_{STEP}$	-3.8	-9.2	-5.4	-1.6	-3.1	-6.3
AA (0.2M)	$\Delta RMSE_{MLR}$		-0.67	-1.0	-0.6	-0.47	-2.5
	$\Delta RMSE_{STEP}$		-2.9	-1.0	-1.3	-0.00	-3.9
Citric acid (1%)	$\Delta RMSE_{MLR}$			-0.3	-1.1	-0.14	-1.2
	$\Delta RMSE_{STEP}$			-1.3	-1.1	-4.3	-4.2
M-III	$\Delta RMSE_{MLR}$				-1.0	-0.15	-1.0
	$\Delta RMSE_{STEP}$				-1.9	-0.00	-1.2
BaCl ₂ ⁱⁱⁱ	$\Delta RMSE_{MLR}$					-3.1	-5.9
	$\Delta RMSE_{STEP}$					-4.5	-6.9
Ambic-I	$\Delta RMSE_{MLR}$						-2.1
	$\Delta RMSE_{STEP}$						-2.1

ⁱ RMSE MLR - RMSE SLR of models including pH_{KCl 1:2.5}, denoted by $\Delta RMSE_{MLR}$

ⁱⁱ RMSE MLR - RMSE SLR of models including different combinations of soil properties as regressors, denoted by $\Delta RMSE_{STEP}$

ⁱⁱⁱ Compulsive exchange extraction method

Appendix C: IR spectral model appendices

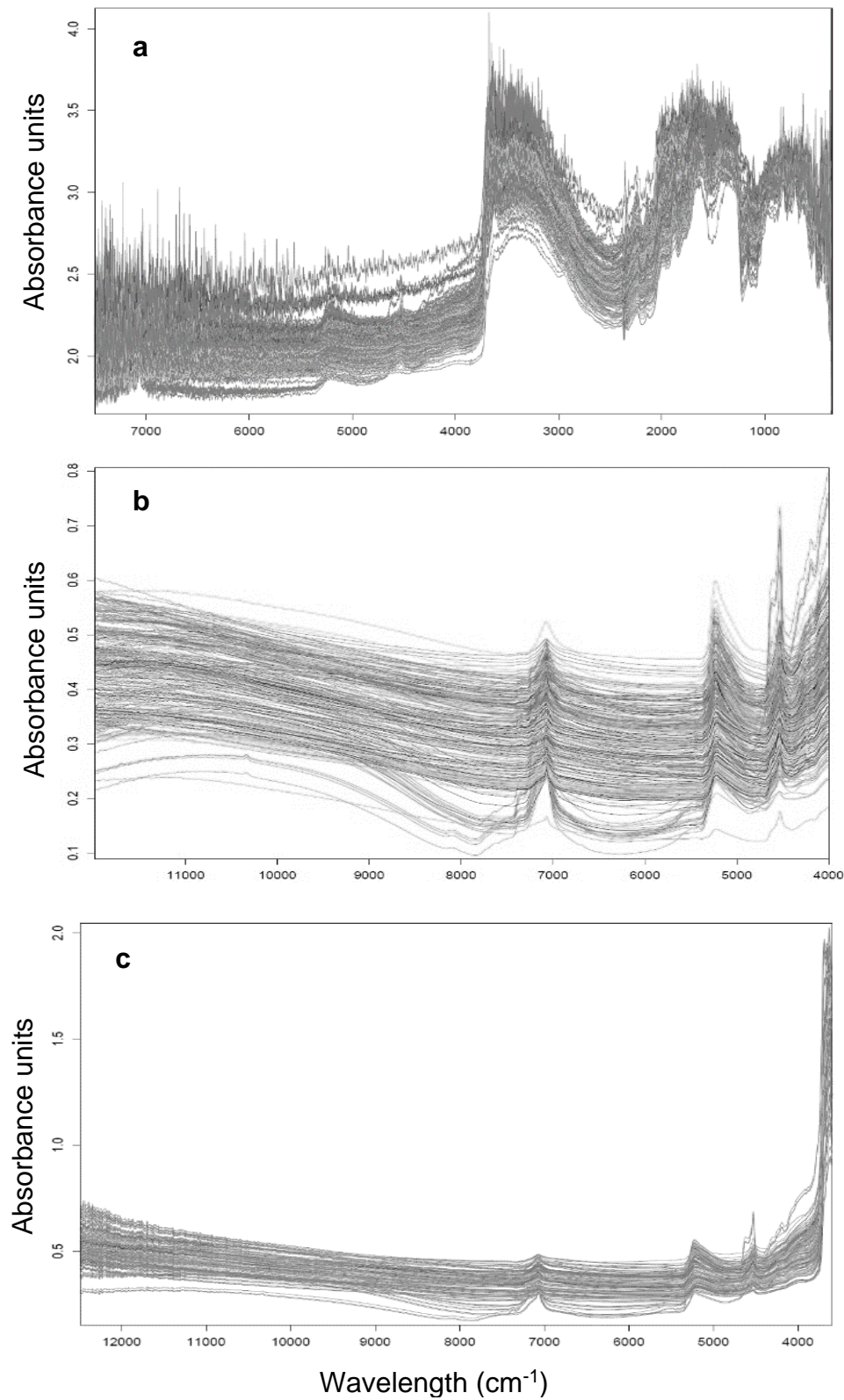


Figure C- 1 Infrared radiation wavelength (cm⁻¹) versus absorbance units for scans of the full soil cation dataset with the Alpha (a), Matrix (b) and MPA (c) spectroscopic instruments. Full soil pH dataset was highly similar, differing by replacement of 6 samples only

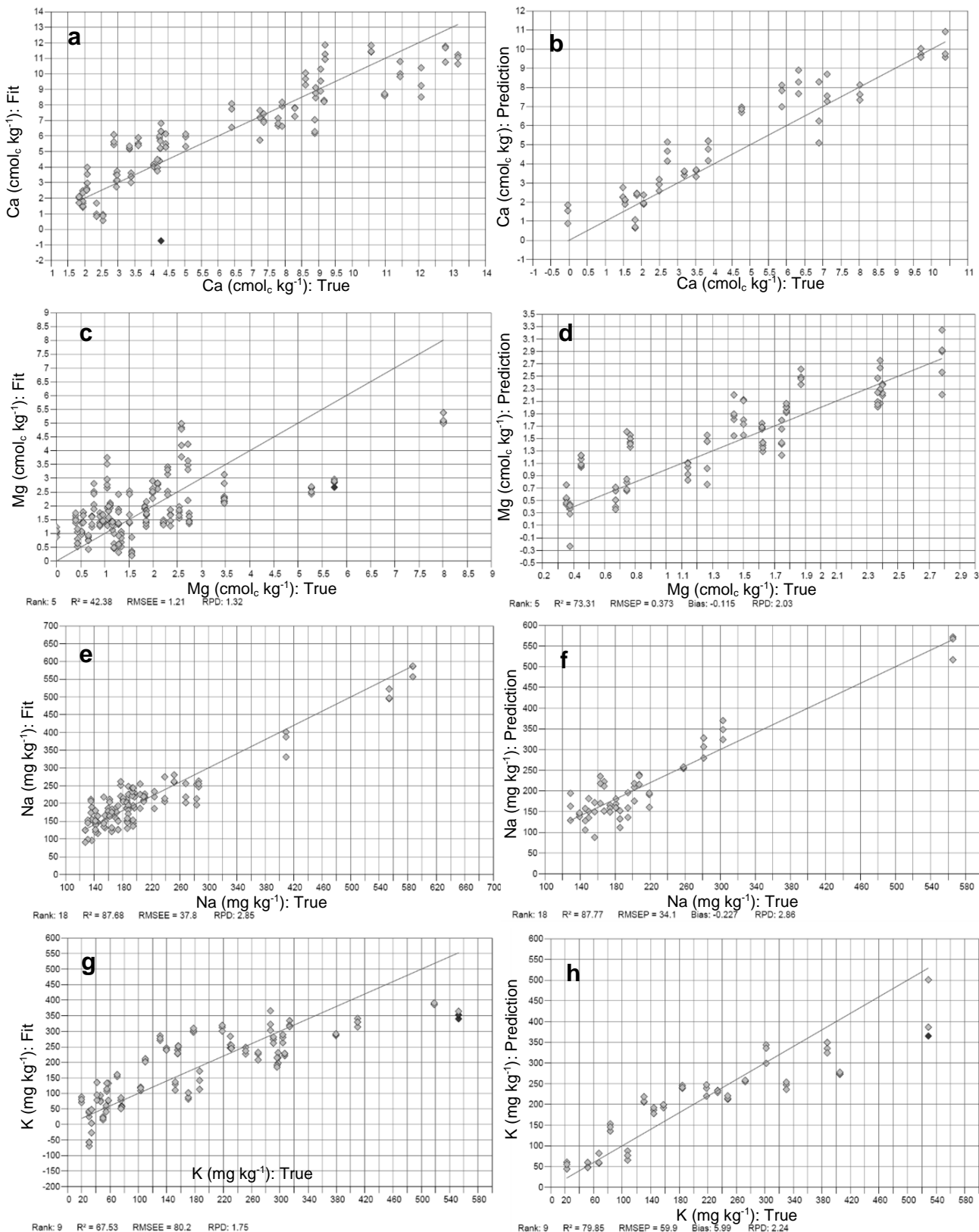


Figure C-2 Graphs for calibration & validation of infrared spectral-based models to determine of Ca (a & b), Mg (c & d), Na (e & f) and K (g & h) content

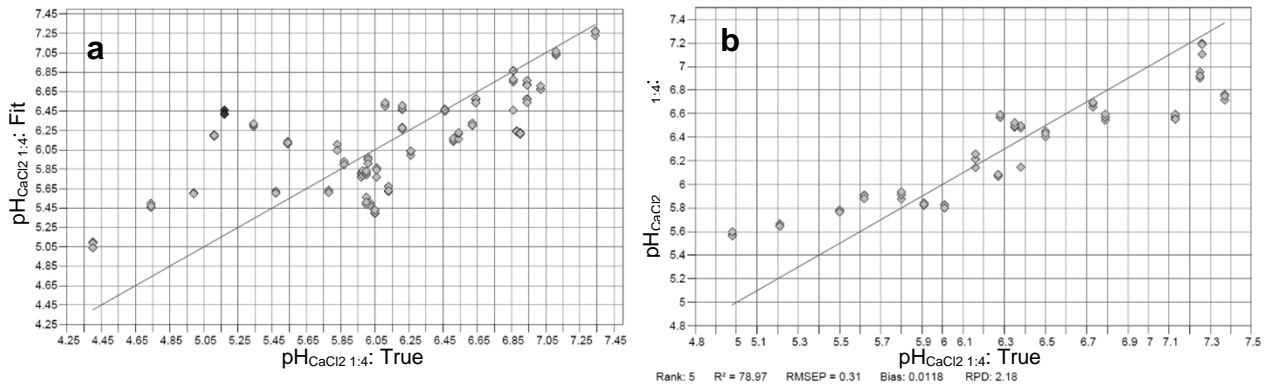


Figure C- 3 Graphs for calibration (a) & validation (b) of infrared spectral-based models to determine of soil pH

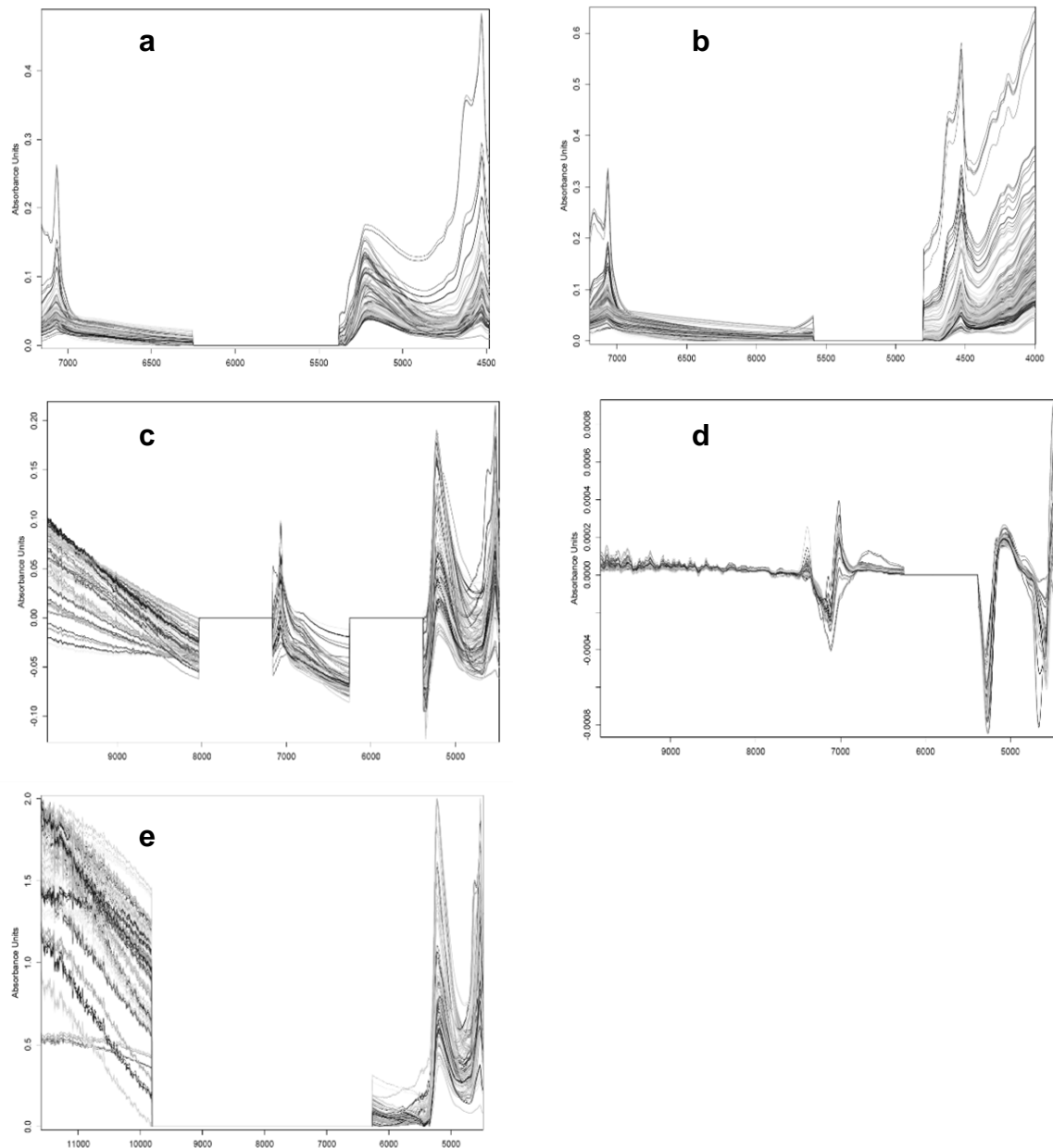


Figure C- 4 Spectra after pre-processing, used for calibration of models predicting exchangeable Ca (A, 1st der. / SLS), Mg (B, COE), Na (C, SNV) and K (D, 1st der. / MSC) and soil pH (E, Nmin-max)

Table C- 1 A review of validation statistics of models for prediction of soil chemical properties based on spectral response in IR regions of the electromagnetic spectrum

Soil chemical property	R ² val	RPD/RPIQ*	RMSE ⁱ	IR Range	Regression mode	No. samples	Soil type	CV (%) ⁱⁱ	Reference
Ca	0.89		0				Unspecified	0	(Janik <i>et al.</i> , 1998)
Ca (0.01M AgTU)	0.38		0.7	MIR	PLSR	49	Chromosols and kurosols in Australia	34	(Viscarra Rossel <i>et al.</i> , 2006)
Ca (0.01M AgTU)	0.67	1.7	0	UV-VIS-NIR	PCR	161	New South Wales and South East Queensland	90	(Kamrunnahar <i>et al.</i> , 2003)
Ca (0.01M AgTU)	0.28	0.2	0	NIR	PCR	207	New South Wales and South East Queensland	130	(Pirie <i>et al.</i> , 2005)
Ca (1M KCl)	0.94		2.2	VIS-NIR	MARS	739	Eastern and Southern Africa	0	(Shepherd & Walsh, 2002)
Ca (1M NH ₄ Cl)	0.98		0	NIR	MPLS	50	Central and northern Germany	0	(Chodak <i>et al.</i> , 2004)
Ca (1M NH ₄ OAc)	0.5	1.9	4	NIR	PCR	744	Major land resource areas in the US	68	(Chang <i>et al.</i> , 2001)
Ca (1M NH ₄ OAc)	0.8		0	NIR	PLSR	100	Fine alluvial smectitic rice field in California	8	(Van Groenigen <i>et al.</i> , 2003)
Ca (1M NH ₄ OAc)	0.56		0	MIR	PLSR	100	Fine alluvial smectitic rice field in California	8	(Van Groenigen <i>et al.</i> , 2003)
Ca (1M NH ₄ OAc)	0.65	1.7	1.6	VIS-NIR	PLSR	359	Entisol in Northern Turkey	9	(Bilgili <i>et al.</i> , 2010)
Ca (1M NH ₄ OAc)	0.86	2.7	0	NIR	PLSR	550	Alluvial and aeolian deposits in South-Eastern Australia	48	(Dunn <i>et al.</i> , 2002)
Ca (1M NH ₄ OAc)	0.89	3	6.9	MIR	PLSR	36854	Various soils of the US	0	(Dangal <i>et al.</i> , 2019)
Ca (1M NH ₄ OAc)	0.84		11	MIR	PLSR	38068	Various soils of the US	0	(Sanderman <i>et al.</i> , 2020)
Ca (Ba/ Mg)	0.76	1.9	15	NIR	PLSR	140	Soils from Czech, Denmark, Netherlands & UK	54	(Marín-González <i>et al.</i> , 2013)
Ca (M-III)	0.8	2.2	2.8	NIR	PCR	784	Major land resource areas in the US	0	(Chang <i>et al.</i> , 2001)
Ca (M-III)	0.75	3.8*	0	NIR-MIR	PLSR	273	Rice fields in sub-Saharan Africa	110	(Johnson <i>et al.</i> , 2019)
Ca (M-III)	0.21		0	MIR	RF	1907	Sub-Saharan African soils, AfSIS database	0	(Sila <i>et al.</i> , 2016)
K	0.47		0.2	NIR	PLSR		Chromosols and kurosols in Australia	0	(Viscarra Rossel <i>et al.</i> , 2006)
K	0.8	2.4	0.2	MIR	CNN	650	Scottish soils, NSIS database	150	(Haghi <i>et al.</i> , 2021)
K	0.33		0				Unspecified	0	(Janik <i>et al.</i> , 1998)
K (0.01M AgTU)		0.8	0	UV-VIS-NIR	PCR	161	New South Wales and South East Queensland	68	(Kamrunnahar <i>et al.</i> , 2003)
K (1M NH ₄ Cl)	0.91		0	NIR	MPLS	50	Central and northern Germany	0	(Chodak <i>et al.</i> , 2004)
K (1M NH ₄ OAc)	0.5	1.4	0.4	NIR	PCR	742	Major land resource areas in the US	67	(Chang <i>et al.</i> , 2001)
K (1M NH ₄ OAc)	0.25	1.1	0.3	VIS-NIR	PLSR	359	Entisol in Northern Turkey	30	(Bilgili <i>et al.</i> , 2010)
K (1M NH ₄ OAc)	0.61	1.6	0	NIR	PLSR	550	Alluvial and aeolian deposits in South-Eastern Australia	34	(Dunn <i>et al.</i> , 2002)
K (1M NH ₄ OAc)	0.36		0.7	MIR	PLSR	37702	Various soils of the US	0	(Sanderman <i>et al.</i> , 2020)
K (M-III)	0.5		0.2	MIR	PLSR	952	Various soils of the US	0	(Sanderman <i>et al.</i> , 2020)
K (modified Olsen)	0.66		0.3	VIS-NIR	MARS	739	Eastern and Southern Africa	0	(Shepherd & Walsh, 2002)
Mg	0.83	2.4	0	MIR	CNN	650	Scottish soils, NSIS database	140	(Haghi <i>et al.</i> , 2021)
Mg	0.76		0				Unspecified	0	(Janik <i>et al.</i> , 1998)

Soil chemical property	R ² val	RPD/RPIQ*	RMSE ⁱ	IR Range	Regression mode	No. samples	Soil type	CV (%) ⁱⁱ	Reference
Mg (0.01M AgTU)	0.63	1.7	0	UV-VIS-NIR	PCR	161	New South Wales and South East Queensland	78	(Kamrunnahar <i>et al.</i> , 2003)
Mg (0.01M AgTU)	0.69	1.7	0	MIR	PCR	207	New South Wales and South East Queensland	130	(Pirie <i>et al.</i> , 2005)
Mg (1M KCl)	0.91		0.8	VIS-NIR	MARS	739	Eastern and Southern Africa	0	(Shepherd & Walsh, 2002)
Mg (1M NH ₄ Cl)	0.99		0	NIR	MPLS	50	Central and northern Germany	11	(Chodak <i>et al.</i> , 2004)
Mg (1M NH ₄ OAc)	0.5	1.8	1.3	NIR	PCR	743	Major land resource areas in the US	65	(Chang <i>et al.</i> , 2001)
Mg (1M NH ₄ OAc)	0.9		0	NIR	PLSR	100	Fine alluvial smectitic rice field in California	12	(Van Groenigen <i>et al.</i> , 2003)
Mg (1M NH ₄ OAc)	0.61		0	MIR	PLSR	100	Fine alluvial smectitic rice field in California	12	(Van Groenigen <i>et al.</i> , 2003)
Mg (1M NH ₄ OAc)	0.43	1.3	0.5	VIS-NIR	PLSR	359	Entisol in Northern Turkey	19	(Bilgili <i>et al.</i> , 2010)
Mg (1M NH ₄ OAc)	0.85	2.7	0	NIR	PLSR	550	Alluvial and aeolian deposits in South-Eastern Australia	50	(Dunn <i>et al.</i> , 2002)
Mg (1M NH ₄ OAc)	0.68		3.1	MIR	PLSR	38122	Various soils of the US	0	(Sanderman <i>et al.</i> , 2020)
Mg (Ba/ Mg)	0.88	2.6	0	NIR	PLSR	140	Soils from Czech, Denmark, Netherlands & UK	63	(Marín-González <i>et al.</i> , 2013)
Mg (M-III)	0.75	3.8*	0	NIR-MIR	PLSR	274	Rice fields in sub-Saharan Africa	130	(Johnson <i>et al.</i> , 2019)
Na	0.33		0				Unspecified	0	(Janik <i>et al.</i> , 1998)
Na (0.01M AgTU)	0.34	1.2	0	UV-VIS-NIR	PCR	161	New South Wales and South East Queensland	120	(Kamrunnahar <i>et al.</i> , 2003)
Na (1M NH ₄ Cl)	0.87		0	NIR	MPLS	50	Central and northern Germany	0	(Chodak <i>et al.</i> , 2004)
Na (1M NH ₄ OAc)	0.5	0.9	0.1	NIR	PCR	439	Major land resource areas in the US	100	(Chang <i>et al.</i> , 2001)
Na (1M NH ₄ OAc)	0.09	1	0	VIS-NIR	PLSR	359	Entisol in Northern Turkey	33	(Bilgili <i>et al.</i> , 2010)
Na (1M NH ₄ OAc)	0.64	1.7	0	NIR	PLSR	550	Alluvial and aeolian deposits in South-Eastern Australia	83	(Dunn <i>et al.</i> , 2002)
Na (1M NH ₄ OAc)	0.64		0.1	MIR	PLSR	16 259	Various soils of the US	0	(Sanderman <i>et al.</i> , 2020)
pH	0.45	1.4	0.1	MIR	PLSR	30	Alfisols, Aridisols and Inceptisols, Iran	0	(Aghamir <i>et al.</i> , 2019)
pH	0.89	3.2*	0.3	MIR	SAM-PLSR	1456	Paddy soils in Nanjing, China	0	(Ma <i>et al.</i> , 2019)
pH (CaCl ₂)	0.55	1.5	0.6	NIR	PCR	743	Major land resource areas in the US	23	(Chang <i>et al.</i> , 2001)
pH (CaCl ₂)	0.86		0.1	MIR	PLSR	118	Chromosols and kurosols in Australia	6	(Viscarra Rossel <i>et al.</i> , 2006)
pH (CaCl ₂)	0.67		0	MIR	PLSR	184	Unspecified	0	(Janik <i>et al.</i> , 1998)
pH (CaCl ₂)	0.8	2.3	0	NIR	PLSR	550	Alluvial and aeolian deposits in South-Eastern Australia	14	(Dunn <i>et al.</i> , 2002)
pH (CaCl ₂)	0.76		0.3	MIR	PLSR	1935	Mollisols of the US	0	(Seybold <i>et al.</i> , 2019)
pH (H ₂ O)	0.56	1.5	0.6	NIR	PCR	743	Major land resource areas in the US	12	(Chang <i>et al.</i> , 2001)
pH (H ₂ O)	0.73	2.7*	0.5	MIR	PLSR	2285	Rice fields in sub-Saharan Africa	16	(Johnson <i>et al.</i> , 2019)
pH (H ₂ O)	0.6	1.5	0	NIR	PLSR	49	Silt loams in Gansu Province, China	0	(Lu <i>et al.</i> , 2013)
pH (H ₂ O)	0.75		0.1	MIR	PLSR	119	Chromosols and kurosols in Australia	5	(Viscarra Rossel <i>et al.</i> , 2006)
pH (H ₂ O)	0.71	1.8	0	UV-VIS-NIR	PCR		New South Wales and South East Queensland	16	(Kamrunnahar <i>et al.</i> , 2003)
pH (H ₂ O)	0.56		0	MIR	PLSR	183	Unspecified	0	(Janik <i>et al.</i> , 1998)
pH (H ₂ O)	0.83		0.3	VIS-NIR	MARS		Eastern and Southern Africa	0	(Shepherd & Walsh, 2002)
pH (H ₂ O)	0.73		0.3	NIR	PLSR	179	Silt loams in Maryland	10	(Reeves <i>et al.</i> , 1999)

Soil chemical property	R ² val	RPD/RPIQ*	RMSE ⁱ	IR Range	Regression mode	No. samples	Soil type	CV (%) ⁱⁱ	Reference
pH (H ₂ O)		1	0.1				Unspecified	2	(Janik <i>et al.</i> , 1998)
pH (H ₂ O)	0.67	1.7	0	NIR	PCR	171	New South Wales and South East Queensland	21	(Pirie <i>et al.</i> , 2005)
pH (H ₂ O)	0.74	1.9	0.5	MIR	PLSR	35297	Various soils of the US	20	(Dangal <i>et al.</i> , 2019)
pH (H ₂ O)	0.7		0.6	MIR	PLSR	37 123	Various soils of the US	0	(Sanderman <i>et al.</i> , 2020)
pH (H ₂ O)	0.86		0.3	MIR	PLSR	1935	Mollisols of the US	0	(Seybold <i>et al.</i> , 2019)
pH (H ₂ O)	0.86	2.4	0.3	NIR	PLSR	140	Soils from Czech, Denmark, Netherlands & UK	12	(Marín-González <i>et al.</i> , 2013)
pH (H ₂ O)	0.85	3	0.3	NIR	CNN	650	Scottish soils, NSIS database	18	(Haghi <i>et al.</i> , 2021)
pH (H ₂ O)	0.52		0.5	MIR	RF	1907	Sub-Saharan African soils, AfSIS database	0	(Sila <i>et al.</i> , 2016)

i Units of RMSE in $\text{cmol}_c \text{kg}^{-1}$ for all exchangeable cations and pH units for pH

ii Coefficient of variation (%) of reference data

Table C- 2 Figures of merit for models predicting soil chemical properties from NIR spectra generated by the MPA instrument

Method	Property	R ² cal	RMSEE	Rank	R ² val	RMSEP	Bias	RPD	RPIQ	Validation procedure	Pre-processing method	Wavelength regions
AA (1M)	Ca	0.71	2.8	6	0.89	1.8	0.1	3	5	S2	None	12489.4-11594.5 10715.1-8925.4 7158.8-3594.8
	Mg	0.43	1.2	3	0.17	0.7	-0.2	1.2	2.3	S2	2nd der.	12489.4-8038.2 6264-5369.1
	Na	0.82	68	12	0.57	71	-2.4	1.5	1.1	A1	1st der. + SLS	8046-7151.1 5376.8-4482
	K	0.54	92	5	0.50	91	-4.6	1.4	2.2	S1	N _{min-max}	8933.1-7151.1 6264-4482
AA (0.2M)	Ca	0.51	3.1	2	0.66	1.6	-1	2.1	3	S2	1st der.	11602.2-9820.2 6264-5369.1
	Mg	0.51	0.9	4	0.50	0.6	-0.3	1.6	1.5	S2	SLS	8933.1-8038.2
	Na	0.91	48	18	0.84	55	3.7	2.5	2.3	A2	SNV	8933.1-8038.2 5376.8-4482
	K	0.79	70	22	0.69	83	6.5	1.8	3.1	S2	SLS	4489.7-3594.8
Citric acid (1%)	Ca	0.57	7.3	4	0.50	7.3	0.3	1.4	1.1	A1	None	12489.4-11594.5 8933.1-7151.1 6264-5369.1 4489.7-3594.8
	Mg	0.21	2	2	0.42	1.1	0.1	1.3	1.1	A1	1st der. + SLS	10715.1-9820.2 8933.1-8038.2
	Na	0.80	76	13	0.47	87	-7.7	1.4	0.8	A1	1st der. + SLS	8046-7151.1 5376.8-4482
	K	0.74	58	24	0.71	73	23	2	2.9	S2	None	4489.7-3594.8
Ambic-I	Ca	0.78	1.7	3	0.80	1.2	-0.4	2.5	4.4	S2	1st der. + SNV	8933.1-6256.3
	Mg	0.35	1	4	0.53	0.5	-0.2	1.6	2.3	S2	1st der. + SNV	9828-8038.2
	Na	0.46	3.8	8	0.70	3.8	-22	2.1	1.9	A2	1st der. + MSC	626-4482
	K	0.59	85	5	0.70	63	-3.4	1.8	3	A2	COE	12489.4-11594.5 8933.1-8038.2 4489.7-3594.8
M-III	Ca	0.42	5.3	2	0.50	2.6	-0.9	1.5	2.2	S2	MSC	12489.4-8925.4 7158.8-6256.3 4489.7-3594.8
	Mg	0.16	1.6	2	0.34	0.6	-0.2	1.3	2.1	S2	SLS	12489.4-11594.5 10715.1-9820.2 8046-7151.1
	Na	0.88	38	18	0.88	34	-0.2	2.9	1.8	A2	SNV	9828-8038.2 7158.8-6256.3 5376.8-4482
	K	0.68	80	9	0.80	6	6	2.2	3.2	A2	1st der. + MSC	9828-6256.3 5376.8-4482
BaCl ₂	Ca	0.82	1.5	8	0.84	1.2	-0.6	3	4.4	S2	1st der. + SLS	10715.1-8038.2 6264-3594.8

Method	Property	R ² cal	RMSEE	Rank	R ² val	RMSEP	Bias	RPD	RPIQ	Validation procedure	Pre-processing method	Wavelength regions
	Mg	0.91	0.4	17	0.61	0.6	-0.1	1.6	2	A1	N _{min-max}	8046-7151.1 6264-5369.1
	Na	0.24	86	6	0.50	63	-2.2	1.4	1.6	A2	N _{min-max}	9828-8038.2 5376.8-4482
	K	0.81	53	20	0.58	78	1.7	1.5	2.4	A2	N _{min-max}	8933.1-8038.2 6264-4482
Bray-II	Ca	0.42	5.5	6	0.60	2.7	-0.1	1.6	2.7	S2	SLS	11602.2-8038.2 7158.8-6256.3
	Mg	0.10	1.5	2	0.03	0.9	-0.4	1.1	2	S2	MMN	11602.2-10707.4 8046-7151.1 6264-5369.1
	Na	0.91	46	18	0.17	88	19	1.1	1.2	A1	COE	5376.8-4482
	K	0.74	59	12	0.60	72	12	1.6	2.5	S1	MMN	9828-8038.2 5376.8-4482
H ₂ O 1:1	pH	0.44	0.6	3	0.72	0.4	0	1.9	2.9	A2	1st der. + MSC	12489.4-8038.2 7158.8-6256.3 5376.8-3594.8 12489.4-8038.2
H ₂ O 1:5		0.74	0.4	11	0.67	0.4	0.1	1.8	2.5	A2	SNV	11602.2-9820.2 7158.8-4482
KCl 1:2.5		0.56	0.6	3	0.60	0.5	0.1	1.6	2	A1	SNV	12489.4-4482
KCl 1:5		0.63	0.6	6	0.67	0.5	-0.1	1.8	2.3	A2	MSC	12489.4-9820.2 8046-7151.1 4489.7-3594.8
CaCl ₂ 1:2.5		0.75	0.4	9	0.70	0.4	0.1	1.9	3.5	S2	N _{min-max}	12489.4-9820.2 8933.1-8038.2 4489.7-3594.8
CaCl ₂ 1:4		0.51	0.5	5	0.79	0.3	0	2.2	3.2	S2	N _{min-max}	11602.2-9820.2 6264-4482
CaCl ₂ 1:5		0.49	0.5	5	0.65	0.4	0.1	1.7	2.7	S2	COE	8046-6256.3 5376.8-4482

ⁱ Units of RMSE in $\text{cmol}_c \text{kg}^{-1}$ for exchangeable Ca / Mg, mg kg^{-1} for exchangeable Na / K and pH units for pH

ⁱⁱ Pre-processing methods defined in Section 5.2.2

Table C- 3 Figures of merit for models predicting soil chemical properties from NIR spectra generated by the Matrix-F instrument

Method	Property	R ² cal	RMSEE	Rank	R ² val	RMSEP	Bias	RPD	RPIQ	Validation procedure	Pre-processing method	Wavelength regions
AA (1M)	Ca	0.48	3.8	4	0.83	2.1	-0.4	2.5	2.8	S1	N _{min-max}	9596.4-8786.4 7991.9-3995.9
	Mg	0.42	1.2	5	0.73	0.4	-0.1	2	3	S2	COE	7197.3-5592.8 4798.2-3995.9
	Na	0.83	63	12	0.38	85	-13	1.3	0.9	A1	COE	9596.4-8786.4 5600.5-4790.5
	K	0.62	85	9	0.63	77	6.9	1.7	2.6	S1	None	8794.1-7991.9 6395-5592.8 4798.2-3995.9
AA (0.2M)	Ca	0.39	3.4	3	0.55	1.9	-1	1.8	2.6	S2	None	11987.8-10383.2 9596.4-8786.4 4798.2-3995.9
	Mg	0.63	0.7	8	0.50	0.6	-0.1	1.4	1.5	S2	N _{min-max}	11193.2-7991.9 7197.3-5592.8 4798.2-3995.9
	Na	0.75	85	8	0.50	82	2.2	1.4	0.8	A1	SLS	5600.5-4790.5
	K	0.65	84	8	0.67	84	-3.4	1.7	2.9	S1	SNV	8794.1-7991.9 6395-5592.8 4798.2-3995.9
Citric acid (1%)	Ca	0.32	10	3	0.00	4.6	-2.2	1.2	1.1	S2	SNV	9596.4-7189.6 6395-5592.8 4798.2-3995.9
	Mg	0.36	1.8	6	0.16	0.8	-0.3	1.2	2	S2	None	7197.3-5592.8 4798.2-3995.9
	Na	0.75	73	16	0.64	83	-31	1.8	1.6	A2	None	10391-9588.7 7991.9-7189.6 5600.5-3995.9
	K	0.24	100	9	0.65	67	-9.7	1.7	1.9	S1	None	7991.9-6387.3
Ambic-I	Ca	0.79	1.6	5	0.79	1.2	-0.7	2.7	4.2	S2	1st der. + MSC	11193.2-9588.7 8794.1-6387.3 5600.5-4790.5
	Mg	0.42	1	7	0.64	0.4	0	1.7	2.6	S2	COE	7197.3-5592.8 4798.2-3995.9
	Na	0.85	43	19	0.49	62	-22	1.5	1.4	A2	None	5600.5-3995.9
	K	0.66	77	6	0.67	66	-0.5	1.8	2.9	A2	SNV	8794.1-7189.6 5600.5-3995.9
M-III	Ca	0.48	3.8	4	0.83	2.1	-0.4	2.5	2.8	S1	N _{min-max}	9596.4-8786.4 7991.9-3995.9
	Mg	0.42	1.2	5	0.73	0.4	-0.1	2	3	S2	COE	7197.3-5592.8 4798.2-3995.9
	Na	0.83	63	12	0.38	85	-13	1.3	0.9	A1	COE	9596.4-8786.4 5600.5-4790.5

Method	Property	R ² cal	RMSEE	Rank	R ² val	RMSEP	Bias	RPD	RPIQ	Validation procedure	Pre-processing method	Wavelength regions
	K	0.62	85	9	0.63	77	6.9	1.7	2.6	S1	None	8794.1-7991.9 6395-5592.8 4798.2-3995.9
BaCl ₂	Ca	0.80	1.6	10	0.79	1.3	-0.4	2.3	3.8	S2	1st der. + MSC	11987.8-10383.2 9596.4-8786.4 7991.9-6387.3 5600.5-3995.9
	Mg	0.79	0.6	11	0.74	0.4	0	2	2.9	S2	SNV	11987.8-11185.5 10391-9588.7 7991.9-4790.5
	Na	0.26	84	8	0.47	65	-7.3	1.4	1.6	A2	N _{min-max}	11987.8-11185.5 9596.4-8786.4 7197.3-6387.3 4798.2-3995.9
	K	0.65	71	12	0.59	76	-7	1.6	1.9	S2	N _{min-max}	6395-5592.8 4798.2-3995.9
Bray-II	Ca	0.47	5.2	6	0.53	2.9	-0.3	1.5	2.5	S2	SNV	9596.4-8786.4 7197.3-6387.3
	Mg	0.06	1.6	2	0.08	0.8	-0.4	1.2	1.9	S2	SLS	7991.9-7189.6
	Na	0.74	64	15	0.53	77	12	1.5	1.1	A2	MMN	7197.3-6387.3 5600.5-4790.5
	K	0.63	67	8	0.62	71	-5.9	1.6	2.6	S1	SNV	8794.1-6387.3 4798.2-3995.9
H ₂ O 1:1	pH	0.52	0.6	9	0.72	0.4	0	1.9	2.9	A2	None	5600.5-4790.5
H ₂ O 1:5		0.44	0.5	6	0.64	0.4	0.1	1.7	2.7	S2	COE	11193.2-9588.7 4798.2-3995.9
KCl 1:2.5		0.58	0.6	6	0.63	0.6	-0.1	1.7	2.5	A2	None	11987.8-4790.5
KCl 1:5		0.47	0.7	5	0.65	0.6	0.1	1.7	3.1	S2	N _{min-max}	11987.8-10383.2 7197.3-6387.3
CaCl ₂ 1:2.5		0.59	0.5	7	0.66	0.4	0	1.7	3.3	S2	None	11987.8-11185.5 10391-9588.7 7197.3-6387.3 4798.2-3995.9
CaCl ₂ 1:4		0.53	0.5	4	0.71	0.4	0	1.8	2.7	S2	1st der.	11193.2-10383.2 7197.3-6387.3
CaCl ₂ 1:5		0.49	0.5	6	0.61	0.4	0.1	1.6	2.6	S2	None	7991.9-5592.8

ⁱ Units of RMSE in cmol_c kg⁻¹ for exchangeable Ca / Mg, mg kg⁻¹ for exchangeable Na / K and pH units for pH

ⁱⁱ Pre-processing methods defined in Section 5.2.2

Table C- 4 Figures of merit for models predicting soil chemical properties from MIR spectra generated by the Alpha instrument

Method	Property	R ² cal	RMSEE	Rank	R ² val	RMSEP	Bias	RPD	RPIQ	Validation procedure	Pre-processing method	Wavelength regions
AA (1M)	Ca	0.47	4	2	0.79	2.2	0	2.2	2.8	A1	N _{min-max}	6782-6067.1 5352.2-3922.5 2492.8-1777.9
	Mg	0.27	1.4	2	0.18	0.7	-0.3	1.2	1.7	S2	N _{min-max}	2492.8-1777.9
	Na	0.12	110	2	0.19	130	-9.3	1.1	0.5	S1	SLS	2492.8-1777.9
	K	0.58	88	4	0.50	90	-5.4	1.4	2.2	S1	SNV	6067.1-5352.2 4637.4-3207.6
AA (0.2M)	Ca	0.72	2.4	2	0.67	1.6	-0.3	1.8	3.1	S2	MSC	6782-3207.6 1777.9-1063
	Mg	0.32	1	2	0.17	0.7	-0.2	1.2	1.2	S2	N _{min-max}	2492.8-1777.9
	Na	0.51	100	2	0.01	140	20	1	0.9	A2	1st der.	6782-4637.4 2492.8-1777.9
	K	0.68	80	5	0.55	98	15	1.5	2.5	S1	SNV	6067.1-2492.8
Citric acid (1%)	Ca	0.49	8.8	2	0.12	4.3	-1.9	1.2	1.2	S2	1st der. + MSC	3922.5-3207.6 2492.8-1063
	Mg	0.22	1.9	2	0.24	0.8	-0.4	1.3	2.1	S2	MSC	3922.5-3207.6 2492.8-1063
	Na	0.16	130	2	0.16	130	28	1.1	1	A2	1st der. + MSC	6067.1-5352.2 4637.4-3922.5 3207.6-2492.8
	K	0.44	66	3	0.37	81	-8.7	1.3	1.7	A1	None	6782-5352.2 2492.8-348.2
Ambic-I	Ca	0.66	2.1	2	0.73	1.4	-0.5	2.1	3.7	S2	1st der. + SNV	5352.2-3207.6
	Mg	0.36	1	2	0.35	0.6	-0.2	1.4	1.9	S2	MSC	5352.2-4637.4 2492.8-1777.9
	Na	0.14	110	2	0.01	75	-33	1.1	0.7	A1	N _{min-max}	3922.5-1777.9
	K	0.55	89	3	0.58	74	7	1.6	2.6	A2	1st der.	4637.4-3922.5
M-III	Ca	0.58	4.5	2	0.52	2.5	-0.4	1.5	2.5	S2	1st der. + MSC	3922.5-2492.8 1777.9-1063
	Mg	0.30	1.5	2	0.32	1	0	1.2	1	S1	SNV	6067.1-5352.2 2492.8-1777.9
	Na	0.50	72	4	0.32	81	-3	1.2	0.8	A2	COE	6782-4637.4 2492.8-1777.9
	K	0.50	97	3	0.71	72	-4.9	1.9	2.7	A2	MSC	7496.8-4637.4 3922.5-1777.9
BaCl ₂	Ca	0.68	2	2	0.75	1.5	-0.4	2.1	3.5	S2	1st der. + MSC	6067.1-2492.8 1777.9-1063
	Mg	0.51	0.9	2	0.57	0.6	-0.2	1.7	2.2	S2	MSC	5352.2-4637.4 2492.8-1777.9
	Na	0.58	65	7	0.51	63	-2.3	1.4	1.7	A2	None	6067.1-4637.4 3207.6-1777.9
	K	0.64	72	5	0.34	93	3.9	1.2	1.7	S1	SNV	6067.1-5352.2 4637.4-2492.8

Method	Property	R ² cal	RMSEE	Rank	R ² val	RMSEP	Bias	RPD	RPIQ	Validation procedure	Pre-processing method	Wavelength regions
Bray-II	Ca	0.19	6.4	3	0.37	3.4	0.2	1.3	2.8	S2	None	6782-5352.2 1777.9-1063
	Mg	0.23	1.4	2	0.13	1.1	-0.1	1.1	1.4	A1	MMN	7496.8-6782 6067.1-5352.2 3922.5-3207.6
	Na	0.09	110	2	0.16	110	-5.7	1.1	0.5	S1	None	7496.8-6067.1 5352.2-3922.5 3207.6-1777.9
	K	0.63	66	3	0.46	84	14	1.4	2.2	S1	SNV	6067.1-5352.2 4637.4-3207.6
H ₂ O 1:1	pH	0.89	0.3	8	0.72	0.4	0	1.9	3.7	S2	SLS	4637.4-1063
H ₂ O 1:5		0.89	0.2	8	0.67	0.4	0.1	1.8	2.9	S2	SLS	6782-6067.1 3922.5-1063
KCl 1:2.5		0.69	0.5	5	0.73	0.5	0	1.9	2.4	A1	SNV	6067.1-6067.1 4637.4-4637.4 2492.8-2492.8
KCl 1:5		0.89	0.3	8	0.66	0.5	0	1.7	2.1	A1	N _{min-max}	5352.2-4637.4 3207.6-1063
CaCl ₂ 1:2.5		0.88	0.3	8	0.78	0.4	0.1	2.2	4.1	S2	SLS	4637.4-1063
CaCl ₂ 1:4		0.82	0.3	7	0.72	0.4	0	1.9	2.8	S2	SLS	4637.4-1063
CaCl ₂ 1:5		0.88	0.2	8	0.73	0.4	0	1.9	3.1	S2	SNV	6067.1-5352.2 4637.4-3922.5 3207.6-1063

ⁱ Units of RMSE in cmolc kg^{-1} for exchangeable Ca / Mg, mg kg^{-1} for exchangeable Na / K and pH units for pH

ⁱⁱ Pre-processing methods defined in Section 5.2.2

Table C- 5 Figures of merit for restricted-range models predicting soil chemical properties from MIR spectra generated by the MPA instrument

Method	Property	R ² cal	RMSEE	Rank	R ² val	RMSEP	Bias	RPD	RPIQ	Validation procedure	Pre-processing method	No. outliers	Range	Wavelength regions
AA (1M)	Ca	0.71	2.8	4	0.70	1.8	0.7	2	3	S2	None	3	0.2705-9.89	12489.4-11594.5 10715.1-8925.4 8046-7151.1 5376.8-3594.8
	Mg	0.81	0.5	17	0.79	0.6	0.1	2.2	2.4	A2	1st der.	3	0.3545-5.75	8933.1-7151.1 6264-4482
AA (0.2M)	Ca	0.51	2.5	2	0.68	1.6	-0.9	2.1	3.1	S2	1st der.	2	0.4777-10.09	11602.2-9820.2 6264-5369.1
	Mg	0.33	0.9	6	0.60	0.5	-0.2	1.7	1.7	S2	None	3	0.1974-3.497	8933.1-7151.1 6264-5369.1
Citric acid (1%)	Ca	0.51	6.9	6	0.61	6.3	-0.1	1.6	1	S1	None	3	0.51-36.36	12489.4-11594.5 8933.1-8038.2 6264-3594.8
	Mg	0.59	1	7	0.58	0.6	-0.1	1.6	2.9	S2	MSC	3	0.34-3.23	8933.1-8038.2
	Na	0.80	65	13	0.51	84	-5.1	1.4	0.9	A1	1st der. + SLS	2	5-626	8046-7151.1 5376.8-4482
	K	0.56	63	5	0.51	72	-9.5	1.4	2	A1	1st der. + SNV	6	19-458	8046-7151.1
Ambic	Mg	0.26	1	2	0.67	0.4	-0.1	1.8	2.7	S2	1st der. + SNV	2	0.2236-3.2	8933.1-8038.2
M-III	Ca	0.61	3.7	4	0.62	2.2	-0.8	1.8	2.5	S2	2nd der.	6	0.3903-13.09	8046-7151.1 4489.7-3594.8
	Mg	0.17	1.4	2	0.41	0.6	-0.1	1.3	2.2	S2	SLS	3	0.0776-2.997	12489.4-11594.5 10715.1-9820.2 8046-7151.1
BaCl ₂	Mg	0.51	0.8	7	0.72	0.5	-0.2	2	2.8	S2	COE	2	0.08237-3.601	11602.2-10707.4 8933.1-8038.2 7158.8-4482

ⁱ Units of RMSE in $\text{cmol}_c \text{ kg}^{-1}$ for exchangeable Ca / Mg, mg kg^{-1} for exchangeable Na / K and pH units for pH

ⁱⁱ Pre-processing methods defined in Section 5.2.2

Table C- 6 Figures of merit for restricted-range models predicting soil chemical properties from MIR spectra generated by the Matrix-F instrument

Method	Property	R ² cal	RMSEE	Rank	R ² val	RMSEP	Bias	RPD	RPIQ	Validation procedure	Pre-processing method	No. outliers	Range	Wavelength regions
AA (0.2M)	Ca	0.48	2.6	3	0.58	1.8	-0.8	1.7	2.7	S2	N _{min-max}	5	0.4777-10.09	11193.2-9588.7 4798.2-3995.9
Citric acid (1%)	Mg	0.59	0.9	9	0.70	0.5	0	1.8	3.4	S2	COE	8	0.34-3.23	11987.8-11185.5 7197.3-3995.9
Ambic-I	Mg	0.43	0.8	7	0.77	0.4	-0.1	2.1	3.2	S2	None	10	0.2236-3.2	10391-8786.4 7991.9-4790.5
M-III	Ca	0.59	3.8	2	0.66	2.1	-0.6	1.8	2.7	S2	1st der. + SNV	10	0.3903-13.09	9596.4-8786.4 7991.9-7189.6 6395-5592.8

ⁱ Units of RMSE in $\text{cmol}_c \text{kg}^{-1}$ for exchangeable Ca / Mg, mg kg^{-1} for exchangeable Na / K and pH units for pH

ⁱⁱ Pre-processing methods defined in Section 5.2.2

Table C- 7 Figures of merit for restricted-range models predicting soil chemical properties from MIR spectra generated by the Alpha instrument

Method	Property	R ² cal	RMSEE	Rank	R ² val	RMSEP	Bias	RPD	RPIQ	Validation procedure	Pre-processing method	No. outliers	Range	Wavelength regions
AA (1M)	Ca	0.68	3	2	0.69	1.8	-0.5	1.9	2.9	S2	1st der. + MSC	3	0.2705-9.89	3922.5-2492.8 1777.9-1063
	Na	0.38	51	3	0.17	120	20	1.1	0.9	A2	MSC	6	46.17-638.1	7496.8-6782 4637.4-3207.6 2492.8-348.2
AA (0.2M)	Ca	0.69	2.2	2	0.70	1.5	-0.4	1.9	3.2	S2	MSC	2	0.4777-10.09	6782-6067.1 4637.4-3207.6 1777.9-1063
	Mg	0.50	0.6	4	0.60	0.5	-0.1	1.6	1.7	S2	MSC	6	0.1974-3.497	4637.4-3922.5 3207.6-1777.9
	Na	0.25	100	2	0.10	110	-6.8	1.1	0.6	A1	MSC	3	164.5-762.2	7496.8-4637.4 3207.6-1063
Citric acid (1%)	Mg	0.55	1	2	0.36	0.7	-0.2	1.3	2.3	S2	MSC	3	0.34-3.23	4637.4-3207.6 2492.8-1777.9 1063-348.2
Ambic-I	Mg	0.62	0.5	6	0.73	0.4	0.1	1.9	2.9	S2	SNV	9	0.2236-3.2	2492.8-1777.9
	Na	0.20	72	2	0.10	72	-10	1.1	0.7	A1	N _{min-max}	3	152.2-534.8	4637.4-1777.9
M-III	Ca	0.66	3.8 0	2	0.51	2.6 0	0.1 0	1.4 0	2.7 0	S2	1st der. + MSC	3	0.3903-1 3.09	3922.5-3207.6 1777.9-1063
Bray-II	Ca	0.36	4.4	4	0.43	3.2	0.5	1.3	2.3	S2	MMN	3	0.156-12.82	6067.1-5352.2 3207.6-2492.8
	Mg	0.31	1.1	4	0.29	0.7	-0.1	1.2	2.2	S2	COE	3	0.01172-3.098	3922.5-2492.8 1777.9-1063

ⁱ Units of RMSE in $\text{cmol}_c \text{kg}^{-1}$ for exchangeable Ca / Mg, mg kg^{-1} for exchangeable Na / K and pH units for pH

ⁱⁱ Pre-processing methods defined in Section 5.2.2

Table C- 8 Performance of models validated with either local or non-local sample subsets as an indication of the effect of soil texture on predictions of exchangeable base cations and soil pH from spectral data

Property ⁱ	Validation procedure ⁱⁱ	RMSEP ⁱⁱⁱ	Δ RMSEP ^{iv}	Confirmation of effect ^v
Ca	C: V = low clay: high clay	1.6	0.40	Yes
	C: V = high clay: low clay	1.9	0.79	
	Low clay excluded V = Sands only	1.1 1.4	-0.07 0.28	
Mg	C: V = low clay: high clay	1.6	1.2	No
	C: V = high clay: low clay	0.70	0.33	
	Low clay excluded V = Sands only	0.60 0.52	0.23 0.14	
Na	C: V = low clay: high clay	80	46	No
	C: V = high clay: low clay	53	19	
	Low clay excluded V = Sands only	69 58	35 24	
K	C: V = low clay: high clay	97	37	No
	C: V = high clay: low clay	83	23	
	Low clay excluded V = Sands only	48 57	-12 -2.7	
pH	C: V = low clay: high clay	0.36	0.05	Yes
	C: V = high clay: low clay	0.40	0.09	
	Low clay excluded V = Sands only	0.29 0.50	-0.03 0.19	

ⁱ Reference method and spectroscopic analysis instrument as shown in Table 5-3

ⁱⁱ Validation procedure in order as described in Section 5.2.2

ⁱⁱⁱ Units of RMSE in $\text{cmol}_c \text{kg}^{-1}$ for exchangeable cations and pH units for pH

^{iv} Δ RMSEP = $RMSEP_{inv} - RMSEP_{global}$ where $RMSEP_{global}$ refers to RMSEP of the original best model calibrated from the global dataset and $RMSEP_{inv}$ refers to the RMSEP of the corresponding investigative model

^v Higher accuracy for local predictions and poorer accuracy for non-local predictions in all four validation procedures, were deemed consistent indicators of an effect of soil texture on model performance

Table C- 9 Performance of models validated with either local or non-local sample subsets as an indication of the effect of SOM on predictions of exchangeable base cations and soil pH from spectral data

Property ⁱ	Validation procedure ⁱⁱ	RMSEP ⁱⁱⁱ	Δ RMSEP ^{iv}	Confirmation of effect ^v
Ca	C: V = high %C: low %C	1.2	0.05	Yes
	Low %C excluded	0.82	-0.33	
	C: V = high %N: low %N	1.5	0.30	
Mg	Low %N excluded	0.61	-0.54	No
	C: V = high %C: low %C	0.59	0.22	
	Low %C excluded	0.39	0.01	
Na	C: V = high %N: low %N	1.0	0.63	No
	Low %N excluded	0.26	-0.11	
	C: V = high %C: low %C	52	18	
K	Low %C excluded	45	11	No
	C: V = high %N: low %N	93	59	
	Low %N excluded	76	42	
pH	C: V = high %C: low %C	66	6.5	Yes
	Low %C excluded	86	26	
	C: V = high %N: low %N	69	9.0	
	Low %N excluded	91	31	
	C: V = high %C: low %C	0.34	0.03	
	Low %C excluded	0.28	-0.03	
	C: V = high %N: low %N	0.38	0.07	
	Low %N excluded	0.29	-0.02	

ⁱ Reference method and spectroscopic analysis instrument as shown in Table 5-3

ⁱⁱ Validation procedure in order as described in Section 5.2.2

ⁱⁱⁱ Units of RMSE in $\text{cmol}_c \text{kg}^{-1}$ for exchangeable cations and pH units for pH

^{iv} Δ RMSEP = $RMSEP_{inv} - RMSEP_{global}$ where $RMSEP_{global}$ refers to RMSEP of the original best model calibrated from the global dataset and $RMSEP_{inv}$ refers to the RMSEP of the corresponding investigative model

^v Greater accuracy of local predictions and poorer accuracy of non-local predictions in all four validation procedures, were deemed consistent indicators of the effect of SOM on model performance