

**ASSESSING PRODUCTIVE SOIL-
LANDSCAPES IN VICTORIA USING DIGITAL
SOIL MAPPING**

Nathan J. Robinson

B.App.Sci. (Hons)

Thesis submitted to Federation University in total fulfilment of the
requirements for the degree of Doctor of Philosophy

September 2016

Faculty of Science and Technology



PO Box 663

University Drive, Mount Helen

Ballarat, Victoria 3353

Australia

ABSTRACT

Spatial soil information is used to support questions on agriculture and the environment from global to local scales. Historically, soil mapping has been used to inform and guide a multitude of land users with their decisions. Demand for specific spatial soil information is increasing in response from a wider range of users operating across agricultural and environmental domains.

To satisfy these demands, users must be provided with practical and relevant spatial soil information. Novel approaches are required to deal with global deficiencies in available soil information. A major limitation to this is the plethora of incongruent legacy data with poor spatial and temporal coverage.

This research study initially identifies the specific needs of users for spatial soil information with a focus on the requirements of biophysical modellers. Secondly, error sources that hamper Digital Soil Mapping (DSM) are identified, described and assessed using pH in practical and relevant examples. A final aim is to spatially predict soil properties (e.g. clay mineralogy) that underpin soil chemical behaviour. This is achieved by harmonising legacy data in combination with new spectroscopy techniques and a spatial inference approach.

The spatial soil information needs of biophysical modellers in Victoria, Australia were found to be consistent with global needs for information including soil water characteristics, organic carbon and effective rooting depth. To accommodate stochastic and epistemic uncertainties in spatial soil information, uncertainty frameworks proved effective to deal with, and understand the limitations of legacy data in spatial inference

models. Robust and reliable spectroscopic models for properties that are linked to functions and services delivered by soil were achieved and used in 3D spatial models.

These findings will enable a tactical response through the delivery of pertinent spatial soil information that is contemporary, quality assured and sought by users. Learnings presented should enable producers of spatial soil information to be more comprehensive in their delivery of products that are easy to use, accessible and understood by a growing user community.

STATEMENT OF AUTHORSHIP

Except where explicit reference is made in the text of the thesis, this thesis contains no material published elsewhere or extracted in whole or in part from a thesis by which I have qualified for or been awarded another degree or diploma. No other person's work has been relied upon or used without due acknowledgment in the main text and bibliography of the thesis.

A handwritten signature in black ink, appearing to read "Nathan Pakin". The signature is written in a cursive style with a large, looping initial 'N'.

Signature

Date 2nd September 2016

PREFACE

This study presents the results and findings from five journal papers undertaken with the PhD candidate as the lead author on uncertainty assessment and Digital Soil Mapping (DSM), using case studies to demonstrate their implementation for real-world application. Of the five papers, three are published and two have been submitted for review.

The introduction, study region overview and literature review (chapters 1, 2 and 3) provide an overview of the current demand and utility for soil mapping from a Victorian perspective with a focus on the chronological development of soil mapping and digital soil mapping as a discipline in soil science. Current knowledge gaps and research priorities are described as: (i) the requirement for new environmental covariates to improve spatial prediction qualities; (ii) assessment and validation of DSM techniques using different environmental covariates in different physiographic settings requires testing; (iii) understanding user requirements (including representation) and products for user groups is required; (iv) an ongoing need for continuous spatial predictions of soil properties across the globe; (v) integration of expert systems and/or fuzzy systems in soil survey should be supported; (vi) the next evolution of DSM (excluding Digital Soil Assessment) is to predict soil properties in space and time, and (vii) development of spatial soil inference systems is needed. Chapter 2 provides a brief overview of the study region (western Victoria) for chapters 5, 6 and 8.

The first paper, *Soil data for biophysical models in Victorian landscapes: current needs and challenges* (*Geoderma Regional* – 2016; Chapter 4) explore user requirements for soil information with a focus on the needs of biophysical modellers. The changing requirements of modellers over the last 5-years and the soil properties that are required

and impact model sensitivity are discussed. The following papers as chapters 5 and 6: *Identification and interpretation of sources of uncertainty in soils change in a global systems-based modelling process (Soil Research - 2015)* and *Improving information content in soil pH maps: a case study in south-western Victoria* (submitted to *European Journal of Soil Science*) present a conceptual framework to consider implementation of a more holistic assessment of uncertainty by accommodating stochastic and epistemic error sources. A worked example for soil pH in south-western Victoria highlights key aspects of this framework in a Digital Soil Mapping context. Chapter 7 (*Assessment of error sources in measurements of field pH: effect of operator experience, time-of-day and test kit differences* – submitted to *Communications in Soil Science and Plant Analysis*) explores further elements of error and uncertainty in field and laboratory measurement of pH used for soil mapping purposes. The final paper *The 3D distribution of phyllosilicate clay minerals in western Victoria (Geoderma - 2016)* as Chapter 8 presents a DSM implementation that integrates spectroscopic models formed using mid-infrared (MIR) and legacy quantitative XRD measurements with spatial inference models to predict clay mineral (kaolinite, illite and smectite) abundance for agricultural landscapes of western Victoria. Investigations and findings of these papers are condensed with conclusions derived and further work identified in Chapter 9.

The candidate has also co-authored three journal papers that are relevant to the setting and discussion of this thesis and are included in Appendix A. An additional five book chapters have been produced by the candidate and are presented in Appendix B. Appendix C provides a detailed overview for western Victoria that supplements descriptions and details in Chapter 2 and those in Chapter 5, 7 and 8.

LIST OF PUBLICATIONS

This thesis is submitted in accordance with Federation University Australia Regulation 5.1, thesis incorporating published papers. Publications from the candidate undertaken for this degree include five journal papers (led by the candidate) as the foundation of the thesis:

Refereed journal papers

1. Robinson, N.J., Dahlhaus, P.G., MacEwan, R.J., Alexander, J., 2016. Soil data for biophysical models in Victorian landscapes: current needs and challenges. *Geoderma Regional* 7(3), 259-270 <http://dx.doi.org/10.1016/j.geodrs.2016.03.004>.
2. Robinson, N.J., Benke, K.K., Norng, S., 2015. Identification and interpretation of sources of uncertainty in soils change in a global systems-based modelling process. *Soil Research* 53(6), 592-604 <http://dx.doi.org/10.1071/SR14239>.
3. Robinson, N.J., Rees, D.B., Norng, S., Benke, K.K., Davey, M., submitted. Assessment of error sources in measurements of field pH: effect of operator experience, time-of-day and test kit differences. *Communications in Soil Science and Plant Analysis*.
4. Robinson, N.J., Benke, K.K., Norng, S., Crawford, D.M., submitted. Improving information content in soil pH maps: a case study in south-western Victoria. *European Journal of Soil Science*.
5. Robinson, N.J., Kitching, M., 2016. The 3D distribution of phyllosilicate clay minerals in western Victoria. *Geoderma*.
<http://dx.doi.org/10.1016/j.geoderma.2016.07.006>.

Co-authored papers

1. Marchant, B.P., Crawford, D.M., Robinson, N.J., 2015. What can legacy datasets tell us about soil quality trends? Soil acidity in Victoria. *IOP Conf. Series: Earth and Environmental Science* 25 (2015) 012015. <http://dx.doi.org/10.1088/1755-1315/25/1/012015>.

2. Scheffe, C.R., Barlow, K.M., Robinson, N.J., Crawford, D.M., McLaren, T.I., Smernik, R.J., Croatto, G., Walsh, R.D., Kitching, M., 2015. 100 Years of superphosphate addition to pasture in an acid soil—current nutrient status and future management. *Soil Research* 53(6) 662-676.
<http://dx.doi.org/10.1071/SR14241>.
3. Crawford, D.M., Chia, K., Norng, S., Kitching, M., Robinson, N.J., to be submitted. Accounting for measurement errors when harmonising incongruent soil data – a case study.

Book chapters

1. Robinson, N.J., Kitching, M., Rab, M.A., Fisher, P.D., 2012. Prediction of field capacity and permanent wilting point using rapid soil sensing approaches. In: Minasny, B., Malone, B., McBratney, A., (Eds.), *Digital Soil Assessments and Beyond*. CRC Press, The Netherlands.
2. Marchant, B.P., Crawford, D.M., Robinson, N.J., 2012. Spatial and temporal prediction of soil properties from legacy data. In: Minasny, B., Malone, B., McBratney, A., (Eds.), *Digital Soil Assessments and Beyond*. CRC Press, The Netherlands.
3. Robinson, N.J., Benke, K.K., Hopley, J., MacEwan, R.J., Clark, R., Rees, D.B., Kitching, M., Imhof, M.P., Bardos, D., 2014. Multi-source data integration and identification of uncertainties affecting production of a digital soil map. In: Arrouays, D. et al. (Eds.), *GlobalSoilMap, Basis of the global spatial soil information system*. CRC Press, The Netherlands.
4. Hopley, J., Rees, D.B., MacEwan, R.J., Clark, R., Benke, K.K., Imhof, M.P., Robinson, N.J., Bardos, D., 2014. A digital soil map of Victoria – VicDSMv1. In: Arrouays, D. et al. (Eds.), *GlobalSoilMap, Basis of the global spatial soil information system*. CRC Press, The Netherlands.
5. MacEwan, R.J., Robinson, N.J., Imhof, M.P., Rees, D.B., Hopley, J., Crawford, D.M., 2014. Multi-source data integration and identification of uncertainties affecting production of a digital soil map. In: Arrouays, D. et al. (Eds.), *GlobalSoilMap, Basis of the global spatial soil information system*. CRC Press, The Netherlands.

The five papers as the core of this thesis are listed below with the relative contribution from the candidate listed along with the primary activities undertaken.

Chapter	Publication title	Publication status	Nature and extent of candidates contribution
4	Soil data for biophysical models in Victorian landscapes: current needs and challenges	Published <i>Geoderma Regional</i>	Initiation, ideas, set up, data preparation, analysis of results, leading write up. 85%
5	Identification and interpretation of sources of uncertainty in soils change in a global systems-based modelling process	Published <i>Soil Research</i>	Initiation, ideas, set up, data preparation, methodology, analysis of results, leading write up. 50%
6	Improving information content in soil pH maps: a case study in south-western Victoria	Submitted <i>European Journal of Soil Science</i>	Initiation, ideas, set up, data preparation, methodology, analysis of results, leading write up. 60%
7	Assessment of error sources in measurements of field pH: effect of operator experience, time-of-day and test kit differences	Submitted <i>Communications in Soil Science and Plant Analysis</i>	Joint initiation, ideas, set up, data preparation, analysis of results, leading write up. 35%
8	The 3D distribution of phyllosilicate clay minerals in western Victoria	Published <i>Geoderma</i>	Initiation, ideas, set up, data preparation, methodology, analysis of results, leading write up. 95%

ACKNOWLEDGEMENTS

- Jodie, Sebastian and Noah. Without your tolerance and patience none of this would be possible. Thank-you for lasting the journey with me, and now let the good times roll.
- My parents (Ray and Yvonne) have always advocated trying your best and to use your initiative. Thank-you for sustaining and supporting my other pursuits and hobbies in farming and agriculture. Big thanks to Pam and Spud who have always been there for our family when study duties have called. My grandparents have always been supportive of all my pursuits in life, and I hope this thesis is some recognition of their contribution and drive for higher achievement.
- This thesis would not be possible without the guidance and enthusiasm of my supervisors – Pete Dahlhaus and Richard MacEwan. You have made a tremendous impression on me in the time I have studied under, and beside you. I hope this is a continuation of the journey towards making a difference to the longevity and sustainability of our landscapes.
- Colleagues and co-authors: Doug Crawford, David Rees, Mark Imhof, Grant Boyle, Kurt Benke, Sorn Norng, Steve Williams, Jonathan Hopley, Rob Clark, Kohleth Chia, Matt Kitching, Michelle Davey, Jenny Alexander, Sonia Thompson – a big thank-you for sustaining the effort in publishing joint works and contributions.
- Federation University and DEDJTR (previously DEPI) – Without the efforts of Frank Stagniti, Peter Vamplew (associated supervisor) Mark Imhof and Angela Avery for supporting industry and federal scholarships, none of this would be possible. Current and past projects including the Understanding Farming Systems

and Soil, and the Land Knowledge Foundations project have underpinned contributions in this thesis.

- CERDI and Federation University researchers (Peter Vamplew (co-supervisor), Helen Thompson, Andrew Macleod, Stafford McKnight, Kim Dowling) – I hope the steps we have taken towards providing new insight and data on soils and landscapes of Victoria continues to be used in valuable approaches and techniques you are establishing.
- The DEDJTR Macleod team – Bruce Shelley, George Croatto, Maureen DePasquale, Ron Walsh, Stuart Wells, Silvia and Carlos Aguirre and Tom Baker. Apologies for occupying all the space with soil samples and getting you into OH&S troubles as a result.
- CSIRO – Mark Raven, Raphael Viscarra Rossel, Peter Wilson, Linda Karsies, Mike Grundy and Ross Searle. The opportunities to travel to workshops and conferences over the last 5-years have been invaluable and only possible with your support. Thanks for your assistance with archive samples, data and guidance.
- DPIPWE (Tasmania) – Darren Kidd and Mat Webb. A big thank-you for helping me with R coding and implementing techniques you have refined. Always great to chat on how the Blues will fair next year!
- DSM colleagues of the working group – Mark Thomas, John Wilford, Karen Holmes, Craig Liddicoat, Jon Gray, Noel Schoknecht, Jason Hill, Dan Brough, Budiman Minasny, Brendan Malone. Always great to catch up and share our ideas in the pursuit of better soil maps.
- Soil surveyors and geoscientists who through their life-time contribution to soil survey in Victoria and other states, we are all indebted. Thanks to John Martin

who has been a ceaseless ball of enthusiasm and continues to unearth lost treasures.

- International colleagues (Ben Marchant, Allan Lilly, Murray Lark and Helaina Black) who have hosted me while visiting, and spent time in Victoria as part of visits.

TABLE OF CONTENTS

Chapter 1 Introduction	1
Understanding user needs for spatial soil information	3
Accommodating various sources of error in modelling and mapping	4
Implementation of spatial prediction methods to derive novel maps for properties linked to soil functions	6
Overall aim:	7
Summary of research contribution	8
Data and methods summary	9
Thesis structure and linkages	10
References	15
Chapter 2 Western Victoria overview	20
Landscapes of western Victoria	21
Climate	26
Vegetation, land use and agriculture	26
References	28
Chapter 3 Spatial soil information: user needs, new prediction methods and uncertainties	30
3.1 Introduction	32
3.2 Why is spatial soil information needed?	33
3.2.1 Global to local needs for spatial soil information	33
3.2.2 Users of spatial soil information in Australia	34
3.2.3 Inadequacy of user needs assessment	37
3.2.4 Evolving needs of users	40
3.2.5 User needs and participatory action	41
3.2.6 Access and delivery of spatial soil information	44
3.3 Soil mapping, benefits, costs and utility	47
3.3.1 The changing purpose and needs for soil mapping in Victoria	48
3.3.2 Supporting government and policy making	58
3.3.3 Benefit vs costs and the investment logic for soil mapping	60
3.3.4 Utility of soil survey, including legacy maps for new applications	64

3.4	Digital Soil Mapping (DSM).....	66
3.4.1	What is DSM?.....	67
3.4.2	DSM and conventional mapping	68
3.4.3	Digital Soil Assessment and Digital Soil Risk Assessment.....	70
3.4.4	Data and knowledge deficiencies.....	72
3.4.5	Method performance/robustness and user perceptions	73
3.4.6	New environmental covariates.....	75
3.4.7	New soil sites	76
3.5	Uncertainty.....	77
3.5.1	What is uncertainty?	77
3.5.2	Uncertainty and risk.....	78
3.5.3	Support, scale and reducing uncertainty	79
3.5.4	User perception	80
3.5.5	Inclusion and representation of uncertainty with DSM products	81
3.6	Research directions	83
	Acknowledgements.....	85
	References.....	85
	Chapter 4 Soil data for biophysical models in Victorian landscapes: current needs and challenges.....	102
4.1	Introduction.....	104
4.1.1	Model limitations	104
4.1.2	Soil data availability	105
4.1.3	Understanding soil data needs of biophysical models	106
4.1.4	Collection of soil data for modelling	106
4.2	Methods.....	107
4.2.1	Study design and data collection.....	108
4.2.2	Biophysical models.....	109
4.2.3	Modelling scale	110
4.2.4	Industry/land use	114
4.2.5	Soil data requirements of models.....	114
4.2.6	Model sensitivity.....	115
4.3	Results.....	118

4.3.1 Model users	118
4.3.2 What models, scales and industries?.....	119
4.3.3 Model sensitivity to soil properties.....	121
4.4 Discussion	122
4.4.1 Soil properties in models	126
4.4.2 Changes in demand for soil property data	127
4.5 Looking forwards.....	129
4.6 Conclusion	131
References.....	132
Chapter 5 Identification and interpretation of sources of uncertainty in soils change in a global systems-based modelling process	141
5.1 Introduction.....	143
5.2 Types of Uncertainty.....	147
5.2.1 Taxonomy of uncertainty	148
5.3 Sources of uncertainty.....	151
5.4 Example 1: Disaggregation of error sources.....	166
5.5 Example 2: Spatial uncertainty and geostatistics.....	170
5.6 Conclusion	174
Future research.....	175
References.....	176
Chapter 6 Improving information content in soil pH maps: a case study in south-western Victoria	182
6.1 Introduction.....	184
6.2 Methods.....	187
6.2.1 Study area.....	187
6.2.2 Land use	187
6.2.3 Soil data	188
6.2.4 Digital Soil Mapping.....	191
6.3 Results.....	195
6.3.1 Establishing fixed and random effects	195
6.3.2 Spatial prediction of pH _w	197
6.3.3 Probability of limiting soil pH conditions	199

6.4 Discussion	200
6.5 Conclusion	205
6.6 Appendix	206
References	207
Chapter 7 Assessment of error sources in measurements of field pH: effect of operator experience, test kit differences and time-of-day	212
7.1 Introduction	214
7.2 Materials and methods	218
7.2.1 Materials	218
7.2.2 Experimental design	221
7.2.3 Statistical analyses	222
7.3 Results	225
7.3.1 Experiment 1	225
7.3.2 Experiment 2	230
7.4 Discussion	235
7.4.1 Assessor experience	235
7.4.2 Model and bias	237
7.4.3 Time-of-day (light quality)	238
7.4.4 Kit type differences	238
7.4.5 General comments	239
7.5 Conclusion	242
References	243
Chapter 8 The 3D distribution of phyllosilicate clay minerals in western Victoria	246
8.1 Introduction	248
8.1.1 Measurement of clay minerals	249
8.1.2 Prediction of clay minerals using IR (VIS, NIR and MIR)	250
8.1.3 Mapping of clay minerals	250
8.2 Methods	252
8.2.1 Study area	252
8.2.2 Soil sites	253
8.2.3 MIR spectra acquisition	253
8.2.4 XRD calibration samples	254

8.2.5 Spectroscopic calibration models	257
8.2.6 Mapping clay minerals.....	260
8.3 Results.....	263
8.3.1 Clay mineral MIR calibration models.....	263
8.3.2 Key absorption features of clay minerals (kaolinite, illite, smectite)	265
8.3.3 Application of clay mineral spectroscopic models to MIR spectra with sites	267
8.3.4 Mapping of clay mineral abundances	269
8.3.5 Interpretation of prediction variables and relationship to western Victoria (kaolinite, illite, smectite)	272
8.3.6 Comparison of clay mineral predictions against existing studies and investigations	281
8.4 Discussion	285
8.4.1 MIR calibration models	285
8.4.2 Maps of soil clay minerals	287
8.4.3 Looking forwards.....	291
8.5 Conclusion	292
References.....	293
Chapter 9 Conclusions	303
Discussion	303
Summaries of research findings.....	305
Research objective 1: Identify what are users' needs for spatial soil information and how this has changed in Victoria over the last century.....	305
Research objective 2: Develop an approach to accommodate, and illustrate to users of spatial soil information, the various error sources in modelling and mapping....	310
Research objective 3: Investigate the potential use of legacy data supplemented with new spectroscopic predictions to predict the regional distribution of two key soil properties - pH and clay mineralogy – for areas of western Victoria.....	314
Overall research conclusions	320
Future work.....	321
References.....	324
Appendix A.....	327
Appendix B	424
Appendix C	500

LIST OF FIGURES

Figure 1.1. Structure and linkages between chapters of the thesis.	12
Figure 2.1. Western Victoria.....	22
Figure 2.2. Landscape features of Western Victoria (background shading is terrain).....	24
Figure 3.1. Generalised soil information requirements.....	36
Figure 3.2. Documented uses of spatial soils information in Victoria between 2009 and 2014 from spatial soil information requests (source: DEDJTR 2014).	38
Figure 3.3. Questions requiring spatial soils information against the six levels of inquiry for the Steinitz framework (Robinson <i>unpublished results</i>).	42
Figure 3.4. Ecosystem services inventory and questions that will require spatial soil information to address these (Robinson <i>unpublished results</i>).	43
Figure 3.5. Spatial soil information products supplied between 2009 and 2012, shown against 2013-2014.	48
Figure 3.6. Primary reasons for soil survey in Victoria as relative proportions (1890-2015).	50
Figure 3.7. Verbal uncertainties using probabilistic data: from Lark et al. (2014).....	83
Figure 4.1. Agricultural industries and land uses in Victoria (2014).....	114
Figure 4.2. Comparison of soil properties that were considered sensitive in models for the 2009 and 2014 surveys.....	123
Figure 4.3. Soil attributes that models were sensitive to and thresholds to identify soil attributes to focus data provision in support of future model applications.....	128
Figure 5.1. Hierarchy of the Steinitz Framework identifies the six primary iterative questions posed in the modelling process (Steinitz, 1990).	152
Figure 5.2. Sources of uncertainty expressed in a conceptual process modelling representation.....	154
Figure 5.3. (a) Original area of an exemplar soil attribute, A (adapted from Benke et al., 2011), and (b) uncertainty in geometric properties represented by standard deviations of soil attribute, A, its boundary, B, its position (centre-of-mass), C, and rotation, D.	157
Figure 5.4. Example of spatial inaccuracy in georeferencing soil sites (see overlaid red lines).....	160
Figure 5.5. Random error possibilities of the true pixel location for a point in a legacy soil site relative to surrounding pixels..	161
Figure 5.6. Data for pH measurements from the Hopkins River and Curdies River catchments in south-western Victoria for the time period 1989-91.....	172

Figure 5.7. Data for pH measurements from the Hopkins River and Curdies River catchments in south-western Victoria for the time period 1992-94.....	173
Figure 6.1. Soil sites and their space-time distribution in south-western Victoria.....	188
Figure 6.2. (a) Land use in 2000 and (b) 2014 and (c) changes between 2000 and 2014.	190
Figure 6.3. pH distribution for sample depth for the two levels (0 – 10 or 0 – 10+ cm).....	194
Figure 6.4. Seasonal differences in pHw.	196
Figure 6.5. Spatial prediction (left) and variance estimates (right) for Model 1 (top) and Model 2 (bottom) for summer.....	198
Figure 6.6. Spatial prediction (left) and variance estimates (right) for Winter (top) and Spring (bottom).....	199
Figure 6.7. Spatial prediction (left) and variance estimates (right) for Summer (top) and Autumn (bottom).	201
Figure 6.8. (a) Mapped probabilities that pHw was less than, or equal to 6.0 (top), and (b) 5.3 (bottom) for autumn.	202
Figure 7.1. Samples used in experiments with pH level.....	219
Figure 7.2. Trellis plot of field pH versus lab pH by Assessor.....	225
Figure 7.3. Trellis plot of field pH versus lab pH by Experience.....	226
Figure 7.4. Field pH versus lab pH with fitted models using LM (black line) and MLFR (blue line). The red dashed line is the 1:1 line.	229
Figure 7.5. Field pH versus lab pH for the four different scenarios - (Case 1) Experienced and 1PM (top left-hand corner), (Case 2) Experienced and 5PM (top right-hand corner), (3) Inexperienced and 1PM (bottom left-hand corner) and (4) Inexperienced and 5PM (bottom right-hand corner). The black line is the 1:1 line.	231
Figure 7.6. Trellis plots of field pH versus lab pH by Kit by Assessor.	232
Figure 7.7. Trellis plot of field pH versus lab pH for Kit by Assessor Type.....	234
Figure 8.1. Soil sites and quantitative XRD calibration sites against 1 st tier geomorphological divisions of the VGF for western Victoria, Australia.....	255
Figure 8.2. MIR spectra of clay mineral calibration samples.....	256
Figure 8.3. Histogram of clay mineral (kaolinite, illite and smectite) content of samples used in PLSR calibration models.	265
Figure 8.4. Clay mineral calibration models for kaolinite, illite and smectite.....	266
Figure 8.5. VIP and SR scores for the PLSR clay mineral models.	266
Figure 8.6. Clay mineral predictions for kaolinite (a), illite (b), smectite (c) and associated errors for samples.....	268

Figure 8.7. Ternary diagram of kaolinite, illite and smectite predictions from the spectroscopic models.	269
Figure 8.8. Map predictions (lower, mean and upper) from left to right for kaolinite (top), illite (middle) and smectite (lower) for 0-5 cm.	275
Figure 8.9. Map predictions (lower, mean and upper) from left to right for kaolinite (top), illite (middle) and smectite (lower) for 5-15 cm.	276
Figure 8.10. Map predictions (lower, mean and upper) from left to right for kaolinite (top), illite (middle) and smectite (lower) for 15-30 cm.	277
Figure 8.11. Map predictions (lower, mean and upper) from left to right for kaolinite (top), illite (middle) and smectite (lower) for 30-60 cm.	278
Figure 8.12. Map predictions (lower, mean and upper) from left to right for kaolinite (top), illite (middle) and smectite (lower) for 60-100 cm.	279
Figure 8.13. Map predictions (lower, mean and upper) from left to right for kaolinite (top), illite (middle) and smectite (lower) for 100-200 cm.	280
Figure 8.14. Major predictor variables used for kaolinite maps.	281
Figure 8.15. Major predictor variables used for illite maps.	282
Figure 8.16. Major predictor variables used for smectite maps.	283

LIST OF TABLES

Table 3.1. Highly sensitive soil parameters for landscape models (from Robinson et al. 2010).	41
Table 3.2. Selected surveys from Victoria against the phases of soil mapping in Australia. ...	59
Table 3.3. Benefit-cost ratios for spatial soil data (survey).	63
Table 4.1. Identified models that have been applied in Victoria.	111
Table 4.2. Land use, estimated area and value of agricultural commodities in Victoria.	115
Table 4.3. Hydrological, physical, chemical, biological and ancillary properties required for models.	116
Table 4.4. Survey summaries.	118
Table 4.5. Responses from modellers to application of a model at a particular spatial modelling scale for agricultural industries for the 2009 workshop and 2014 survey.	119
Table 4.6. Number of modellers to apply a simulation model for the primary agricultural industries from the 2009 workshop and 2014 survey.	120
Table 4.7. Model applications at the four spatial modelling scales for the 2009 workshop and 2014 survey.	121
Table 5.1. Links between decision-making and uncertainty in the Steinitz Framework.	153
Table 5.2. Comparison of error sources for error-budget model for data quality of clay (indicative data from Nelson et al., 2011).	169
Table 6.1. Soil sites (N) for the collection periods.	189
Table 6.2. Summary statistics for pH measurements.	190
Table 6.3. Environmental (spatial) covariates available for model-based geostatistics.	192
Table 6.4. Verbal scale for likelihood (probability) used for pHw scenarios (≤ 5.3 and ≤ 6.0)	195
Table 6.5. Summary statistics for environmental covariates and factors (e.g. Sample depth) used in the Likelihood ratio test for the LMM (Model 2).	196
Table 6.6. Mean soil pH prediction and variance estimates for Model 2; and Model 1 (in brackets).	197
Table 7.1. Site, sampled depth, ASC order and soil properties.	219
Table 7.2. ANOVA for the absolute difference between lab pH and field pH with two types of assessors and thirteen levels of pH.	227

Table 7.3. Summary of LM and MLFR model parameters between pH data measured in the field (from our experiment) as the response variable (y) and pH data measured in the laboratory as the fixed variable (x).	228
Table 7.4. Summary of MLFR model parameters for field pH (y) and lab pH data (x) for four scenarios: (Case 1) Experienced and 1PM, (Case 2) Experienced and 5PM, (3) Inexperienced and 1PM and (4) Inexperienced and 5PM.....	230
Table 7.5. ANOVA for the absolute difference between lab pH and field pH with two kits and twelve levels of pH. Mean values are presented.	235
Table 8.1. Soil properties for MIR calibration samples.	257
Table 8.2. Environmental covariates used in Digital Soil Mapping.	261
Table 8.3. Summary statistics for samples used for calibration purposes (* number of samples used in PLSR model).	263
Table 8.4. Summary statistics for clay mineral data used in the Cubist model trees.....	270
Table 8.5. Map diagnostic results for kaolinite, illite and smectite..	271

“There are few subjects upon which it is more difficult to make an accurate, and at the same time intelligible report, than upon soils. The difficulty arises partly from the nature of the subject and partly from the vagueness of the terms used in speaking of soils”

(T.C. Chamberlin, 1877)

“The difficulty is that attributes which are relevant are not necessarily mappable, just as those that are mappable are not always relevant.”

(Frank Gibbons, 1981)

Oils ain't oils, and soils ain't soils

(Castrol GTX advertisement)

Chapter 1 Introduction

As a physical medium, soil is the veneer of organic and unconsolidated material on the Earth's surface derived from infinite combinations of chemical, physical, biological and morphological properties and characteristics. Soil is a vital natural resource asset that delivers many ecosystem services to support and sustain flora and fauna. Degradation of this global resource is recognised as a major threat to provisioning and regulating ecosystem services such as storing carbon, filtering water, energy production, regulation of green-house-gas emissions and agri-food production (FAO and ITPS, 2015; Robinson et al., 2012). Soil is essential to the interactions of water, plant and atmosphere domains, from water infiltration and filtering, to engineering, or the provision of shelter for ground burrowing organisms. The demand on soil to deliver many services, often at once, places this asset under tremendous pressure given its inherent qualities, fragility and variable resilience. This competing demand for soil and land by humans is recognised in the Millennium Goals (www.un.org/millenniumgoals) and has instigated global initiatives such as the Global Soil Partnership (www.fao.org/globalsoilpartnership) and is ably supported by endeavours such as the GlobalSoilMap.net project (www.globalsoilmap.net).

Australian soils are pedologically diverse, some being very old and of low soil fertility, others, from more recent landscape formations, are young and fertile. Landscapes can be highly vulnerable to degradation from the threats to sustainable land management including salinization, acidification, structure decline, waterlogging and physio-chemical constraints. These are real threats to soil security (McBratney et al., 2014) and soils long-term provisioning roles including a global need to increase yield and quality of food production (van Ittersum et al., 2013). Increasing food security, addressing the significant

loss of arable land to erosion or pollution and the restoration of productive soils are identified as global priorities (FAO and ITPS, 2015). The potential loss of arable land to rising sea-levels and impacts to soil functions from climate change exasperates the current degraded state of soil for many countries. Since European colonisation over 200 years ago, changes in land use and management have contributed greatly to the degraded state of many soils and landscapes for this continent. For example, acidification in surface soil for Australia is considered to impact upon 50 million hectares of agricultural land with estimated annual production losses of \$1.59 billion. In the state of Victoria alone, acidification is estimated to cost over \$470 million each year (NLWRA, 2002). Changes in land use and the adoption of modern farming system practices are occurring across extensive areas of Victoria in response to changing environmental conditions (e.g. decreased rainfall), volatile global commodity markets and evolving labour markets. Understanding the nuances of agro-ecological interactions for these landscapes is challenging. However, by combining spatial knowledge and information on soil and terrain with developing trends in weather patterns, there is the potential to define and tailor management and farming systems for these landscapes.

The demand for spatial soil information¹ for purposes such as land use planning and to support implementation of global climate and carbon models (e.g. Reich and Hobbie, 2013; Wieder et al., 2013) is increasing (Hartemink, 2008; Sanchez et al., 2009). Land use planning has traditionally been guided by land evaluation techniques that use spatial soil information to identify resource potential and limitations, and prescribe inputs (e.g. fertilizer and chemicals) with management strategies to fulfil production capacity while protecting the environment. Conventionally, user needs for spatial soil information has focused on traditional users and frameworks such as land evaluation. A considerably

¹ used to represent the quantitative expression of maps and soil databases in a spatial domain

wider market of users exists that is either unaware or not exposed to this information (Wilson, 2012). In Australia, land use planning and government policy has been supported by soil survey and mapping. However, major deficiencies exist in current and reliable soil data (often referred to as legacy data) to support endeavours to derive spatial soil information for many nations including Australia.

To respond to this inadequacy of spatial soil information, the key objective of this thesis is to demonstrate the importance and contribution of various error sources to uncertainty in the production of spatial soil information relevant to users and linked to soil functions including primary production.

Understanding user needs for spatial soil information

Soil information is required at adequate resolutions (spatial and temporal), in an acceptable delivery format and has accuracies and uncertainties provided that can assist land managers with decisions on agriculture and the environment. These users of spatial soil information, e.g. broadacre dryland farmers, horticultural managers, rural consultants and government policy makers, all have spatial soil information needs that can be as diverse within, as among, user groups. This added complexity is a potential reason why so little published information exists on the needs of users (Omuto et al., 2013). Understanding the agro-ecological context of spatial soil information needs and how questions posed by users are likely to evolve in response to advances in technology and the rapid expansion of sensors is an ongoing challenge to the pedological community.

It is timely given these rapid advances in technology over the last decade (Roudier et al., 2015) and new approaches to soil data acquisition and sharing from crowd sourcing and citizen science (Rossiter et al., 2015) to review what, when and how this information

should be provided for future users. In Australia (with specific examples for Victoria) there has been considerable change in the motivations for producing spatial soil information over the last century that reflect government priorities surrounding colonisation and settlement, enhancing agricultural production, nature and conservation and urban development. Specific and evolving spatial soil information requirements are valuable for key users (e.g. biophysical modellers) that operate across many agricultural and environmental domains as this will support targeted delivery of spatial soil information for current and future uses. By understanding these specific needs, the current paucity of spatial soil information can be addressed through Digital Soil Mapping (DSM, McBratney et al., 2003) which provides an approach to predict soil properties (e.g. pH, EC, salinity, clay content) at various scales (e.g. paddock to catchment) with remotely and proximally sensed data (e.g. geophysics, terrain derivatives) using spatial inference techniques.

Accommodating various sources of error in modelling and mapping

In the delivery of spatial soil information from DSM approaches, it is important that errors and uncertainties are quantified to communicate to users if the information is appropriate for their needs (Carré et al., 2007; Heuvelink 2014). Uncertainty, which is a lack of assurance or conviction (knowledge) in an observation or model (Goovaerts, 1997) has a negative connotation that suggests unreliability. Often in DSM, uncertainty is defined only from statistical uncertainty whereas it is actually based on a dichotomy of aleatory uncertainty (statistical variability or error) and epistemic uncertainty (lack of information).

In the production of a digital soil map using modelling procedures, sources of error that lead to uncertainty can be documented, quantified and their overall contribution to error propagation quantified. Legacy soil site data for example can be littered with issues including data format, lack of harmonisation, imprecision and inadequate georeferencing (Krol et al., 2008). These data errors all contribute to model error which has proven to be a major factor in the aleatory uncertainty of DSM (Nelson et al., 2011). Few if any studies have accounted for spatial and temporal dependence in their mapping implementation including the temporal uncertainty of parameter values used (Finke, 2012). This is a key consideration in the use of model-based techniques to produce predictive maps for properties of interest that are susceptible to change from climatic or anthropogenic impacts.

In modelling and simulation, generally a two-step process is adhered to with model selection uncertainty (epistemic uncertainty) and then statistical variability which is treated through error propagation or other approaches (aleatory uncertainty). While sources of stochastic and epistemic uncertainty have been identified and discussed (Refsgaard et al., 2007; Benke et al., 2011), there appears to be no instances worldwide in soil modelling or mapping where these have been brought together to enable a comprehensive assessment of uncertainty in the analytical process. Approaches to accommodate uncertainty through-out the analytical process from problem definition to prediction and error, and finally implementation in a decision making process require consideration.

Implementation of spatial prediction methods to derive novel maps for properties linked to soil functions

Current mapping efforts through the GlobalSoilMap project have focused on key soil properties linked to carbon and water storage and production (organic carbon, particle size distribution, bulk density, depth to rock or limiting layer, available water capacity, cation exchange capacity), and degradation issues such as acidification and salinization (pH, EC). These properties are the focus of new maps produced for Australia (Grundy et al., 2015) and Africa (Hengl et al., 2015) for example. There is a need to provide contemporary spatial assessments of soil condition and scenario maps (McBratney et al., 2003) for governments and policy interventions. Embodying the time dimension into assessments of soil condition remains a challenge in DSM (Lagacherie et al., 2008) especially when dealing with legacy data that can be sparse in space and time, and changes in land use and management regimes.

There is an enduring deficiency in spatial information for fundamental soil properties such as clay mineralogy (Grunwald, 2009) that are implicitly connected to functions of soil health (Viscarra Rossel, 2011). Functions such as sustaining life and society; resistance to erosion; providing a physical medium for plants, animals and infrastructure; cycling and storage of matter (e.g. carbon); and storage and filtration of water are central to concepts of soil health. Clay mineralogy and pH as soil properties (focus of this thesis) are connected in provisioning roles affecting the availability of nutrients to plants and animals, supporting soil biological communities, storing of organic matter and filtering of water.

Many countries around the globe can benefit from the wealth of legacy data resources they possess including archived soil samples that can be re-used and analysed using non-destructive spectroscopic techniques for properties such as mineralogy (Viscarra Rossel et

al., 2009). Yielding information contained in legacy databases from methods with qualitative and quantitative determinations could be harmonised with spectroscopic models and spatial inference methods to predict the distribution of key properties linked to ecosystem services embodied in soil health.

Opportunities exist to advance our knowledge on these soil properties (pH and clay mineralogy) by exploiting the resources available and succinctly unravelling the stories that can be told on landscape process and interactions with agriculture and the environment. Here the role of the soil scientist is critical to ensure that pedological realities are not lost (MacEwan et al., 2014) and that opportunities exist to reconstruct profiles from the spatial predictions of soil properties under the guidance and evaluation of pedologists.

Overall aim:

To improve the global knowledge of the various sources of error contributing to uncertainty in Digital Soil Mapping, three objectives were defined that focus on aspects of data inputs to DSM and the modelling techniques implemented to deliver maps of key properties linked to primary production for agricultural landscapes.

The primary objectives are to:

1. Define what spatial soil information is sought by users to support biophysical models for agricultural landscapes.
2. Understand and account for potential error sources as input variables in DSM applications.

3. Produce spatial predictions of soil properties (e.g. pH, clay mineralogy) connected to soil functions supporting agriculture.

The specific research aims of this thesis are to:

- a. Identify what are users' needs for spatial soil information and how this has changed in Australia and Victoria over the last century.
- b. Develop an approach to accommodate, and illustrate to users of spatial soil information, the various error sources in modelling and mapping.
- c. Investigate the potential use of legacy data supplemented with new spectroscopic predictions to predict the regional distribution of two key soil properties - pH and clay mineralogy – for areas of western Victoria.

Summary of research contribution

The research questions posed in this project will have national strategic value and relevance to global initiatives such as the Global Soil Partnership and GlobalSoilMap projects. This research project will identify DSM techniques to predict soil properties for key agricultural landscapes of western Victoria, and to define if available legacy data can be used to predict changes for dynamic soil properties in response to natural and anthropogenic impacts. The research will have wider practical applicability across south-eastern Australia and similar soil-landscapes in Mediterranean climatic environments for these key soil properties. Understanding the relative contributions and defining frameworks to accommodate uncertainty for soil assessments will benefit DSM practitioners and make users of these products aware of DSM error sources. Agriculture and the environment are also expected to benefit from new soil maps for properties that

can be applied in precision agriculture, natural resource management and strategic planning.

Data and methods summary

The datasets used in research investigations have been made available by the Department of Economic Development, Jobs, Transport and Resources (DEDJTR), the Commonwealth Scientific and Industrial Research Organisation (CSIRO) as part of the Soil and Landscape Grid of Australia project (Grundy et al., 2015), and undergraduate and postgraduate studies from Federation University Australia.

Soil site data (discussed further in Appendix C) has been collated from various sources including the Victorian Soil Information System (VSIS), CSIRO National Soil Site Database (<https://data.csiro.au>), published and unpublished reports (e.g. Colwell, 1977; Crawford and Robinson, 2014). Sites have been supplemented with Mid Infrared (MIR) predictions for samples where archive samples exist in either the National Soil Archive (www.clw.csiro.au/aclep/archive/) or Victorian Soil Archive (Johnstone, 2011) and have been linked to soil site databases.

Spatial environmental predictors such as terrain models and their derivatives have also been collated from the DEDJTR and the CSIRO. Organising these covariates using a single coordinate system (VicGrid94), establishing a base grid and resampling all datasets has also been undertaken. Numerous derivative products (e.g. Timesat parameters) from core covariates (e.g. MODIS time-series) were delivered as part of this exercise. Ongoing maintenance of these datasets is also necessary for the different study (spatial) extents and resolution of covariates sought.

Mapping methods applied in this research (Cubist model trees, ordinary kriging and Linear Mixed Models (LMM)) are core techniques used by the DSM community. While the original intent was to evaluate a suite of different model procedures, it became apparent that computational proficiency and practicability made it impossible to apply all available techniques.

Data analysis procedures were undertaken using Matlab (www.mathworks.com/products/matlab), GenStat (www.vsnl.co.uk/software/genstat) or R (www.r-project.org). Analysis of MIR spectra were undertaken using Matlab and the PLS toolbox (www.eigenvector.com/software/pls_toolbox.htm). Spectral models were run iteratively and refined as existing quantitative X-Ray Diffraction (XRD) determinations were sourced or new analysis was undertaken.

Uncertainty estimation procedures were also implemented using all of the aforementioned software, Table Curve 3D (www.sigmaplot.com/products/tablecurve3d/tablecurve3d.php) and @RISK (www.palisade.com/risk).

Thesis structure and linkages

The thesis comprises nine chapters, of which five are journal papers as detailed with the candidate's contribution listed in the Preface. Chapter 1 provides an introductory account of the demand for spatial soil information and understanding user needs, uncertainty assessment to account for the various sources of error in modelling and mapping, and delivery of soil mapping for soil properties (pH and clay mineralogy) that are linked to soil functions and services. Chapter 2 describes the study region of interest for the thesis. A literature review (Chapter 3) provides background context to research investigations presented in chapters 4 to 8. Findings from presented research are synthesized and

summarised in Chapter 9 as a conclusion to this thesis. The linkages between chapters and their organisation for the thesis are illustrated in Figure 1.1. The theme that links these chapters is the importance of understanding user needs for spatial soil information, tailoring this information through integrating legacy data with new methods to provide greater certainty in decisions by users of soil information.

Chapter 2 describes western Victoria and its physiographic setting, climate, geology and geomorphology, soils and land use. While there is inevitable duplication here with regional context information presented in later chapters (5, 6 and 8), this study region overview provides a more detailed and comprehensive description of western Victoria than any of the individual papers. Chapters 3, 4 and 7 are largely based upon data (including soils, industry and natural resource management) for Victoria. Attempts have been made to minimise duplication, however this has been inevitable with the structure of this thesis by incorporating published papers.

Chapter 3 presents a detailed literature review that explores the history of soil mapping in Victoria, its roots in colonisation and changing emphasis over the last century. The review discusses the evolution of conventional soil survey to Digital Soil Mapping and the challenges that exist to its successful widespread implementation (e.g. understanding user needs, uncertainty and its communication to users). Baseline concepts on the evolution of soil mapping and how the needs or priorities for soil mapping for Victoria, Australia have evolved are discussed including the investment logic for soil mapping. The new digital age in soil mapping is detailed with future potential to use a wealth of free and accessible spatial datasets. Topics of this review are central to the research questions posed and research initiated in the following chapters.

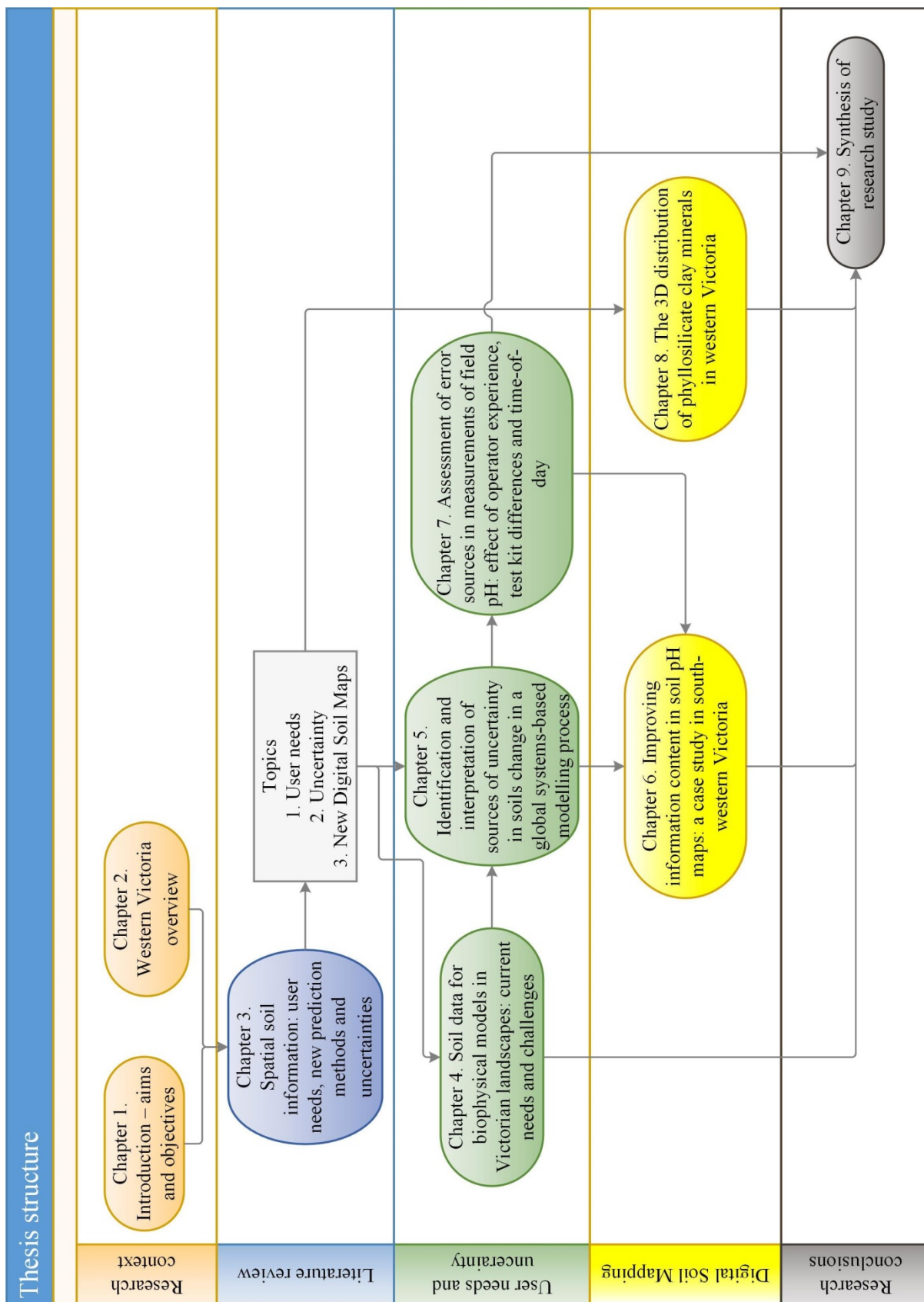


Figure 1.1. Structure and linkages between chapters of the thesis.

Chapter 4 is a review of the soil data requirements for biophysical modellers, a key primary user group of spatial soil information. Changing patterns of model application that reflect government and industry priorities to enhance primary production are discussed with an increasing demand for higher resolution digital soil mapping. Soil properties that affect model sensitivity are identified, although some properties considered important in land evaluation and linked to soil functions and processes (e.g. pH and clay mineralogy) were either overlooked or excluded as part of this review. This may be due to a mismatch between the models and their simplified data needs, inadequate agro-ecological process understandings or failure to deliver soil data at fine scale for properties of interest.

Chapter 5 presents a novel systems-based framework to integrate various sources of error including geometry, position and polygon attributes for modelling and mapping purposes. Two case studies, one of which is linked to changes in soil pH are presented as examples of the Global Representation of Uncertainty in the Modelling Process (GRUMP) framework. The importance of these errors contributing to uncertainty and communicating these to users through maps or models is important to support appropriate use of this information. Identified as a research priority, this chapter explores the many aspects of uncertainty analysis such as statistical variability (aleatory uncertainty) which is often discussed in the absence of epistemic uncertainty (lack of information), and connects these as part of a more fulsome approach to uncertainty analysis than previous examples.

Chapter 6 provides a logical extension and implementation of the GRUMP framework presented in Chapter 5, using an example implementation of the uncertainty framework in Digital Soil Mapping for pH in south-western Victoria. As a threat to primary production and protection of arable land, soil acidification and understanding the baseline on soil pH

is a priority for many countries. Epistemic error sources that contribute to uncertainty in the mapping of soil pH are identified and discussed in detail including temporal variability, harmonisation of legacy observations, integration of expert opinions and model structure adequacy from epistemic learnings. Spatial predictions of pH using conditional simulation and a Linear Mixed Model approach are transformed into more informative products for users of spatial information for agronomic and land management decisions. This is achieved by focusing on critical agronomic thresholds to plant production and likelihood of being below these thresholds.

Chapter 7 continues the focus on soil pH by examining if the prediction accuracy and error in a field pH determination method using different field kits and user experience were significant, and how this would affect the relationship between field and laboratory pH measurements. Given the spatial and temporal paucity of available laboratory pH measurements across large areas of Australia (and Victoria), field pH measurements with greater certainty could be used to populate gaps in available pH observations for mapping and monitoring purposes at regional to national scales. Two experiments were undertaken to examine effects and error due to different test kits, assessor experience and the time of assessment. This enabled the error bounds of the prediction and confidence intervals to be defined and confirmed factors that contribute to field pH uncertainty can be addressed with adequate training and quality assurance procedures to minimise potential errors.

Chapter 8 presents a novel approach using quantitative XRD analysis to calibrate MIR spectroscopy and implement these predictions using model trees to map clay mineral distribution. Soil survey has tended to focus on the collection and measurement of properties that are easy to observe. Clay mineralogy however is expensive and time consuming to acquire, therefore it is rarely measured or observed and hence a likely reason for its exclusion in soil data needs for many users including biophysical modellers.

The implementation of calibration models to predict clay mineral abundance for exhaustive spectral libraries was undertaken for nearly 3000 sites (11,500 samples) in western Victoria. Predictions were harmonised to the six GlobalSoilMap specified depth intervals. Spatial modelling methods including cross-validation procedures and linkages to key soil forming factors from model-tree implementation are described. Kaolinite was found to be the dominant clay mineral across all six depth intervals, followed by smectite then illite. The approach has delivered useful results based upon relatively few calibration samples and can be easily implemented by organisations with available spectral libraries.

Chapter 9 is a summary of the key findings and associated strengths/weaknesses of the research investigations undertaken. The chapter draws together the threads of discussion from research presented across the five papers and identifies future directions for research including areas of further investigation on the topics of understanding user needs, and delivering useful and specific spatial soil information with uncertainties defined. Supplementary information and detail is provided in: Appendix A (selected conference papers and journal papers with the candidate as lead or co-author on Digital Soil Mapping implementation in Victoria, error sources associated with digital soil mapping production and assessment of measurement errors in soil analytical chemistry data; Appendix B (five book chapters with the candidate as lead or co-author on Digital Soil Mapping and uncertainty assessment); and Appendix C (an overview of western Victorian landscapes, their uses and inherent characteristics (e.g. geology, geomorphology, soil) and background soil sites and mapping for this region).

References

Benke, K.K., Wyatt, R.G., Sposito, V.A., 2011. A discrete simulation approach to spatial allocation of commodity production for revenue optimisation over a local region. *Journal of Spatial Science* 56(1), 89-101.

- Carré, F., McBratney, A.B., Mayr, T., Montanarella, L., 2007. Digital soil assessments: Beyond DSM. *Geoderma* 142(1-2), 69-79.
- Colwell, J.D., 1977. National soil fertility project. Volume 2. Soil fertility relationships CSIRO Division of Soils in collaboration with State Departments of Agriculture and the Fertilizer Industry, Adelaide.
- Crawford, D., Robinson, N.J., 2014. A pilot survey of change in soil pH using legacy sites, *Soil Change Matters*. Department of Environment and Primary Industries, Bendigo, Victoria.
- FAO, ITPS, 2015. Status of the World's Soil Resources (SWSR) - Technical Summary. Food and Agriculture Organisation of the United Nations and Intergovernmental Technical Panel on Soils, Rome, Italy.
- Finke, P.A., 2012. On digital soil assessment with models and the Pedometrics agenda. *Geoderma* 171–172, 3-15.
- Goovaerts, P., 1997. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, New York.
- Grundy, M.J., Rossel, R.A.V., Searle, R.D., Wilson, P.L., Chen, C., Gregory, L.J., 2015. Soil and Landscape Grid of Australia. *Soil Research* 53(8), 835-844.
- Grunwald, S., 2009. Multi-criteria characterization of recent digital soil mapping and modeling approaches. *Geoderma* 152(3-4), 195-207.
- Hartemink, A.E., 2008. Soils are back on the global agenda. *Soil Use and Management* 24(4), 327-330.
- Hengl, T., Heuvelink, G.B.M., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Shepherd, K.D., Sila, A., MacMillan, R.A., Mendes de Jesus, J., Tamene, L., Tondoh, J.E., 2015. Mapping Soil Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current Predictions. *PLoS ONE* 10(6), e0125814.
- Heuvelink, G.B.M., 2014. Uncertainty quantification of GlobalSoilMap products. In: Arrouays, D. et al. (Eds.), *GlobalSoilMap, Basis of the global spatial soil information system*. pp. 335-340. CRC Press, The Netherlands.

- Johnstone, P., Robinson, N., Imhof, M., Rees, D., Hunter, D., Crawford, D., Wells, S., Walsh, R., Faderson, A., Williams, S., Kitching, M., 2010. Systems integration: enhancing soil information delivery to Victoria. 19th World Congress of Soil Science, Soil Solutions for a Changing World, Brisbane, Australia, pp. 11-14.
- Krol, B.G.C.M., 2008. Towards a Data Quality Management Framework for Digital Soil Mapping with Limited Data. In: Hartemink, A.E., McBratney, A.B., Mendonça-Santos, M.L. (Eds.), *Digital Soil Mapping with Limited Data*. Springer, pp. 137-149.
- Lagacherie, P., 2008., Digital Soil Mapping: A State of the Art. In: Hartemink, A.E., McBratney, A.B., Mendonça-Santos, M.L. (Eds.), *Digital Soil Mapping with Limited Data*. Springer, pp. 3-14.
- MacEwan, R.J., Robinson, N.J., Imhof, M.P., Rees, D.B., Hopley, J., Crawford, D.M. 2014. Reflections on GlobalSoilMap.net, DSM and the legacy of soil survey information in Victoria, Australia – the pedology matters. In: Arrouays, D. et al. (Eds.), *GlobalSoilMap, Basis of the global spatial soil information system*. CRC Press, The Netherlands, pp. 145-148.
- McBratney, A.B., Field, D.J., Koch, A., 2014. The dimensions of soil security. *Geoderma* 213, 203-213.
- McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. *Geoderma* 117(1-2), 3-52.
- National Land and Water Resources Audit, 2002. *Australians and Natural Resource Management*. National Land and Water Resources Audit, Commonwealth of Australia.
- Nelson, M.A., Bishop, T.F.A., Triantafyllis, J., Odeh, I.O.A., 2011. An error budget for different sources of error in digital soil mapping. *European Journal of Soil Science* 62(3), 417-430.
- Omuto, C., Nachtergaele, F., Rojas, R.V., 2013. State of the art report on global and regional soil information: Where are we? Where to go? FAO.

- Refsgaard, J.C., van der Sluijs, J.P., Højberg, A.L., Vanrolleghem, P.A., 2007. Uncertainty in the environmental modelling process - a framework and guidance. *Environmental Modelling & Software* 22(11), 1543-1556.
- Reich, P.B., Hobbie, S.E., 2013. Decade-long soil nitrogen constraint on the CO₂ fertilization of plant biomass. *Nature Clim. Change* 3(3), 278-282.
- Robinson, D.A., Hockley, N., Dominati, E., Lebron, I., Scow, K.M., Reynolds, B., Emmett, B.A., Keith, A.M., de Jonge, L.W., Schjøning, P., Moldrup, P., Jones, S.B., Tuller, M., 2012. Natural Capital, Ecosystem Services, and Soil Change: Why Soil Science Must Embrace an Ecosystems Approach. *Vadose Zone Journal* 11(1).
- Rossiter, D.G., Liu, J., Carlisle, S., Zhu, A.X., 2015. Can citizen science assist digital soil mapping? *Geoderma* 259–260, 71-80.
- Roudier, P., Ritchie, A., Hedley, C., Medyckyj-Scott, D., 2015. The rise of information science: a changing landscape for soil science. *IOP Conference Series: Earth and Environmental Science* 25(1), 012023.
- Sanchez, P.A., Ahamed, S., Carré, F., Hartemink, A.E., Hempel, J., Huising, J., Lagacherie, P., McBratney, A.B., McKenzie, N.J., Mendonça-Santos, M.L., Minasny, B., Montanarella, L., Okoth, P., Palm, C.A., Sachs, J.D., Shepherd, K.D., Vagen, T-G., Vanlauwe, B., Walsh, M.G., Winowiecki, L.A., Zhang, G-L., 2009. *Science*, 325, 680-681. <http://www.sciencemag.org/content/325/5941/680.short>.
- van Ittersum, M.K., Cassman, K.G., Grassini, P., Wolf, J., Tittonell, P., Hochman, Z., 2013. Yield gap analysis with local to global relevance - a review. *Field Crops Research* 143, 4-17.
- Viscarra Rossel, R.A., 2011. Fine-resolution multiscale mapping of clay minerals in Australian soils measured with near infrared spectra. *J. Geophys. Res.* 116(F4), F04023.
- Viscarra Rossel, R.A., Cattle, S.R., Ortega, A., Fouad, Y., 2009. In situ measurements of soil colour, mineral composition and clay content by vis-NIR spectroscopy. *Geoderma* 150(3-4), 253-266.

Wieder, W.R., Bonan, G.B., Allison, S.D., 2013. Global soil carbon projections are improved by modelling microbial processes. *Nature Clim. Change* 3(10), 909-912.

Wilson, P., 2012. Improving Australian soil data and information governance. CSIRO, Canberra.

Chapter 2 Western Victoria overview

This chapter provides an overview of the physical geography for the landscapes of western Victoria. The use of these landscapes for agricultural production, nature conservation and colonisation since European settlement is discussed. The clearance of native vegetation from land and changes in land use over the last 180 years is described in this chapter. Summaries are provided of the geological and current geomorphological setting which when connected with climate and vegetation influence the distribution of soil types and their characteristics. A comprehensive overview of the physiographic, climatic, geological, geomorphological, soil and land use setting for the region is provided in Appendix C.

To enable a thorough evaluation of Digital Soil Mapping and research techniques developed as part of this thesis, Western Victoria and areas within were selected for chapters 5, 6 and 8 due to the diversity and complexity of landscapes within. The region is noted for its highly productive landscapes from farming systems including wool, lamb, grains (cereal and pulses), beef, dairy and horticulture. Interactions of land use with soil, and changing uses for land were considered in the selection of this region. The findings from this research are anticipated to have wider applicability with the soil mapping community in landscapes of south-eastern Australia and like Mediterranean environments.

Study areas for investigations presented in this thesis are all located within the bounds of Western Victoria which comprises over 135,000 km² of agricultural and environmental landscapes (Figure 2.1). Soil management issues known to limit production capacity in western Victorian are summarised in assessments for the Primary Production Landscapes of
Victoria

(http://vro.agriculture.vic.gov.au/dpi/vro/vrosite.nsf/pages/primary_prod_landscapes)

including the effects of soil acidity in surface and subsoil, limitations to plant roots caused by shrink-swell clays, soils that can be prone to compaction, and the resilience of soils when challenged with physical and chemical changes due to management or climate impacts. Changing climatic conditions and global agricultural commodity volatility have resulted in large structural adjustment (i.e. include more agricultural commodities in their income stream) for primary producers in western Victoria. Changes in land use and regional de-population trends are set against a backdrop of growing global demand for food and loss of arable land (FAO and ITPS, 2015) that is likely to require soils to become more productive.

Landscapes of western Victoria

Western Victoria has erosional and depositional landscapes as reflections of their diverse geological and climatic evolution. Landscapes are spatially delineated using a hierarchical system called the Victorian Geomorphology Framework (VGF; Rees et al., 2010) that combines areas of common geological, landform, climate, soils and vegetation. For western Victoria, with five tier-one divisions of the VGF have been delineated including the North Western Dunefields and Plains, Northern Riverine Plains, Western Uplands, Western Plains and Southern Uplands. The north of the region is bound by the Murray River as part of the Murray Basin of south eastern Australia (Figure 2.2). Nested within the Murray Lowlands are landscapes of the Northern Riverine Plain and the North Western Dunefields and Plains (otherwise known as the Victorian Mallee and Wimmera Plain). Trending east-west across western Victoria is the Great Dividing Range with highlands that terminate at the western margin of the Dundas Tableland - the Glenelg

River (Hills, 1975). Plateaus, strike ridges and valleys of the Western Uplands; the low elevation ranges from faulted and tilted sandstone blocks of the Southern Uplands; and the extensive volcanic and sedimentary deposits of the Western Plains form the southern part of western Victoria.

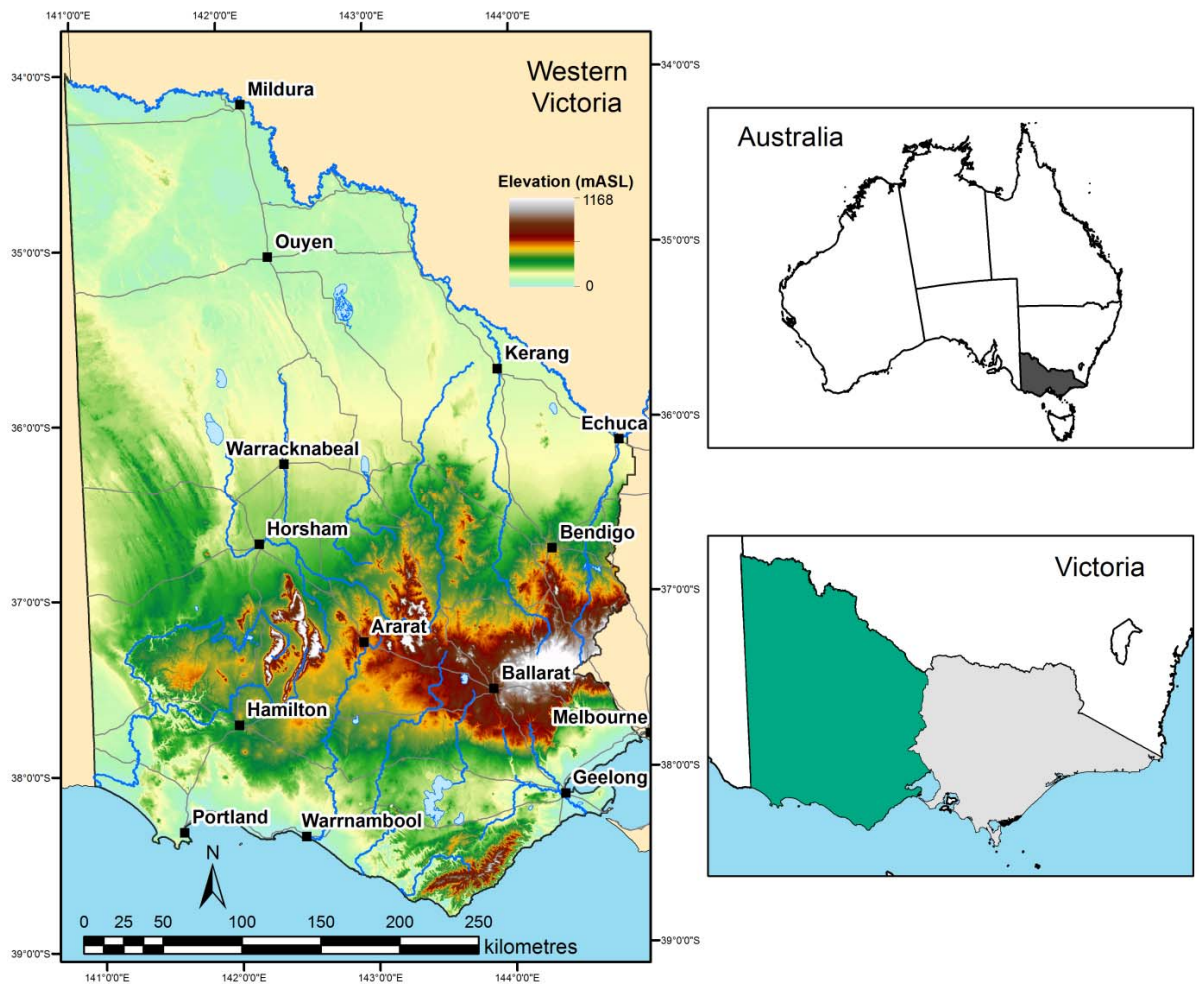


Figure 2.1. Western Victoria.

The Northern Riverine Plain includes a series of modern major tributaries (Campaspe, Loddon, Richardson, Avoca and Wimmera rivers) that begin as streams in the Western Uplands and flow north towards the Murray River. Lakes and basins (often bordered with lunettes) are scattered across the Riverine Plains and fill from periodic inflows during

floods, localised rainfall events and groundwater discharge. Red texture contrast soils (Sodosols; Isbell, 2002) associated with the prior stream complexes (Butler, 1950) are common with Calcarosols and Vertosols less so.

The Victorian Mallee north western Victoria have formed from a series of arid phases in the mid to late Pleistocene period (Bowler et al., 2006). Calcareous deposits from lacustrine, aeolian and alluvial sediments blanket much of the underlying geology including a series of NNW/SSE trending stranded beach ridges with intervening depressions and flats. Gradational or uniform soils (Vertosols and Calcarosols) with texture contrast soils (Sodosols) occur across the aeolian landscapes. Parallel and parabolic dunes from erosion of these ridges form extensive siliceous sand sheets with sandy soils (Tenosols and Rudosols) of the Big, Little and Sunset deserts (Lawrence, 1966).

Further south, the self-mulching to poorly structured clay plains of the Wimmera are extensive and overlay the series of stranded beach ridges to the north of the West Victorian Uplands. To the south of the Wimmera Plain is the West Victorian Uplands (also known as the Western Highlands; Hills, 1940). The highest peaks include Mt William (1168 m) and the Major Mitchell Plateau (over 1100 m) of the Grampians Ranges. The Grampians Ranges have abundant sandy soils that have weak to no pedological development. To the west of the Grampians Ranges is the Dundas Tableland as a low elevation plateau (320-360 m). This tableland has been deeply weathered, tilted and faulted giving it a domed topography. Texture contrast soils (Chromosols and Sodosols) with abundant ferruginous nodules in bleached horizons above strongly mottled (often as tiger mottles) dense clay subsoils are widespread across this landscape.

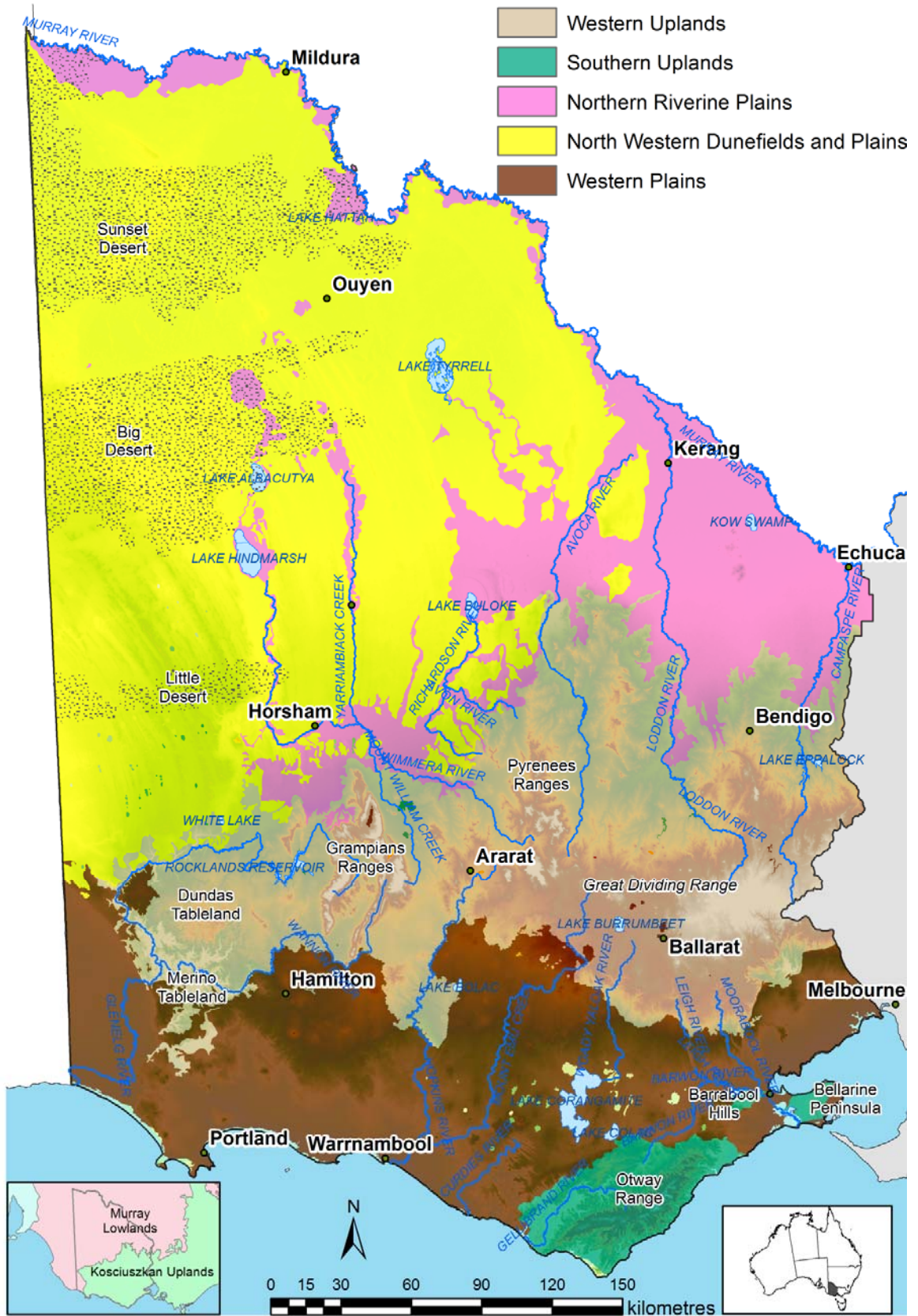


Figure 2.2. Landscape features of Western Victoria (background shading is terrain).

Ridges, plateaux and hills east of the Grampians Ranges include the peaks of Mt Macedon (1001 m) and the Camels Hump (1011 m), Mount Buangor (966 m) and volcanic cones of Mt Buninyong (745 m) and Mt Warrenheip (741 m). The upland bedrock residuals comprise volcanic, sedimentary, metamorphic rocks and granitic plutons of Cambrian and Ordovician age (Joyce et al., 2003). Undulating hills and valleys with alluvial systems occur on northern and southern slopes of the Great Dividing Range. Texture contrast soils are common and occur with sodic subsoils (red, brown and yellow); non-sodic subsoils in slightly higher rainfall environments (Chromosols); and occasionally with acidic subsoils (Kurosols). Dermosols (gradational texture profiles) including iron rich soils (Ferrosols) are found on greenstone ridges and basalt flows. The northern slopes are gentle and asymmetrical while southern slopes tend to be shorter and deeply dissected (Hills, 1975).

South of the West Victorian Uplands are the West Victorian Plains encompassing the Western District Volcanic Plains (Hills, 1940). Volcanic eruptions have produced an extensive veneer of basalt, scoria and ash that extends into South Australia. Basalt flows are layered with inter-mixed ash and scoria that have buried palaeosols. Sodosols, Chromosols and Vertosols are found in various associations depending upon the age of the parent material, its formation and climatic history. Dermosols and Ferrosols are associated with more recent volcanic deposits. Stream systems are poorly developed due to the disruptive volcanic history with stony rises representing the most recent volcanic activity and some of the youngest landforms in Australia (Stone et al., 1977).

Acidic texture contrast soils (Kurosols) and sandy soils of various pedological organisation (Podosols and Rudosols) are found in the dune, swamp and plain series of the Millicent Plain extending south towards the Victorian coastline where strongly

structured red and friable loams (Ferrosols and Chromosols) have formed on the limestone calcareous dunes.

In the southern-most areas of western Victoria are the coastal plains and elevated fault blocks of the South Victorian Uplands including the Otway Range, Barrabool Hills and Bellarine Peninsula. The plateaux and gentle slopes include strongly structured gradational soils (Dermosols) that can be deep and organic-rich, with shallow and stony soils occurring on steeper slopes.

Climate

Western Victoria includes the driest and wettest landscapes in Victoria, from the semi-arid Murray Dunefield in the north to the temperate plateau and valleys of the Otway Range in the south. Paleoclimates have been key determinants of landscape evolution including soil distribution and formation, e.g. aeolian-arid phases in the Mallee. Winter and spring with July and August are traditionally the wettest months from modern records. Generally rainfall in south-western Victoria exceeds evapotranspiration. On the plateau of the Otway Range, Weeaprounah has Victoria's highest mean annual rainfall of 1936 mm (records from 1901 to 2014). In northern Victoria, the climate story is considerably different. Evapotranspiration here can be 5 times greater than annual rainfall and has significant implications for primary production due to short growing seasons. Ouyen in the central Mallee has a mean annual rainfall of 331 mm (1911 to 2015) with October as the wettest month (mean monthly rainfall of 34 mm).

Vegetation, land use and agriculture

Since European settlement, selective clearance of native vegetation has occurred largely for agricultural purposes. The intensification of land use and impacts on the natural resources have altered many of these ecosystems for human purposes (MacEwan et al.,

2010). As a result, there has been considerable change to the vegetation of some bioregions (e.g. Victorian Volcanic Plains) while other bioregions (e.g. Otway Range) remain largely intact (Dahlhaus, 2012).

Colonisation by early European squatters began in the 1830's where large areas of land were taken for purposes of grazing by livestock. This settlement period and the discovery of gold in the 1850's had widespread impacts on land, its use and how it was managed. Significant clearance of native vegetation begun in earnest to support the gold rush that started at Ballarat and Buninyong in 1851. Native forests and grasslands were rapidly modified or cleared for mining and agricultural purposes to support the influx of migrants from overseas. Agriculture continued to thrive in the following two decades with the decline in mining and development of new railway lines between Melbourne and Ballarat and Bendigo. Land clearance and removal of native vegetation continued to support these developments (Nathan, 1999). Dryland agriculture had spread further north into the drier parts of the state including the first pastoral leases in the Mallee country. The first wave of soldier settlement schemes were enacted to support returning servicemen post the First World War, leading to higher density agriculture including the formation of irrigation systems in the north near Kerang and along the Murray River (Powell, 1970). Clearance of Mallee landscapes was initiated with the modernization of farming through mechanical harvesters, new cereal varieties and land management practices, e.g. fallowing. The first 30-years of the 20th century coincided with rapid advances in agriculture, but also considerable damage to soil from erosion due to wind and water from mis-management (Victorian Institute of Surveyors, 1940). Effectively agriculture and the mining of soil was the next resource boom post the gold rush.

The end of the Second World War saw a second wave of soldier settlements developed as part of rural recovery efforts across Australia (Powell, 1970). Replacement of native

pastures with 'improved pastures' and increasing knowledge on the role of soil fertility and production lead to some rapid increases in production from land including that which was previously considered worthless to agriculture (e.g. northern Mallee). Population centres such as Ballarat, Geelong, Bendigo, Warrnambool and Mildura have continued to expand and grow (resulting in land use change) with the decline in rural populations outside these growth centres. Significant increases in land value prices over the last two decades with changing climatic conditions has favoured the expansion of grains production in southern Victoria where it was previously considered too wet to do so (Myers, 1963).

References

- Bowler, J.M., Kotsonis, A., Lawrence, C.R., 2006. Environmental evolution of the Mallee region, Western Murray Basin. *Proceedings of the Royal Society of Victoria* 118(2), 161-210.
- Butler, B.E., 1950. Theory of prior streams as a causal factor in the distribution of soils in the Riverine Plain of southeastern Australia. *Australian Journal of Agricultural Research* 1, 231-252.
- Dahlhaus, P.G., 2012. Salinity risk in the Corangamite Region, Australia, Flinders University.
- FAO, ITPS, 2015. Status of the World's Soil Resources (SWSR) - Technical Summary. Food and Agriculture Organization of the United Nations and Intergovernmental Technical Panel on Soils, Rome, Italy.
- Hills, E.S., 1940. *The Physiography of Victoria*. Whitcombe & Tombs, Melbourne.
- Hills, E.S., 1975. *Physiography of Victoria. An Introduction to Geomorphology*. Whitcombe & Tombs, Melbourne.
- Isbell, R.F., 2002. *The Australian Soil Classification. Revised Edition*. Australian Soil and Land Survey Handbooks Series 4. CSIRO Publishing, Melbourne.

- Joyce, E.B., Webb, J.A., Dahlhaus, P.G., Grimes, K.G., Hill, S.M., Kotsonis, A., Martin, J.J., Mitchell, M.M., Neilson, J.L., Orr, M.L., Peterson, J.A., Rosengren, N.J., Rowan, J., Rowe, K., Sargeant, I.J., Stone, T., Smith, B.L., White, S., Jenkin, J.J., 2003. Geomorphology, the evolution of Victorian landscapes. In: Birch, W.D. (Ed.), *Geology of Victoria*. Geological Society of Australia (Victoria Division), Melbourne, pp. 533-561.
- Lawrence, C.R., 1966. Cainozoic stratigraphy and structure of the Mallee region, Victoria. *Proceedings of the Royal Society of Victoria* 79, 517–554.
- MacEwan, R.R., Reynard, K., Robinson, N.J., Imhof, M., Morse-McNabb, E., McAllister, A., 2010. Land use change: understanding and managing soil dynamics. *Proceedings of the Royal Society of Victoria* 122(2), 29-42.
- Myers, D.J., 1963. A reconnaissance survey of soil waterlogging in Southern Victoria. Victorian Department of Agriculture, Melbourne.
- Nathan, E.L., 1999. Dryland Salinity on the Dundas Tableland: a historical appraisal. *Australian Geographer* 30(3), 295-310.
- Powell, J.M., 1970. *The Public Lands of Australia Felix*. Oxford University Press, Melbourne.
- Stone, J., Peterson, J.A., Fifield, L.K., Cresswell, R.G., 1997. Cosmogenic chlorine-36 exposure ages for two basalt flows in the Newer Volcanics province, western Victoria. *Proceedings of the Royal Society of Victoria* 109, 121–131.
- Victorian Institute of Surveyors, 1940. *Soil Erosion in Victoria*. A Symposium arranged by the Victorian Institute of Surveyors. Department of Lands and Survey, Melbourne. 176p.

Chapter 3 Spatial soil information: user needs, new prediction methods and uncertainties

Why do we produce soil maps? Aren't they just pretty wall paper from a stamp collecting phase? Throughout human history there has been a desire to better understand the occurrence of natural phenomena such as flora, fauna, hydrology, climate and geology and their spatial interactions. Pedology and pedodiversity represent one such natural phenomenon which is largely a manifestation of many interacting phenomena known as soil forming factors (Dokuchaev, 1886; Jenny, 1941). There has been a desire to understand ecology and agricultural dynamics throughout civilization. This has led to a perceived need for spatial soil information for purposes such as land evaluation to enhance primary production from soil without compromising its longevity. Challenges remain to ensure that these 'pretty maps' aren't just static statements in time, rather they become the spatial template from which we communicate with a wider audience the latest knowledge on the issues of food security and production, carbon storage, water storage and infiltration and human health.

This review strives to establish some baseline concepts on the evolution of soil mapping and how the needs or priorities for soil mapping from a retrospective glance for Victoria, Australia have evolved. The roles of soil science and policy making will be touched upon including the investment logic for soil mapping and failings in existing paradigms such as willingness to pay, user pays and benefit-cost. With the rapid uptake of technology over the last few decades, the production of soil maps has evolved greatly leading to a new digital age in soil mapping (Digital Soil Mapping). These advances are encouraging and allow the exploitation of a wealth of available spatial environmental predictors, e.g. terrain models. However, the promise of this new suite of seamless and easily updatable

maps raises fundamental issues that exist for any soil map: who are the users, what format do they want the information in and is the map fit for their required purpose? A decline in new soil survey and the potential loss of existing pedological knowledge is also a potential issue.

The review briefly defines some of these concerns and raises opportunities for the advancement of modern soil mapping practices. This includes the importance of understanding error sources and their contribution to uncertainty in spatial soil information (topic of Chapters 5, 6 and 7). Ideas are synthesized and discussed at the completion of this chapter setting the scene for the following chapters and the research questions posed (listed in Chapter 1).

3.1 Introduction

Soil mapping is fundamental to land evaluation, understanding the interactions of physical, chemical and biological processes in the pedosphere (Bouma, 1989) and the variability of these phenomena in space and time. Soil information is recognised as one of the five pillars of action for the Global Soil Partnership (www.fao.org/globalsoilpartnership) and one of the five goals of the National Soil Research, Development and Extension Strategy in Australia (Department of Agriculture 2014).

Previous works have summarised for Australia the history of: soil mapping (Taylor, 1970; Gibbons, 1983); soil classification (Isbell, 1992) and Digital Soil Mapping (Bui, 2006). Recent productions by Minasny and McBratney (2015) and Brevick et al. (2015) paint a historical and future scene for soil mapping and pedology. This review explores the history of soil mapping in Victoria, Australia as context for issues on: understanding user needs for spatial soil information, the evolution of soil mapping to the current soil mapping paradigm – Digital Soil Mapping (DSM), and the importance of quantifying and communicating uncertainty in spatial soil information to users. The review is structured around the topics:

- Why is spatial soil information needed?
- Soil mapping, benefits, costs and utility,
- Digital Soil Mapping, and
- Uncertainty.

For this review, the terms *spatial soil information* is used to quantitatively represent the maps and soil databases in a spatial domain, and *soil attributes* is used to represent soil

properties (e.g. pH, organic carbon, mineralogy, texture, clay %, total phosphorus) or classes (e.g. soil order, structure, permeability, drainage or colour).

3.2 Why is spatial soil information needed?

3.2.1 Global to local needs for spatial soil information

There is an acknowledged global demand for spatial soil information (Hartemink 2008; Sanchez et al., 2009) to support global climate and carbon models (e.g. Reich and Hobbie, 2013; Wieder et al., 2013). Climate and carbon models (Amundson et al., 2015) require soil information to represent soil processes embedded in these models.

There has been a lack of reliable high-resolution spatial soil information to support requirements at global (Sanchez et al., 2009), national (McKenzie et al., 2008), state and territory (Robinson et al., 2010) scales. Recent efforts to address this deficiency in spatial soil information at a global scale include SoilGrids1km (Hengl et al., 2014) and organic carbon change assessments (Stockmann et al., 2015). At a national scale, the Soil and Landscape Grid of Australia (Grundy et al., 2015) using three-dimensional spatial modelling (Viscarra Rossel et al., 2015) has produced soil maps for 10 soil properties including organic carbon, pH and particle size fractions (clay, sand and silt). Tasmania (Kidd et al., 2015) and Western Australia (Holmes et al., 2015) are examples of recent state-based soil mapping efforts.

In Australia, there still remains little reliable information suitable for decision-making even post the pleas of McKenzie (1991). This is largely due to most soil data used in producing spatial soil information being outdated, incomplete, inconsistent or unavailable. The deficiency in reliable spatial soil information coincides with an increasing and diverse set of uses for soil mapping, and has led to an evolution of

traditional approaches to adapt modern techniques in analysis and presentation (Basher, 1997).

3.2.2 Users of spatial soil information in Australia

The National Soil Research, Development and Extension Strategy for Australia (Department of Agriculture, 2014) identified the primary users of soil information as broadacre dryland farmers, tree plantation managers, irrigation farmers, rural consultants (e.g. agronomists), infrastructure managers, catchment management authorities, government authorities at local, state and national levels, and the scientific community (researchers). These user groups are consistent with those identified by Omuto et al. (2013) as part of a global user needs assessment.

What information do users want

There is a variety of soil properties required for modelling, research and decision making purposes including soil moisture availability, nutrition, toxicity and physical constraints. The major physical and chemical properties are the most frequently sought as they are linked to global programs to increase food production. All users would prefer high resolution data, but it is unclear how many users actually need data of this resolution to make informed judgements. There remains a preference for site data especially by modellers and researchers, whereas in contrast, those in policy and extension are more interested in continuous and consistent products for extension purposes at coarse resolution (Wood and Auricht, 2011). Biophysical modellers represent a large user group seeking digital soil maps. Modellers' preference for soil information is varied though, depending upon the agricultural industry and their operational scale (e.g. site, paddock, region or state). Access to primary soil information (i.e. data) will remain important as

uses for soil information will continue to evolve through adaptation for purposes linked to agriculture and the environment (Alexander et al., 2015). Given the uses of data have changed and will continue to do so, it will be important to have access to all the primary data that is available, allowing adaptation to a variety of purposes.

Within user groups, information needs can be diverse exposing the multidimensional utility of soil information. Scale requirements vary among the user groups, e.g. a farmer generally requires point, paddock or farm scale information whereas for government and policy making purposes, information is generally required at a district, region, state or national level (Omuto et al., 2013). Soil information products that users seek are classified into four areas: primary (soil site observations); soil map (general or specific purpose map with attributed soil properties or classes that reflect the original utility of the map); derived map (maps that have been re-interpreted or harmonised to predict soil properties of interest) and interpreted service (soil information including maps that have been used in either risk assessments or simulation models to support their decision making, e.g. wind erosion threat index). Generalised user requirements for spatial soil information are contrasted against delivery scale (e.g. resolution of information sought) in Figure 3.1 to illustrate the diverse and overlapping requirements of user groups across map scales.

To support future requirements for soil mapping, promotion of the specific benefits and actions (such as targeted investment and understanding production risks) that can be achieved due to this information are required (Cook et al., 2008). A shift from soil class assignment to mapping of soil properties that support implementation of simulation models has occurred in response to the need for quantification and uncertainty estimation.

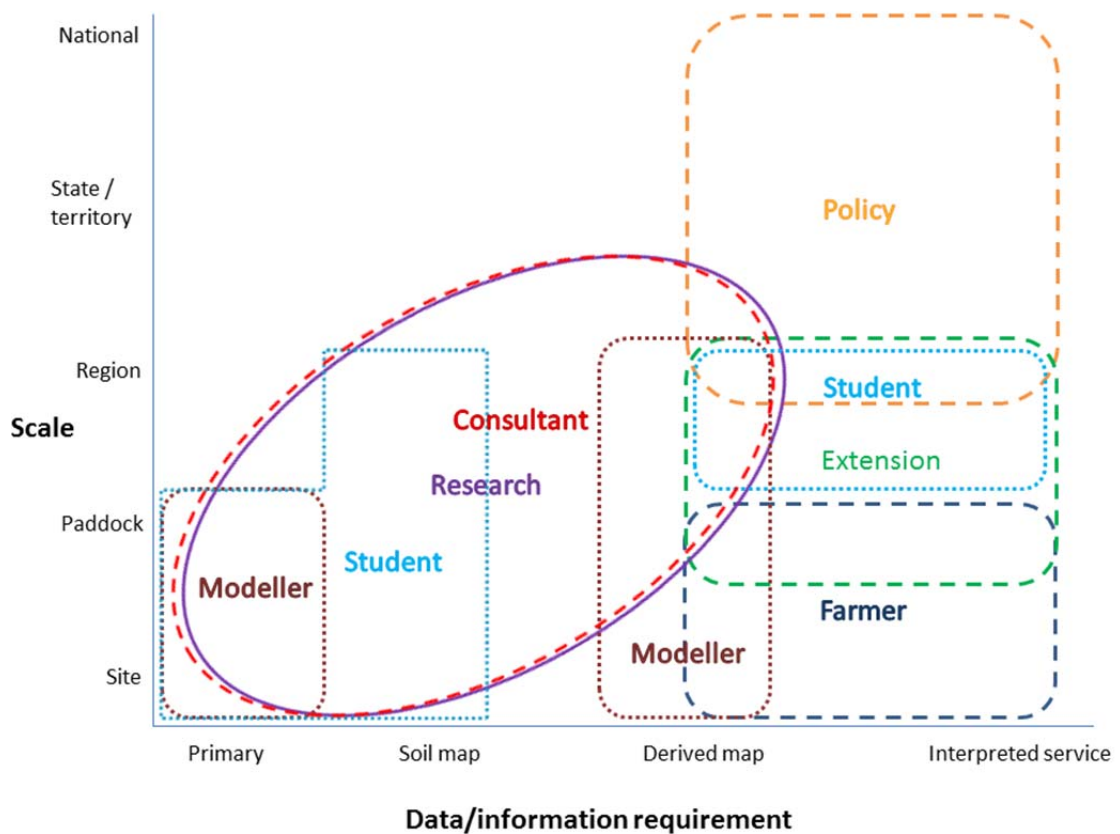


Figure 3.1. Generalised soil information requirements.

Case study – Victoria

In 2014, a series of workshops and surveys was undertaken in Victoria to identify the spatial soil information requirements of users, the key datasets or products they required, and the issues they encountered in accessing and using this information (Alexander et al., 2015). Key findings of this engagement with users were that:

- Not all users are aware of the soil data and information that is already available,
- Users need easy access to soil information,
- Users find that the soils information and data on the web is difficult to access (navigation of web),

- Needs of users and within user groups are quite varied, some requiring primary or raw data, others needing interpreted products for their specific purposes,
- Currency of soil information is important to users – products must be kept up to date,
- Confidence in the information is necessary for users (e.g. precision, accuracy, independent, trustworthy, fit-for-purpose and current).

Further findings identified that user preference was for the information to be provided in a consistent and structured manner via the web or smartphone application. Access to soil mapping is important to enable users to apply these maps where gaps in user soil information and knowledge exist (Alexander et al., 2015). Favourable responses were received from extension providers (knowledge brokers) for the potential of new soil maps to support advisers and farmers. Ongoing engagement with users to further and maintain understandings of their requirements for soil information is valuable to ensure that relevant and specific soil information to land management is delivered by government.

3.2.3 Inadequacy of user needs assessment

There remains an absence of detailed evidence and justification of the ‘need’ to produce spatial soils information (Omuto et al., 2013). Few studies (e.g. Robinson et al., 2010; Wood and Auricht, 2011) have assessed the requirements for users of spatial soil information systems, e.g. Australian Soil Resource Information System (www.asris.csiro.au). The use of spatial soils information has tended to focus on the needs of traditional users, yet there remains a considerably wider user market that is unaware or not exposed to this information (Wilson, 2012). Three examples of user needs analysis in Victoria follow.

User needs for spatial soil information – Victorian examples

A local survey of users of spatial soil information in Victoria from documented information supply requests for DEDJTR² between 2009 and 2014 identified that mapping was used to address questions on: management issues and production constraints; for engagement with stakeholders or inventory; and resource assessment reporting purposes (Figure 3.2). Some of the uses of existing soil maps include land capability assessment, land use planning and stakeholder engagement using visualisation examples.

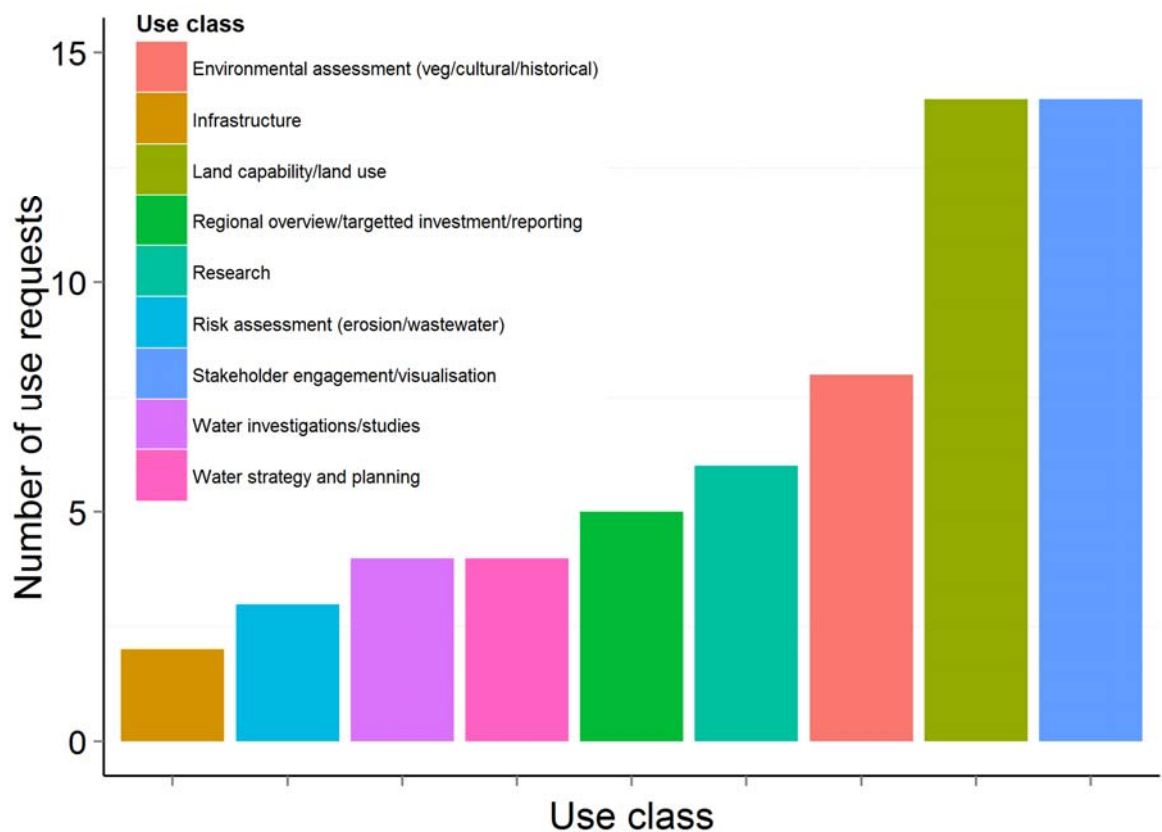


Figure 3.2. Documented uses of spatial soils information in Victoria between 2009 and 2014 from spatial soil information requests (source: DEDJTR 2014).

² DEDJTR - Department of Economic Development, Jobs, Transport and Resources is the custodian of soil and land information for the Victorian government. Note that the number of data requests over this time period was undertaken using a manual recording process through a nominated data supplier. Data supplies are now largely automated and have increased considerably since 2014 through electronic delivery systems such as Data.vic.gov.au.

The second user needs example is for biophysical modellers that use numerous tools (such as agronomic models) to support users with management decisions at various spatial scales. Models used for production and environmental purposes are directed towards quantitative simulation rather than qualitative or semi-quantitative interpretations (Bouma et al., 1986). These agronomic models require soil information to develop simulations and rely on the model structure and the quality of the input information to produce meaningful and reliable results (Zaks and Kucharik, 2011). Finer-scale and quality controlled soil information are desired by biophysical modellers as they require this information to refine, develop and run simulation and process models to address questions on production constraints and environmental impacts (e.g. rainfall, physical or chemical soil limitations or variety suitability). Hydrological and ecological model domains require specific and sometimes interpreted soil information to operate. Key soil properties used in biophysical models in Victoria have been defined by Nichol (2006) and refined by Robinson et al. (2010) and are listed in Table 3.1.

The third user-needs example involved a review of landscape analysis questions from two catchment management authority regions in south western Victoria (Glenelg Hopkins and Corangamite). This revealed that significant dependencies exist for spatial soils information to address Natural Resource Management issues. Questions from specialists, managers and individuals (Shanks, 2006) were assessed to identify if they had a spatial soil information requirement. Of nearly 300 questions collated including those from strategic reports relevant to the management of agricultural landscapes, 32% had a likelihood of requiring spatial soils information to answer the question. This subset of questions requiring spatial soil information was assessed against the Steinitz framework (Steinitz, 1990) to identify where the questions corresponded with a 'level of inquiry' in the framework. The six levels of this landscape and environmental design framework are:

representation, process, evaluation, changes, impact and decisions. Spatial soil information was found to be useful across all of the six stages (Figure 3.3). When questions were tested against an ecosystems services inventory (Binning et al., 2001), maintenance of soil health is overwhelmingly recognised as the primary ecosystem service where spatial soils information can contribute (Figure 3.4).

3.2.4 Evolving needs of users

User demand for soil maps continue to evolve as the data on specific themes of the Earth's surface (e.g. vegetation, hydrology, farmed land) have increased in spatial and temporal definition. The advent of new technologies such as GIS has equipped today's resource scientists with capabilities to handle complex and continuously variable information that is 'outstripping' the capacity of conventional maps (Cook et al., 1996). Numerous authors including Bouma (1986), Cook et al. (1996) and McKenzie et al. (2008) highlight that user requirements continue to evolve as the environmental questions change and the resolution of soil mapping sought increases.

Historically there has been a perceived requirement for soil survey, rather than clearly defined proven needs (Cook et al., 2008). External demand for soil survey during the last 40-years has not been strong and a supply-driven mantra has existed. Lack of demand by potential users including planners and farmers (Basher, 1997; Manderson and Palmer, 2006) are reported. A failing of a supply-driven process is that potential users may not be aware that spatial soil information exists in a format that can support and influence their decisions (Alexander et al., 2015). A favoured approach from users and stakeholders is for participatory-type approaches to support decision-making (Bouma, 2001).

Table 3.1. Highly sensitive soil parameters for landscape models (from Robinson et al. 2010).

			Model domain			
		Hydrological processes	Crops, nutrients and pastures (agricultural production)	Carbon and greenhouse	Forestry and biodiversity	
Hydrological	Air-dry moisture content	✓	✓	✓	✓	
	PWP, CLL	✓	✓	✓	✓	
	FC, DUL	✓	✓		✓	
	K_{sat}	✓	✓		✓	
	Infiltration rate	✓	✓		✓	
	Rooting depth	✓	✓		✓	
	Moisture retention	✓	✓		✓	
Physical	Clay %	✓	✓	✓	✓	
	Sand %	✓			✓	
	Silt %	✓			✓	
	Stones %	✓	✓	✓	✓	
	Bulk density	✓	✓	✓	✓	
	Soil structure				✓	
	Soil texture				✓	
Chemical	C/N ratio	✓	✓	✓	✓	
	CEC			✓	✓	
	NH_4	✓	✓	✓	✓	
	NO_3	✓	✓	✓	✓	
	Organic carbon	✓	✓	✓	✓	
	P	✓	✓	✓	✓	
	Total N	✓	✓	✓	✓	
	Carbon fractions			✓	✓	
Other	Soil depth	✓			✓	

3.2.5 User needs and participatory action

Current decision-making processes are more comprehensive than in the past due to the increased complexity and diversity of stakeholders involved and uses to be addressed (Christian, 1978; Bouma, 2001). Beckenstein et al. (1996) define stakeholders in environment and natural resource issues as citizens, companies and their share-owners, employers, customers, communities and policy makers. While stakeholders aren't necessarily direct users of spatial soils information, or have contributed to the production of such information, they do have a desire to be consulted and engaged in the participatory phases of a decision-making process. Engagement of stakeholders in this

process can be defined according to the complexity of the problems linked to land use, and the mechanisms or capacity of the spatial soils information to explain these.

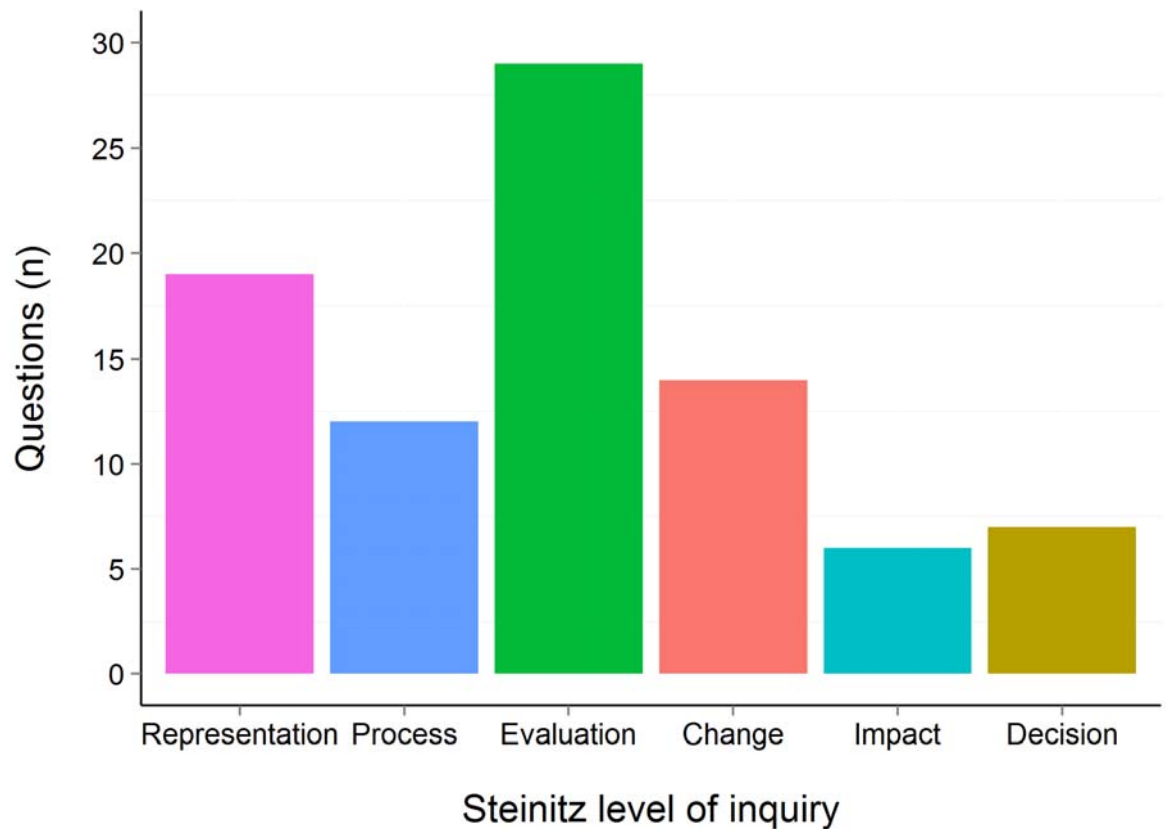


Figure 3.3. Questions requiring spatial soils information against the six levels of inquiry for the Steinitz framework (Robinson *unpublished results*).

To solve decision-making issues and recognising the role of spatial soils information in this process, engagement and participatory action with stakeholders is a key activity.

Elements to be considered in this process include (adapted from Bouma, 2001):

- Negotiate – be involved in the planning phase (where practical) of the problem definition and proposed approach, e.g. multi-disciplinary teams.

- Understand the ‘best-fit’ of spatial soils information to answer the specific question posed. Consideration of scale, quality/utility and the process in which the information will be used, e.g. catchment hydrology mapping. The phrase – ‘fit-for-purpose’ - applies here.
- Recognise that there are possible research and development deficiencies in knowledge and explicitly what assumptions have been made and what research is required. This includes base soil science and understanding of human impacts to ecosystem services.
- Inter-disciplinary teams provide the opportunity for enhanced understanding and learning collectively, and as individuals. This should lead to better outcomes in the decision making process.

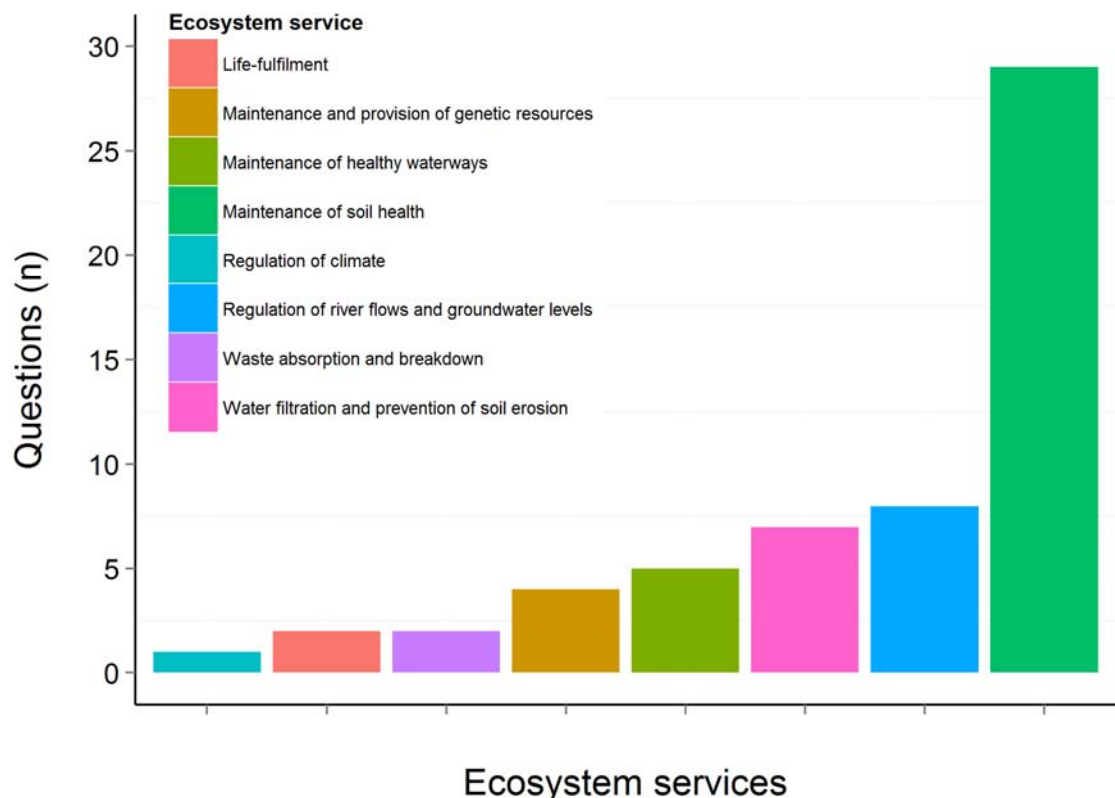


Figure 3.4. Ecosystem services inventory and questions that will require spatial soil information to address these (Robinson *unpublished results*).

Engagement with users and the development of solutions as part of multi-disciplinary teams is critical to answering these modern questions (Bouma, 2001). It is here where soil scientists can make their greatest contribution (Bouma, 2015). The application of soil mapping in tandem with latest soil science research provides a unique opportunity to guide and solve these environmental issues. Innovative solutions should be based on the application of new technology and approaches to translate legacy soil mapping into what is required (Bouma, 1997). Soil scientists need to be supportive, responsive and clear with their messages to support policy makers and the questions they pose. Unfortunately, and all too often, soil science is neglected and therefore conclusions are often simplifications of reality with a weak foundation.

3.2.6 Access and delivery of spatial soil information

Cartographically, maps today are represented either as a raster or vector data structure. Modellers and research scientists preference is for spatial soil information to be available in these formats (Wood and Auricht, 2011; Omuto et al., 2013). Nationally, modellers would like direct access to available soil site data online to exploit for their purposes (Wood and Auricht, 2011) while farmers and consultants have expressed a desire for information available in hard copy such as reports (Omuto et al., 2013). Modellers often have the benefit of technological expertise and therefore are able to transform soil information into specific inputs for their modelling requirements.

The widespread use of smartphones and applications has seen a major shift in the way we provide digital information to users. The SoilMapp application provides users with access to Australian soil property information for simulation modelling purposes on crop growth (Thomas et al., 2012). Making users aware of where and what soil sites exist through a Google service has been recommended (Wood and Auricht, 2011; Alexander et al., 2015). The open interfaces and protocols defined by OpenGIS specifications (Open

Source Geospatial Foundation (www.osgeo.org)) enable interoperable solutions including the web, wireless and location-based services that empower technology developers to make complex digital spatial information and services accessible to numerous applications. Internet services such as Web Feature Service (WFS) and Web Coverage Service (WCS) (Open Geospatial Consortium (www.opengeospatial.org)) can provide users with direct links between data and information services. New technologies and web delivery standards such as WFS and WCS enable the delivery of soil data and information for expedient consumption and use by service providers and end-users (<http://www.opengeospatial.org/projects/initiatives/soildataie>). Integration of this data in real-time with proximal and remote sensing systems represents the next significant advancement to support agricultural production. Online access to reports from systems such as Victorian Resources Online (VRO) provides valuable contextual information and knowledge captured in web delivered pages (Imhof et al., 2011) to link sites with delivered maps.

Users of spatial soil information are also requesting supporting metadata including error and uncertainty estimates and details on the information delivered, e.g. map making method and details on soil sites used (Omuto et al., 2013; Alexander et al., 2015). Spatial soil information should also be easily downloadable with fitness-for-purpose statements to guide users (Wood and Auricht, 2011; Alexander et al. 2015).

The current paradigm to make spatial soil information available and accessible for reuse by the public has gained considerable momentum (Zuiderwijk and Janssen, 2014). Globally this view is shared amongst all inter-government organisations and initiatives such as GlobalSoilMap and the Global Soil Partnership. In Australia access to spatial soil information has been limited but this situation is rapidly changing. A goal of the National Soil Research, Development and Extension Strategy (Department of Agriculture, 2014) is

to ‘*Improve quality, availability and access to soil data and information*’ that includes providing users with ‘*Maps of functional properties of soils at appropriate resolutions*’. These goals are consistent with the few accounts of user needs for soil information (e.g. Wood and Auricht, 2011; Robinson et al., 2010; Omuto et al., 2013).

Accessibility and availability of spatial soils information in Victoria since 2013 has changed with introduction of an open access policy and the implementation of the Victorian Government Data Directory (www.data.vic.gov.au). Data requests have increased dramatically as a result of this open access policy (Figure 3.5). Open access to spatial soils information is expected to support the delivery of new services to community and business, increase productivity, improve research outcomes and establish more effective management of spatial soils information. It is anticipated that open access to soil information will contribute significantly in the future to eResearch and the increasing demand for international connectivity (Wilson 2012).

Nevertheless, challenges exist to make soil information freely accessible including: establishment and implementation of governance roles and responsibilities; development of metadata standards; formalisation of copyright and licensing; data quality statements and data provision channels. These can be addressed through commitment, management and collaboration between public and private organisations. There are other untold issues that should be considered in open access to spatial soils information that depend on the ‘type’ of licencing agreement with the jurisdiction and the opportunity for dialogue between users and producers (e.g. privacy conditions and terms). Issues that need to be addressed in the provision of spatial soil information via open access include:

- How to identify what spatial soil information was used for and is the information being used correctly?

- What question or issue is the information being used to address?
- How can the producers of this information improve it for future users and uses?
- What deficiencies and issues (limitations) associated with the product is identified by the user?
- How should producers guide users on the use, and the limitations of the information (is metadata appropriately detailed)?
- Does a process exist where newly derived products by users can be accessed by other users and used to inform and update existing products?
- What resources are necessary to support the process between the spatial soil information producer and user including responding to feedback and further support requests from users?
- What process improvement strategy is provided for information sharing and what level of financial support is required to deliver this?
- What resourcing is necessary to maintain and reticulate products?
- In what form, and by what channels, should information be made available to users?

3.3 Soil mapping, benefits, costs and utility

This section briefly describes the evolution of soil mapping in Victoria with reference to the changing purposes for survey (needs) and the motivations for these changes (e.g. land degradation). The utility of these maps for existing and new applications, benefits, costs and the utility of soil mapping are also discussed to highlight the value proposition of soil mapping to users.

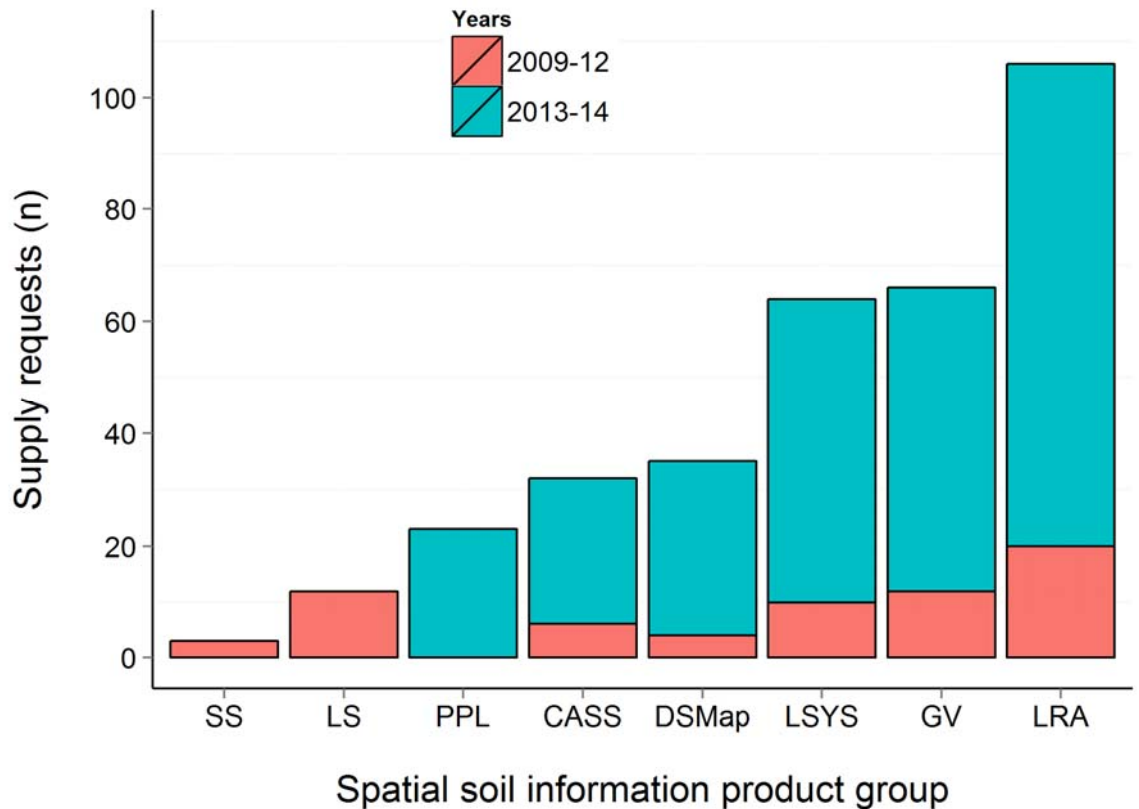


Figure 3.5. Spatial soil information products supplied between 2009 and 2012, shown against 2013-2014 (CASS – Coastal Acid Sulphate Soils, DSMap – Digital Soil Map of Victoria version 1, GV – 1:250,000 Geomorphology of Victoria, LRA – Land Resource Assessments ($\leq 1:100,000$), LSYS – 1:250,000 Land systems, LSs – Local soil map surveys ($< 1:100,000$), SS – sites from the Victorian Soil Information System (VSIS), PPL – Primary Production Landscapes of Victoria).

3.3.1 The changing purpose and needs for soil mapping in Victoria

Traditionally, the purpose of soil and land surveys has been considered in a ‘categorical perspective’ as either ‘General’ purpose (providing a broad range of soil information products for many different uses by one or more clients who may not be defined) or ‘Special’ purpose – interchangeably used with ‘Specific’ purpose (where data products, i.e. single soil property maps, are defined and targeted to meet the particular need of a user) (Beckett and Burrough, 1971; Dent and Young, 1993; Schoknecht et al., 2008).

Behind every soil map is a set of underlying principles, philosophies or objectives for undertaking that survey. As Gibbons (1981) states, in Australia we like to ‘have it both ways – general purpose and specific purpose, all used side by side’.

Through the phases of soil mapping in Victoria, the purpose of surveys appears to have alternated, signifying the changing requirements of users and the evolving need for knowledge on land resources. Since the 1840’s, the primary reasons for soil mapping in Victoria has reflected concerns and priorities including conservation (environment), productive agriculture (e.g. irrigation development), settlement (including colonization and land clearance primarily for dryland agricultural purposes) and urban development. Figure 3.6 presents a subjective assessment of the evolving purposes for soil mapping in Victoria. The agriculture versus environment paradigm for soil mapping (Bouma, 1989) has endured through mapping programs in Victoria for nearly a century.

Surveys before 1927 include broad-scale soil assessments of Victoria’s agricultural landscapes as part of an exploratory phase of soils in Australia (Taylor, 1970). Here the priority was the identification of relatively suitable land for agricultural settlement (Gibbons, 1983). This included initial chemical and mechanical assessment of soils from different geologies. In the early 1900’s there was a desire to correlate soil analysis with field trials but soil surveys were considered ‘not feasible’ at that stage (Martin, 1998). The impetus for soil survey was lost until the 1920’s when governments of the Pacific were encouraged to push ahead with soil surveys to support growing of pasture and crops (The Argus, 22/08/1923). During the late 1920’s and early 1930’s, expansion of irrigation schemes resulted in production issues including salinity and waterlogging that saw significant loss of vines and trees. Thus a requirement for large-scale specific surveys to guide and remedy these issues for settlements along the Murray River was established with soil surveys. At this time, Leeper et al. (1936) undertook a detailed survey at Mt

Gellibrand to identify soil types and their agronomic limitations as the first of the modern field surveys (Gibbons, 1981).

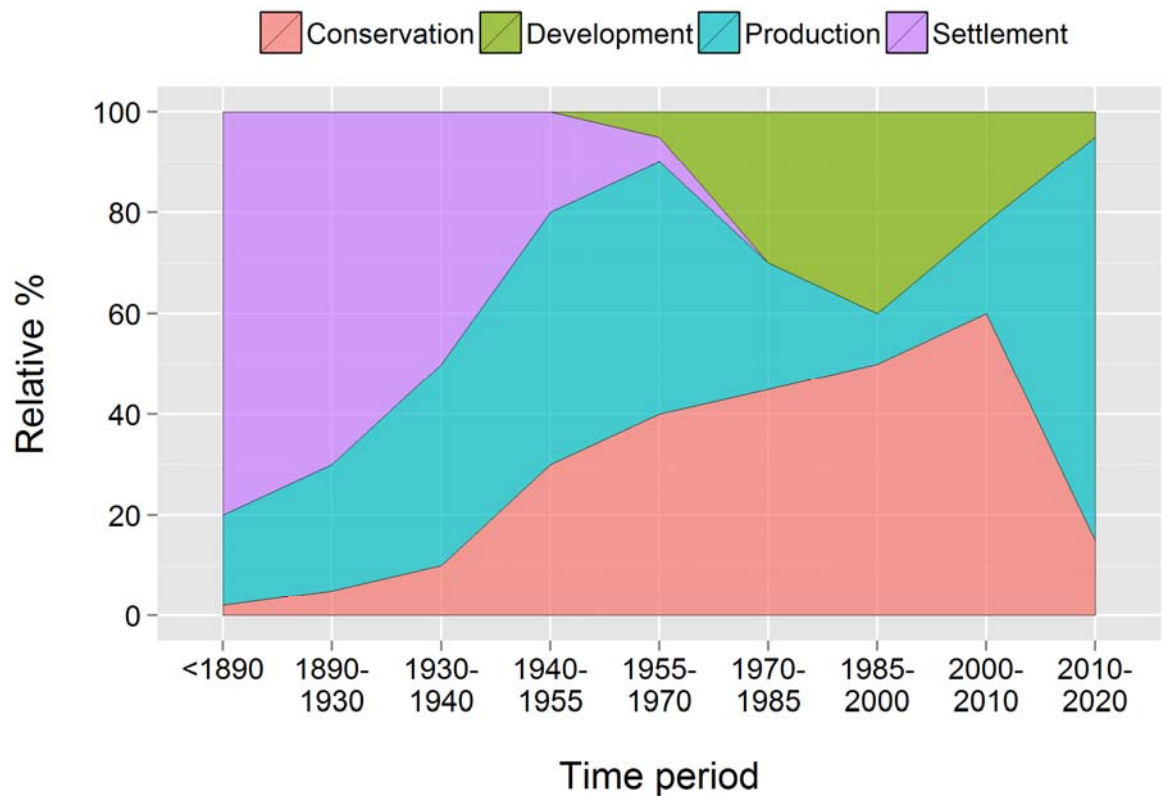


Figure 3.6. Primary reasons for soil survey in Victoria as relative proportions (1890-2015).

A growing concern was the siltation of reservoirs that coincided with the droughts and wind erosion events of this period. In the late 1930's, soil conservation and the importance of soil erosion to the national economy was recognised (Scott and Olley, 2003). Low wheat prices combined with record global wheat production and severe droughts resulted in significant depopulation trends of the Mallee region. Establishment of the Sand-Drift Committee in 1933 and the Mallee Research Station in 1935 had an emphasis on the provision and education of improved farming methods to reduce soil drift (Ballinger, 2012). This conservation movement was enhanced with the establishment of

government organisations responsible for conservation practices including the Soil Conservation Board in Victoria (later known as the Soil Conservation Authority) in 1940 from passing of the Soil Conservation Act.

Surveys for conservation purposes also began with the first recognised study mapping soil in the Dookie district (NE Victoria) by R.G. Downes (Thompson, 1979). Downes defined ‘units of land husbandry’ to raise and determine the optimum level of production while maintaining the ecological equilibrium. Soil erosion was also mapped and recommendations on land use provided (Downes, 1949). This initiated a sequence of surveys across Victoria with an emphasis on ecology, land use and primary production. These land-system surveys continued until the early 1990’s with two-thirds of Victoria mapped as part of this program.

The period from 1940 to 1955 produced many large-scale soil surveys that were used for farm planning, irrigation and water allocation, and identification of district problems (Martin, 1998). Labelled the ‘Golden Age of Soil Mapping in Australia’ (Taylor, 1970), this period was responsible for high resolution soil maps as pre-requisites in rural reconstruction schemes nationally (Gibbons, 1981). Significant advances in soil science disciplines saw a wave of newly adapted and implemented analytical chemistry techniques to support these specific purpose surveys. Examples include: relations between particle size and field texture (Marshall, 1947), field pH determination (Raupach and Tucker, 1959) and soil classification (Prescott, 1944; Stephens, 1953).

The post Second World War period saw a rapid expansion in the Australian economy, significant foreign capital investment and enhanced export opportunities for Australian produce (Edgar, 1966). Increased world demand for agricultural produce drove substantial increases in primary production from irrigation and dryland regions including

wheat, wool and beef. The rapid production gains were largely due to provision of agricultural extension services to support farmers and the enhanced agricultural research that was supported through statutory levies and industry supported augmentation of research between universities, state departments of agriculture and the CSIRO (Edgar, 1966).

From 1955 to 1970, a diverse range of general and special purpose surveys were undertaken with priorities including farm planning and layout of research stations, land suitability for various irrigated crops and parallel investigations, and small-scale ecosystem surveys with a conservation emphasis, e.g. reducing erosion and distinguishing edaphic (crop response) needs. However, the merits of soil mapping and its usefulness to meet these requirements were being questioned (Butler, 1958; Leeper, 1956). The major concerns shared by Leeper and Bulter were that the prioritization of soil classification and genetic origins had negated the mapping of properties and soil attributes that were relevant to production, as it is today. Gibbons (1981) identifies this phase as ‘the rationalisation of soil-land mapping’. Within this phase there are five proclaimed areas of rationalisation including:

- Formulation and implementation of regional models of soil distribution (e.g. soil association, land system, ecosystem),
- Knowledge of what users want soil maps for and how to provide it,
- Special-purpose surveys where the use of the soil and land is considered in the design and implementation of the survey,
- The development of the land system approach,
- Cost and worthiness of the survey.

The philosophy on the ‘usefulness and predictive ability’ of soil mapping, changed in the late 1960’s and 1970’s (Gibbons, 1983). At this time there was a decline in general purpose surveys due to limited application of these surveys for large areas with nonspecific location recommendations (Olson, 1976). While these general purpose maps had proved an effective approach at introducing soil information to users ‘quickly and efficiently’, there was a greater demand for detailed information that addressed specific land use recommendations. Reviews of this time (Hallsworth, 1978; Olson, 1976; Beckett and Bie, 1978) suggest this change in soil mapping was part of a significant national shift from productivity to conservation focus. This transition included the delivery of specific purpose soil maps with requirements for soil/land properties used in classification systems, e.g. land capability assessment.

Research in the dryland of the Wimmera investigated the edaphic links to surface soil physical properties following specific seasonal conditions as part of a broader survey project (Martin, 1974). This was followed over the next decade by soil-landform mapping in cereal production regions of Victoria and was complemented by high resolution surveys to support production trials on research farms and cereal breeding programs.

Conservation surveys were also changing to meet evolving user preferences in a shift from general purpose surveys to detailed surveys for urban development (e.g. land capability assessment). This need for resource inventory to support expanding urban and suburban areas was recommended by Olson (1976) in a raft of recommendations including the need for government agencies to support the knowledge on soil productivity and future research to complement this direction. Olson (1976) highlighted this changing role of users and stakeholders in soil information and the failure of past surveys and research as this was often not provided in a format easily used. The emphasis of soil

surveys had shifted to address problems of environmental management (Cook et al., 2008).

Rowe et al. (1988) and Lindsay and Rowe (1990) developed land evaluation techniques including land capability guidelines that were refined and enhanced to support the implementation of surveys with assessments that could be readily interpreted and understood by planners and managers. The guidelines provide information on soil and land in a simple and systematic format which could be easily integrated with other information tied to a municipal planning process. The objective of these guidelines was to ‘prevent ecological degradation and retain productive capacity’ where implemented (Lindsay and Rowe, 1990). In recognition of this growing emphasis to service local government, a survey was undertaken by Lorimer (1990) to establish attitudes towards and understandings of land inventory and capability information to support municipal planning, and to identify where there was a need or ‘willingness’ to use land information. A survey of 128 rural shires in Victoria (with a 93% response) identified a good general awareness of land capability assessment (78% of respondents) with 70% of all shires possessing a planning or development strategy and 89% believing that land capability information was important. The survey identified that 22% of shires were in the midst of developing a new strategy that would benefit significantly from access to land capability information. Understanding the needs of this specific user group in local government provided the justification for a program of land capability studies over the succeeding 15-years.

The uniformity and variance of soils was gaining recognition at this time and the associated costs of soil survey (Bie and Beckett, 1970; Beckett and Burrough, 1971). Spatial variability as an issue resulted in the conventional thematic choropleth model being described as *‘too simple to describe the reality of soil variation adequately’*

(Webster, 1985). Beckett and Webster (1971) emphasized the importance of lateral variability and associated variances for soil classes and properties. The prediction of soil properties and the high variability of soils over short distances have resulted in a poor overall predictive ability of soil mapping based upon soil classes due to the weakness of the soil classification scheme used for mapping (Gibbons, 1983).

Changes in the type of soils data collected and how information was delivered to users was occurring also. Maps were produced to provide general information on the capacity of land to support major agricultural uses in dryland landscapes. Large-scale specific-purpose surveys were centred on land capability for pasture, annual crops and deep-rooted perennials, and investigations of grapevine vigour decline (Martin, 1998). New techniques in sampling, spatial interpolation and stochastic modelling of soil variation such as geostatistics were pioneered during this period.

The trend of surveys favouring conservation purposes and urban development continued for the next three decades. This represented a significant decline in soil survey requests and changing client demands (Martin, 2006). The National Soil Conservation Program (NSCP) that began in 1983 and ran until 1992 with a modest budget to support soil conservation, supported states and territories to undertake regional surveys to advise users on the constraints (e.g. soil pH and acidification), limitations and potential of agriculture. The South Western Victoria Soils and Landform survey (Maher and Martin, 1987) is an example of this work where the priority was the expansion of grains production in the high rainfall zone of southern Victoria.

The pioneering of new survey techniques and application in different regions at various scales in the late 1980's and 1990's was driven by the widespread adoption of computer technology and Geographic Information Systems (GIS) with remote sensing (Bui, 2006).

Surveys of this period echo a transition from conventional soil survey into a new age where soil mapping is supported by rapid advances in information technology. As a consequence, numerous specific purpose surveys were undertaken with detailed assessments of land capability and land suitability that were a hybrid of legacy and evolving mapping methods. Few general purpose surveys were undertaken at this time with the continued rationalisation and reduction in resources available for conventional soil survey programs.

The current emphasis of the state and federal government priorities in Australia is productivity (<https://federation.dpvc.gov.au/>) and security of supply. The expansion of agriculture to meet global food requirements into the future (Alexandratos and Bruinsma, 2012) with the promotion of sustainable agriculture is a central issue of global initiatives such as the United Nations Sustainable Development Goals (SDG) (<https://sustainabledevelopment.un.org/focussdgs.html>). Expansion of agriculture and increased production is anticipated from the competing demand for land to support the expanding energy and water sectors (Godfray et al., 2010). Closing the yield gap and identifying dryland farmland that is not realising its potential productivity (Alexandratos and Bruinsma, 2012) is a priority to support the future demand for food in a sustainable manner.

Investment priorities such as the Food to Asia Action Plan (2013) for Victoria (<http://agriculture.vic.gov.au/agriculture/food-and-fibre-industries/exports-to-asia>) aim to build markets with Asia on premium food and beverage products. A key action of this plan is the targeting of research and development, extension and innovation to grow primary production. This represents a re-direction away from the conservation and sustainability paradigm which has been the key driver for delivery of spatial soil information over the last three decades.

Today there are over 350 documented soil and land surveys (and studies) undertaken by various state and federal government organisations in Victoria. Of these, 106 surveys were considered in the production of a state coverage of soil and land surveys for future landscape modelling and monitoring programs (Robinson et al., 2010). These surveys have been grouped according to evolutionary phases of soil mapping (Table 3.2), and generally, there is agreement with the survey purpose of these periods. This includes:

- 51 surveys at a scale finer than 1:35,000;
- 37 surveys at a scale between 1:35,000 and 1:100,000;
- 18 surveys at a scale coarser than 100,000.

The surveys range in scale from 1:10,000 (fine scale soil survey) to 1:512,000 (broad scale soil-landform and land systems). Within surveys the range of soils, soil site density, and scale of map production all strongly influence the quality of the delivered map. For each survey, the lead organisation has a major bearing on the purpose, style and methodologies used. For example, the Department of Agriculture (1874-1996) was responsible for most surveys concerning irrigation development and food production, whereas the Soil Conservation Authority (SCA) were focussed on land protection and undertook small scale land system studies with an ecological focus. Scale was also a factor in the selection of surveys as those that were finer than 1:35,000 had land components mapped and described whereas broad scale surveys may or may not have land components defined, let alone spatially delineated.

Justification for the implementation of these surveys range from underlying environmental phenomena impacting upon agriculture (e.g. salinity, water erosion), population growth and the need to support urban expansion, nature and water conservation and aspirational research questions to understand soil distribution and

function, performance and origins. Survey mapping approaches include land capability (detailed, erosion risk, broad assessment), soil-landform and land systems (mostly general purpose), soil series and associations, land suitability including irrigation development and investigation of district problems, e.g. salinity.

Of the 106 surveys, only 6 have been undertaken in the last 35-years with links to process and land use questions, coinciding with the perceived gradual decline of pedology as a science discipline (Basher, 1997).

3.3.2 Supporting government and policy making

To support claims for spatial soil information, the policies for soil survey and soil science need a clear mandate supported by the citizens rather than purely scientific considerations (Bouma and Drooge, 2007). A connection of research (including soil survey) with real-world solutions is required. Bouma (2009) states that there is *'usually little information about the context within which research studies have been made let alone about their relevance in that particular context. In contrast, a positive approach can be followed by prioritising soil functions that are universal in character and can be studied more exclusively by soil scientists.'* This presents an opportunity for spatial soils information to act as a medium for effective communication of environmental threats and land use possibilities.

Table 3.2. Selected surveys from Victoria against the phases of soil mapping in Australia.

Period	Distinguishing features including focus of delivering products	Soil Survey (General or Specific)	No. of published surveys	Soil Classification Scheme [#]	Survey purposes	Survey type
Pre 1890	Soil map sketches by surveyors for land settlement purposes	General		Local		
1890 to 1927	Reconnaissance soil data collection with an emphasis on topsoil and rock.	General		Local		
1927 to 1940	“The Science of Soils” phase – large-scale soil surveys that initially targeted irrigation areas and associated problems, e.g. salinity, and later in dryland.	Specific	7	Prescott	Agriculture (irrigation) – salinity, rising groundwater, suitable soil types	Soil series
1940 to 1955	“The golden age of Soils Mapping in Australia” - flexible soil survey techniques that recognised landscape patterns and soil associations. Mapping was at variable scales supporting re-settlement post the 2 nd world war.	Specific	11	Prescott, Stephens	Agriculture (irrigation) – salinity, rising groundwater; Conservation – erosion, land use	Soil series
1955 to 1970	Beginning of the “Rationalisation of soil-land mapping” - Broad-scale mapping on ecosystem concepts and inter-relationships with the environment and conservation outcomes, e.g. soil erosion and edaphic studies	General and Specific	17	Stephens, Stace	Agriculture (irrigation) – suitable soils; Agriculture (dryland); Conservation - erosion, land use, resource condition; Urban development	Soil series Soil associations Land systems
1970 to 1985	Pedomorpholith and pedogenetic models as the basis of soil association and land systems	General	33	Northcote, Stace	Agriculture (irrigation) – suitable soils; Agriculture (dryland); Conservation - erosion, land use, resource condition; Urban development	Soil associations Land systems Land capability
1985 to 2000	Pioneering of Digital Soil Mapping techniques and application in different regions at various scales	General and Specific	24	Isbell and Northcote	Agriculture (irrigation) – suitable soils; Agriculture (dryland); Conservation - erosion, land use, resource condition; Urban development	Land systems Land capability Soil-landforms
2000 to 2010	Global adaptation of Digital Soil Mapping – further refinement and enhancement of techniques with enhanced availability of covariates and legacy site data for implementation at various scales	General and Specific	14	Isbell	Conservation - erosion, land use, resource condition Urban development	Land capability Soil-landforms
2010 to 2015	Digital Soil Assessment and Global provision of fine-resolution digital soil information products	General	2	Isbell	Urban development; coastal development	Land capability Soil-landforms

[#] Prescott (1944); Stephens (1953); Stace (1968); Northcote (1979); Isbell (2002)

Fisher and Crawford (2015) present a seven step process to answer policy problems. The role that soil survey plays in soil science and policy making is linked to the ‘what’, ‘when’, ‘where’ and contribute to the ‘why’ for questions that are posed. Within this process framework, two important questions that require the contribution of spatial soil information are:

- Establish the nature and significance of the problem or opportunity (the ‘what’), and
- Design and conduct annual reporting and tailored evaluations of the research, development and extension being undertaken to address the problem.

Given technological advances in soil mapping and development of risk assessment concepts, opportunities exist for the delivery of spatial soil information that can be readily and flexibly linked to ‘spatial scenarios’. For example, spatial soil information can be used to support:

- a spatial representation of problems by region and industry, and
- how is the problem changing over time and has it responded to past interventions?

3.3.3 Benefit vs costs and the investment logic for soil mapping

For a soil map to achieve specified outcomes in response to an issue or question there must be well-founded investment logic for this to occur. ACIL (1996) implemented an economic framework to assess the cost-benefit of land assessment projects and identified 18 general categories for use of soil information to support outcomes including increased economic production, avoided environmental damage and industry development. Clearly the value proposition for soil survey should consider both current and future uses and that overall responsibility for control and financing of soil survey lies with state and national authorities (Martin, 1980). This has proven difficult to articulate to investors and potential

users given the decline in soil survey and advent of neoliberalist user-pays philosophies (Basher, 1997).

The existing rationale for investment in soil survey tends to reflect the magnitude of the issue and benefits, costs, needs of users and their proposed actions as a result of the survey. An example of this is the contrast that exists between a farmer with needs for contemporary soil information to make real-time management decisions linked to production (e.g. fertilizer application, liming rates, ground-cover/erosion potential) against those needs of government to implement good governance through scientifically rigorous policy on resource use and environmental degradation. History tells us that there is always a need for spatial soils information to support analysis of an issue and formulation of a decision, irrespective of scale, cost or current political persuasion. Spatial soil information due to its long-term value and diversity of users requires enduring financial support to deliver collective benefits to all users.

Where the investment in spatial soil information has been successful is when delivered information contributed to sound decisions on soil and land management. The benefits of soil survey (a primary spatial soil information source) are complex to evaluate but far outweigh the investment cost as determined from benefit-cost analyses (ACIL, 1996). Initial studies by Klingebiel (1966) and Bie and Beckett (1970) identified the dis-benefits and negative scenarios (e.g. avoided costs by informed and changed land use or management) for the economic justification of soil survey. Hallsworth (1978) identified that it was difficult to quantify benefits due to intangible social and community accruals that relate to the avoided adversity experienced by collective users rather than an individual. Beckett and Burrough (1971) prescribe that the ability of the survey and resultant map to answer questions posed by the user was fundamental to the map utility and the benefits derived.

Investigations by ACIL (1996) and Odeh and McBratney (1996) expanded the economic reasoning for soil survey by not only emphasizing the benefits due to avoided costs, but also the benefits due to adoption of new technologies and information. Both studies accommodate uncertainties in the application of data and the incremental flow of benefits that stem from the survey investment. Other economic considerations include the quantification of 'intangible' benefits and costs, value due to risk reduction (decreased likelihood) and structured linking of quantified 'likely' or 'actual' benefits.

Between 1964 and 1978 less than 2% of survey related publications in Australia mention benefits to users (Hallsworth, 1978). Review studies (Table 3.3) have identified significant benefit-cost ratios (BCR) that support the conclusion that soil survey is cost-effective (Manderson and Palmer, 2006). Where historical surveys are used in new synthesis studies or soil maps, establishing the BCR is difficult due to the absence (or poor documentation) of financial details for the legacy survey and that benefits to original beneficiaries may be unclear. Building upon the base soil information for synthesis studies (Bui et al., 2008) or new soil mapping, will incrementally add value through a value change proposition (Craemer and Barber, 2007). The value of these surveys are only realised once used in a decision-making context, and therefore it is desirable to understand user needs, monitor their use and tailor products to support high frequency and widespread application for their purposes.

User-pays and knowledge services

Existing investment paradigms have used the concept of user-pays, and willingness to pay (WTP) for soil survey (Diafus et al., 2013). The value of soil survey programs remains a contentious issue for all nations including those with existing programs (Giasson et al., 2000). Given the benefits of existing and new soil survey (from empirical evidence) that leads to higher returns to farmers (Diafus et al., 2013), it appears counter-intuitive that

government is reluctant to invest in new soil survey given the benefits derived (Martin, 1980). Manderson and Palmer (2006) counter the private funding (user-pays) argument by stating that ‘targeting farmers is simply not a sound business decision’, and that the priority should be well-resourced organisations, e.g. government, agribusiness. Basher (1997) identified the competitive science funding model and the change in institutional roles towards problem-orientated pedological research in the ‘Rationalisation of soil-land mapping’ period. Other confounding perceptions on the value proposition for soil survey include (i) why private funding would not be forthcoming, (ii) BCR’s are relatively low in comparison to those achieved in other sciences, e.g. > 100 in health sciences are common, and (iii) BCR overestimate benefits (Craemer and Barber 2007).

Table 3.3. Benefit-cost ratios for spatial soil data (survey).

Study	BCR	Details
Klingebliel (1966)	46:1 to 123:1	Low, medium and high survey intensity (USA)
ACIL (1996)	3:1 to 115:1	3 case studies (AUS)
Odeh and McBratney (1996)	17:1	1 case study in northern NSW (AUS)
Carrick et al., (2010)	6:1 to 13:1	1 case study in Southland New Zealand

The concepts of WTP, BCR and market failure, demonstrate an existing paradigm that values information services. The emergence of information and knowledge management as wealth creation processes are not adequately factored into the intrinsic value delivered by soil survey. The current demand for open access to spatial soil information will increase the value of these existing surveys as they are transformed using knowledge into new products (Benkler, 2006). Online knowledge and information management systems such as Victorian Resources Online (Imhof et al., 2011) are key services that capture

legacy knowledge, make it freely accessible to a diverse network of next users and support spatial soil information products.

3.3.4 Utility of soil survey, including legacy maps for new applications

A significant argument for the investment in soil survey is that the information delivered improves user's ability to answer questions on production and the environment. Beckett and Burrough (1971) using soil class mapping, grouped these questions into categories: what are they – proportions and area with related soils and properties; what is the soil class at a particular location and the properties of that soil; and where can soil classes with diagnostic properties be found. It is the joint expression of 'where and what' that define the need of a soil map. Beckett and Burrough (1971) summarise that the utility of the map is dependent upon the completeness of the profile classes, correct prediction of profile classes and definition of the soil property sampling distribution, and the presence of an accuracy assessment of the mean/model predictions for the profile classes.

Clearly there are contemporary uses that drive the initial investment in soil survey (Beckett and Burrough 1971). However there may be multidisciplinary or future benefits that are extremely difficult to foresee and assess. For some time, there has been a tenuous dependence on legacy site information to produce maps due to a paucity of reliable and contemporary data (McKenzie 1991). The philosophy of collect once, use twice (or more) adheres to soil data and information as it can be used numerous times for many different purposes (Craemer and Barber, 2007). In most surveys there is a trade-off between resolution (scale) and necessary detail to mirror the purpose, intended use and perceived cost-benefits (Bouma 1989). Surveys may need to be augmented with predictions and/or observations of dynamic soil processes that are linked to production and conservation.

A compounding issue is the belief that soil surveys can be used indefinitely and with unsound principles due to misunderstandings of the context, scale, utility and purpose of the map (Gibbons, 1961). The advent of GIS and the increasing application of remote sensing in soil mapping potentially increase the inappropriate use of maps as users are generally not trained or informed on the implicit quality criteria used in map creation (McKenzie, 1991).

Map makers are often forced to use existing legacy data and translate this into information products that will support ‘sound decisions’ by users (Dent and Young, 1993; McKenzie et al., 2008). The incomplete coverage of spatial soil information with contemporary assessments of dynamic soil properties (Tugel et al., 2005) was identified as an issue in Australia (McKenzie, 1991). Today, soil mapping efforts are plagued by incomplete data curation from past and current soil survey programs. As a consequence, this high dependency on the application of legacy surveys for general or specific purpose synthesis studies persist.

Changes in land use, management systems and climate since the time of data collection represent further sources of uncertainty in mapping (Lagacherie, 2008). Resourcing of mapping and monitoring programs to support environmental monitoring and modelling is recognised as a global issue (Global Soil Partnership, 2014).

Pertinent questions to consider are: will increased access to spatial soils information lead to improved knowledge and decision-making or, are we constrained by the limitations that exist in our legacy data and inadequate contemporary data curation? An article by Angela Hsu from the Yale Centre for Environmental Law and Policy (www.huffingtonpost.com/angel-hsu/does-the-environment-need_b_3568529.html) asked the question ‘Does the environment need big data? Hsu stated “*Despite the data*

available, we are still woefully plagued with gaps in knowledge, imperfect data, and uncertainty. We lack, for example, global datasets for national recycling rates, waste management, and toxic chemicals. That leaves us frequently creating indicators based on incomplete or imperfect data. These indicators are meant to provoke policymakers to act on an environmental issue. One danger in creating these proxy measures is that issues with data gaps are often ignored because the underlying problems are masked.” Bouma and McBratney (2013) ask a question along similar lines, *“Why not focus on the indicator itself, rather than on a proxy value?”*

3.4 Digital Soil Mapping (DSM)

Soil maps have traditionally been presented as a 2-D object, however supporting technologies such as GIS have enabled 3-D representation of soil-landscapes (Grunwald et al., 2001; Grunwald and Barak, 2003) and in the future the fourth dimension for applications such as change detection (McBratney et al., 2003). Already applications and tools such as virtual reality have been used in soil sciences to understand and convey processes and models of soils and landscapes to a new generation of users (Grunwald et al., 2000). Benefits due to accessible, available, cheap and temporally current data have been considerable in the advancement of new soil mapping programs. The advent of new technologies and desktop efficiencies represent a new age in soil mapping that were identified as aspirational goals by Gibbons (1961) and Butler (1963). Burrough (1987) discussed these new tools and technologies for land evaluation and described them as the ‘state-of-the-art’ GIS. This age could be encapsulated and described as the pedometric period (Table 3.3).

As a discipline, Digital Soil Mapping (DSM) is many things, but at its core is the prediction of soil in space and time. Conceptually, DSM is still wedded to the cause of soil survey in the delivery of spatial soil information to support better decision making on land use. The principles identified by McBratney et al. (2003) still hold true and have been supplemented by a plethora of recent publications.

3.4.1 What is DSM?

DSM aims to “create and populate geographically referenced soil databases generated at a given resolution by using field and laboratory observation methods coupled with environmental data through quantitative relationships” (definition as used by the International Union of Soil Scientists Digital Soil Mapping Working Group). The major advantage of DSM (otherwise known as predictive soil mapping; Scull et al., 2003) over conventional soil mapping is that it provides a continuous prediction of a soil property and is capable of deriving uncertainties and error propagation can be tracked in the mapping process. Harmonisation of methods and observations across space and time is also an advantage of these quantitative approaches. DSM supports flexible yet quantifiable approaches to predict soil properties (e.g. pH, EC, organic carbon, clay content) at various scales (e.g. paddock to catchment) with remotely and proximally sensed data (e.g. geophysics, terrain derivatives) using spatial inference techniques.

Scull et al. (2003) and Behrens and Scholten (2006) describe the primary goals of predictive soil mapping as:

1. To produce models for predicting soil properties from spatial covariates to efficiently and effectively collect soil data, e.g. state-factor CLORPT approaches and pedotransfer functions.
2. To present soil continuity maps, e.g. geostatistics.

3. To incorporate expert knowledge in predictive mapping and to understand soil variance.

McBratney et al. (2003) reviewed approaches used to creating digital soil maps, the methods and data used to populate information systems. Based on the review, a framework for predicting soil properties (evolution of Hans Jenny's Soil Forming Factors function published in 1941) across regions of interest with seven factors was presented:

$$S_p[x, y, t] = f(s[x, y, \sim t], c[x, y, \sim t], o[x, y, \sim t], r[x, y, \sim t], p[x, y, \sim t], a[x, y, \sim t], n) + \varepsilon$$

where S_p is the predicted soil attribute, s is soil information from a prior map, remote or proximal sensing, or from expert knowledge, c represents the climate at a point, o is the organisms, r is the topography/landscape attributes, p is parent material, a equals time (age) and n is the spatial position and ε is the residuals (unexplained error). x, y are the precise spatial coordinates and t is at an approximate time.

3.4.2 DSM and conventional mapping

A transition from qualitative to quantitative mapping procedures has enabled soil survey to be more adaptable and responsive to modern-land use questions than in past surveys. DSM as an approach, or tool (Behrens and Scholten, 2006), is recognised for its quantitative underpinnings in contrast to conventional soil survey processes that are based on qualitative theories and models (McBratney et al., 2003). This transformation from conventional mapping to DSM across the globe is due to the accepted advantages of digital soil mapping over traditional survey practices including cost-benefits, uncertainty assessment, consistency, explicit mapping methodologies and the ability to readily produce a soil map with the advent of newly available soil or covariate data (Carré et al., 2007). Here lie distinct advantages of DSM over traditional approaches (McBratney et al., 2003).

Possibly the greatest asset of a map derived using DSM in comparison to a conventional soil map is the ability to recreate and manipulate the map of soil properties or classes in digital form using a GIS. This provides an opportunity for the map to be ‘self-updating’ where, as new observations or insights into processes are collated and accessible in a digital system, the production process of map making can begin again. This is essentially the philosophy behind the soil-landscape inference system (Robinson et al., 2010) and the ability to store and manage the model (Heuvelink et al., 2010) used to create the map. As Heuvelink et al. (2010) suggest, using a model management approach will enable flexibility in the various DSM iterations (e.g. spatial and temporal bounds, support and resolution), can save on storage and supports data sharing, enables maps to be easily updated (and archived) and supports multiple realizations using uncertainty propagation methods. Fundamentally it is about storing models rather than maps.

Like many current areas of soil science, DSM is transitioning from a research phase where methods are being developed and tested to an operational phase where it is used consistently by soil surveyors to improve and efficiently produce soil maps (Boettinger et al., 2010; Grunwald et al., 2011; Wilson and Thomas, 2012; Minasny and McBratney, 2015). Published DSM examples at various scales include: global – SoilGrids (Hengl et al., 2014); national – Soil and Landscape Grid of Australia (Viscarra Rossel et al., 2015); state/territory – (Gray et al., 2016) and region – South Australia’s agricultural zone (Liddicoat et al., 2015).

However, there are imperfections that exist in DSM. Perceived deficiencies in data, methods, connection with user requirements and the lack of standard DSM methodologies for widespread operationalization do limit the wider adoption of DSM (Hempel et al., 2008). The contemporary value of a digital soil map should not differ from that of a conventional soil map in principle. However, it is very difficult to foresee what the

advances in technology will be and how users will want spatial soil information delivered as a result of these advances.

Map ‘representation’ and ‘presentation’ should be considered in delivery of spatial soil information to avoid disappointment and disharmony from user groups. Accompanying metadata will enable users to know exactly what the map is, how it was produced, and guiding principles on its application (e.g. fit-for-purpose statement) that will benefit adoption of digital soil maps. It remains to be seen if users will want tailored purpose maps (e.g. land capability assessments, erosion hazard risk assessments) in contrast to numerous general purpose maps that can be derived using DSM. A further implication is that maps produced today using DSM may only have a relatively short life expectancy given the expedient and flexible approaches available to producing these maps today. Perhaps it will be the process of map creation and the learnings from this with qualitative guidance that are the truly valuable aspects of DSM to capture for future map production purposes.

3.4.3 Digital Soil Assessment and Digital Soil Risk Assessment

The evolution of DSM as a field of soil science has been propelled by the need for specific quantification of threats to soil and soil functions (Carré et al., 2007). Known as Digital Soil Assessment (DSA), the outputs of DSM (soil property/class spatial prediction) are used as inputs to answer problems raised by users to protect soil functions and suppress threats to soil. DSA as a natural progression of DSM is analogous to specific purpose survey. The maturation of DSM to a DSA is viewed as essential otherwise we run the risk ‘*of expiring on a mountain of unused digital maps*’ (McBratney et al., 2012).

A progression of DSA is Digital Soil Risk Assessment (DSRA). The primary objective in DSRA is the implementation of management scenarios and interventions to guide policy

development or management interventions. This is achieved through the integration of various data and information sources on socio-economics and the environment where scenarios can be tested and threats impacting the soil resource analysed. The inclusion of accuracy and prediction risk with uncertainty estimates aims to support users in their decision making processes (Carré et al., 2007).

For DSRA and DSA to succeed, both rely upon quality DSM outputs. As McBratney et al. (2012) asked “*To what degree can we continue to produce DSM information without first considering its end use?*” Collection and production of soil information supporting multidisciplinary approaches to sustainable land use will need to be guided by which disciplinary information (e.g. hydrological, ecological, social, soil) is most critical to address the question posed. The value proposition of DSRA and DSA will become evident to users when the underlying soil information from DSM is geared towards soil functions and the threats that are posed (Bouma, 2001). Attention on soil functions provides direct links to ecosystem services and concepts such as natural capital (Bouma, 2009). A further question posed by McBratney et al. (2012) was “*to what extent do we soil scientists need to step up and help development of assessment methods?*” This is a relevant extension of an earlier suggestion by Bouma et al. (1986) that called for soil scientists with their survey interpretations to become intermediates offering insight into the soil process and issues as part of multidisciplinary teams.

The integration DSA and DSRA has tremendous benefits from the reduced costs, formalised and consistent application of standard methods, and models that can be easily updated including assessment of error propagation as uncertainty estimation. This can be undertaken at all stages of the assessment process (Carré et al., 2007; Grunwald et al., 2011). Questions remain on how to address, or, accommodate large uncertainties if they occur in the assessment process. Potential sources of error in risk assessment are difficult

to nullify, however uncertainty approaches to identify, assess and quantify error sources in DSRA can be constructed. Published examples of DSA and DSRA for different uses include: irrigation development (land capability assessment) in Tasmania (Kidd et al., 2012, 2014), soil contamination risk assessment in the Nor-Pas-de-Calais region of France (Caudeville et al., 2012), soil natural capital in New Zealand (Hewitt et al., 2012), soil quality assessment for Hong Kong (Sun et al., 2012) and delineation of food production zones such as terrons in New South Wales (Hughes et al., 2012).

3.4.4 Data and knowledge deficiencies

McBratney et al. (2003) identified the inadequacies in available and quality soil site data to fit spatial inference models. Opportunities to develop and implement new sampling methods, refine and apply new sensors for rapid and cheap soil data acquisition and use of legacy soil data should be at the forefront of supporting DSM (Lagacherie, 2008). Legacy data is hampered by issues including format, lack of harmonisation, imprecision and inadequate georeferencing (Krol et al., 2008) and it may be unsuitable as it was collected for a specific purpose with no further use in mind.

All DSM methods (as has been the practice in existing survey) are dependent upon data from field and laboratory analysis to develop conceptual models (Lagacherie, 2008). A paucity of soil site data to undertake independent validation as a quality assessment and adequately fit such models using stochastic methods (McBratney et al., 2003) remain issues for the DSM community.

DSM has a high reliance on environmental data (covariates) to populate the spatial inference systems. To date proximal sensing has achieved better results than remotely sensed primary or secondary derivatives in digital soil mapping applications (Mulder et al., 2011). This is attributed to the coarser resolution of remote sensing leading to reduced

pixel purity. Vegetation indices have been used with mixed success (Mulder et al., 2011) but new passive and active satellite sensors such as Sentinel 1 and 2 (<https://earth.esa.int/web/guest/missions/esa-operational-eo-missions>), combined with existing platforms such as MODIS and Landsat offer better signal-to-noise ratios, high temporal frequency of image collection and band settings that favour improved correlation with soil property measurement.

Remote sensing of soil properties at depth is still a knowledge gap for DSM (Minasny et al., 2008a). Issues such as how to predict buried horizons may be possible in the future through synergies of geochronological data with models of soil-landscape evolution (Minasny and McBratney, 1999; Vanwalleghem et al., 2013) and newly available active and passive sensor data.

3.4.5 Method performance/robustness and user perceptions

There is an ongoing need for a meta-analysis to establish why some DSM methods perform better than others and their overall robustness (Scull et al., 2003; Grunwald, 2009). This analysis should contrast performance of global and local assessments and the requirement for successful local implementation to ensure global studies are credible. Further investigations are necessary to define if ‘performance’ or ‘predictive power’ should be attributed to the mapping technique, available soil site data, available covariates and qualities of these relative to the soil properties of interest, calibration and validation techniques, uncertainty or landscape complexity. Operational complexity, data handling capabilities, model simplicity and purity also should be considered when choosing which DSM method is a ‘best fit’ for implementation.

Grunwald (2009) in a review identified a deficiency of ‘intrinsic’ soil properties such as biological, morphological and mineralogical used in DSM studies. Although these

properties are recognised as important to understanding soil functions and processes, there has been a marked decline in soil mineralogy, soil morphology and soil genesis research in comparison to pedometrics (Hartemink et al., 2001). The production and use of DSM products in some branches of soil and environmental science (e.g. watershed and hydrological modelling) has been sparse (Terribile et al., 2011; Thompson et al., 2012). This may be due to global efforts to deliver general use soil maps to understand the inventory and stocks of terrestrial organic carbon in soil (Grunwald et al., 2009; Robinson et al., 2012). Opportunities exist to exploit the new spatial soil information systems around the globe at various scales and for mapping many different properties linked to production and the environment.

With a global pre-occupation on the development, evaluation and the application of digital soil mapping methods and approaches, a potential oversight has been the scant detail on the commissioning of, implicit value and usefulness of information contained in digital soil maps (Lark and Knights, 2015). How users will interpret and understand the implications of uncertainty attached to a map product are not clear. Lark et al. (2014) discusses the importance of using probabilities as determined from a geostatistical model with clear messages on the root causes of uncertainty (e.g. spatial variability) and outcomes from management scenarios. They note that this approach will *'indicate where particular interventions are likely to be required by the land manager, and also where further soil sampling is required in order to resolve uncertainty about local conditions and make a more robust decision'*. This example has real-world relevance with inferred risk linked to mapping and implications of management interventions.

Users including soil scientists must also be aware of the potential uses and limitations of DSM. This includes the time-soil variation relations that may be obscure, or not defined in DSM products. Users must be educated in the use and limitation of these maps in lieu

of classical soil maps (Hempel et al., 2008). Likewise, as DSM practitioners, we need to be responsive and paying close attention to the evolving needs of user groups.

3.4.6 New environmental covariates

Traditionally maps have provided a snapshot of historic soil conditions as information used in the map creation process are often constrained to legacy data of historic conditions (Grunwald et al., 2011). The growing demand for contemporary information on soil condition and how, where and when changes occur to the soil resource are of primary interest.

The current data deluge in science (Roudier et al., 2015) heralds new opportunities to undertake DSM with a plethora of available environmental covariates at fine and coarse resolution over different time and spatial scales. New satellite platforms that will deliver new covariates include the Sentinel (1 and 2) sensors. Sentinel-1 is a synthetic aperture radar satellite that operates in C-band and provides continuous imagery regardless of environmental conditions.

Advances in space-time modelling will be necessary to understand Anthropocene changes to the soil resource (Adewopo et al., 2014) and for future projections (Grunwald et al., 2011). Current baseline soil condition may benefit from historical interactions between land cover and management using remote sensing (Sheffield and Morse-McNabb, 2015). Understanding the resilience, capacity (and where this has been compromised), and performance of agro-ecosystems will be important to guide policy decisions on land use and management to improve processes and functions delivered by soil (Crawford and Fisher, 2014).

There has been considerable effort by research groups to improve the quality of spatial information on land use history (Sinclair et al., 2012), current land use and land cover

(Morse-McNabb et al., 2015), and ground cover (Sheffield et al., 2015). This new data may prove useful to interpret production constraints caused by soil and interactions between soil and plant where a yield gap is occurring (van Ittersum et al., 2013). Further development of spatial datasets on land management would be beneficial as management appears to be a primary factor to explain significant changes in soil where long term use of multiple practices has occurred (Robertson et al., 2015).

3.4.7 New soil sites

New soil site data is necessary to support time-series analysis for properties that are dynamic and respond to management practices. A surge in the collection of sensor data and availability of this data has occurred (Kshetri, 2014), but gaps remain in the current patchwork of currently available data to enable broad assessment of soil and land condition and impacts due to agriculture (Zaks and Kucharik, 2011). New methods of collecting soil information from the private sector including precision agriculture are also encouraged through crowd sourcing and citizen science (Rossiter et al., 2015). Ongoing maintenance, governance, and resourcing of reliable infrastructure to support this sharing of soil data and information remain troublesome in Australia. Access to data using creative commons (<http://creativecommons.org.au/>) and interoperable solutions are advocated including the roles and responsibilities for custodianship (Wilson, 2012).

Wireless sensor networks that are connected through the internet can deliver real-time soil data in a format to support calibration and validation of biophysical models for agricultural decision making (Zaks and Kucharik, 2011). The advances in information technology and rapid expansion of sensors have seen an exponential increase in data being generated on soil (Roudier et al., 2015). Data mining techniques and the advent of high performance computing can support the integration of these various data sources into spatial soil information systems.

While new soil sites should be a priority to improve DSM outputs for users, the importance and intrinsic value of legacy soil site data should not be forgotten as it is impossible to go back in time and sample soil to benchmark its condition. Many sites with associated data remain in notebooks and as sites sheets in archives and filing cabinets. This data often fills spatial gaps in the soil inventory but also provides snap shots in time to make assessments on changes in ecosystems, soil and the functions and processes they deliver. Prior to 1980 in Victoria, there was considerable resourcing for field survey which appears unlikely to be repeated in the near future. In Victoria between 80,000 and 90,000 sites were surveyed (MacEwan et al., 2014). Development of methods to expediently capture, collate and harmonise this legacy data with analytical methods and procedures of today (e.g. infrared spectroscopy) will support assessments to understand soil change and impacts to soil caused by agriculture.

3.5 Uncertainty

For a map to be useful for a specific purpose, errors and uncertainties should be quantified to communicate if the map is appropriate (Heuvelink, 2014). In this section, a brief overview of the elements of uncertainty analysis is provided and links to risk described. User perceptions on uncertainty and conveying uncertainty on a map is presented and potential options to reduce uncertainty in maps are discussed.

3.5.1 What is uncertainty?

Uncertainty is used interchangeably with reliability, accuracy, precision, error and confidence (Minasny and Bishop, 2008). It is described as a lack of assurance or conviction (knowledge) in an observation or model (Goovaerts, 1997). Uncertainty is based on a dichotomy of aleatory uncertainty (statistical variability or error) and

epistemic uncertainty (lack of information). In DSM these sources of uncertainty can be documented, quantified and their overall contribution to error propagation determined. Sources of stochastic and epistemic uncertainty have been enumerated by Refsgaard et al. (2007) and Benke et al. (2011), enabling assessment of uncertainty in analytical processes (Mowrer, 1999).

In DSM, uncertainty is often represented as statistical (aleatory) uncertainty (Heuvelink 2014). This is the lack of confidence in an estimated value equalling a true value. A simple calculation of the 95% prediction interval (PI) is by subtracting and adding 1.96 times the kriging standard deviation to the kriging prediction. However, the 95% confidence required for scientific proof can be beyond the practical and financial capabilities of many science domains (Lemons, 1996). Heuvelink (1996) identified a mixed model of spatial variation for uncertainty estimation as this technique combines discrete and continuous components in the one approach. In contrast, data input errors to models are considered more significant in model studies than in mapping as errors don't just affect the initial model state and some of its processes, but may impact its boundary conditions including timelines (Finke, 2012). Another confounding issue is the 'completeness' of data to meet requirements of the model. Where deficiencies exist in input data and predictive functions are used to increase available data for models, there is likely to be an overall increase in error (Finke, 2012).

3.5.2 Uncertainty and risk

Uncertainty is considered an inherent part of risk (Mowrer, 1999), however, estimating risk by integrating 'risk in the decision' and the 'risk acceptable to the decision-maker' is challenging (Agumya and Hunter, 1999). Methods to define acceptable risk thresholds can be determined via expert judgement, boot-strapping procedures or formal cost-benefit analysis. In cost-benefit examples, benefit is the reduction in risk with cost representing

the financial expense required to achieve that benefit (Agumya and Hunter, 1999). Cost-benefit and economic decisions is valuable as it establishes a critical baseline that defines what level of information is required to meet an acceptable risk (Lowrance, 1976; Griffith et al., 1999).

For a risk based approach, error propagation and uncertainty should be accommodated to define acceptable risk thresholds (Agumya and Hunter, 1999). Attitudes of users and user groups to risk can vary depending upon the risk attitude of the user. Information uncertainty represents the amount of risk a user is prepared to accept, although risk is rarely considered in the assessment of fitness-for-use. Assessment of fitness-for-use compares risk-in-a-decision with the risk acceptable to the decision maker. This approach is desirable as it is simple and cheap to implement, benefits from available information contained in metadata and can be easily understood by users for their particular application (Agumya and Hunter, 1999). It is essential for users to assess the ‘fitness-for-use’ principle for a map (Agumya and Hunter, 1999), but how to establish a ‘fitness’ statement on a DSM product is still to be conceived.

3.5.3 Support, scale and reducing uncertainty

In DSM, there are usually many different types of spatial data used with different scales resulting in different spatial autocorrelation models. Scale issues have led to unbalanced models being derived. The models can be applied at different scales (supports) leading to errors in input parameters for DSA and DSRA purposes (Finke, 2012). The concept of support is important to convey to map users (Heuvelink, 1999) as this can significantly influence the outputs of error propagation assessments. Support can be temporal and/or spatial in nature. Point based methods, e.g. kriging at point support, assume point observations are 100% correct. Here the measurement errors aren’t interpolated or accommodated in these approaches (Fisher, 1999). Uncertainty can be calculated at the

point support and associated block (defined area). Within block uncertainty is averaged while the aggregation of larger blocks will reduce uncertainty at the expense of resolution (Heuvelink, 1999). Validation of these soil maps should adhere to the spatial support of the covariate data, and ultimately, the maps are better if this is the case (Bishop et al., 2015).

To reduce uncertainty in a map, an immediate solution is to collect further information (e.g. additional soil sites) to obtain improved estimates. Other suggestions include the implementation of model (map) ensembles where maps from separate operations can be combined to derive a mixture map. This mixture map ideally should integrate the strengths of all maps and address weaknesses (high uncertainty), enabling better spatial predictions (Griffith et al., 1999; Finke, 2012; Malone et al., 2014). Non-discriminatory approaches to apply model ensembles are recommended where inputs are weighted by error/uncertainty contribution using Bayesian Model Averaging (Finke, 2012).

3.5.4 User perception

Uncertainty has a negative connotation, and as McBratney et al. (2003) suggest a preference is to use the phrase ‘certainty’ instead. There remains scepticism on the value and importance of uncertainty assessment (Mowrer, 1999), however, the DSM community has recognised the importance of quantifying uncertainty and is leading numerous other environmental mapping programs across the globe (Heuvelink, 2014). Unfortunately, unfounded uncertainty often outweighs the scientific or technical aspects of uncertainty. Conveying the importance of uncertainty analysis to those in the policy sphere is valuable but there has been little understanding to date on how uncertainty affects the decision making process (Mowrer, 1999). Convincing policy makers of the relative importance of technical uncertainty to decision making is a priority for the DSM community (Carré et al., 2007).

Clearly there are considerable investigations to be undertaken to understand what level of knowledge users require to deploy such maps (Hengl and Toomanian, 2004). Simplicity of interpretation and use is the key reason for adoption of uncertainty representation (Agumya and Hunter, 1999). Users require education on the information qualities contained in a digital soil map and how the associated uncertainty will impact their use (Carré et al., 2007). There are few tools and techniques available to make users (such as policy makers) aware of the uncertainties in a model (Finke 2012). Producers of DSM need to be more aware of end-user requirements to tailor products including uncertainty estimates to suit and support user needs (Gascoigne and Wadsworth, 1999; Carré et al., 2007). The uncertainty measures provided to users should be independent and represent uncertainty by a method understood and agreed to by the target data users. Ideally uncertainty estimation should include the decision making process that may result in a loss of information for the particular problem (Gottsegen et al., 1999).

3.5.5 Inclusion and representation of uncertainty with DSM products

For users who base decisions on DSM outputs, consequences (e.g. financial, social, risk to human life) could arise from incorrect conclusions based upon information that was imprecise, inaccurate, unreliable and uncertain. Consequences can be expressed as risk (Agumya and Hunter, 1999). Maps need to expose the magnitude of the errors, their spatial distribution (Griffith et al., 1999) and how error is propagated as a result. The uncertainty should be quantified or estimated with the map representing the errors and uncertainties and easily communicated to users (Heuvelink, 2014).

Users of maps need confidence in the map from a DSA or DSRA to enable reliable decisions to be made (Carré et al., 2007). An estimate of uncertainty, regardless of its magnitude, is considered better than no estimate at all (Gascoigne and Wadsworth, 1999). Finke (2012) suggests that if the uncertainty was high in a DSA, application of evidence

filters (Finke et al., 2008) for different evidence levels to clarify responsibility for a decision should be considered. The linking of DSM, DSA and DSRA enable the estimation of uncertainty at all stages and has benefits for producers and users in their efforts to minimise error and uncertainty (Carré et al., 2007).

Over the last decade, there has been useful research on representations of map uncertainty. Hengl and Toomanian (2004) incorporated uncertainty into the map through a hybrid static visualization technique that used whitening and pixel mixing to represent uncertainty. They advocated that it is crucial to define what end users perceive by looking at maps (with included uncertainty) and what visualizations mean to their decision making process.

Previous studies have emphasized where probability values fell below a threshold (Webster 1994). The study by Lark et al. (2014) is a valuable example of a DSRA with users in mind and support for their management decisions. The authors use of a standard ‘verbal scale for probabilistic information’ developed for the IPCC (Mastrandrea et al., 2010) with ‘language intensifiers’ as well as the corresponding probabilities (as percentages) describe the certainty associated with the map (Figure 3.7). For policy makers and government, this information is useful as it enables government to strategically invest (should it choose to do so) to address production or environmental issues in a risk approach. The approach developed by Lark et al. (2014) can also be used for decision making where users can make judgements on the financial imperatives of management interventions.

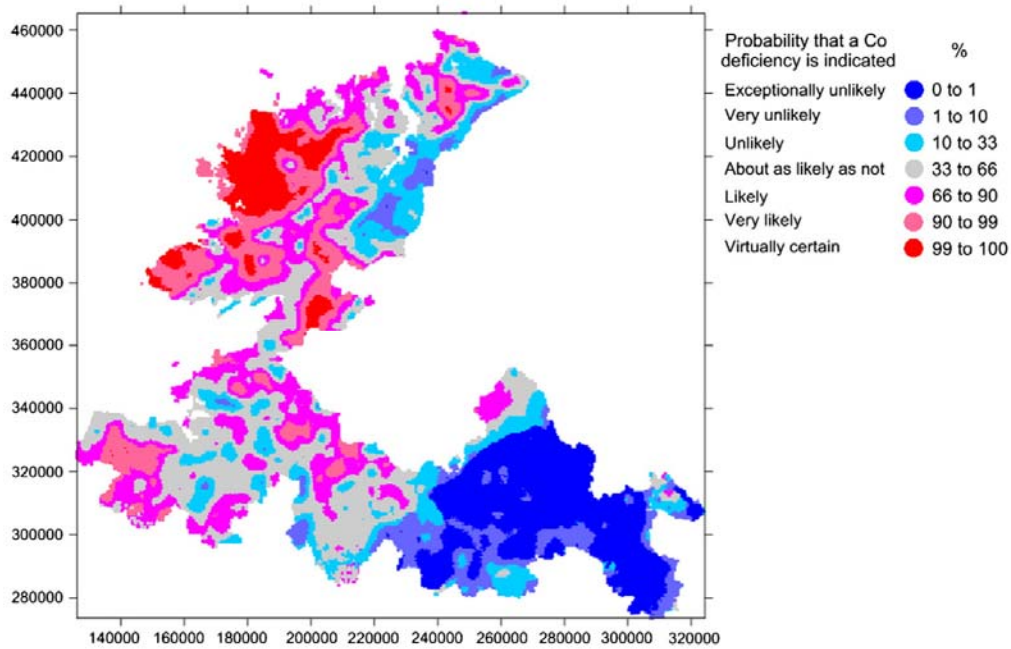


Figure 3.7. Verbal uncertainties using probabilistic data: from Lark et al. (2014).

3.6 Research directions

The core themes touched on in this review include understanding user needs for spatial soil information with connection to past and current soil mapping priorities. The priorities and investment logic for soil mapping have been steered by the economic and environmental issues at that time, e.g. regional reconstruction efforts from 1940-1955 to coincide with the end of the Second World War and returning servicemen. While the focus has ebbed and flowed between production and environmental for soil mapping (Section 3.3.1), understanding the current needs for users of spatial soil information (with reflection on past survey purposes) has been identified in this review as a research and knowledge gap. A case study focussing on spatial soil information needs of biophysical modellers as a key next user group is presented as Chapter 4.

To support these needs, there is a clear desire to have high resolution and contemporary soil information for decision making purposes. The costs and benefits of soil mapping are

clear from examples presented for Victoria, and from published benefit-cost studies (Section 3.3.3). Government is reluctant to invest in soil mapping programs even with the ability to leverage legacy survey information for new purposes and reducing the overall cost. This can be supplemented by the advances in information technology and methodologies pioneered in DSM towards providing soil information linked to decisions and soil management issues (e.g. pasture and crop health).

A complication to this synergy of legacy spatial soil information with DSM is the dependency on legacy data that is plagued by incomplete or incongruent data. Soil pH is one such property where extensive collections of pH measurements exist for field and laboratory based assessments. The ability to leverage this legacy data for mapping and modelling, supported by knowledge on the error sources that can be linked should enable more widespread use of this spatial data for mapping and monitoring purposes. This forms the basis for research investigations of Chapter 7.

Important to supporting this provision of spatial soil information is the implementation of uncertainty approaches that guide and inform users on the errors, assumptions and completeness of the information for use in decisions. The review of uncertainty analysis has identified that the DSM community has been proactive in the communication of uncertainty and leads many environmental mapping programs across the globe. Opportunities exist though to advance knowledge on error sources not commonly factored into error propagation approaches (e.g. epistemic uncertainty). This is the focus of Chapter 5 with an emphasis on soil change.

The advancement in uncertainty approaches accommodating epistemic uncertainties can be used in DSA or DSRA to provide greater certainty to users enabling more reliable management decisions. Changes in land use and management systems are one such

example. The approach presented by Lark et al. (2014) provides a useful template on which to base the communication of uncertainty in spatial soil information. A case study that integrates a more wholesome uncertainty analysis supported with communication of this to users is presented in Chapter 6.

The final research topic of this thesis combines legacy soil data and samples, scanned using infrared spectroscopy for predicting expensive and time consuming properties such as clay mineralogy (Chapter 8). While this review and the following chapter fail to identify these as priority or 'highly sensitive' properties to landscape models, their use in land evaluation assessment and other soil research is acknowledged. An opportunity to exploit the plethora of available data for mapping these difficult to measure properties by combining legacy data with spectral models is considered a research gap.

Acknowledgements

Murray Lark for access to the map (Figure 3.7) published in Geoderma (<http://dx.doi.org/10.1016/j.geoderma.2014.03.002>) under Creative Commons (<http://creativecommons.org/licenses/by/3.0/legalcode>).

References

- ACIL, 1996. The development of an economic framework and instruments for assessing the benefits and costs of land resource assessment in Australia. Report to the Australian Collaborative Land Evaluation Program. ACIL Economics & Policy Pty Ltd.
- Adewopo, J.B., VanZomerem, C., Bhomia, R.K., Almaraz, M., Eggleston, E., Judy, J.D., Lewis, R.W., Lusk, M., Miller, B., Moorberg, C., Hodges, Snyder E., Tiedman,

- M., 2014. Top-ranked priority research questions for soil science in the 21st century. *Soil Science Society of America Journal* 78, 337-347.
- Agumya, A., Hunter, G.J., 1999. Assessing “Fitness for Use” of geographic information: what risk are we prepared to accept in our decisions? In: Lowell, K., Jaton, A., (Eds.), *Spatial accuracy assessment: land information uncertainty in natural resources*. Ann Arbor Press, Michigan, 35-43.
- Alexander, J., MacEwan, R., Turner, G., Heemskerk, G., 2015. Spatial data and information management sub-program 2: User needs for soil data and information. Department of Economic Development, Jobs, Transport and Resources. State of Victoria, Epsom, Victoria, pp. 35.
- Alexandratos, N., Bruinsma, J., 2012. *WORLD AGRICULTURE TOWARDS 2030/2050. The 2012 Revision*. ESA Working Paper No. 12-03. Agricultural Development Economics Division, Food and Agriculture Organization of the United Nations.
- Amundson, R., Berhe, A.A., Hopmans, J.W., Olson, C., Sztein, A.E., Sparks, D.L., 2015. Soil and human security in the 21st century. *Science* 348, 1261071-1-6.
- Ballinger, R., 2012. Mildura (Former Shire of Walpeup) Heritage Study Stage 1. Vol. 1 Mildura Rural City Thematic Environmental History. Context Pty Ltd.
- Bardgett, R.D., van der Putten, W.H., 2014. Belowground biodiversity and ecosystem functioning. *Nature* 515, 505-511.
- Basher, L.R., 1997. Is pedology dead and buried? *Australian Journal of Soil Research* 35, 979-994.
- Beckenstein, A.R., Long, F.J., Arnold, M.B., Gladwin, T.N., 1996. *Stakeholder negotiations: exercises in sustainable development*. Boston. Richard Irwin Publishing.
- Beckett, P.H.T., Bie, S.W., 1978. Use of soil and land system maps to provide soil information in Australia. CSIRO Australia Division of Soils Technical Paper No. 33.

- Beckett, P.H.T., Burrough, P.A., 1971. The relation between cost and utility in soil survey. V. The cost-effectiveness of different soil survey procedures. *Journal of Soil Science* 22, 481-489.
- Beckett, P.H.T., Webster, R., 1971. Soil variability: a review. *Soils and Fertilizers* 34(1), 1–15.
- Behrens, T., Scholten, T., 2006. Digital soil mapping in Germany – a review. *Journal of Plant Nutrition and Soil Science* 169, 434-443.
- Benke, K.K., Pettit, C.J., Lowell, K.E., 2011. Visualisation of spatial uncertainty in hydrological modelling. *Journal of Spatial Science* 56, 73–88.
- Benkler, Y., 2006. The Wealth of Networks. http://www.benkler.org/Benkler_Wealth_Of_Networks.pdf.
- Bie, S.W., Beckett, P.H.T., 1970. The costs of soil survey. *Soils and Fertilizers* 33(3), 203-217.
- Binning, C.E., Cork, S.J., Parry, R., Shelton, D., 2001. Natural Assets: An Inventory of Ecosystem Goods and Services in the Goulburn Broken Catchment Canberra.
- Bishop, T., Horta, A., Karunaratne, S., 2015. Validation of digital soil maps at different spatial supports. *Geoderma* 241-242, 238-249. <http://dx.doi.org/10.1016/j.geoderma.2014.11.026>.
- Boettinger, J., Howell, D.W., Moore, A.C., Hartemink, A.E., Kienast-Brown, S., 2010. Digital soil mapping, Bridging research, environmental application, and operation. Springer.
- Bouma, J., 1986. Using soil survey information to characterize the soil-water state. *Journal of Soil Science* 37, 1-7.
- Bouma, J., van Lanen, H.A.J., Breeuwsma, A., Wösten, H.J.M., Kooistra, M.J., 1986. Soil survey data needs when studying modern land use problems. *Soil Use and Management* 2(4), 125-130.
- Bouma, J., 1989. Using soil survey data for quantitative land evaluation. In: Stewart, B.A., (Ed), *Advances in Soil Science* 9, 177-213. Springer-Verlag.

- Bouma, J., 2001. The role of soil science in the land use negotiation process. *Soil Use and Management* 17, 1-6.
- Bouma, J., 2009. Soils are back on the global agenda: Now what? *Geoderma* 150(2), 224-225.
- Bouma, J., Drooge, P., 2007. Translating soil science into environmental policy: A case study on implementing the EU soil protection strategy in The Netherlands. *Environmental Science & Policy* 10(5), 454-463.
- Bouma, J., McBratney, A.B., 2013. Framing soils as an actor when dealing with wicked environmental problems. *Geoderma*, 200-201, 130-139.
- Bouma, J., 2015. Policy and effective action for soil security: a need for reframing the soil story. *IOP Conference Series: Earth and Environmental Science* 25: 1-7. <http://dx.doi.10.1088/1755-1315/25/1/012002>.
- Brevick, E.C., Calzolari, C., Miller, B.A., Pereira, P., Kabala, C., Baumgarten, A., Jordán, A., 2015. Soil mapping, classification, and pedologic modelling: History and future directions. *Geoderma* 264, 256-274.
- Bui, E.N., 2006. A review of digital soil mapping in Australia. In: Lagacherie, P., McBratney, A.B., Voltz, M. (Eds.), *Digital Soil Mapping: An Introductory Perspective*. *Developments of Soil Science: Volume 31*, 25-38.
- Bui, E.N., McKenzie, N.K., Jacquier, D.W., Gregory, L.J., 2008. Synthesis studies: making the most of existing data. In: McKenzie, N.J., et al. (Eds.), *Guidelines for Surveying Soil and Land Resource*. 2nd edition. CSIRO Publishing, Collingwood, Victoria.
- Burrough, P.A., 1987. Mapping and map analysis: new tools for land evaluation. *Soil Use and Management* 3(1), 20–25.
- Butler, B.E., 1958. The diversity of concepts about soils. *The Journal of the Australasian Institute of Agricultural Science*, 14-20.
- Butler, B.E., 1963. *Can pedology be rationalized?* Australian Soil Science Society, Publication No. 3, Canberra.

- Carré, F., McBratney, A.B., Mayr, T., Montanarella, L., 2007. Digital soil assessments: Beyond DSM. *Geoderma* 142, 69-79.
- Caudeville, J., Goovaerts, P., Carré, F., Bonnard, R., Arrouays, D., Jolivet, C.C., Saby, N.B.A., Boudet, C., 2012. Spatial modelling of human exposure to soil contamination – an example of digital soil assessment. In Minasny, B., et al. (Eds.), *Digital Soil Assessments and Beyond*, pp. 19-24, CRC Press.
- Christian, C.S., 1978. Co-ordination of land resource survey and evaluation. In: Vol. Research Report No. 8. Department of Environment, Housing and Community Development.
- Cook, S.E., Corner, R.J., Grealish, G., Gessler, P.E., Chartres, C.J., 1996. A rule-based system to map soil properties. *Soil Science Society of America Journal* 60, 1893–900.
- Cook, S.E., Jarvis, A., Gonzalez, J.P., 2008. A new global demand for digital soil information. In: Hartemink, A.E., McBratney, A.B., Mendonça-Santos, M.L. (Eds.), *Digital soil mapping with limited data*, pp. 31-41. Springer.
- Craemer, R., Barber, M., 2007. Building a business case for investment in soil information. National Land and Water Resources Audit, Canberra.
- Dent, D., Young, A., 1993. *Soil survey and land evaluation*. E & FN Spon, London.
- Department of Agriculture, 2014. *The National Soil Research, Development and Extension Strategy, Securing Australia's Soil, For profitable industries and healthy landscapes*, Canberra.
- Diafus, I., Panagos, P., Montanarella, L., 2013. Willingness to pay for soil information derived by digital soil maps: a choice experiment approach. *Vadose Zone Journal* 12(4). <http://dx.doi.10.2136/vzj2012.0198>.
- Dokuchaev, V.V., 1886. *Materials on Land Evaluation of the Nizhni Novgorod Governorate. Natural and Historical Part: Report to the Nizhni Novgorod Governorate Zemstvo. Vol. 1: Key Points in the History of Land Evaluation in the European Russia, with Classification of Russian Soils*. Tipogr. Evdokimova, St. Petersburg.

- Downes, R.G., 1949. A soil, land-use and erosion survey of parts of the counties of Moira and Delatite, Victoria. CSIRO, Melbourne.
- Edgar, G., 1966. Australia agriculture in the post war era. Presentation on acceptance of the Farrer Memorial Medal for 1966. V.C.N. Blight, Government Printer.
- Finke, P.A., Meylemans, E., Van de Wauw, J., 2008. Mapping the possible occurrence of archaeological sites by Bayesian inference. *Journal of Archaeological Science* 35, 2786–2796.
- Finke, P.A., 2012. On digital soil assessment with models and the Pedometrics agenda. *Geoderma* 171-172, 3-15.
- Fisher, J., Crawford, M., 2015. The policy and science of soil change – a Victorian perspective. *IOP Conference Series: Earth and Environmental Science* 25: 1-10. <http://dx.doi.10.1088/1755-1315/25/1/012003>.
- Fisher, P.F., 1999. Set theoretic considerations in the conceptualization of uncertainty in natural resource information. In: Lowell, K., Jaton, A., (Eds.), *Spatial accuracy assessment: land information uncertainty in natural resources*. Ann Arbor Press, Michigan, 147-149.
- Gascoigne, J., Wadsworth, R., 1999. Mapping misgivings: Monte Carlo modelling of uncertainty and the provision of spatial information for international policy. In: Lowell, K., Jaton, A., (Eds.), *Spatial accuracy assessment: land information uncertainty in natural resources*. Ann Arbor Press, Michigan, 53-59.
- Giasson, E., Van Es, C., van Wambeke, A., Bryant, R.B., 2000. Assessing the economic value of soil information using decision analysis techniques. *Soil Science* 165(12), 971-978.
- Gibbons, F.R., 1961. Some misconceptions about what soil surveys can do. *Journal of Soil Science* 12, 96-100.
- Gibbons, F.R., 1981. *Soil mapping in Australia*. Soil Conservation Authority, Victoria.
- Gibbons, F.R., 1983. Chapter 21, *Soil mapping in Australia*. Soils an Australian viewpoint, pp. 267-276.

- Global Soil Partnership, 2014. Plan of action for pillar four of the Global Soil Partnership. Adopted by the GSP Plenary Assembly. FAO, Rome.
- Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York.
- Gottsegen, J., Montello, D., Goodchild, M., 1999. A comprehensive model of uncertainty in spatial data. In: Lowell, K., Jaton, A., (Eds.), Spatial accuracy assessment: land information uncertainty in natural resources. Ann Arbor Press, Michigan, pp. 175-182.
- Griffith, D.A., Haining, R.P., Arbia, G., 1999. Uncertainty and error propagations in map analyses involving arithmetic and overlap options: inventory and prospects. In: Lowell, K., Jaton, A., (Eds.), Spatial accuracy assessment: land information uncertainty in natural resources. Ann Arbor Press, Michigan, pp. 11-25.
- Grundy, M.J., Viscarra Rossel, R.A., Searle, R.D., Wilson, P.L., Chen, C., Gregory, L.J., 2015. Soil and Landscape Grid of Australia. Soil Research 53(8), 835-844. doi.org/10.1071/SR15191.
- Grunwald, S., 2009. Multi-criteria characterization of recent digital soil mapping and modelling approaches. Geoderma 152, 195-207.
- Grunwald, S., Barak, P., 2003. 3D geographic reconstruction and visualization techniques applied to land resource management. Trans. GIS 7(2), 231-241.
- Grunwald, S., Barak, P., McSweeney, K., Lowery, B., 2000. Soil landscape models at different scales portrayed in Virtual Reality Modeling Language (VRML). Soil Science 65, 598-614.
- Grunwald, S., McBratney, A.B., Thompson, J.A., Minasny, B., Boettinger, J.L., 2012. Digital Soil Mapping in a Changing World. In Minasny, B., et al. (Eds.), Digital Soil Assessments and Beyond, pp. 301-306, CRC Press.
- Grunwald, S., McSweeney, K., Rooney, D.J., Lowery, B., 2001. Soil layer models created with profile cone penetrometer data. Geoderma, 103, 181-201.

- Grunwald, S., Thompson, J.A., Boettinger, J.L., 2011. Digital soil mapping and modelling at continental scales: finding solutions for global issues. *Soil Science Society of America Journal* 75(4), 1201-1213.
- Hallsworth, E.G., 1978. Purposes and requirements of land resource survey and evaluation. Commonwealth and state government collaborative soil conservation study 1975-1977, Report 3. Department of Environment, Housing and Community Development, Canberra.
- Hartemink, A.E., McBratney, A.B., Cattle, J.A., 2001. Developments and trends in soil science: 100 volumes of *Geoderma* (1967–2001). *Geoderma* 100, 217–268.
- Hartemink, A.E., 2008. Soils are back on the global agenda. *Soil Use and Management* 24, 327-330.
- Hempel, J.W., Hammer, R.D., Moore, A.C., Bell, J.C., Thompson, J.A., Golden, M.L., 2008. Challenges to digital soil mapping. In: Hartemink, A.E., McBratney, A.B., Mendonça-Santos, M.L. (Eds.), *Digital soil mapping with limited data*, pp. 81-90, Springer.
- Hengl, T., Toomanian, N., 2004. Maps are not what they seem: representing uncertainty in soil property maps. In: Caetano, M., Painho, M., (Eds.), *7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*.
- Hengl, T., de Jesus, J.M., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B.M., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J.G.B., Walsh, M.G., Ruiperez Gonzalez, M., 2014. SoilGrids1km – Global soil information based on automated mapping. *PLoS ONE* 9(8), e105992. <http://dx.doi.10.1371/journal.pone.0105992>.
- Heuvelink, G.H.M., 1999. Aggregation and error propagation in GIS. In: Lowell, K., Jaton, A., (Eds.), *Spatial accuracy assessment: land information uncertainty in natural resources*. Ann Arbor Press, Michigan, pp. 219-225.
- Heuvelink, G.B.M., 1996. Identification of field attribute error under different models of spatial variation. *International Journal of Geographical Information Systems* 10, 921-936.

- Heuvelink, G.B.M., 2014. Uncertainty quantification of GlobalSoilMap products. In: Arrouays, D., McKenzie, N.J., Hempel, J., de Forges, A.R., Mcbratney, A.B. (Eds.), *GlobalSoilMap - Basis of the global soil information system*. CRC Press, The Netherlands, pp. 335-340.
- Heuvelink, G.B.M., Brus, D.J., De Vries, F., Vašát, R., Walvoort, D.J.J., Kempen, B., Knotters, M., 2010. Implications of digital soil mapping for soil information systems. *Proceedings of the 4th Global Workshop on Digital Soil Mapping*, Rome.
- Heuvelink, G.B.M., van Egmond, F.M., 2010. Space–Time Geostatistics for Precision Agriculture: A Case Study of NDVI Mapping for a Dutch Potato Field. In: Oliver, M.A., (Ed.), *Geostatistical Applications for Precision Agriculture*. pp. 117-137, Springer, The Netherlands.
- Hewitt, A., Dominati, E., Webb, T., Cuthill, T., 2012. Soil natural capital quantification by the stock adequacy method. *Geoderma* 241-242, 107-114.
- Holmes, K.W., Griffin, E.A., Odgers, N.P., 2015. Large-area spatial disaggregation of a mosaic of conventional soil maps: evaluation over Western Australia. *Soil Research* 53(8), 865-880. <http://dx.doi.org/10.1071/SR14270>.
- Hughes, P., McBratney, A.B., Malone, B.P., Minasny, B., 2012. Development of terrons for the Lower Hunter Valley wine-growing region. In: Minasny, B., et al. (Eds.), *Digital Soil Assessments and Beyond*, pp. 31-36, CRC Press.
- Imhof, M., Cox, M., Faderson, A., Harvey, W., Thompson, S., Rees, D., Pettit, C., 2011. Natural resource knowledge and information management via the Victorian Resources Online website. *Future Internet* 3, 248-280.
- Isbell, R.F., 1992. A brief history of national soil classification in Australia since the 1920s. *Soil Research* 30(6), 825-842.
- Isbell, R.F., 2002. *The Australian Soil Classification. Revised Edition*. Australian Soil and Land Survey Handbooks Series 4. CSIRO Publishing, Melbourne.
- Jenny, H., 1941. *Factors of soil formation*. McGraw-Hill, New York, NY.

- Kidd, D.B., Webb, M.A., Grose, C.J., Moreton, R.M., Malone, B.P., McBratney, A.B., Minasny, B., Viscarra Rossel, R.A., Cotching, W.E., Sparrow, L.A., Smith, R., 2012. Digital Soil Assessment: Guiding Irrigation expansion in Tasmania, Australia. In: Minasny, B., et al. (Eds.), *Digital Soil Assessments and Beyond*, pp. 3-8, CRC Press.
- Kidd, D., Webb, M., Malone, B., Minasny, B., McBratney, A.B., 2015. Digital soil assessment of agricultural suitability, versatility and capital in Tasmania, Australia. *Geoderma Regional* 6, 7-21.
- Kidd, D.B., Malone, B.P., McBratney, A.B., Minasny, B., Webb, M.A., 2014. Digital mapping of a soil drainage index for irrigated enterprise suitability in Tasmania, Australia. *Soil Research* 52(2), 107-119.
- Klingebiel, A.A., 1966. Costs and returns of soil surveys. *Soil Conservation* 32, 3-6.
- Kshetri, N., 2014. The emerging role of BiG Data in key development issues: opportunities, challenges and concerns. *Big data and society* July-December, 1-20.
- Lagacherie, P., 2008. Digital Soil Mapping: A state of the art. In: Hartemink, A.E., McBratney, A.B., Mendonça-Santos, M.L. (Eds.), *Digital Soil Mapping with Limited Data*, Springer.
- Lark, R.M., Ander, E.L., Cave, M.R., Knights, K.V., Glennon, M.M., Scanlon, R.P., 2014. Mapping trace element deficiency by cokriging from regional geochemical soil data: A case study on cobalt for grazing sheep in Ireland. *Geoderma* 226–227, 64-78.
- Lark, R.M., Knights, K.V., 2015. The implicit loss function for errors in soil information. *Geoderma* 251-252, 24-32.
- Leeper, G.W., 1956. THE CLASSIFICATION OF SOILS. *Journal of Soil Science* 7(1), 59-64.
- Leeper, G.W., Nicholls, A., Wadham, S.M., 1936. Soil Studies in the Mt. Gellibrand Area. *Proceedings of the Royal Society of Victoria* 49(1), 1936.
- Lindsay, A., Rowe, K., 1990. Land evaluation: the Victorian land capability assessment approach *Proceedings of the Ecological Society of Australia* 16, 475-490.

- Lemons, J., 1996. *Scientific Uncertainty and Environmental Problem Solving*. Blackwell Science: Cambridge, MA.
- Lorimer, M.S., 1990. *A Survey of the Current Attitude of Shire Councils in Victoria to Land Inventory and Land Capability Information*. Department of Conservation & Environment: Bendigo.
- Lowrance, W.W., 1976. 'Of Acceptable Risk: Science and the Determination of Safety.' (William Kaufmann: Los Altos, California).
- MacEwan, R.J., Robinson, N.J., Imhof, M.P., Rees, D.B., Hopley, J., Crawford, D.M. 2014. Reflections on GlobalSoilMap.net, DSM and the legacy of soil survey information in Victoria, Australia – the pedology matters. In: Arrouays, D., McKenzie, N.J., Hempel, J., de Forges, A.R., Mcbratney, A.B. (Eds.), *GlobalSoilMap, Basis of the global spatial soil information system*. CRC Press, The Netherlands, pp. 145-148.
- Maher, J., Martin, J.J., 1987. *Soils and Landforms of South-Western Victoria, Part 1*. State Chemistry Laboratory Melbourne.
- Malone, B.P., Minasny, B., Odgers, N., McBratney, A.B., 2014. Using model averaging to combine soil property rasters from legacy soil maps and from point data. *Geoderma* 232-234, 34-44.
- Manderson, A., Palmer, A., 2006. Soil information for agricultural decision making: a New Zealand perspective. *Soil Use and Management* 22, 393-400.
- Marshall, T.J., 1947. Mechanical composition of soil in relation to field descriptions of texture. Council for Scientific and Industrial Research, Bulletin No. 224, Melbourne.
- Martin, J.J., 1974. Soil type, mapping unit and wheat production in the Wimmera. Australian Soil Science Conference, La Trobe University, Melbourne.
- Martin, J.J., 1980. Justification and financial responsibility for soil survey. Victoria Department of Agriculture.
- Martin, J.J., 1998. A Brief History of the Work of the Soils Unit of the Department of Agriculture. <http://www.vic.dofaghistory.org/Soils.html>.

- Martin, J.J., 2006. History of Department of Agriculture Soil Surveys. http://vro.agriculture.vic.gov.au/dpi/vro/vrosite.nsf/pages/soil_dept_agriculture_surveys.
- Mastrandrea, M.D., Field, C.B., Stocker, T.F., Edenhofer, O., Ebi, K.L., Frame, D.J., Held, H., Kriegler, E., Mach, K.J., Matschoss, P.R., Plattner, G., Yohe, G.W., Zwers, F.W., 2010. Guidance note for lead authors of the IPCC fifth assessment report on consistent treatment of uncertainties. IPCC Cross-Working Group Meeting on Consistent Treatment of Uncertainties, Jasper Ridge, California, 6-7 July 2010.
- McBratney, A.B., Field, D.J., Koch, A., 2014. The dimensions of soil security. *Geoderma* 213, 203-213.
- McBratney, A.B., Minasny, B., Wheeler, I., Malone, B.P., van der Linden, D., 2012. Frameworks for digital soil assessment. In: Minasny, B., et al. (Eds.), *Digital Soil Assessments and Beyond*, pp. 9-14, CRC Press.
- McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. *Geoderma* 117, 3-52.
- McKenzie, N.J. 1991. A strategy for coordinating soil survey and land evaluation in Australia. Divisional report No. 114, CSIRO Division of Soils, Canberra.
- McKenzie, N.J., Grundy, M.J., Webster, R., Ringrose-Voase, A.J., 2008. *Guidelines for Surveying Soil and Land Resource*. 2nd edition. CSIRO Publishing, Collingwood, Victoria.
- Minasny, B., Bishop, T.F.A., 2008. Analysing uncertainty. In: McKenzie, N.J., et al. (Eds.), *Guidelines for Surveying Soil and Land Resources*, 2nd edition. pp. 383-394, CSIRO Publishing: Collingwood, Melbourne.
- Minasny, B., McBratney, A.B., 1999. A rudimentary mechanistic model for soil production and landscape development. *Geoderma* 90(1-2), 3-21.
- Minasny, B., McBratney, A.B., 2015. Digital soil mapping: A brief history and some lessons. *Geoderma* 264, 301-311.

- Minasny, B., McBratney, A.B., Lark, R.M., 2008. Digital soil mapping technologies for countries with sparse data infrastructures. In: Hartemink, A.E., McBratney, A.B., Mendonça-Santos, M.L. (Eds.), *Digital Soil Mapping with Limited Data*, pp. 15-30 Springer.
- Morse-McNabb, E., Sheffield, K., Clark, R., Lewis, H., Robson, S., Cherry, D., Williams, S., 2015. VLUIS, a land use data product for Victoria, Australia, covering 2006 to 2013. *Scientific Data* 2:150070. <http://dx.doi: 10.1038/sdata.2015.70>.
- Mowrer, H.T., 1999. Accuracy (re)assurance: selling uncertainty assessment to the uncertain. In: Lowell, K., Jaton, A., (Eds.), *Spatial accuracy assessment: land information uncertainty in natural resources*. Ann Arbor Press, Michigan, pp. 3-10.
- Mulder, V.L., de Bruin, S., Schaepman, M.E., Mayr, T.R., 2011. The use of remote sensing in soil and terrain mapping — A review. *Geoderma* 162, 1-19.
- Nichol, G., MacEwan, R., Pettit, C., Dorrrough, J., Suter, H., Hossain, H., Cherry, D., Beverly, C., Cheng, X., Melland, A., Huang, Z., McNeill, J., Hood, A., Rab, A., Sheffield, K., Duncan, D., 2006. *Landscape analysis models and frameworks – a review*. Our Rural Landscape 1.1 – New dimensions for agricultural landscapes. Department of Primary Industries, Bendigo, Victoria.
- Odeh, I.O.A., McBratney, A.B., 1996. Estimating the value of soil information: an example. *ACLEP newsletter* 5(4), 6-9.
- Olson, G.W., 1976. *Opportunities for soil survey interpretation*. A report to the Government of Victoria. Soil Conservation Authority, Kew, Victoria.
- Omuto, C., Nachtergaele, F., Vargas Rojas, R., 2013. *State of the art report on global and regional soil information: Where are we? Where to go?* Global Soil Partnership Technical Report. Food and Agriculture Organization of the United Nations, Rome, Italy.
- Prescott, J.A., 1944. *A soil map of Australia*. C.S.I.R. Aust. Bull. No. 177.
- Raupach, M., Tucker, B.M., 1959. The field determination of soil reaction. *Journal of the Australian Institute of Agricultural Science* 25, 129-133.

- Refsgaard, J.C., van der Sluijs, J.P., Højberg, A.L., Vanrolleghem, P.A., 2007. Uncertainty in the environmental modelling process – A framework and guidance. *Environmental Modelling & Software* 22, 1543-1556.
- Reich, P.B., Hobbie, S.E., 2013. Decade-long soil nitrogen constraint on the CO₂ fertilization of plant biomass. *Nature Climate Change* 3, 278-282.
- Robinson, N.J., MacEwan, R.J., Crawford, D.M., Rees, D.B., Rab, A.M., Imhof, M., 2010. Capturing soil landscape parameters to underpin regional biophysical modelling: Final report for MIS 08533. Department of Primary Industries, Epsom.
- Robinson, D.A., Hockley, N., Dominati, E., Lebron, I., Scow, I., Reynolds, B., Emmett, B.A., Keith, A.M., de Jong, L.W., Schjøningg, P., Moldrup, P., Jones, S.B., Tuller, M., 2012. Natural Capital, Ecosystem Services, and Soil Change: Why Soil Science Must Embrace an Ecosystems Approach. *Vadose Zone Journal* 11(1). <http://dx.doi.org/10.2136/vzj2011.0051>.
- Robertson, F., Armstrong, R., Partington, D., Perris, R., Oliver, I., Aumann, C., 2015. Effect of cropping practices on soil organic carbon: evidence from long-term field experiments in Victoria, Australia. *Soil Research*. <http://dx.doi.10.1071/SR14227>.
- Rowe, R.K., Howe, D.F., Alley, N.F., 1988. *Soil Conservation Practice, Manual of Guidelines for Land Capability Assessment in Victoria*. Department of Conservation, Forestry and Lands.
- Rossiter, D.G., Liu J., Carlisle, S., Zhi, A-X., 2015. Can citizen science assist digital soil mapping? *Geoderma* 259-260, 71-80.
- Roudier, P., Ritchie, A., Hedley, C., Medyckyj-Scott, D., 2015. The rise of information science: a changing landscape for soil science. *IOP Conference Series: Earth and Environmental Science* 25: 1-10. <http://dx.doi.10.1088/1755-1315/25/1/012023>
- Sanchez, P.A., Ahamed, S., Carré, F., Hartemink, A.E., Hempel, J., Huising, J., Lagacherie, P., McBratney, A.B., McKenzie, N.J., Mendonca-Santos, M.L., Minasny, B., Montanarella, L., Okoth, P., Palm, C.A., Sachs, J.D., Shepherd, K.D., Vagen, T-G., Vanlauwe, B., Walsh, M.G., Winowiecki, L.A., Zhang, G-L., 2009. *Science* 325, 680-681. <http://www.sciencemag.org/content/325/5941/680.short>.

- Schoknecht, N., Wilson, P.R., Heiner, I., 2008. Survey specification and planning. In: McKenzie, N.J., et al. (Eds.), *Guidelines for Surveying Soil and Land Resource*. 2nd edition. CSIRO Publishing, Collingwood, Victoria.
- Scott, A., Olley, J., 2003. *Settlement, erosion and muddy waters. Lessons from the past*. Murray Darling Basin Commission.
- Scull, P., Franklin, J., Chadwick, O.A., McArthur, D., 2003. Predictive soil mapping: a review. *Progress in Physical Geography* 27(2), 171-197.
- Shanks, A., 2006. Analysis of the landscape questions of Victorian natural resource managers. *Our Rural Landscape 1.1: New dimensions in agricultural landscapes*. Department of Primary Industries.
- Sheffield, K., Morse-McNabb, E., 2015. Using satellite imagery to assess trends in soil crop productivity across landscapes. *IOP Conference Series: Earth and Environmental Science* 25: 1-10. <http://dx.doi.10.1088/1755-1315/25/1/012013>.
- Sheffield, K., Morse-McNabb, E., Clark, R., Robson, S., Lewis, H., 2015. Mapping dominant annual land cover from 2009 to 2013 across Victoria, Australia using satellite imagery. *Scientific Data* 2, 150069. <http://dx.doi.10.1038/sdata.2015.69>.
- Sinclair, S.J., White, M.D., Medley, J., Smith, E., Newell, G.R., 2012. Mapping the past: Constructing a digital land-use history map for Victoria, Australia. *Proceedings of the Royal Society of Victoria* 124, 193-206.
- Stockmann, U., Padarian, J., McBratney, A., Minasny, B., de Brogniez, D., Montanarella, L., Hong, S., Rawlins, B.G., Field, D.J., 2015. Global soil organic carbon assessment. *Global Food Security* 6, 9-16.
- Steinitz, C., 1990. A framework for theory applicable to the education of landscape architects (and other environmental design professionals). *Landscape Journal* 9, 136–143.
- Stephens, C.G., 1953. *A Manual of Australian Soils*. CSIRO, Melbourne.
- Sun, X.L., Zhao, Y.G., Wu, Y.J., Zhao, M.S., Wang, H.L., Zhang, G.L., 2012. Spatio-temporal change of soil organic matter content of Jiangsu Province, China, based on digital soil maps. *Soil Use and Management* 28(3), 318-328.

- Taylor, J.K., 1970. The development of soil survey and field pedology in Australia, 1927-1967. CSIRO, Melbourne.
- Terrible, F., Coppola, A., Langella, G., Martina, M., Basile, A., 2011. Potential and limitations of using soil mapping information to understand hydrology. *Hydrology and Earth System Sciences* 15, 3895-3933.
- The Argus., 22/08/1923. <http://trove.nla.gov.au/newspaper/page/425804>.
- Thomas, M., Dalglish, N., Wilson, P., Hubert, K., Jacquier, D., Swan, G., Holzworth, D., Birks, M., Brunner, L., Davis, P., 2012. SoilMapp – mobilising Australian soil data. "Capturing Opportunities and Overcoming Obstacles in Australian Agronomy". Edited by I. Yunusa. Proceedings of 16th Australian Agronomy Conference 2012, 14-18 October 2012, Armidale, NSW.
- Thompson, G.T., 1979. A brief history of soil conservation in Victoria, 1834-1961. Soil Conservation Authority, Melbourne.
- Thompson, J.A., Roecker, S., Grunwald, S., Owens, P.R., 2012. Digital soil mapping: interactions with and applications for hydrology. *Hydrology*, 1st edition, 2012, 665-709. [http://dx.doi.org/10.1016.B978-0-12-386941-8.00021-6](http://dx.doi.org/10.1016/B978-0-12-386941-8.00021-6).
- Tugel, A.J., Herrick, J.E., Brown, J.R., Mausbach, M.J., Puckett, W., Hipple, K., 2005. Soil change, soil survey, and natural resource decision making: a blueprint for action. *Soil Science Society of America Journal* 69, 738-747.
- van Ittersum, M.K., Cassman, K.C., Grassini, P., Wolf, J., Tittonell, P., Hochman, Z., 2013. Yield gap with local to global relevance – a review. *Field Crops Research* 143, 4-17.
- Vanwalleghem, T., Stockmann, U., Minasny, B., McBratney, A.B., 2013. A quantitative model for integrating landscape evolution and soil formation, *Journal of Geophysical Research* 118, 1-17. <http://dx.doi.org/10.1002/JGRF.20026>.
- Viscarra Rossel, R.A., Chen, C., Grundy, M.J., Searle, R., Clifford, D., Campbell, P.H., 2015. The Australian three-dimensional soil grid: Australia's contribution to the GlobalSoilMap project. *Soil Research* 53(8), 845-864. <http://dx.doi.org/10.1071/SR14366>.

-
- Webster, R., 1985. *Advances in Soil Science*, Volume: 3. Springer-Verlag, pp. 1-70.
- Webster, R., 1994. The development of pedometrics. *Geoderma* 62(1-3), 1-15.
- Wieder, W.R., Bonan, G.B., Allison, S.D., 2013. Global soil carbon projections are improved by modelling microbial processes. *Nature Clim. Change* 3(10), 909-912.
- Wilson, P., 2012. Improving Australian soil data and information governance. For the Australian Collaborative Land Evaluation Program. CSIRO, Canberra.
- Wilson, P., Thomas, M., 2012. Digital soil mapping in ACLEP 2012 and beyond. CSIRO, Australian Collaborative Land Evaluation Program.
- Wood, B., Auricht, C., 2011. ASRIS / ACLEP user needs analysis. Australian Collaborative Land Evaluation Program.
- Zaks, D.P.M., Kucharik, C.J., 2011. Data and monitoring needs for a more ecological agriculture. *Environmental Research Letters* 6, 1-10.
- Zuiderwijk, Janssen, M., 2014. Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly* 31, 17-29.

Chapter 4 Soil data for biophysical models in Victorian landscapes: current needs and challenges

The literature review (Chapter 3) emphasized the inadequacy of detailed information on the needs of users for spatial soil information. Recent attention paid to the topic of user needs has touched on the diversity within and among user groups of spatial soil information. A critical prerequisite of future soil mapping efforts is to engage and understand the specific requirements of users to support their application of this information. The aim of this research was to *identify what are users' needs for spatial soil information and how this has changed* by focussing on the requirements of biophysical modellers as a key user group. Therefore the objective of this chapter was to:

Define what spatial soil information is sought by users to support biophysical models for agricultural landscapes.

This chapter reviews the demand for, and trends in, soil property data for various models being used to support government policies and programs in the State of Victoria, Australia, over the 2009-2014 period. The use of biophysical models to support agricultural industries management for increased food production and environmental protection is on the rise.

I examined perceptions of the parameters that affect model sensitivity and error through surveys, workshops and interviews of public sector biophysical modellers. Although the data requirements to support the models and their sensitivities have remained similar over the 5-year period, there has been a decrease in the diversity of models used. There is evidence of increased application of models at point/site scale to support grains, dairy and livestock production industries in Victoria. This narrowing of model selection and soil

data requirements has occurred at a time when input data for models has never been more prolific. The vast array of available data sources will require evaluation and harmonization as part of solutions that integrate error sources through uncertainty approaches (Chapter 5). Opportunities are identified to deliver finer scale soil data from digital soil mapping to better meet modelling requirements across different scales for agricultural industries in Victorian landscapes.

4.1 Introduction

Since the seminal modelling of global population and resources by the Club of Rome (Meadows et al., 1972), increased computing power has led to more sophisticated biophysical models that are used to support agricultural industries' management for increased food production and environmental protection. Such biophysical models simulate the biological, chemical and physical processes of agricultural systems (Keating and Grace, 1999; Boote et al., 2010) and are increasingly implemented as tools to model agricultural landscapes and support decision making processes by farmers and their advisers (Bergez et al., 2010). These biophysical models enable users to test and answer important questions on land use and condition as well as management and production scenarios.

4.1.1 Model limitations

Models must become more robust to represent scenarios that can include critical changes in climate, management practices and farming systems in the future (Sinclair and Seligman, 1996; Asseng et al., 2013). Successful modelling relies on available and accurate topographic, climatic, land use and soil data (Bouma et al., 1986). Soil data may represent steady state and/or dynamic processes depending on the complexity of the model. Data from soil survey and mapping has focused on static properties rather than those that change (Bouma et al., 1986). While static properties have an important role in modelling, dynamic properties must also be modelled for many soil processes and interactions between biosphere, hydrosphere and pedosphere (Wagenet et al., 1991). There are likely consequences as estimation of soil properties may introduce considerable error into models.

Soil scientists need to understand the role and importance of soil data in the modelling process to enable the delivery of available, current, reliable and plausible soil data for

these models. Model developers understand the soil data required to support their model, including error and uncertainty from parameter estimation, systematic bias and sensitivity. Baker (1996) suggests that model developers need to be honest about the limitations of models and the research required to address these. Making end-users (e.g. land managers) aware of these limitations in soil data or the model itself is central to the ongoing success and utility of farming systems models for decision making and management (Keating and McCown, 2001).

4.1.2 Soil data availability

Due to the rapid expansion and use of soil data in digital form provided by sensors, conventional soil maps have become largely unsuitable for many users who wish to view soil data at finer scales (Bouma, 1989). Advances in technology and development have seen a global surge in sensing and acquisition of data, its collection, management and availability.

Referred to as the ‘New Digital Age’ (Schmidt and Cohen, 2013), or ‘Era of Big Data’ (Boyd and Crawford, 2012; Mayer-Schonberger and Cukier, 2013), the current period provides unprecedented opportunities for an improved understanding of our global environments including agroecosystems. The use of volunteered geographic information and citizen science is also contributing substantially to the volume of soil (Rossiter et al., 2015) and environmental data (Fienen and Lowry, 2012; Werts et al., 2012; Sui et al., 2013). As governments adopt open data policies (Zuiderwijk and Janssen, 2014) this emerging collaboration of large data arrays and analytical procedures with progressive and complex modelling will potentially enhance management philosophies of agricultural industries globally.

4.1.3 Understanding soil data needs of biophysical models

Research into users' soils data needs is scarce (Omuto et al., 2013). Wagenet et al. (1991) discuss the data requirements of simulation models and how existing soil survey plus predictive functions (pedotransfer functions) can supply a minimum dataset that includes dynamic soil properties that respond to land management change or climatic impacts such as flooding. Nichol et al. (2006) in a review of models and methods for landscape analysis defined the key model sub-domains that require soil and land attributes such as: hydrological, plant growth (crop, pasture or forestry), carbon and climate change, ecology, and biodiversity. This review of qualitative and quantitative biophysical models identified their soil data requirements and examples of where they have been implemented.

A complementary study by Robinson et al. (2010b) collated modellers' opinions regarding the key soil properties affecting sensitivity for these same biophysical models. Wood and Auricht (2011) defined current and future soil information requirements for the Australian Soil Resource Information System (ASRIS) based on interviews with selected modellers and the responses given to requests for data and information from ASRIS. This review identified a suite of physical, chemical, hydrological, biological and site characteristics at various scales that were sought by ASRIS users.

4.1.4 Collection of soil data for modelling

The synthesis and delivery of soil data to support modelling is subject to government priorities (MacEwan et al., 2014), advances in research, and changes in user needs for soil data to address questions posed. There is a constant need to adapt and enhance soil survey information as new questions arise (Bouma, 1989). Questions relate to systems that operate at different scales, requiring soil data at different levels of detail (Bouma, 2001).

Given the multiple challenges of scale, evolving needs of users and the availability of soil data in various formats, it is timely to ask if the right soil data to support sustainable agricultural development is being provided. This should then focus delivery of soil data on properties of direct relevance to improve model predictions and consequent decisions. In this paper, we present an example for the state of Victoria, Australia, that identifies (i) the simulation models used in agricultural industries, and the application scale at which these models are implemented, to support government policies and programmes, (ii) the soil data that modellers perceive as affecting model sensitivity and uncertainty, and (iii) any changes and trends in the demand for soil property data in the last 5 years. Future challenges in soil data and information provision to support modelling are discussed, including the context of demand, availability of soil data in various formats and how this will assist in the parametrization process of biophysical models for optimising agriculture management.

4.2 Methods

The study uses qualitative and quantitative data from surveys, focus groups and unstructured interviews summarised from an expert workshop in 2009 (Robinson et al., 2010) and follow up survey in 2014. The workshop was conducted in March 2009 to establish what biophysical models were applied and used soil data, what were the sources of the data, how sensitive were these models to the data, and what the future requirements for data in modelling applications were. The 2014 survey was undertaken to investigate changes in demand for soils data in models and included modellers that attended the 2009 workshop. Responses from 2009 and 2014 were collected using different evaluation techniques and it is recognised that participants respond differently between questionnaire

and interview prompts (Oei and Zwart, 1986). While focus groups enable thorough and engaging dialogue on complex topics, and surveys enable objective assessment of responses, a desirable approach is to combine the two approaches that enable qualitative and quantitative responses to be collated. Sound quantitative data analysis and interpretations can be explained and reinforced by qualitative responses. This supports the utilisation of these two evaluation techniques in the workshop in 2009 and justified a comparison with those of the 2014 survey.

4.2.1 Study design and data collection

Researchers from the former Victorian state government agencies (Department of Primary Industries and Department of Sustainability and Environment) and the University of Melbourne participated in the study. Participants include 23 model developers and practitioners in 2009 and 31 in 2014, operating in a diversity of model domains and sub-domains including agricultural production, ecological sciences, catchment hydrology, environmental pollution and nutrient flow.

These modellers were chosen as they are recognised as specialists in operating these models for landscape modelling and assessment (Nichol et al., 2006).

Modellers that participated in the workshop were assigned to four modelling sub-domains that use soil data, including:

- Forestry and biodiversity (FB)
- Carbon and greenhouse (CG)
- Crops, pastures and nutrients (CPN)
- Hydrological processes (HP).

Responses from participants were recorded using a survey questionnaire and focus groups as part of the workshop representing these modelling sub-domains. This approach enabled exploration of questions and associated issues further with all workshop participants.

The knowledge gained from responses at this workshop was used to refine questions for the online survey in 2014. This online web-survey was conducted using Survey Monkey[®] (www.surveymonkey.com). The questions that were developed for this study include:

1. What models are being used, at what spatial modelling scale are they applied and what soil data are being used to run these?
2. To what industry/land use are the models applied?
3. What are the key soil data for the models applied including the spatial scale of the input soil data?
4. Are the applied models sensitive to the soil property data?

This study synthesizes results from both the 2009 and 2014 surveys. After the completion of the 2014 survey, additional follow-up interviews were conducted with selected modellers to test initial conclusions and to identify logic for changes observed between the surveys.

Responses recorded from the online survey and workshop included whether a modeller applied a model at a particular scale (not how frequently). Model implementation has been reported as an ‘application’ and no specific time-constraints were stipulated to respondents on this application of the model.

4.2.2 Biophysical models

Many different types of models exist, including computer, conceptual, descriptive, deterministic, empirical, mathematical, mechanistic and stochastic. Although these have different operational architectures, nearly all are used in landscape analysis. Biophysical sciences and systems including ecology; soil and water/hydrology; solid earth processes; hydrogeology; agricultural production; environmental pollution and nutrient flow; and

land-use change interact, are modelled and used in ensemble biophysical simulation modelling approaches.

For this study, the models were selected from the review by Nichol et al. (2006) and listed as being used by participants at the 2009 workshop. This suite of models used, either individually or as part of a model ensemble, is assigned against the modelling sub-domain groups of the 2009 workshop and are briefly described in Table 4.1.

4.2.3 Modelling scale

For modelling, there is a scale that is distinct from both time and spatial scale, known as the ‘modelling’ (or working) scale (Blöschl and Sivapalan, 1995). The modelling scale generally reflects the process of interest and the design of the model being applied. For this study, the spatial modelling scales described by Dooge (1982; 1986) have been used with slight modifications to reflect the application of these models to agriculture:

- Local scale (point or site) – 1 to 30 m,
- Paddock scale – generally 100 to 1000 m,
- Catchment scale – 10 km,
- Regional scale – 1000 km

Table 4.1. Identified models that have been applied in Victoria.

Modelling domain	Model	Description
Hydrological processes	SWAT	The Soil Water and Assessment Tool (Neitsch et al., 2001) is a continuous time model that can be used to assess impacts of management and climate on water supplies and watersheds. The model is often applied in large river basins.
	CAT	Catchment Analysis Tool (CAT) is a multilayer hydrological model that links biophysical data and models at a range of spatial and temporal scales (Beverly 2007). This model ensemble uses existing models for crop growth, forest growth, grazing systems, water balance and groundwater.
	WaterCAST	The Water and Contaminant Analysis and Simulation Tool for catchment modelling (Argent et al., 2008) is a hydrological tool that selects and links component models to predict flow and constituent loads (e.g. sediment, nutrients, salt) at defined points in a river network over different time steps and scales. Core processes operating include runoff generation, constituent generation and filtering.
	HYDRUS	The model can simulate water movement in multiple dimensions (1, 2 or 3) with heat and solutes in variably saturated media (Šimůnek et al., 2011). The Richards equation is used to simulate the transport mechanisms from hydraulic properties estimated by van Genuchten (1980) functions for soil textural classes.
	Howleaky?	This model is a decision support system that assesses the impacts of land uses, soil conditions, management and climate on water balance and quality (McClymont et al., 2011). Largely based on PERFECT (Littleboy et al., 1992), the model operates in 1 dimension at various scales on a daily time step.
	PERFECT	The Productivity Erosion Runoff Functions to Evaluate Conservation Techniques (Littleboy et al., 1992) model is used for cropping and pasture systems to predict water balance for management sequences. Operating on a daily time step, soil water is updated daily in this one dimensional model from crop/pasture sequence criteria and management parameters.
	CatchMODS	The Catchment Scale Management of Diffuse Sources (CatchMODS) framework simulates the effects of different management actions on nutrients into surface water systems. CatchMODS (Newham et al., 2004) applies biophysical datasets including stream networks, soil properties, rainfall distribution, land use and climate in a decision support framework with additional models able to be linked.
	RUSLE	The Revised Universal Soil Loss Equation is a modification of the USLE to account for effect of slope steepness and length on erosion. This soil erosion model uses a combination of terrain, soil and vegetation inputs to predict hillslope erosion (sheet and rill) potential.

Table 4.1 (continued)

Modelling domain	Model	Description
	DRAINMOD	DRAINMOD (Skaggs 1980) simulates the hydrology of poorly drained, high water table soils at high temporal frequency. Input variables including rainfall and evapotranspiration, rooting depth of crops and pastures and soil properties are used to predict impacts of management on soil water and crop yield. The model is applied at paddock or watershed scales.
Crops, pastures and nutrients (Growth)	APSIM (and Yield Prophet)	Agricultural Production Systems Simulator is a modelling systems framework (Keating et al., 2003) to integrate biophysical modules that simulate processes in farming systems with management scenarios. Primarily concerned with plant production, the model also has economic and environmental models linked to support users with decisions at point and paddock scale.
	FNLI	Farm Nutrient Loss Index is a decision support tool concerned with nutrient loss through hydrological and atmospheric pathways. The model is applied generally at paddock or 'group of paddocks' scale to identify key factors in availability and transport on nutrients (Melland et al., 2004).
	CROPSYST	Cropping Systems Simulation Model uses biophysical data inputs including climate, soil properties, crop details and management operation data to understand effects of these factors on production and environment. Modules can be included or excluded for various scenarios. The model is a multi-year, multi-crop, daily time step crop growth simulation model (Stockle and Nelson 1996).
	DAIRYMOD / SGS	This model operates on a daily time-step for pasture growth scenarios in dairy farming systems (Johnson, 2013). Several modules including pasture growth, water (rainfall and irrigation), soil organic matter and nitrogen dynamics, animal attributes including growth and lactation, stock movement and management.
	GrassGro	This decision support tool is used in sheep and beef industries to quantify the variability in pasture and animal production, and associated risks. Climate, soil properties, management, pasture composition are all inputs. Operating on a daily time step, daily weather data is used to model processes of pasture growth and animal production (Simpson et al., 2002).
	FAO-56	This soil water evaporation model is able to identify evaporation and transpiration components in irrigated and rainfed crop settings. A crop coefficient is derived from crop details and averaged soil evaporation data. The model can be run on a daily time step (Allen et al., 1998).

Table 4.1 (continued)

Modelling domain	Model	Description
Carbon and greenhouse	RothC	This model operates on a monthly time step to simulate the turnover of organic carbon in non-waterlogged surface soil. Outputs from the model include an estimate of total organic carbon, microbial biomass and $\Delta 14C$ calculated on a years to centuries timescale. Climate, soil properties and management inputs are the primary data used to model the breakdown of organic carbon inputs into active components (www.rothamsted.ac.uk/sustainable-soils-and-grassland-systems/rothamstedcarbon-model-rothc).
	FullCAM	A carbon accounting system, FullCAM supports the estimation of carbon stock change on forest and agricultural systems. The model is comprised of two sub-modules that simulate C in live vegetation, soil, debris and products, and can account for management practices and interventions. The system integrates various models including 3-PG and RothC and can accommodate sensitivity and uncertainty analysis to provide prediction error estimates (http://www.environment.gov.au/climate-change/greenhouse-gas-measurement/land-sector).
	CENTURY	The CENTURY model attempts to simulate the plant-soil environment including carbon and nutrient dynamics for different types of ecosystems (grasslands, forest, crops, and savannahs). Crop and grassland modules are used with a soil organic matter submodel to simulate flow and pools (organic and inorganic) of carbon, nitrogen, phosphorus and sulphur (Metherell et al., 1993).
Forestry and biodiversity	3PG+	The 3PG+ forest growth model (based on 3PG) calculates dry mass production from the photosynthetically active radiation (PAR). Biophysical input factors including temperature, soil properties including soil water availability and salinity are used in empirical relationships to predict yield (Landsberg and Waring, 1997). The model has a multilayer soil water balance calculated on a daily time step (Morris and Baker, 2002) and has been integrated into the CAT model.

4.2.4 Industry/land use

The dominant agricultural land uses where simulation models have been implemented in Victoria (Figure 4.1; Morse-McNabb et al., 2015) are listed (Table 4.2), together with statistics for estimated area and value of agricultural commodities (Australian Bureau of Statistics for 2012–13 — www.abs.gov.au).

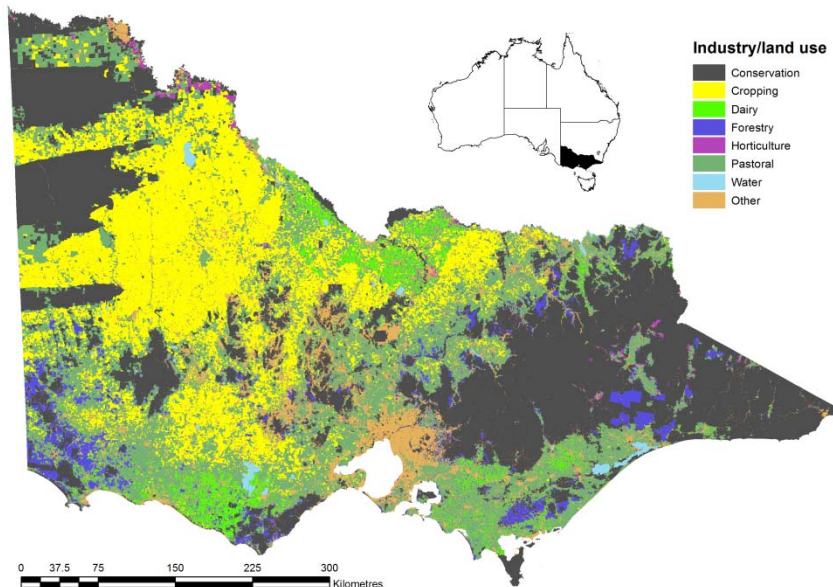


Figure 4.1. Agricultural industries and land uses in Victoria (2014) – from <https://www.data.vic.gov.au/data/dataset/victorian-land-use-information-system-2014>.

4.2.5 Soil data requirements of models

The reviews of Wagenet et al., (1991), Nichol et al., (2006), Robinson et al., (2010) and Wood and Auricht (2011) with the GlobalSoilMap (GSM) twelve priority properties (www.GlobalSoilMap.net) were used to establish a preliminary set of soil hydrological, physical, chemical, biological and ancillary properties required for models (Table 4.3). This was used to develop targeted questions for participants in the 2014 web-survey.

Table 4.2. Land use, estimated area and value of agricultural commodities in Victoria.

Land use	Area (hectares)	Commodity value (\$ bn)
Dairy	602 764	\$3.7
Cropping	4 256 006	\$3.24
Horticulture	197 069	\$2.07
Other pastoral	6 251 002	\$3.57
Forestry	517 744	
Conservation	3 638 186	

4.2.6 Model sensitivity

The term 'sensitivity' is sometimes used interchangeably with 'important', 'correlated', 'effective' or 'influential'. Crick et al., (1987) defined 'important' parameters to models as those whose uncertainty contributes greatly to the uncertainty in assessment results, while sensitive parameters are those that have a significant influence on model results.

The 2009 workshop used simple descriptive classes for model sensitivity to soil data including: (1) Highly sensitive (requires high precision and error terms), (2) Class or category (e.g. loam, in-lieu of particle size distribution), or (3) Qualitative estimate. Here the term 'sensitivity' was used; firstly, to identify if a parameter was viewed as 'critical', or necessary in the implementation of the model, and secondly, if the particular attribute required precision and accuracy to achieve reliable model results. The 2014 online survey adopted the same definition for sensitivity but respondents were asked only to identify whether or not the model was sensitive to a soil property.

4.3 Results

The online survey achieved an 83% response rate with 31 modellers that were distributed among the modelling sub-domains (FB=5, CG=6, CPN=10, HP=10) completing the survey. Of the 23 modellers that participated in the 2009 workshop, 18 completed the 2014 online web survey thus enabling changes in soil data priorities to be analysed (Table 4.4). Those that didn't complete the survey were either no longer working in government or had moved to another position that didn't include modelling.

Table 4.4. Survey summaries.

Survey	Year	Evaluation technique	No. of respondents	modeller	Overall response rate (%)
Workshop	2009	Focus group and survey	23		88*
Online survey	2014	Questionnaire	31		83

* at least three recognised modellers that were unable to attend the workshop

4.3.1 Model users

Some change has been observed in the application of models at various scales by modellers between 2009 and 2014 (Table 4.5). For the dairy industry, DAIRYMOD and CAT achieved the highest number of users in 2014 (three each). Other models applied include plant production focused models, including APSIM and GrassGro, and models linked to nutrients and hydrology, including SWAT, PERFECT and Howleaky?. The cropping industry also has a diversity of models (12 in 2009 and 9 in 2014) applied at the four modelling scales. APSIM (13 users) and CAT (six users) registered the most users in 2014 while CROPSYST was also used with a production focus and RothC implemented for carbon accounting and quantification purposes.

In the other pastoral industries (primarily lamb, wool and beef), up to 13 models have been applied, reflecting a diversity of production issues including nutrient loss and catchment hydrology impacts. RothC has been used for carbon modelling in Victorian pastoral systems. In horticulture, modellers have applied up to six models in 2009 and 2014 including CAT, CROPSYST and FAO-56. The forestry industry has seven different models applied with 3PG+ and CAT the most common. The forestry industry models have focused on catchment hydrology responses including water yield and harvest in response to plantations of native and introduced species. The carbon accounting model FullCAM has also been used for estimation of terrestrial ecosystem carbon stocks on forested land.

Table 4.5. Responses from modellers to application of a model at a particular spatial modelling scale for agricultural industries for the 2009 workshop and 2014 survey.

Survey	Agricultural industry*modelling scale				
	Dairy	Cropping	Other pastoral	Horticulture	Forestry
2009	13	20	18	7	9
2014	17	30	22	7	12

4.3.2 What models, scales and industries?

Simulation models most widely applied at any scale for the primary agricultural industries in 2009 include SWAT, CAT and HYDRUS and in 2014 CAT, APSIM and PERFECT (Table 4.6). In 2009, a modeller averaged 3.6 model applications (model x modelling scale) while in 2014 this had increased to 5.1 (n=153).

Table 4.6. Number of modellers to apply a simulation model for the primary agricultural industries from the 2009 workshop and 2014 survey.

Model	Dairy		Cropping		Other pastoral		Horticulture		Forestry and biodiversity	
	2009	2014	2009	2014	2009	2014	2009	2014	2009	2014
SWAT	1	2	2	1	2	1	2		3	2
HYDRUS	1		1	2	1	1	1	1	2	
PERFECT	1	2	2	3	2	2				4
GrassGro	1	1			1	2				1
3PG+									2	6
APSIM		2	4	13	1	3				1
RothC		1	2	2	1	2	1		2	1
CAT	2	3	2	5	3	5		2	1	8
Howleaky?	1	2	1	1	1	2				
CROPSYST			2	2				2		
DAIRYMOD	1	3				2				
FAO-56	1	1	1	1	1	2	1	2	1	1
WaterCAST										1
FullCAM			1		1		1		2	4
RUSLE	1		1		1					
FNLI	1									
CENTURY									1	
DRAINMOD	1				1					
CatchMODS	1		1		1					

In 2009 there were 19 different models used, but only 14 in the 2014 survey. An increase was observed for the plant growth modelling sub-domain from 17 applications in 2009 to 47 in 2014 and for forestry and biodiversity from 7 in 2009 to 13 in 2014. The hydrological processes modelling sub-domain accounted for the most model applications

(44 in 2009 and 76 in 2014) while a decrease was observed in carbon modelling from 12 applications in 2009 to eight in 2014.

At the local modelling scale (point/site), there has been an increase from 22 applications in 2009 to 52 in 2014 (Table 4.7). This can be partly attributed to modellers using multiple models for comparison purposes or in ensembles. The most widely used point/site models from responses in 2014 are the APSIM and CAT models.

Table 4.7. Model applications at the four spatial modelling scales for the 2009 workshop and 2014 survey.

Survey	Spatial modelling scale			
	Point/site	Paddock	Catchment	Region
2009	22	31	20	10
2014	52	45	38	23

More models were applied at the paddock scale than at a site/point scale in 2009, although this did reverse in the 2014 survey. For catchment modelling, nearly twice as many model applications were identified in 2014 as in 2009. In 2009, over 75% of responses for catchment scale modelling were from the hydrological processes model subdomain where 10 different models were used. The least frequent combination of model application by modelling scale is at the regional scale. The number of model applications has increased from 10 in 2009 to 23 in 2014 where 11 models were used.

4.3.3 Model sensitivity to soil properties

The 2009 workshop identified 23 soil properties that were considered to impact upon model sensitivity (Robinson et al., 2010b). Across all the model sub-domains, soil properties that contributed to model sensitivity include: air-dry moisture content (%), critical lower limit (CLL)/permanent wilting point (PWP) (m^3/m^3), clay content (%), bulk

density (Mg/m^3), ammonium (NH_4^+), nitrate (NO_3^-), organic carbon (%), total phosphorus (Total P) and total nitrogen (Total N).

Responses from the 2014 online survey indicate no further additions to the key soil properties identified in 2009 (Figure 4.2). Soil depth and carbon fractions achieved the highest response for contribution to model sensitivity in 2009 with 87% of survey respondents in 2014 identifying soil depth as being important. Other properties that were considered important in their contribution to model sensitivity for 2009 and 2014 include organic carbon, rooting depth, bulk density, clay content, CLL/PWP and drained upper limit (DUL)/field capacity (FC). Soil properties that recorded a decrease in contribution to model sensitivity include cation exchange capacity (CEC), Total P and carbon fractions.

Ammonium, nitrate and carbon to nitrogen ratio (C/N) did not achieve the same level of recognition in 2014, on impacts to model sensitivity, as compared with those of the 2009 workshop. No significant change was identified for soil properties considered to affect model sensitivity between 2009 and 2014 except for soil depth which was overlooked by participants in 2009 (in 2014 this achieved the highest response of any soil property). A decrease in response to model sensitivity by Total P and coarse fragments (>2 mm fraction) was recorded for the carbon and greenhouse, and forestry/biodiversity model subdomains.

4.4 Discussion

Consistent with previous research by Heemskerk and MacEwan (2007), this study confirms that a diversity of models with soil data requirements are used to understand landscape processes linked to management, production and the environment. However, there has been a reduction in the number of models used for landscape analysis in

Victoria between the two surveys of this study. This may reflect priorities of government investment for agricultural industries, a loss of personnel within modelling based programmes, and perceived issues in developing skills with models.

While a decrease in the range of models used was observed, there was a substantial increase in the number of users and applications of models. This is attributed to the increasing need for modellers to apply multiple models at various scales and coupling of models in ensembles (e.g. CAT) that link models within a landscape framework (Beverly et al., 2005).

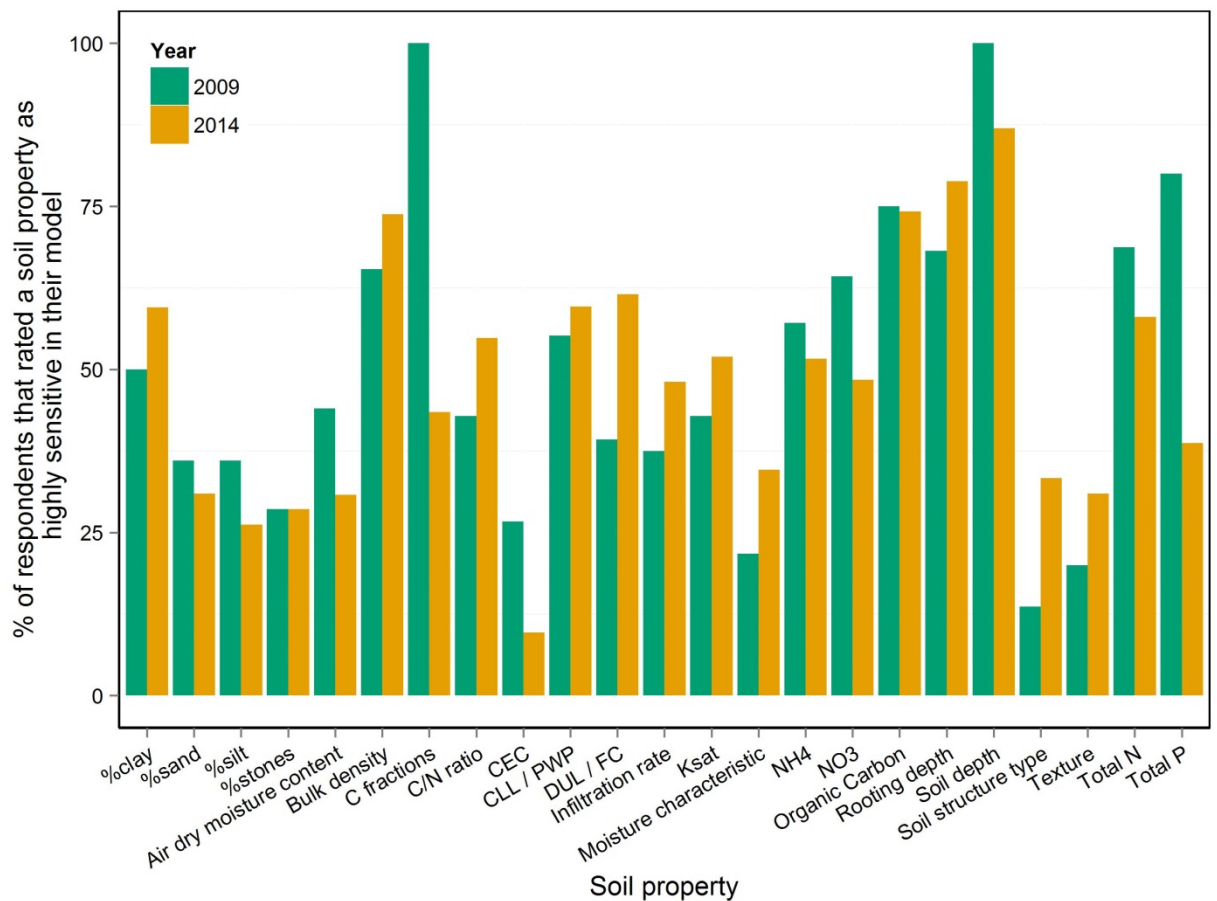


Figure 4.2. Comparison of soil properties that were considered sensitive in models for the 2009 and 2014 surveys.

The increase in model users also corresponds with an increase in the application of point-based models such as APSIM and its online commercially available interface Yield Prophet. Increased use and diversity of models to support the certain industries can be attributed to the changed government focus for research and development investment in key agricultural industries, such as grains, dairy and red meat production. Such point-based models are likely to be more focused on productivity outcomes directly linked to a site rather than catchment or regional scale model applications that are often directed towards understanding impacts of land use management and impacts on hydrology and biodiversity.

Due to growing local, national and international focus on food and fibre production (Tilman et al., 2002) and a change in agro-ecological conditions in Victoria with the end of the Millennium Drought (van Dijk et al., 2013), there appears to be less emphasis on climate change and use of carbon models and assessment (including RothC, FullCAM and CENTURY). This also corresponds with the completion of significant national and state research programs such as the Soil Carbon Research Program (SCaRP) to quantify, assess and understand composition of soil carbon stocks for Australia (Baldock et al., 2013).

A correlation between the area of agricultural industry and the frequency of model uses is evident. Industries that use the largest portion of agricultural land in Victoria such as cropping and other pastoral - beef, lamb and wool (10.3 Mha) also have the greatest model utilization. Other industries including horticulture and dairy appear to be focused largely on technology developments to increase production and therefore have a reduced emphasis on modelling.

Across agricultural industries, there are many different organisations and individuals that deliver modelling services. This study has centred on the needs of a specific group of public sector biophysical modellers. While the sample size in this study was small, the participants are specialised and representative of professionals in model development and use worldwide.

With regards to the satisfaction or 'fit for purpose' requirement for soil data, some modellers suggest that high spatial resolution soils data is not required for modelling (Heemskerk and MacEwan, 2007). A response from the 2014 survey supports this view and illustrates further issues between models and the provision of soil data: "even though we would always prefer finer resolution and more detail, it often isn't really necessary". This is consistent with responses to cost and accuracy issues on data access of the Global Soil Partnership survey (Omuto et al., 2013) where modellers prefer data accessible online and can make do with less accurate data.

These responses are symptomatic of a mismatch that currently exists between the models, agro-ecological process understandings and the current ability to deliver fine scale soil data. It is unclear if this mismatch is due to models being systematically reduced through trial and error where 'tuning' and 'matching' models to one set of conditions results in a low dependency on soil data, or a minimum soil data requirement due to an absence of suitable data (Bouma, 1989). For agro-ecosystems with water and nitrogen limits, as is often the case for many regions of Australia and Sub-Saharan Africa (Sinclair and Rufty, 2012; Zhang et al., 2016), deterministic biophysical growth models often have a large bias and uncertainties for soil characteristics (Aggarwal, 1995). There has been a tendency to refine and make plant growth models more sophisticated, although the parameters linked to soil and climate are static and therefore potentially cause the model capability and data requirement to be unbalanced (Bouma, 2001). The exclusion of soil

data may reveal if the model is unbalanced in terms of sensitivity, uncertainty and relative contribution of different biophysical factors to the design (Aggarwal, 1995; Bouma and McBratney, 2013).

4.4.1 Soil properties in models

A simple view of model sensitivity to soil data and the importance of soil attributes to the model performance were applied in this study. Results from the 2009 workshop identified 23 hydrological, physical and chemical properties that were highly sensitive in respect to one model or another. Properties relating to plant growth, including effective rooting depth and soil depth (depth to rock), were also included.

Between 2009 and 2014 there was some change in the proportion of respondents that rated a soil property as highly sensitive to their model application. For example, in the 2009 survey 13 % of respondents identified that soil depth impacted model sensitivity, and this increased 87% of respondents in the 2014 survey. This increased recognition of the model sensitivity to soil depth may be due to participants simply overlooking this as an ‘assumed’ requirement in biophysical models. Another theory is that due to further model refinement and research since 2009 confirming the importance of soil depth to understand soil water dynamics with links to catchment yield and plant production (X Cheng, Personal communication).

Respondents in the 2014 survey did not include pH and EC as sensitive properties in their models (there was also a low recognition of model sensitivity to these properties in 2009). This may be due to other functions such as ‘root growth exploration’ accounting for these properties in the model. This finding conflict with the use of these two soil properties extensively for on-ground decision making linked to land use, land classification and management, in particular understanding potential limitations to crop and pasture

production. Alkalinity and acidity affect micro and macronutrient availability and toxicity to plants, and salinity, which increases the osmotic potential around plant roots, restricts water uptake and reduces plant health (Rowell, 1994). This indicates that current process models are either incomplete in consideration of soil and landscape processes (e.g. drainage) as compared to land evaluation models, or that these properties are relatively unimportant in the tuning of biophysical models compared to other properties such as availability of water and nitrogen.

4.4.2 Changes in demand for soil property data

Little has changed since McKenzie (1991) identified that there will always be a requirement to collect new data and information to support research in agricultural production. A priority is to determine what soil properties to focus on immediately, as resource constraints often preclude collection and provision of soil data for all properties. The workshop and survey indicate six to eight key properties that should be of immediate focus to support modelling efforts using the current models. Using different thresholds to represent response rates to the 2009 and 2014 results (Figure 4.3), these properties are: critical lower limit/permanent wilting point (CLL/PWP); drained upper limit/field capacity (DUL/FC); hydraulic conductivity (Ksat); clay proportion (clay%); bulk density; organic carbon; soil depth; and effective rooting depth. Of these, organic carbon and effective rooting depth are potentially the most dynamic properties due to climate-land use-management-plant variety interactions. All properties are ranked in the top 12 property requests from the USDA-NRCS Web Soil Survey for 2011 (Thompson et al., 2012) with only hydraulic conductivity not currently sought for the GlobalSoilMap project. Physical (and hydrological) properties including bulk density are relatively sparse in soil information systems as measurements are time-consuming to acquire but may be predicted within practical limits via pedotransfer functions (Sequeira et al., 2014).

Chemical properties, however, are dynamic and require direct measurement to account for their temporal and spatial variability (Wagenet et al., 1991). These properties, including Total N, Total P, nitrate and ammonium, are essential inputs to plant growth models and are flagged as required for soil information systems such as the ASRIS.

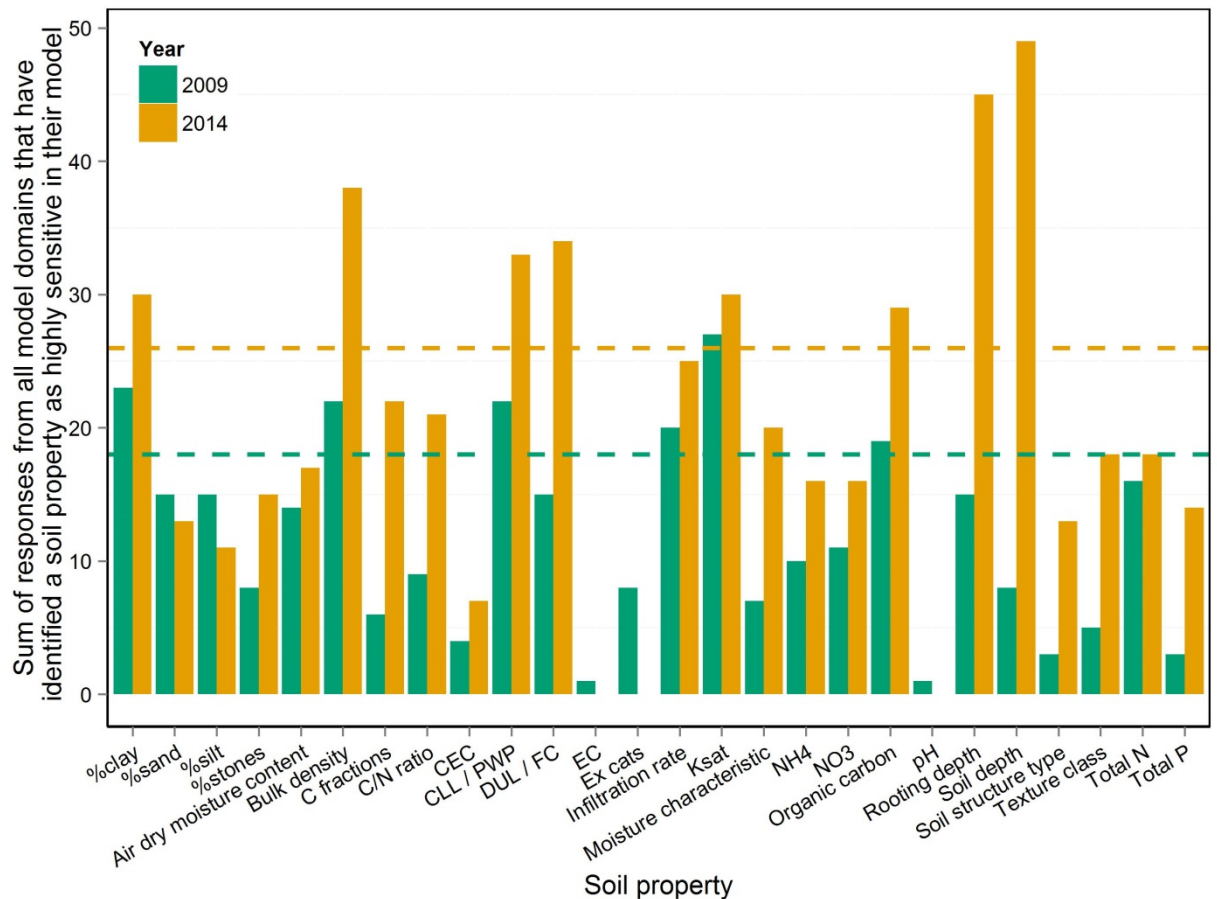


Figure 4.3. Soil attributes that models were sensitive to and thresholds to identify soil attributes to focus data provision in support of future model applications.

Future soils data needs to support models will depend on political cycles and the policy questions posed (Fisher and Crawford 2014). This includes response to environmental triggers such as drought, floods, disease outbreak and land contamination and degradation. Currently there is a trend for crop growth models to focus on plant

phenology and evaporation without the balance from soil data and research to improve models and eliminate assumptions (Bouma, 2010).

4.5 Looking forwards

Over the last century there has been considerable collection of soil data for physical, chemical and biological properties, processes and functions. Over that timeframe there have been many changes in technology and methods. There is significant growth in new data collection on plant and soil, where static and dynamic properties are inferred from latest sensing technologies (Zaks and Kucharik, 2011). Harmonisation of legacy data with that generated by contemporary methods as well as incorporation of new soil properties is therefore a critical area of work that will affect the reliability of modelled outputs.

New soil sensing techniques such as diffuse reflectance spectroscopy in the visible, near-infrared and mid-infrared spectral ranges can provide rapid and cost-effective predictions of soil properties to support data harmonisation. The integration of these sensor data with legacy data will require methods to harmonise results from different methods for soil measurements, e.g. plant available water characteristics used in crop models from geophysical sensors (see Robinson et al., 2010a). Reference soil data that is contemporary and precise for local and global calibration purposes will be critical to harmonization approaches.

Proximal and remote sensing technologies in agriculture are generating large volumes of spatially dense data (Gebbers and Adamchuk, 2010; Schimmelpfenning and Ebel, 2011). A proposed approach to deal with the plethora of high spatial and temporal resolution data is multisensor data fusion (Adamchuk and Rossel, 2010). The authors identify that complementary data from multisensor platforms will provide users with: greater

confidence in comparison to a single data source; timeliness of data acquisition for use in models and management decisions; and increased certainty.

Given the development of new sensors and vast data arrays that result in high volume and variety data being generated continuously (Kitchin, 2013), opportunities exist for the integration of ‘big data’ (gartner.com) with biophysical models where sensor data are direct inputs into these systems (Roudier et al., 2015). This will support modelling of environmental flux and space–time relationships between these biophysical systems. The linking of real-time data sources with biophysical models should enable recalibration and tuning of models to environmental conditions.

The inclusion of harmonized data into biophysical models can be supplemented with uncertainty frameworks to provide greater transparency and certainty in modelling. Such frameworks enable the identification of error sources and assumptions in models that can be refined (Robinson et al., 2015). Informing users of model assumptions and uncertainties is important to retain confidence in model outputs for decision making purposes.

Models have continued to evolve to reflect improvements based on progressive discovery and over simplification (Keating and McCown, 2001). A possible scenario as suggested by Siegel (2013) is that hundreds of different algorithms (as model ensembles) can be applied with data from various sources to determine a single, or ensemble, product that best explains the system of interest and supports new theories and understandings (Kitchin 2014).

There is a need for model developers and soil data providers to work more closely in interdisciplinary teams aimed at identifying approaches where data solutions can be implemented in order to resolve the problems identified (Bouma, 2001; Bouma and

McBratney, 2013). The Agricultural Model Intercomparison and Improvement Project (AGMIP; www.agmip.org) is attempting to improve the interoperability of models and data for large scale assessment of climate change and impacts to agriculture in regions including Sub-Saharan Africa (AGMIP, 2014). Global challenges including the availability of skilled and trained personnel and appropriate operating environments to support these new advances for less economically favoured nations remain.

4.6 Conclusion

Biophysical models use soil data at various spatial modelling scales for agricultural production purposes. This study has established that the soil data requirements of public-sector modellers operating in agricultural and environmental sciences have remained similar over the last 5-years. A decrease in the diversity of biophysical models used reflects a changing focus of government research and development towards agricultural industries using more established and recognised models. New data are also currently being collected and stored. In the last 5 years there has been increased application of models at point/site scale for the grains, dairy and livestock production industries in Victoria.

Six to eight soil properties are priority items due to their contribution to model sensitivity and uncertainty. These are CLL/PWP, DUL/FC, Ksat, clay%, bulk density, organic carbon, soil depth and effective rooting depth. Fine scale Digital Soil Mapping and validation, supported by contemporary soil data and uncertainty frameworks, provides opportunities to evaluate the relative importance of these properties. This may involve new techniques such as combination of community sourced data and accessing tacit

knowledge; to precision agriculture systems and physical collection of new soil samples to calibrate the products (Rossiter et al., 2015).

With the significant growth anticipated in new data on soil and plant dynamics, there will be a need for biophysical models to accommodate these new sources of data in their architecture. The influx of new data provides opportunities for improved scenario modelling for better understandings of interactions in the environment and how these can be best managed to optimise delivery of ecosystem services. Data harmonisation and uncertainty assessments will be important to ensure future relevance and accuracy of models. Close interactions between model developers, soil scientists and end-users is needed to direct the refinement of agricultural production and landscape process models to support global issues such as food, water and energy security.

References

- Aggerwal, P.K., 1995. Uncertainties in plant, soil and weather inputs used in crop growth models: implications for simulation outputs and their applications. *Agricultural Systems*, 48, 361-384.
- Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56. FAO - Food and Agriculture Organization of the United Nations, Rome.
- Amundson, R., Berhe, A.A., Hopmans, J.W., Olson, C., Sztein, A.E., Sparks, D.L., 2015. Soil and human security in the 21st century. *Science*, 348, 1261071-1-6.
- Argent, R.M., Brown, A., Cetin, L.T., Davis, G., Farthing, B., Fowler, K., et al., 2008. *WaterCAST User Guide*. eWater CRC, Canberra.
- Asseng S., Ewert, F., Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., 2014. Uncertainty in simulation wheat yields under climate change. *Nature Climate Change*, 3, 827-832.

- Baker, J.M., 1996. Use and abuse of crop simulation models. *Agronomy Journal*, 88, 689.
- Baldock, J., MacDonald, L., Sanderman, J., 2013. Foreward to 'Soil Carbon in Australia's Agricultural Lands'. *Soil Research*, 51, 1-2.
- Bell, G., Hey, A., Szalay, A. 2009. Beyond the data deluge. *Science*, 323, 1297-1298.
- Bergez, J.E., Colbach, N., Crespo, O., Garcia, F., Jeuffroy, M.H., Justes, E et al., 2010. Designing crop management systems by simulation. *European Journal of Agronomy*, 32, 3-9.
- Beverly, C., 2007. Technical manual models of the catchment analysis tool (CAT1D Version 22). Department of Sustainability and Environment, Victoria.
- Blöschl, G., Sivapalan, M., 1995. Sale issues in hydrological modelling: a review. *Hydrological Processes* 9, 251-290.
- Boote, K.J., Jones, J.W., Hoogenboom, G., White, J.W., 2010. The Role of Crop Systems Simulation in Agriculture and Environment. *International Journal of Agricultural and Environmental Information Systems*, 1, 41-54.
- Bouma, J., 1989. Using soil survey data for quantitative land evaluation. In: *Advances in Soil Science* (ed. BA Stewart), pp. 177-213. Springer-Verlag, New York.
- Bouma, J., 2001. The role of soil science in the land use negotiation process. *Soil Use and Management*, 17, 1-6.
- Bouma, J., 2010. Implications of the knowledge paradox for soil science. *Advances in Agronomy*, 106, 143-171.
- Bouma, J., Lanen, H.A.J., Breeuwsma, A., Wösten, H.J.M., Kooistra, M.J., 1986. Soil survey data needs when studying modern land use problems. *Soil Use and Management*, 2, 125-130.
- Bouma, J., McBratney, A.B., 2013. Framing soils as an actor when dealing with wicked environmental problems. *Geoderma*, 200-201, 130-139.
- Boyd, D., Crawford, K., 2012. Critical questions for big data. *Information, Communication & Society*, 15, 662-679.

- Cook, S.E., Jarvis, A., Gonzalez, J.P., 2008. A new global demand for digital soil information. In: *Digital Soil Mapping with Limited Data* (eds A.E. Hartemink, A. McBratney, M.L. Mendonça Santos), 31-41. Springer.com.
- Craemer, R., Barber, M., 2007. Building a Business Case for Investment in Soil Information. National Land & Water Resources Audit, ACIL Tasman Pty Ltd, Canberra, Australia.
- Crick, M.J., Hill, M.D., 1987. The role of sensitivity analysis in assessing uncertainty. In: *NEA workshop on uncertainty analysis for performance assessments of radioactive waste disposal systems*; Seattle, USA; 24-26 Feb 1987. Nuclear Energy Agency of OECD (NEA), pp. 116-129.
- Dooge, J.C.I., 1982. Parameterization of hydrologic processes. In: *Land Surface Processes in Atmospheric General Circulation Models* (ed. P.S. Eagleson), pp. 243-288. Cambridge University Press, London.
- Dooge, J.C.I., 1986. Looking for hydrologic laws. *Water Resources Research*, 22, 46S-58S.
- Fienen, M.N., Lowry, C.S. 2012. Social.Water - A crowdsourcing tool for environmental data acquisition. *Computers & Geosciences*, 49, 164-169.
- Fisher, J., Crawford, M., 2014. Policy and science of soil change – a Victorian perspective. *IOP Conference Series: Earth and Environmental Science*, 25, Soil Change Matters 2014, Bendigo, Victoria, Australia. Doi:10.1088/1755-1315/25/1/012002.
- Gebbers, R., Adamchuk, V.I., 2010. Precision Agriculture and Food Security. *Science*, 327, 828-831.
- Heemskerk, G., MacEwan, R.J., 2007. An evaluation of DPI science capability for landscape analysis. Department of Primary Industries, Primary Industries Research Victoria, Bendigo, Australia.
- Johnson, I.R. 2013., DairyMod and the SGS Pasture Model: a mathematical description of the biophysical model structure. IMJ Consultants, Dorrigo, NSW, Australia.

- Keating, B.A., Carberry, P.S., Hammer, G.L., Probert, M.E., Robertson, M.J., Holzworth, D. et al., 2003. An overview of APSIM, a model designed for farming systems simulation. *European Journal of Agronomy*, 18, 267-288.
- Keating, B., Grace, P., 1999. Using Biophysical Simulation Models to Assess the Performance of Agricultural Systems in Risky Environments. In: *Risk Management for Maize Farmers in Drought-Prone Areas of Southern Africa; Proceedings of a Worskhop held at Kadoma Ranch, Zimbabwe, 1-3 October 1997*. Page 67-82. The International Maize and Wheat Improvement Center (CIMMYT), Mexico.
- Keating, B. McCown, R.L., 2001. Advances in farming systems analysis and intervention. *Agricultural Systems*, 70, 555-579.
- Kitchin, R., 2013. Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography*, 3, 262-267.
- Kitchin, R., 2014. Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, April-June 2014, 1-12.
- Landsberg, J.J., Waring, R.H., 1997. A generalised model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance and partitioning. *Forest Ecology and Management*, 95, 209-228.
- Littleboy, M., Silburn, D.M., Freebairn, D.M., Woodruff, D.R., Hammer, G.L., Leslie, J.K., 1992. Impact of soil erosion on production in cropping systems. I. Development and validation of a simulation model. *Australian Journal of Soil Research*, 30, 757-74.
- Lynch, C., 2008. Big data: How do your data grow? *Nature*, 455, 28-29.
- MacEwan, R.J., Robinson, N.J., Imhof, M.P., Rees, D.B., Hopley, J., Crawford, D.M., 2014. Reflections on GlobalSoilMap.net, DSM and the legacy of soil survey information in Victoria, Australia – the pedology matters. In: *GlobalSoilMap, Basis of the global spatial soil information system* (eds D. Arrouays, N. McKenzie, J. Hempel, A.C. Richer de Forges & A. McBratney). CRC Press/Balkema.

- Mayer-Schonberger, V., Cukier, K., 2013. *Big Data: A revolution that will transform how we live, work and think*. Houghton Mifflin Harcourt publishers, New York.
- McClymont, D., Freebairn, D.M., Rattray, D.J., Robinson, J.B., White, S. 2011., *Howleaky V 5.40.10 : Exploring water balance and water quality implication of different land uses*. (At: <http://www.howleaky.net/>. Accessed: 18/06/2015).
- McKenzie, N.J., 1991. *A strategy for coordinating soil survey and land evaluation in Australia*. Divisional report No. 114, CSIRO Division of Soils, Canberra.
- McKenzie, N.J., Grundy, M.J., Webster, R., Ringrose-Voase, A.J. 2008. *Guidelines for surveying soil and land resources*. CSIRO Publishing, Collingwood, Victoria, Australia.
- Meadows, D.H., Meadows, G., Randers, J., Behrens III, W.W. 1972. *The Limits to Growth*. Universe Books, New York.
- Melland, A.R., Gourley, C.J.P., Smith, A.P., Tarbottom, I.S., Peverill, K.I. 2004. *Developing a farm nutrient loss index for grazed pastures in Australia*. SuperSoil 2004: 3rd Australian New Zealand Soils Conference, 5 – 9 December 2004, University of Sydney, Australia.
- Metherell, A.K., Harding, L.A., Vernon Cole, C., Parton, W.J. 1993., *CENTURY - Soil Organic Matter Model, Agroecosystem Version 4.0*. GPSR Technical Report No. 4, United States Department of Agriculture, Agricultural Research Service, Great Plains Systems Research Unit.
- Morse-McNabb, E., Sheffield, K., Clark, R., Lewis, H., Robson, S., Cherry, D., Williams, S. 2015. *VLUIS, a land use data product for Victoria, Australia, covering 2006 to 2013*. *Scientific Data* 2, 150070.
- Morris, J.D., Baker, T.G., 2002. *Using a process-based forest model to estimate the potential productivity of Eucalyptus plantations in southern China*. pp. 325-337. In R. Wei and D. Xu (eds) *Proceedings of the International Symposium Eucalyptus Plantations: Research, Management and Development*. World Scientific, Singapore.

- Neitsch, S.L., Arnold, J.G., Kiniry, J.R., Williams, J.R., 2001. Soil and Water Assessment Tool User's Manual, Version 2000. U.S. Department of Agriculture, Agricultural Research Service, Temple, TX.
- Nelson, M.A., Bishop, T.F.A., Odeh, I.O.A., Triantifilis, J., 2011. An error budget for different sources of error in digital soil mapping. *European Journal of Soil Science*, 62, 417-430.
- Newham, L.T.H., Letcher, R.A., Jakeman, A.J., Kobayashi, T., 2004. A framework for integrated hydrologic, sediment and nutrient export modelling for catchment-scale management. *Environmental Modelling & Software*, 19, 1029–1038.
- Nichol, G., (ed) 2006. Landscape analysis models and frameworks – a review. Victorian Department of Primary Industries, Epsom, Victoria.
- Oei, T.I., Zwart, F.M., 1986. The assessment of life events: self-administered questionnaire versus interview. *Journal of Affective Disorders*, 10, 185-190.
- Omuto, C., Nachtergaele, F., Vargas Rojas, R., 2013. State of the art report on global and regional soil information: Where are we? Where to go? Global Soil Partnership Technical Report. Food and Agriculture Organization of the United Nations, Rome, Italy.
- Porter, J.H., Hanson, P.C., Lin, C.C., 2012. Staying afloat in the sensor deluge. *Trends in Ecology & Evolution* 27, 121-129.
- Robinson, N.J., MacEwan, R.J., Crawford, D., Rees, D., Rab, A., Imhof, M., 2010. Capturing soil landscape parameters to underpin regional biophysical modelling. Victorian Department of Primary Industries, Bendigo.
- Robinson, N.J., Benke, K.K., Norng, S., 2015. Identification and interpretation of sources of uncertainty in soils change in a global systems-based modelling process. *Soil Research*, 53(6), 592-604. Doi: 10.1071/SR14329.
- Rossiter, D.G., Liu J., Carlisle, S., Zhi, A-X., 2015. Can citizen science assist digital soil mapping? *Geoderma*, 259-260, 71-80.

- Roudier, P., Ritchie, A., Hedley, C., Medyckyj-Scott, D., 2015. The rise of information science: a changing landscape for soil science. IOP Conference Series: Earth and Environmental Science 25: 1-10. doi:10.1088/1755-1315/25/1/012023
- Rowell, D.L., 1994. Soil Science, Methods & Applications. Addison Wesley Longman Limited, Harlow, England.
- Schimmelpfennig, D., Ebel, R., 2011. On the Doorstep of the Information Age: Recent Adoption of Precision Agriculture. EIB-80, U.S. Dept. of Agriculture, Economic Research Service.
- Schmidt, E., Cohen, J., 2013. The New Digital Age: Reshaping the future of people, nations and business. John Murray (Publishers), London.
- Sequeira, C.H, Wills, S.A., Seybold, C.A., West, L.T., 2014. Predicting soil bulk density for incomplete databases. Geoderma, 213, 64-73.
- Siegel, E., 2013. Predictive analytics: The Power to Predict Who Will Click, Buy, Lie or Die. Wiley, Hoboken, New Jersey.
- Silvertown, J., 2009. A new dawn for citizen science. Trends in Ecology and Evolution, 24, 467-471.
- Simpson, R.J., Salmon, L., Moore, A.D., Graham, P., 2002. GrassGro - supporting management of grazing enterprises for profit and sustainability. Proceedings of the Australian Society of Animal Production, 24,451-455.
- Šimůnek, J., van Genuchten, M.Th., Šejna, M., 2011. The HYDRUS Software Package for Simulating Two- and Three Dimensional Movement of Water, Heat, and Multiple Solutes in Variably Saturated Porous Media, Technical Manual, Version 2.0. PC Progress, Prague, Czech Republic.
- Sinclair, T.R., Seligman, N.G., 1996. Crop Modeling: From Infancy to Maturity. Agronomy Journal, 88, 698-704.
- Skaggs, R.W., 1980. DRAINMOD reference report. USDA-SCS South National Technical Center.

- Stockle, C.O., Nelson, R.L., 1996. Cropsyst User's manual (Version 2.0). Biological Systems Engineering Department, Washington State University, Pullman, WA, USA.
- Sui, D., Goodchild, M., Elwood, S., 2013. Volunteered Geographic Information, the Exaflood, and the Growing Digital Divide. In: Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice (eds D. Sui, M. Goodchild & S. Elwood). Springer, Dordrecht.
- Tilman, D., Cassman, K.G., Matson, P.A., Naylor, R., Polasky, S., 2002. Agricultural sustainability and intensive production practices. *Nature*, 408, 671-677.
- Thompson, J.A., Roecker, S., Grunwald, S., Owens, P.R. 2012., Digital Soil Mapping: Interactions with and Applications for Hydropedology. In: Hydropedology (ed. H. Lin). Elsevier B V.
- van Genuchten, M.T., 1980. A Closed-form Equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal*, 44, 892-898.
- Voss, C.I., 2011. Editor's message: Groundwater modelling fantasies – part 1, adrift in details. *Hydrogeology Journal*, 19, 1281-1284.
- Wagenet, R.J., Bouma, J., Grossman, R.B., 1991. Minimum data sets for use of soil survey information in soil interpretative models. In: Spatial variabilities of soils and landforms. SSSA Special Publication, 28, 161-182.
- Werts, J., Mikhailova, E., Post, C., Sharp, J., 2012. An Integrated WebGIS Framework for Volunteered Geographic Information and Social Media in Soil and Water Conservation. *Environmental Management*, 49, 816-832.
- Wood, B., Auricht, C., 2011. ASRIS / ACLEP user needs analysis. Australian Collaborative Land Evaluation Program.
- van Dijk, A.I.J.M., Beck, H.E., Crosbie, R.S., de Jeu, R.A.M., Liu, Y.Y., Podger, G.M. et al., 2013. The Millennium Drought in southeast Australia (2001–2009): Natural and human causes and implications for water resources, ecosystems, economy, and society. *Water Resources Research*, 49, 1040-1057. Doi: 10.1002/wrcr.20123.

Zaks, D.P.M., Kucharik, C.J., 2011. Data and monitoring needs for a more ecological agriculture. *Environmental Research Letters* 6, 1-10.

Zuiderwijk, A., Janssen, M. 2014. Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly*, 31, 17-29.

Chapter 5 Identification and interpretation of sources of uncertainty in soils change in a global systems-based modelling process

In soil mapping and modelling there are many sources of potential uncertainty that can affect the precision and accuracy of the delivered information. It is recognised that uncertainties should be quantified and communicated to enable appropriate use of the delivered map or model information (Heuvelink, 2014). Concepts of uncertainty are largely defined as either stochastic or epistemic in nature and are identified in the literature review (Chapter 3) and are discussed in the previous chapter (Chapter 4). A large emphasis has been placed on the stochastic aspects of uncertainty, yet epistemic uncertainty due to imperfect knowledge can be considerable and detrimental to the delivery of useful information. The aim of the following research is to develop a holistic approach to accommodate and illustrate to users of spatial soil information, the various error sources in modelling and mapping. The primary objectives here were to:

Summarise, characterise and interpret sources of uncertainty in the assessment of soil change;

Describe applications of uncertainty analysis in soil-change research;

Represent uncertainty analysis in soil change in the form of a systems-based process model that captures the major sources of uncertainty in the modelling process.

In the past, uncertainty analysis in soil research was often reduced to consideration of statistical variation in numerical data relating to model parameters, model inputs or field measurements. The simplified conceptual approach used by modellers in calibration studies can be misleading, because it relates mainly to error minimisation in regression

analysis and is reductionist in nature. In this study, a large number of added uncertainties are identified in a more comprehensive attention to the problem. Uncertainties in soil analysis include errors in geometry, position and polygon attributes. The impacts of multiple error sources are described, including covariate error, model error and laboratory analytical error. In particular, the distinction is made between statistical variability (aleatory uncertainty) and lack of information (epistemic uncertainty). Examples of experimental uncertainty analysis are provided and discussed, including reference to error disaggregation and geostatistics, and a systems-based analytic framework is proposed. These concepts are applied in implementations of uncertainty analysis and Digital Soil Mapping in the following chapters (6 and 7). It is concluded that a more comprehensive and global approach to uncertainty analysis is needed, especially in the context of developing a future soils modelling process for incorporation of all known sources of uncertainty.

5.1 Introduction

Changes in soil with respect to behaviour, function and condition occur in response to land use and farming systems. These changes can seriously compromise future capacity for primary production and the provision of ecosystem services, such as vegetation and water supply. Farming systems have contributed to soil degradation from chemical and physical processes, including acidification, erosion, salinization, structure decline, carbon decline and loss of fertility to varying degrees. Understanding the nature and rate of such changes is critical to designing appropriate farming systems that maintain, regulate and enhance the services delivered by soil.

Soil changes are driven by complex interactions involving physics, chemistry and biology that may be natural or anthropogenic in origin and occur on overlapping temporal scales, varying between millennia and contemporary human influence (Young and Crawford, 2004; Richer and Yaalon, 2011). Often change is not readily predictable, or measureable, as the change can occur over a significant timeframe, or strategic questions are not adequately considered and addressed in establishment of long-term soil experiments (Richter et al., 2007). To address this, the development of frameworks that support theoretical development on soil change processes and functions is critical (MacEwan 1997).

Such frameworks may include *functions* (soil formation and genesis) as embodied in *The Factors of Soil Formation* by Jenny (1941), and the *processes* (additions, removals, transfers and transformation) and their interactions that leave an ‘imprint on soil character’ (Simonon, 1959). These interactions between functions and processes are summarised in the expression of the ‘pedon’ with attributes (properties) and qualities that

can be defined (MacEwan, 1997). These properties are also interrelated in this dynamic system and should one property change, another may also change.

The present interest in soil change studies relates to the quantification of dynamic soil properties through measurement and observation of soil attributes in response to human induced impacts (Tugel et al., 2005). To quantify and understand changes in soil requires various techniques, including repeated soil surveys, long-term soil experiments and space-for-time substitutions, which are necessary to reduce error and uncertainty in scientific conclusions.

Monitoring of change in dynamic soil properties needs to incorporate uncertainty in estimates due to the large number of potential error sources (Saby et al., 2008). Sources of uncertainty in the analysis of soil change include the following:

- context (environmental constraints)
- measurement error (see Box 1)
- error propagation in models (see Box 2)
- expert opinion (judgements, estimation, interpretation)
- decision making under uncertainty (see Box 3)
- framing (problem boundaries)
- implementation error (numerical approximations)
- model inputs (soil, hydrology, climate)
- model structure
- model parameter uncertainty
- resolution (spatial and temporal)
- software problems (verification, validation, bugs)

Further limited discussion of some of these categories is provided in the literature associated with uncertainty in pedology and hydrology (e.g. Refsgaard et al., 2007; Benke

et al., 2007; Hopley et al., 2014; Robinson et al., 2014). In the context of modelling and simulation, there is a two-step process of model selection uncertainty (epistemic uncertainty) followed by statistical variability in prediction, which is addressed by estimates of error propagation and other numerical approaches (aleatory uncertainty).

Box 1: Measurement Error

For a specific measurement of a soil property, y_i , there are many extraneous factors $x_1 \dots x_n$ adding uncertainty. Compensation for these factors will reduce epistemic uncertainty, although there is ultimately, still a residual random error ε_i that is irreducible.

$$y_i = x_1 + x_2 + \dots + x_n + \varepsilon_i$$

This type of error is often regarded as epistemic uncertainty and reducible in the literature because the factors that are the major sources of uncertainty are reducible (e.g. instrumentation technology, scale, precision, human error).

Box 2: Error Propagation

Errors can propagate from inputs or parameters through the model $y=f(x)$ and into the outputs (predictions). Two common methods of modelling error (e.g. variance, σ^2) and its propagation through the model are the Taylor method (differential calculus), if the model is smooth, continuous and differentiable,

$$\sigma^2(y) = \sum_{i=1}^n \left[\frac{\partial f(\mathbf{x}_0)}{\partial x_i} \right]^2 \sigma^2(x_i)$$

Or, alternatively, using Monte Carlo simulation, where the output is a probability distribution, from which variance is computed.

Box 3: Decision-making under uncertainty

The problem of decision-making under uncertainty is analogous to comparing estimates of the mean (or expectation values) between two data sets, **such height difference between males and females**, where uncertainty is the variance, σ^2 , with degrees of freedom, n ,

$$z = \frac{\mu_1 - \mu_2}{\sqrt{\sigma^2_1 / n_1 + \sigma^2_2 / n_2}}$$

This model assumes a null hypothesis of no significant difference and normal distributions for each data set. When the difference, z , is large enough, subject to a test criterion, z_{crit} , it is statistically significant.

Uncertainty analysis is traditionally undertaken towards the end of the analytical process, with the focus on experimental replications, or model calibration. Ideally, uncertainty should be considered in entirety through-out the analytical process from problem definition and assumptions, to prediction accuracy and error. This is the traditional view of uncertainty, however, decision-maker views of uncertainty include balancing outcomes with objectives and priorities in the context of policy response to soil change analysis (Walker et al., 2003).

The objectives of this paper are (a) to summarise, characterise and interpret sources of uncertainty in the assessment of soil change, (b) describe applications of uncertainty analysis in soil change research, (c) represent uncertainty analysis in soil change in the form of a systems-based process model that captures the major sources of uncertainty in the modelling process, and (d) provide illustrative examples in the context of error disaggregation and geostatistics.

5.2 Types of Uncertainty

Uncertainty in the past was sometimes regarded in a negative sense and associated with a lack of assurance or conviction in an observation or outcome. The analysis of uncertainty, however, is now deemed very important because a purely deterministic approach provides a prediction without indication of error or uncertainty, i.e. there is no indication of confidence in the answer provided. In fact, Hastings and McManus (2004) highlight the fact that uncertainty ‘is not always a negative to be mitigated’ and that robust and flexible systems can be derived to mitigate these uncertainties while providing additional value to users. The reality is that all experimental science and modelling processes are associated with errors due to range of uncertainties existing in the real-world (Refsgaard et al., 2007).

In soil science, there is a widespread culture of thinking about uncertainty as merely being statistical variability, relating to model parameters and inputs, or the scatter plot associated with a variogram. There are other types of uncertainty, many of which are identified and discussed in the following sections. In different contexts, many of these other uncertainties may be more important in impact than statistical variability in model calibration or experimental replications.

Three dimensions to uncertainty have been defined (see Walker et al., 2003). These dimensions are the location of the uncertainty (where the uncertainty occurs in a model), the level of uncertainty (how the magnitude of the uncertainty contributes to the overall) and the nature (taxonomy) of uncertainty (if the uncertainty is due to incomplete knowledge, variability or ambiguity). Isolating where and the level of the uncertainty in an analytical model process are considered in the various sources of uncertainty discussed later.

5.2.1 Taxonomy of uncertainty

The primary dichotomy that characterises the taxonomy of uncertainty is readily apparent from published studies (e.g. Walker et al., 2003; Wagner and Gupta, 2005; Benke et al., 2007):

(i) Aleatory Uncertainty (statistical variability)

(ii) Epistemic Uncertainty (lack of knowledge)

The process of decision-making under uncertainty, including human inability to understand decision objectives, and observer vagueness and linguistic ambiguity, have all been regarded as sources of epistemic uncertainty (Baecher and Christian, 2000; Refsgaard et al., 2007).

Decision-making under uncertainty may also be viewed in the context of comparison of two soil property measurements (see Box 3). The original source of uncertainty is particularly relevant to soil change as the decision-making process is critical in interpretation and implementation of a management response. The decision-making framework devised by Steinitz (1990, 2012) is one scheme that could provide a systematic process for analysis of decision-model uncertainty.

Aleatory Uncertainty (statistical variability)

Aleatory uncertainty is not reducible, as it relates to innate or natural variability in environmental modelling (and is sometimes referred to as stochastic uncertainty). It can be characterised by probability distributions (e.g. the normal distribution) and can be quantified in Monte Carlo simulation. Increasing sample size will not decrease the standard deviation of the variable, but will decrease the standard error relating to the sampling distribution of the means. Note that, in contrast, measurement error includes additional contributions from other sources of error that are reducible, see Box 1 (Iman

and Helton, 1988; McBratney 1992; Refsgaard et al., 2007; Benke et al., 2008; Robinson et al., 2014). The computation of error propagation from inputs and parameters in a model reveals whether the input error distribution is affected by the structure of the model, and can also quantify the uncertainty in the output distribution, which is the information delivered to users.

Epistemic Uncertainty (lack of information)

This category of uncertainty is due to imperfect knowledge that is generally reducible through the collection of more data and additional studies (but not always). This relates to lack of information (ignorance, or incomplete knowledge of systems and processes) as opposed to statistical variability, which is the dominant approach in soil science. An example of epistemic uncertainty is linguistic (or semantic) uncertainty that can be reduced by resolving ambiguities.

Epistemic Uncertainty - Type 1

This category of uncertainty refers to *known unknowns* and includes linguistic ambiguity, data transcription errors and software bugs. Also included are context, framing and expert opinion, which involves judgement and also sensory performance confined by environmental factors (e.g. colour matching with pH soil test kits in the field). An example of Type 1 uncertainty is the so-called millennium bug in legacy software for computer calendars limited the date format to two digits only, e.g. '99'. When the clock ticked over for the new millennium in Year '2000', time was initialised to zero for accounting software. In accounting spreadsheets, time may be used as a variable, resulting in error propagation and uncertainty in results.

Epistemic Uncertainty - Type 2

This category of uncertainty refers to *unknown unknowns* and is the most serious type of uncertainty. Examples include black swan events – which are unforeseen, very rare, and

very disruptive. For example, a sudden and extreme flood event in the local landscape, such as the soil salinity impacts from the 2010 Victorian floods (http://en.wikipedia.org/wiki/2010_Victorian_floods). The main issue is that Type 2 uncertainty is unexpected and tends to be very rare with potentially major impacts.

Decision-making under uncertainty

Understanding soil change and the different types and sources of uncertainty provides the basis for better communication and final decision-making by land managers, industry and government. It has been noted in the past that uncertainty is pervasive and a fact of life and quantification would lead to greater acceptance in results provided by scientists in decision support (Walker et al., 2003). The decision-making process, when automated in software, ideally should include uncertainty in prediction, as well as prediction accuracy, to improve confidence in *expert systems* for decision support, especially in spatial analysis (Sposito et al., 2010).

Risk is defined as the probability of an adverse event and its consequence, and represents one aspect only of the uncertainty framework specific to the question posed (Hastings and McManus, 2004). Risk is a point estimate, as distinct from uncertainty which is an interval estimate, as described by the confidence interval (Pelizaro et al., 2011). The link between risk, the quantitative model, and uncertainty, and how these interact with the decision-making process is rarely, if ever, considered in modelling (Walker et al., 2003; Wagener and Gupta, 2005).

To support a decision-making process, a framework provides a useful context to enable problem definition, testing of hypotheses and accommodation of uncertainty. MacEwan (2014) identified the Steinitz framework (Steinitz, 1990, 2012) as a valuable tool to

iteratively pose questions that support decision-making (see Figure 5.1). The six primary iterative questions in a soil change context are:

- What data and information do we have to support a representation of the soil?
- What processes are linked to soil attributes in question?
- Do we have the necessary data to answer the questions posed?
- What are the potential drivers of change, e.g. land use change?
- What are the likely impacts of change?
- What needs to be done to achieve the outcomes sought?

Against these questions there are associated uncertainties that may be considered in decision-making. Sources of uncertainty can be aligned to the Steinitz Framework to illustrate the direct link between the decision-making process and uncertainty (Table 5.1).

5.3 Sources of uncertainty

Sources of error and uncertainty can be found in all steps in a modelling process from input through to the final output. The sources of uncertainty in environmental modelling proposed by Walker et al. (2003), and extended by others (e.g. Gupta and Wagener, 2005; Refsgaard et al., 2007; Benke et al., 2007) include model inputs, model structure, model implementation error, parameter uncertainty, measurement error, context and framing. In this paper we have expanded the discussion of error sources to include many other aspects, such as expert opinion, legacy data issues and software operations.

Many sources of uncertainty may be expressed in a process modelling representation. Figure 5.2 shows a framework for incorporation of aleatory and epistemic uncertainties, the so-called Global Representation of Uncertainty in the Modelling Process (GRUMP). Wherever possible, epistemic uncertainties are enumerated through psychophysical experiments and codified categorical classifications. This allows both aleatory and

epistemic uncertainties to be considered in error propagation through the model. Note also that in Figure 5.2, some uncertainties are difficult to classify and have elements of both aleatory uncertainty and epistemic uncertainty, e.g. legacy data and some categories of expert opinion relating to linguistic ambiguity and cognitive performance.

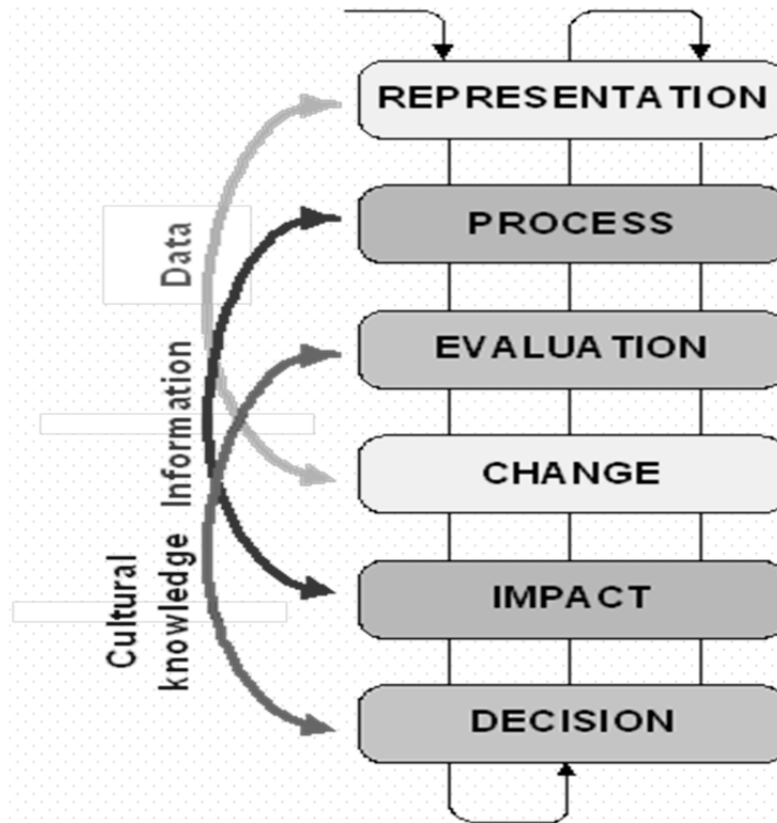


Figure 5.1. Hierarchy of the Steinitz Framework identifies the six primary iterative questions posed in the modelling process (Steinitz, 1990).

Attribute uncertainty

Examples of attribute uncertainty relate to soil types and chemical concentrations, such as soil organic carbon (SOC) content. Attribute uncertainty may also be subject to a description by a probability density function (PDF) and covers (a) the nature of the measurement scale used, and (b) time-space variation. Heuvelink et al., (2007) suggested four classes for the measurement scale, i.e.

- continuous numerical scale (e.g. chemical concentration in soil)
- discrete numerical scale (number of plant species)
- categorical scale (e.g. soil type)
- descriptive text (e.g. history of soil type)

In a similar manner, space-time variability is divided into four classes, with attributes that are:

- constant in space and time (e.g. universal gas constant)
- constant in space but vary in time (e.g. national interest rate)
- constant in time but vary in space (e.g. some geographic/geological features)
- vary in both time and space (e.g. temperature)

Table 5.1. Links between decision-making and uncertainty in the Steinitz Framework.

Question	Example uncertainty source
Representation	Lack of knowledge, e.g. high error in existing spatial prediction of soil pH
Process	Measurement error, e.g. SOC bias and precision in laboratory analysis
Evaluation	Parameter error, e.g. bulk density pedotransfer function parameters
Change	Model structure adequacy, e.g. Linear Mixed Model space-time prediction of SOC stock change
Impact	Scenario analysis and expert opinion, e.g. future land use change propositions
Decision	Inability to understand decision objectives that may be due to data and information deficiencies derived from previous questions

Spatial uncertainty can be quantified by PDFs – including soil type boundaries and polygon representations. Uncertainty in maps of regions and polygons can be modelled in a probabilistic framework using Monte Carlo simulation. For example, multiple enumerations of each object or polygon (by simulation), can be overlaid to produce fuzzy edges or boundaries that visually reflect the degree of uncertainty in these boundaries and can be linked to an uncertainty metric, such as standard deviation (Heuvelink et al., 2007,

Benke et al., 2010). Uncertainty across a boundary of an area or polygon with a specific soil attribute, such as pH, can be represented visually by the cumulative distribution function (CDF) across a fuzzy edge as depicted in Figure 5.3(a). The standard deviation may be used as a metric of uncertainty in properties, including mean value, position and rotation, as shown in Figure 5.3(b).

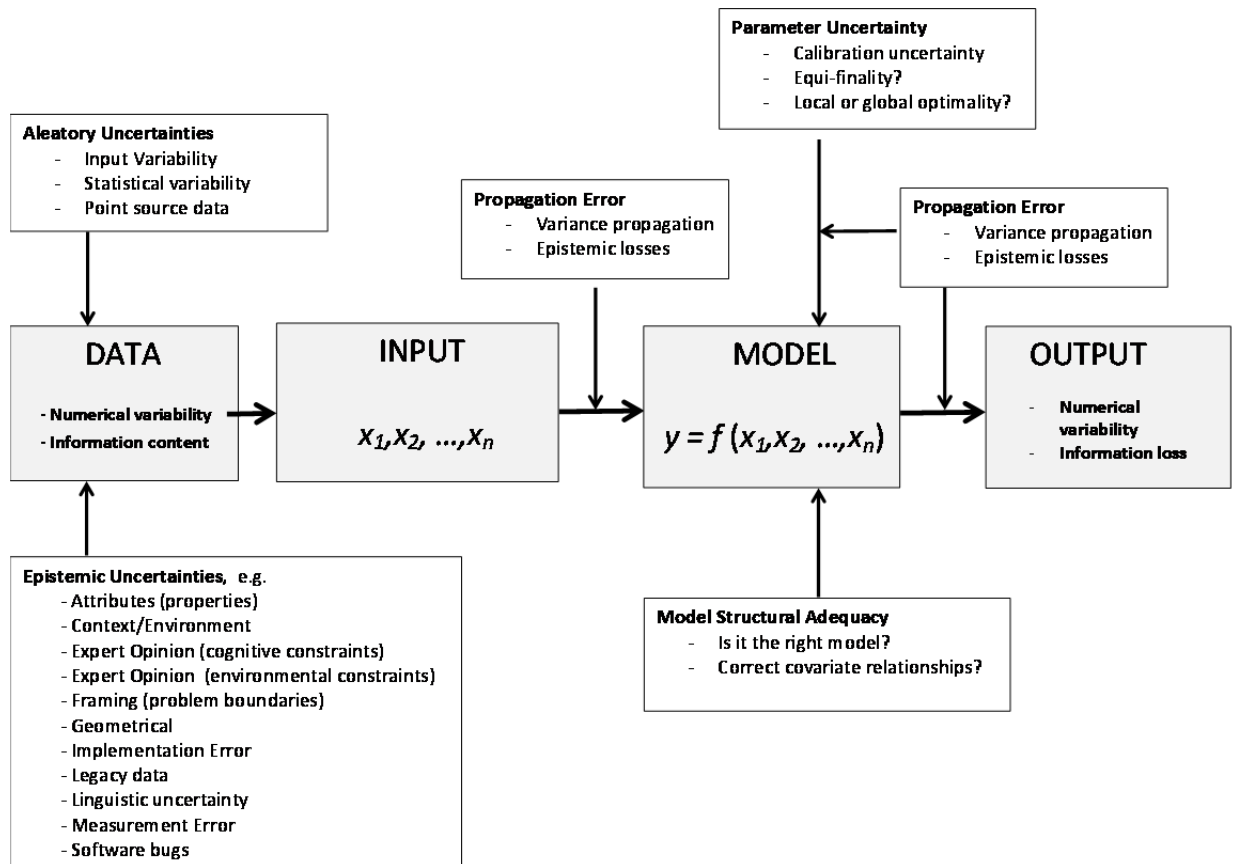


Figure 5.2. Sources of uncertainty expressed in a conceptual process modelling representation. The framework incorporates aleatory and epistemic uncertainties in the so-called Global Representation of Uncertainty in the Modelling Process (GRUMP). Epistemic uncertainties may be enumerated through psychophysical experiments and codified categorical representations. This allows both aleatory and epistemic uncertainties to be considered in error propagation through the model. In the special case of Monte Carlo simulation, the output probability distribution represents uncertainty and its median represents the prediction.

If there is statistical independence in space and time, the joint PDF from the model output is the product of the marginal PDFs -- and can be produced by estimating the separate marginal PDFs (Aerts et al., 2003; Heuvelink et al., 2007). If dependencies exist between variables, these must be determined together with the marginal PDFs. Heuvelink et al. (2007) noted that if dependency exists, the joint PDF is often assumed to be the multivariate normal distribution where the covariance matrix is used for correlated variables. For positional and attribute uncertainty, under some conditions, the covariance depends on the distance between locations and is computed from the variogram (Heuvelink et al., 2007).

Context (environment)

Identification of the correct context, i.e. environment, conditions and circumstances, reduces uncertainty in derived models and predictions. External considerations (e.g. economic, social and political) should be considered in the identification of the hypothesis, which represents the *model* or *question posed* for testing. Working in the wrong context may introduce significant uncertainties, leading to incorrect conclusions and significant cost.

An example to demonstrate tightly defining the context, to reduce uncertainty, is the monitoring of soil pH changes under dryland pasture in Victoria. The context of this study was to improve process knowledge on soil acidification occurring in 'pastoral agricultural land in Victoria'. The *aim* of the investigation was to report changes in surface soil pH and how these changes relate to soil site characteristics (Crawford et al., 1994). The conclusion was that acidification was pronounced where (a) reference site pH was moderately to slightly acidic, (b) no change where strongly acid reference sites were observed, and (c) there were pH increases where the reference site was strongly acid

where perennial pastures were improved. However, a shortcoming of this study was that these environmental conditions were confounded with other environmental conditions, e.g. condition (a) occurs in landscapes where there is less rainfall than (b), and management factors, i.e. subterranean clover based pastures are sown in the lower rainfall conditions associated with (a) while perennial pastures based on white clover are sown in the high rainfall landscapes of (b).

Expert opinion- cognitive constraints

Often, in soil classification, expert ratings of soil types are required for categorical data types, but this process can be subject to uncertainty due to observer age, origin, training or culture bias. Examples include: subjective classification of a soil type between Dermosol or Kandosol (Isbell, 2002), or, description and definition of very fine soil structure that can be interpreted as apedal. In addition, there is uncertainty due to data interpretation and linguistic ambiguity. The latter issue relates to communication and the uncertainty is reducible with further verbal or written elaboration.

Expert opinion – physiological constraints

Many field tests are undertaken by experts, but are limited due to sensory limitations in the observer arising from (a) the ambient illumination and (b) the visual response of the observer. For example, pH matching in the field using soil test kits and cards is limited by spectral illumination. Light scattering in the atmosphere is responsible for the blue colour of the sky during the day (and strong red hue in the evening due to increased path length of travel). The spectral content of daylight is also measurably different in the northern hemisphere than the southern hemisphere (Dixon, 1978).

The pH matching process is also affected by the fact that up to 10% of males are classified as colour blind, so that there is approximately 1:10 chance of an anomalous result (e.g. Pettijohn, 1998). Strictly speaking, colour deficiency occurs in the red-green range due to defects on the X chromosome and this can be checked by the Ishihara Colour test chart (as used for car licence testing).

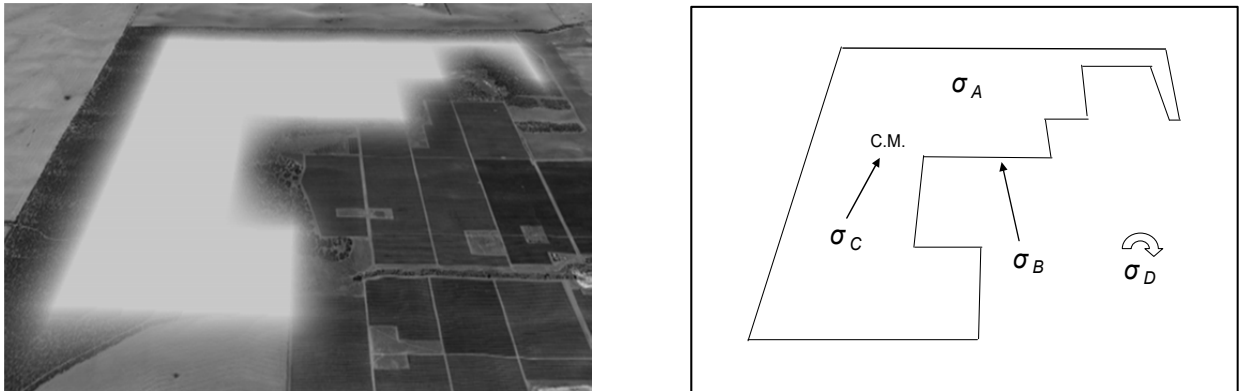


Figure 5.3. (a) Original area of an exemplar soil attribute, A (adapted from Benke et al., 2011), and (b) uncertainty in geometric properties represented by standard deviations of soil attribute, A, its boundary, B, its position (centre-of-mass), C, and rotation, D.

Framing (problem boundaries)

Framing the problem determines the trade-off between a systems-based approach with externalities or, alternatively, a reduced simple and confined case study that is limited by its reductionist nature. Framing issues are extremely common in ecosystems research and environmental modelling where insufficient funding often leads to reductionist approaches, ultimately a poor sample design, and therefore inconclusive results.

The re-sampling study of the National Soil Fertility Project (Colwell, 1977), by Crawford and Robinson (2014), focussed on assessing the change in SOC for defined agricultural regions in Australia Victoria (see Colwell, 1977). The emphasis of the original

investigation was on the examination of the relationships between yield response to fertiliser, soil fertility and environmental factors, such as soil types, while the most recent investigation focussed on the magnitude of changes that had occurred in SOC at these sites.

The limitations of this study are that the observations are from a reduced set of original sites. This weakens the potential to find statistically significant differences in relation to factors such as soil type. Furthermore, the study was confined to soil types originally sampled and failed to consider how SOC may have changed over time for other soil types. Also, only three phases of sampling were undertaken during this time period, and there may be considerable temporal fluctuations in SOC that were not captured in the study, within those sampling times. This could however be mitigated by comparison with suitably designed long-term soil experiments where sample timeframes are considerably less.

Geometrical uncertainty

For an extended region or a polygon in a map, spatial uncertainty in the boundary can be represented as a fuzzy edge (Figure 5.3(a)). The uncertainty metric at the edge is standard deviation, and the edge profile may be represented visually by the CDF. A visualisation scheme may be a useful addition to future digital soil maps. Uncertainty in the attribute, A , with mean, μ_A (from measurements or model predictions) is represented by the metric σ_A for the region, or a polygon, as shown in Figure 5.3 (see also Benke et al., 2011). For the purpose of visualisation, the uncertainty metric for A may also be represented by a pseudo-colour encoding scheme in addition to an assigned numerical value for σ_A (e.g. a spectrum from blue to red, representing low to high uncertainty, respectively).

Uncertainty at the fuzzy edge (boundary) enclosing attribute A is represented by σ_B , which can also be derived from sampling measurements or multiple model realisations using Monte Carlo simulation. Positional uncertainty in the homogeneous region or polygon would be represented by σ_C , representing the variability in the centre-of-mass (C.M.), assuming no shape deformation, which again would be elicited from multiple realisations from Monte Carlo simulation. Finally, rotational uncertainty about an angle θ would be represented by σ_D .

Implementation error (numerical approximation)

Application of algorithms often involves approximations to analytic models, which may be based on differential equations. In the case of interpolation and prediction, non-linear models in particular may have rapid changes in some regions, or local discontinuities that may produce spurious results. Sampling interval size in time-series analysis is a well-known source of error in calibration accuracy. Similarly, assumption of stationarity in statistical properties may be incorrect over time.

Legacy data (e.g. digitisation errors)

Old records of soil properties are often in the form of notebook data and need to be digitised as computer records. Errors are introduced by manual key entry mistakes, numerical precision errors, and changes in measurement methodology over time. Legacy land resource assessment maps often failed to include an indication of accuracy (extent to which an estimated value approaches a true value) and the precision (i.e. dispersion of observed values around the mean – measure of the standard deviation).

Traditionally, soil surveys were undertaken for ‘general-purpose’ or ‘special-purpose’ interpretation regarding soil and land resources. This was further complicated by scale

and recommended use of the survey. Soil sampling sites were often chosen on the best judgement of the surveyor who balanced importance, representativeness and ease of access against the financial and time resources constraining the survey. In many instances sites and samples are opportunistic, e.g. road cuttings and other soil exposures. Today, it is viewed as a requirement that a map should have a quantitative estimate of the uncertainty as this is a fundamental input to biophysical simulation modelling.

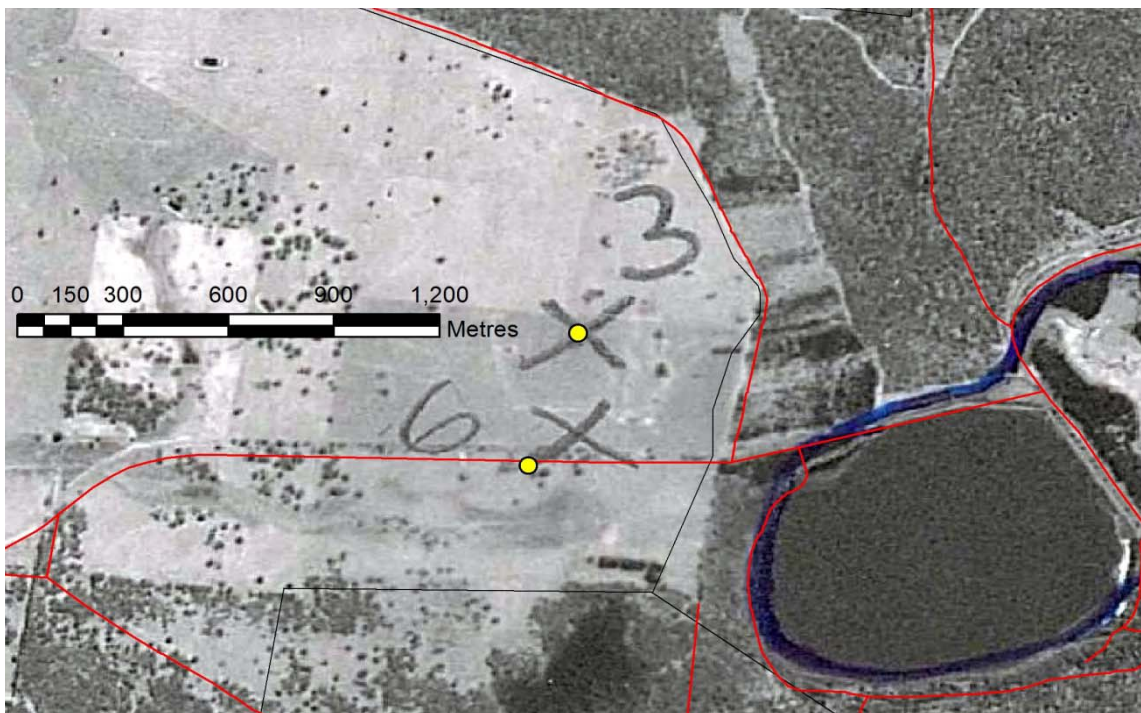


Figure 5.4. Example of spatial inaccuracy in georeferencing soil sites (see overlaid red lines). Note that it may be difficult to ascertain precision level in site locations on air-photos or maps.

Mapping uncertainty

Uncertainties in polygons of soil maps are derived from (i) the measurement process, (ii) spatial variability (positional and attribute uncertainties) of soil and covariates, and assigned membership to a soil body, (iii) method used for spatial

aggregation/generalization, and (iv) uncertainties in the control parameters that these methods use (Burrough, 1993). McBratney (1992) suggested uncertainty with soil information has dimensions that are stochastic (statistical and probability theory), deterministic (chaos theory) and semantic (fuzzy theory). It does not, however, cover adequately the uncertainty defined as ‘epistemic’ in nature, which is a subject of this paper. Accounting for prediction uncertainty through error propagation in models has not been realised in conventional soil mapping to date (although some progress has been made in related land use studies - see, for example, Pelizaro et al., 2011).

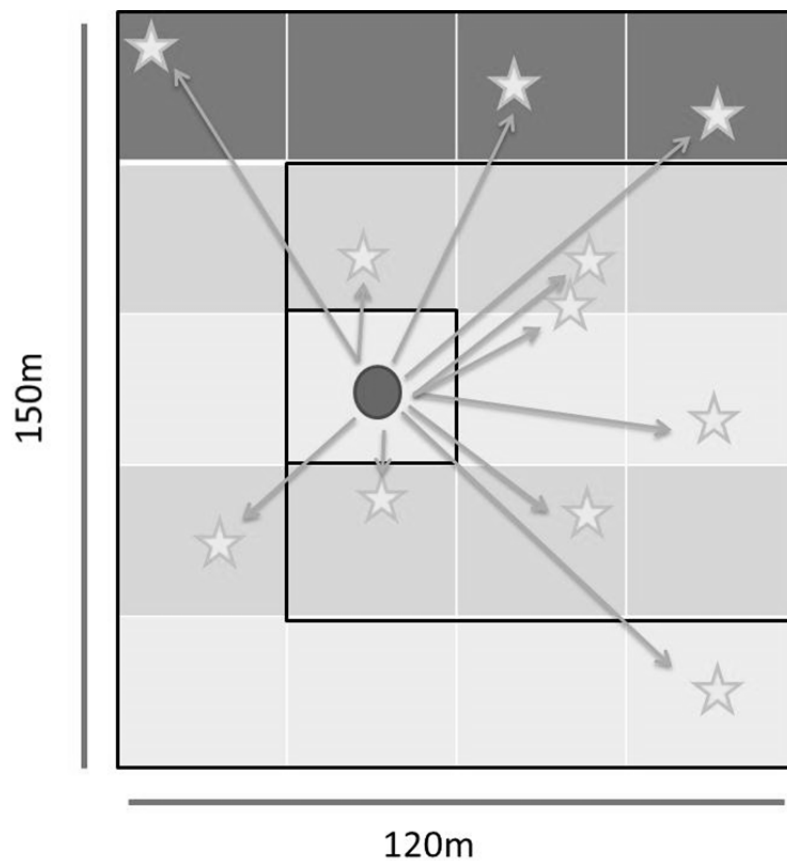


Figure 5.5. Random error possibilities of the true pixel location for a point in a legacy soil site relative to surrounding pixels. Note that a vertical displacement of one pixel results in a 30 m error in this case.

Uncertainty in site location arises where *imprecise* site descriptions are associated with the point-source data; a general reference to a landowner's paddock, or native vegetation, may be all that exist. In these instances uncertainty may be greater than several hundred metres. For broad-scale survey purposes such imprecision is not as critical but may become important when soil-landscape modelling is carried out at fine spatial resolution, e.g. 10 to 30 metres. This is illustrated in Figure 5.4 and Figure 5.5, where the spatial uncertainty can have significant impact on the spatial covariates assigned to that site. Here the site potentially can be assigned to 1 of 20 possible pixels and the covariates that underlie that pixel.

Where marks have been made on maps or pinholes made in aerial photographs to indicate locations, uncertainties may still be in the order of 20 - 500 metres (1 mm is equal to 100 m on a 1:100,000 scale map, and 40 m on an average air photograph). In the 1990s GPS locations had low precision with errors of at least 100 m. From May 2000 civilian GPS accuracy improved to better than 20 m when the US government stopped degrading the satellite data for civilian use. More recently GPS have become more reliable and they consistently provide accuracies of a few metres or even centimetres when differential systems are used. Location coordinates from sites using a GPS prior to May 1st 2000 should be regarded as less accurate than subsequent data unless a DGPS was used.

Measurement error

Measurement of soil properties are often replicated to provide an improved estimate of mean value. Measurement error is classified as epistemic uncertainty because it is reducible (see Box 1). Repeated measurements also provide information for the error distribution of the model output, providing information for a box-and-whisker plot, and uncertainty metrics, such as variance and the confidence interval. Examples in the

literature include Goidts et al. (2009) and the critical inclusion of measurement error sources in the detection of SOC changes; Cayley et al. (2002) and the application of functions to predict soil pH in 0.01 M CaCl₂ from analytes including SOC and EC; Holmes et al. (2011) identified the minor influence of bulk density in determination of SOC stock; Tirez et al. (2014) identified that laboratory measurement was the principle source of error in SOC monitoring; Damasceno et al. (2006) implemented Monte Carlo simulation to derive uncertainty estimates for laboratory observation of pH; Leito et al. (2002) define a deterministic approach to identify separate sources of error in measurement; and Slattery and Burnett (1992) identified potential issues with storage and measurement of pH due to changes with time.

Model input uncertainty

Inputs to models are all subject to errors which can be represented by probability distributions. These errors propagate through the model and contribute to output error. The model inputs include numerical data and information that represent the system or process under investigation.

Model parameter uncertainty

Calibration of a model involves iterative error minimization between model predictions and measurements. The residual error on completion of the parameter estimation process indicates less than perfect fitting. Parameters used in models include exact parameters (universal constants), fixed parameters that have been determined by previous investigations, or calibrated parameters that are determined from calibration attempts to minimize prediction error. These parameters will be point estimates with an associated uncertainty in the estimation. The Bayesian paradigm actually treats the parameters as

probability distributions from which population statistics, such as variance, may be derived as metrics of uncertainty.

Model structure adequacy

Most statisticians and soil scientists are interested in model prediction accuracy, often using a linear or polynomial regression for prediction (as distinct from using the theoretically correct relationship between covariates). Unfortunately this approach often results in curve fitting only. For example, a simple linear regression equation is limited by the linear approximation. A robust model requires the correct relationship between the covariates. Only then can theoretical behaviour be understood and be part of the system analysis. This approach, which replaces the statistical model by a physical model, is referred to as ensuring model structural adequacy. In practice, introducing the research hypothesis, and combining it with inductive reasoning, will lead to improvements in an iterative sense although the absolutely correct model may not be achieved. Calibration may adjust parameters in an effort to compensate for structural inadequacy.

As mentioned in the previous section, a physical model is preferred over a statistical model, if possible. This means that the soil scientist should work in concert with the statistician, otherwise problems can occur. An illustrative example is the simple case of a scatterplot for Newton's law of motion, $F=ma$, where acceleration is proportional to force applied to a mass. A statistician, on observing the scatter plot, fits a straight line with slope, m , referring to m as the gradient. The physical scientist disagrees by saying 'no, m is the mass and is a physical quantity – it has volume, density and gravity and affects the surrounding environment by its presence'. A purely statistical model would miss the physical consequences revealed by the physical model.

Using a multiple linear regression model for prediction introduces uncertainty simply by the fact that nature is not linear, as illustrated by chaos theory, nonlinear dynamics, quantum theory, etc. The fact is that there are no straight lines in nature. The linear approximation used by statisticians is simply to make the arithmetic analysis simple and tractable. The linear approximation adds further uncertainty to analysis.

Positional uncertainty

Positional uncertainty occurs with soil sampling points, polygon map a boundary, linear transects and raster transformations. Positional uncertainty is described by a probability density function (PDF) and relates to objects comprising multiple points with structure that may or may not change under uncertainty (e.g. rigid objects and deformable objects). Positional uncertainty of a point object leads to a shift in its 4-D status $P(x,y,z,t)$ subject to enumeration by the PDF. A rigid body is subject to geometric transformations, such as translation and rotation about an axis, with specific enumerations also subject to a PDF. Deformable objects may be altered by positional uncertainty due to independence of the primitive points.

Resolution (spatial and temporal)

Spatial error is always present in maps due to the effect of sampling errors in scatterplots and interpolation procedures using kriging methods. Time stepping resolution may also introduce calibration errors due to the trade-off between cost and high sampling rates, and autocorrelation. Also, stationarity in statistical properties, such as mean and standard deviation, is often assumed but may change in time and affect calibration studies. The process of analog-to-digital conversion introduces quantisation errors in digitised data.

Software problems (verification, validation, bugs)

Software must be *verified* to check it implements the model correctly. Subsequently, the model must be *validated* against test data. In both cases, insufficient testing over range and sample size may introduce uncertainty. Lack of exhaustive testing of software may fail to identify bugs that may invalidate future results for specific combinations of inputs.

5.4 Example 1: Disaggregation of error sources

Using the SCORPAN model as an example (McBratney et al., 2003), there are four main steps involved in digital soil mapping with uncertainty (Minasny et al., 2010). First, data input for the region of interest requires production of the digital map, using covariates of interest in the study, which may include terrain attributes, multispectral satellite imagery, land use data, geological information and possibly legacy soil maps. Second, estimates of soil properties, including uncertainties, are produced from relationships between point soil measurements and spatially covered covariates, i.e.

$$S = f(s, c, o, r, p, a, n) + \varepsilon (1)$$

where S is the soil property, attribute or class of interest, f is the model incorporating covariates s (other soil properties), c (climate properties), o (organisms), r (topography), p (parent material), a (age or time factor), n (spatial position absolute and relative), and ε is the error. Third, spatially inferred soil properties are used to predict other soil functions, such as soil water content, carbon density, and phosphorus (see Minasny et al., 2010). Thus, the prediction uncertainty of the SCORPAN model combines uncertainties in input data, spatial inferences and soil properties and functions. The fourth step includes completion of a digital soil assessment for use by policy makers and land use managers, including evaluation of soil functions such as biomass production and buffering capabilities (Carré et al., 2007). Note that the SCORPAN approach is specialised to soil,

but is a subset of the GRUMP conceptual model that also incorporates many epistemic uncertainties. The GRUMP framework is suitable for application to any predictive model (Figure 5.2).

A strategy for *disaggregation of error sources* in digital soil mapping using the SCORPAN approach has been suggested recently by Nelson et al. (2011). The approach combines a geostatistical model and Monte Carlo simulation to estimate underlying errors. A Linear Mixed Model (LMM) was used to produce a digital soil map of clay content and prediction error.

Nelson et al. (2011) considered four major sources of error including,

- covariate error
- model error
- analytical error (of soil properties)
- positional error

In the first category of covariate error, environmental covariates, such as mean annual rainfall, are aggregated into a common grid in the model. The principal source of error related to measurement error, except where sensor data is introduced, or low sample rates are involved, or both, leading to the further inclusion of interpolation error. In the second category of model error, sources of error included incorrect model, parameter error, redundant covariates, and interpolation error. Note that all of these sources can inflate the error variance in model prediction, especially given incorrect assumptions on statistical parameters, such as stationarity in first and second order moments.

For example, in the LMM, errors in fixed effects coefficients are assumed subject to the normal distribution. Variation not explained by the model is quantified by the *nugget* and *sill* variance in geostatistics. Effectively, digital soil mapping is a process distinguished

by interpolation of low density soil observations into a dense grid of prediction locations. Model error is subsequently quantified by the error variance for these predictions. This is often executed by the process of bootstrapping, i.e. a model fitting exercise involving multiple realisations of the dataset, which may be obtained from probabilistic simulations of the original whilst retaining its statistical properties (such as first and second order moments).

In the third category of analytic error, the primary consideration is the quantitative error in measurement of soil properties. In the case of soil properties, such as organic carbon content, laboratory methods are expensive but more accurate with lower dispersion than remote sensing methods. In the fourth category of positional error, samples taken near class boundaries produced greater errors than samples from class interiors (Figures 5.5). Historical samples, referred to as legacy data, are associated with larger errors than current data from accurate GPS technology (Grimm and Behrens, 2009; Carré et al., 2007).

In the case of model-based geostatistics, LMM can be used with parameters estimated using the residual maximum likelihood (REML), which was recommended over the standard regression-kriging approach (Lark and Cullis, 2004), coupled with interpolation by the empirical best linear unbiased prediction (E-BLUP). Regression-kriging produces estimates of parameters and spatial correlation separately, which may lead to bias and errors in variable selection (Nelson et al., 2011).

In the case study conducted by Nelson et al. (2011), the general form of the LMM was fitted to clay data using REML for estimation. Error sources were ranked by variance, i.e. contribution to mean square error (MSE) for four data quality scenarios (Table 5.2). Model error (parameter error, interpolation error, etc.) accounted for two thirds of the

total variance in prediction for all four scenarios. Position error accounted for less than 1% of variance and was related to the grid size of interpolation relative to the covariates.

It appears that sources of error are often analytic and covariate in nature, whilst the least error occurs with positional uncertainty and measurement error (Heuvelink and Brown 2007; Nelson et al., 2011). The advantages of the so-called error budget approach is that it resolves the total error in the digital map of clay content into separate proportional contributions from different error sources. Further elaboration on various error categories is provided in the literature (e.g. Refsgaard et al., 2007; Benke et al., 2007; Robinson et al., 2014).

Table 5.2. Comparison of error sources for error-budget model for data quality of clay (indicative data from Nelson et al., 2011). Table shows proportion of variance contribution to MSE of predictions.

Error Source	SCENARIO			
	<i>Good</i>	<i>Spectroscopic</i>	<i>Legacy</i>	<i>Poor</i>
Model	69%	72%	69%	72%
Analytic	< 0.5%	3.5%	< 0.5%	3.55%
Positional	<-0.5%	<-0.5%	<-0.5%	< -0.5%
Covariate	1.7%	3.0%	1.4%	2.7%

The SCORPAN error budget approach is essentially a static model that can be implemented at different time periods discretely. A framework incorporating time dependence explicitly is the STEP-AWBH conceptual model for soil evolution, as proposed by Grunwald et al. (2011). The framework includes anthropogenic and natural forcing's which determine and modulate soils and space-time interactions. The model addresses temporal factors correlating with soil change, including land use change and

climate change in temperature and precipitation, and can be implemented by stochastic simulation methods or deterministic approaches, such as regression trees. At present the model is conceptual in nature and there is not yet sufficient research published on possible practical implementations.

5.5 Example 2: Spatial uncertainty and geostatistics

Geostatistical methods for analysis of spatial data, requiring spatial interpolation by kriging, are widely used in mapping, and more recently for prediction of spatio-temporal change in soil properties. Users of spatial data need to be mindful that models representing dynamic phenomena are subject to uncertainty in inputs and outputs (Cressie and Wikle, 2011). The process of spatial data discovery, treatment and transformation, together with analysis and derived predictions using geostatistics is described in the literature (e.g. Webster and Oliver 2007; Oliver and Webster, 2014). Techniques that use *likelihood-based methods* are now preferred over *method-of-moments* due to more efficient estimation of unknown parameters and assessment of uncertainty in the spatial predictions (Diggle and Ribeiro, 2007; Stein, 1999).

Sources of uncertainty in assessments of soil change were investigated based on soil fertility test data for the period 1989 to 1994 from the Hopkins River and Curdies River catchments in south-western Victoria. The specific objective was spatio-temporal assessment of change in soil pH for 0 to 10 cm (refer to method 4A1 in Rayment and Lyons, 2011). Soil acidification in this region is recognised as a significant land degradation issue and has been reported in numerous studies, including Crawford et al. (1994). Further details on the sample collection method and processing are provided in Marchant et al. (2014).

For the region shown in Figure 5.6, pH measurements were taken for the period 1989-91 and compared with the period 1992-4 to check for evidence of acidification (Figure 5.7). A crude non-spatial comparison of field measurements of pH between the two time periods produced a difference between mean values of 0.265 (95% CI = 0.007, 0.522) for the raw data in the region, according to the t-test between means. The pH data from both time periods were then interpolated using ordinary kriging using the VESPER software package (Minasny et al., 2005). The paired predicted values were then compared (without their associated errors) using the *paired difference* t-test. The test produced a value for the mean difference of 0.222 between the kriged data sets (95% CI = 0.221, 0.223), which was highly significant ($\alpha = 0.001$). This suggested that there was a significant change in pH between the two time periods. However, when the kriged means were compared with the inclusion of their associated errors, there was no significant difference between the two predicted surfaces for any of the grid points. These results illustrate that a simple statistical analysis may give a misleading conclusion because it may not account for all the possible epistemic uncertainties.

Sources of epistemic uncertainty include measurement error, spatial uncertainty of point locations used in the formation of the variogram and its parameters, temporal uncertainty of pH observations, and the exponential variogram model used (other possibilities included spherical, normal, matern or power distributions). In addition, sampling size may have been small, sparse and not representative, and some sites were not resampled in the second time period, or the timescale may not have been appropriate, or relevant, or both. Analytical test results were georeferenced according to descriptions accompanying the data and also represent a potential source of uncertainty. Other potential issues include bias, where farmers focused their analysis on regions where poor growth and production

were expected. It is clear that statistical variability in data analysis must be viewed in the context of possible constraints due to epistemic uncertainties and their effects.

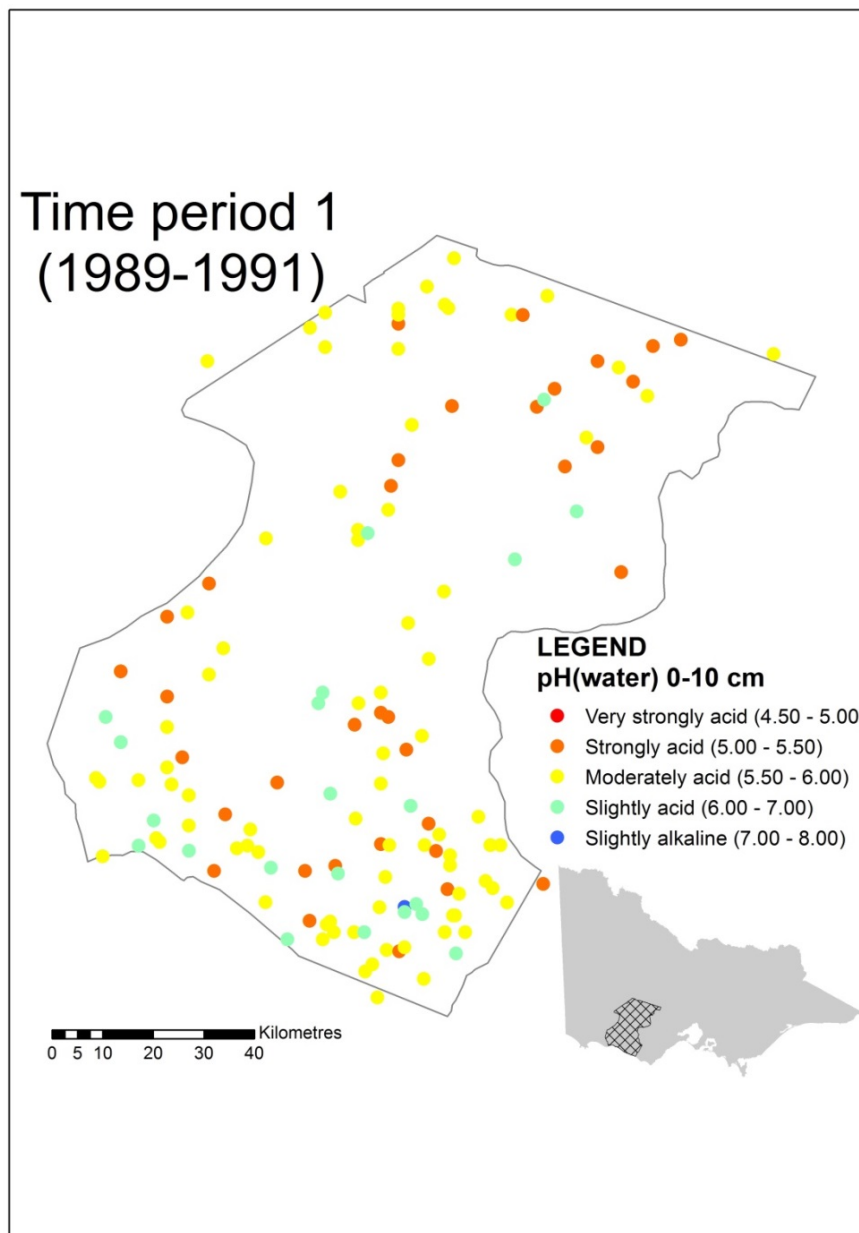


Figure 5.6. Data for pH measurements from the Hopkins River and Curdies River catchments in south-western Victoria for the time period 1989-91.

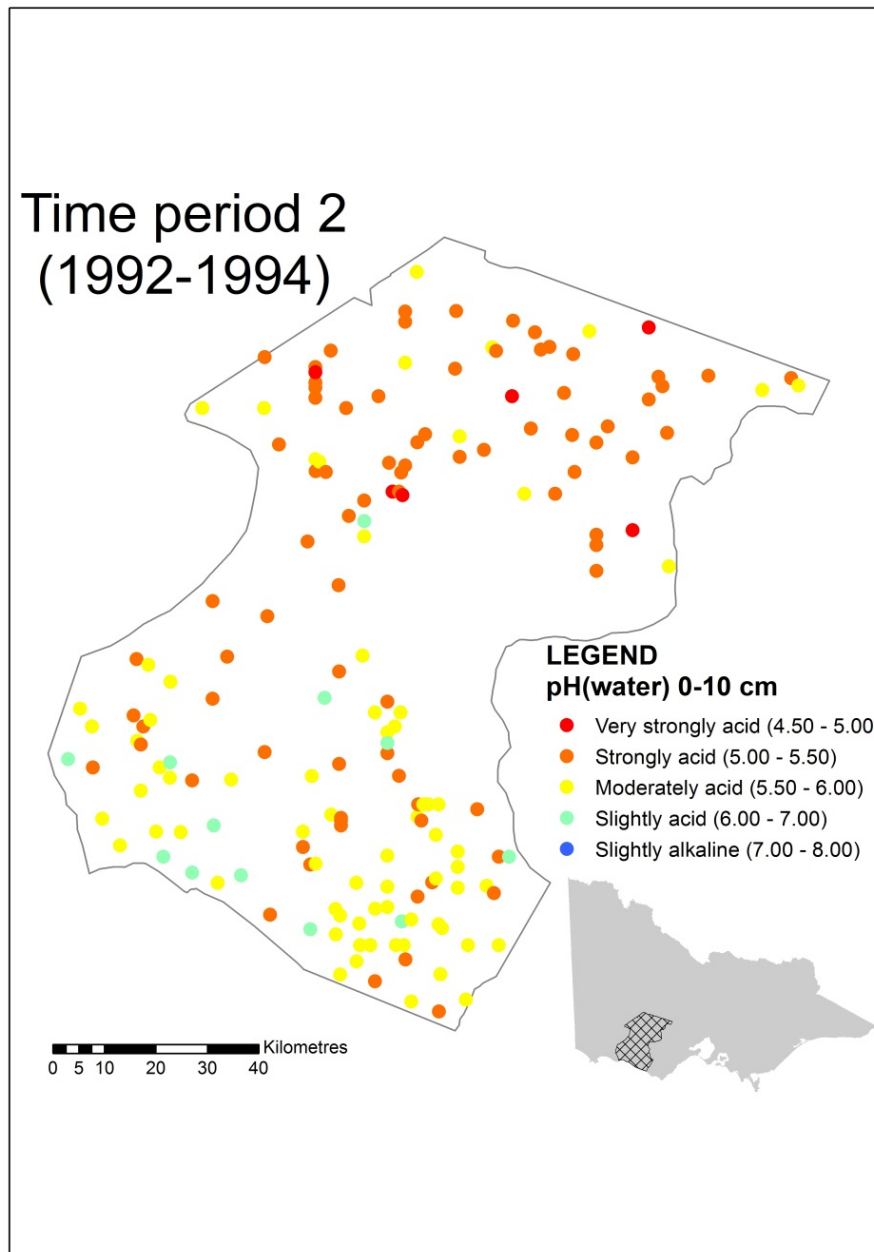


Figure 5.7. Data for pH measurements from the Hopkins River and Curdies River catchments in south-western Victoria for the time period 1992-94.

The GRUMP framework may be used to explore and identify the uncertainties in the modelling process. A new and detailed experiment is currently being designed to address the issues raised above and will use GRUMP to apply an approach proposed by Benke et al. (2008). They applied a biophysical model in multiple Monte Carlo simulation

experiments as the special case of a GRUMP implementation. All inputs were replaced by probability distributions and then each distribution was progressively replaced with the mean value, and a corresponding new Monte Carlo simulation was executed in a manner analogous to step-wise multivariate regression involving forward selection and backward elimination (see Figs 10 and 11 in Benke et al., 2008). The output distribution for each separate simulation experiment was represented by an uncertainty metric, such as the variance or confidence interval, and then compared with the corresponding input distributions in the form of a stochastic sensitivity plot. The inputs responsible for the greatest output uncertainty were consequently identified and ranked.

The approach described above applied the PERT probability distribution, which is used for modelling statistical variables and expert opinion in risk analysis (Vose, 2000). The importance of the PERT distribution is that it can be used to codify data from epistemic uncertainties for use alongside statistical variables in the same simulation experiment. For a proposed error budget, the GRUMP framework with a biophysical nonlinear model represents an alternative approach to a linear statistical model for prediction uncertainty. Its potential advantage is that it adds the benefits of model structural adequacy, best fit, and identification of nonlinear relationships between the covariates.

5.6 Conclusion

In soils analysis, uncertainties include many errors that are statistical, geometrical and epistemic in nature. In this paper, a large number of uncertainties are highlighted in a more comprehensive attention to the problem than traditional consideration of statistical variability only. The effect of multiple error sources of uncertainty is reviewed, including covariate error, model error, laboratory analytical error, and positional error. In particular,

the distinction is made between statistical variability (aleatory uncertainty) and lack of information (epistemic uncertainty).

A global analytic framework for uncertainty was proposed and used to organise contributions from error sources into a process modelling approach. Examples of uncertainty analysis were provided in the case of error disaggregation and geostatistics. It was concluded that a more comprehensive and multi-factor approach to uncertainty analysis is necessary in future, especially in the context of developing a soils modelling process for incorporation of known sources of uncertainty.

Future research

There are a number of issues that require further research in soil science modelling and uncertainty analysis. Prominent topics for further research include the following:

- Strategies are required for managing epistemic uncertainties, including errors in expert opinion, transcription errors from legacy data, and resolution of digitisation.
- Model calibration uncertainty is affected by time stepping resolution. Time stepping issues need more research, such as comparison of time stepping resolution vs prediction accuracy vs computational expense.
- Representation of uncertainty on digital maps requires more research.
- Research is required on visualisation of multi-dimensional data.
- Systems-thinking is needed and not more reductionist-thinking, i.e. error propagation through complete systems, not isolated models, and including also information loss as well estimates of statistical variability. A soil does not exist in isolation - it is part of an ecosystem that includes the atmosphere, climate and hydrology.
- Implementation of the GRUMP framework to practical examples and new cases.

References

- Aerts, J.C.J.H., Heuvelink, G.B.M., Goodchild, M.F., 2003. Accounting for Spatial Uncertainty in Optimization with Spatial Decision Support Systems. *Transactions in GIS* 7, 211–230.
- Baecher, G.B., Christian, J.C., 2000. Natural Variation, Limited Knowledge, and the Nature of Uncertainty in Risk Analysis. *Proceedings of the 25th Anniversary Engineering Foundation Conference on Risk and Uncertainty in Water Resources Engineering*, Santa Barbara.
- Benke, K.K., Hamilton, A.J., Lowell, K.E., 2007. Uncertainty analysis and risk assessment in the management of environmental resources. *Australasian Journal of Environmental Management* 14, 243–249.
- Benke, K.K., Lowell, K.E., Hamilton, A.J., 2008. Parameter uncertainty, sensitivity analysis and prediction error in a water-balance hydrological model. *Mathematical and Computer Modelling* 47, 1134–1149.
- Benke, K.K., Pelizaro, C., 2010. A spatial-statistical approach to the visualisation of uncertainty in land suitability analysis. *Journal of Spatial Science* 55, 257–272.
- Benke, K.K., Pettit, C.J., Lowell, K.E., 2011. Visualisation of spatial uncertainty in hydrological modelling. *Journal of Spatial Science* 56, 73–88.
- Bishop, T.F.A., McBratney, A.B., Whelan, B.M., 2001. Measuring the quality of digital soil maps using information criteria. *Geoderma* 103, 95–111.
- Cayley, J.W.D., McCaskill, M.R., Kearney, G.A., 2002. Changes in pH and organic carbon were minimal in a long-term field study in the Western District of Victoria. *Australian Journal of Agricultural Research* 53, 115–126.
- Crawford, D.M., Baker, T.G., Maheswaren, J., 1994. Soil pH changes under Victorian pastures. *Australian Journal of Soil Research* 32, 105–115.
- Crawford, D.M., Robinson, N.J., 2014. Using legacy sites to assess changes in soil organic carbon. *Soil Change Matters, papers and abstracts*. Bendigo, Australia.

- Carré, F., McBratney, A.B., Mayr, T., Montanarella, L., 2007. Digital Soil assessments: beyond DSM. *Geoderma* 142, 69–79.
- Colwell, J.D., 1977. National soil fertility project. Volume 1: objectives and procedures. CSIRO, Canberra, Australia.
- Cressie, N., Wikle, C.K., 2011. Statistics for spatio-temporal data. John Wiley & Sons Inc, Singapore.
- Damasceno, J.C., Borges, R.M.H., Couto, P.R.G., Ordine, A.P., 2006. Estimation of primary pH measurement uncertainty using Monte Carlo simulation. *Metrologia* 43, 306–310
- Diggle, P.J., Ribeiro jr., P.J., 2007. Model-based geostatistics. Springer, New York, USA.
- Dixon, E.R., 1978. Spectral distribution of Australian daylight, *Journal of the Optical Society of America* 68, 437–450. <http://dx.doi.org/10.1364/JOSA.68.000437>.
- Goidts, E., Van Wesemael, B., Crucifix, M., 2009. Magnitude and sources of uncertainties in soil organic carbon (SOC) stock assessments at various scales. *European Journal of Soil Science* 60, 723–739.
- Grimm, F., Behrens, T., 2009. Uncertainty analysis of sample locations within digital soil mapping approaches. *Geoderma* 155, 154–163.
- Grunwald, S., Reddy, K.R., Newman, S., DeBusk, W.F., 2004. Spatial variability, distribution and uncertainty assessment of soil phosphorus in a south Florida wetland. *Environmetrics* 15, 811–825.
- Grunwald, S., Thompson, J.A., Boettinger, J.L., 2011. Digital soil mapping and modelling at continental scales – finding solutions for global issues. *Soil Science Society of America Journal* 75, 1201–1213.
- Hastings, D., McManus, H., 2004. A Framework for Understanding Uncertainty and its Mitigation and Exploitation in Complex Systems. Engineering Systems Symposium, Massachusetts Institute of Technology, Cambridge, Massachusetts.
- Heuvelink, G.B.M., 2014. Uncertainty quantification of GlobalSoilMap products. In: D. Arrouays, N. McKenzie, J. Hempel, A.R. de Forges, A.B. McBratney (Eds.),

- GlobalSoilMap - Basis of the global soil information system. CRC Press, Netherlands, pp. 335-340.
- Heuvelink, G.B.M., Brown, J.D., van Loon, E.E., 2007. A probabilistic framework for representing and simulating uncertain environmental variables. *International Journal of Geographical Information Science* 21, 497–513.
- Holmes, K.W., Wherrett, A., Keating, A., Murphy, D.V., 2011. Meeting bulk density sampling requirements efficiently to estimate soil carbon stocks. *Soil Research* 49, 680–695.
- Hopley, J., Rees, D.B., MacEwan, R.J., Clark, R., Benke, K., Imhof, M.P., Bardos, D.C., 2014. A digital soil map of Victoria—VicDSMv1. p. 185. In Ed. Arrouays, D, et al., *GlobalSoilMap: Basis of the global spatial soil information system*, CRC Press, Boca Raton, Florida, USA.
- Iman, R.L., Helton, J.C., 1988. An Investigation of Uncertainty and Sensitivity Analysis Techniques for Computer Models, *Risk Analysis* 8, 71–90.
- Isbell, R.F., 2002. *The Australian Soil Classification. Revised Edition.* CSIRO Publishing, Collingwood, Victoria.
- Jenny, H., 1941. *Factors of soil formation.* McGraw-Hill, New York, NY.
- Lark, R.M., Cullis, B.R., 2004. Model-based analysis using REML for inference on systematically sampled data on soil. *European Journal of Soil Science* 55, 799–813.
- Leito, I., Strauss, L., Koort, E., Pihl, V., 2002. Estimation of uncertainty in routine pH measurement. *Accreditation and Quality Assurance* 7, 242–249.
- MacEwan, R.J., 1997. Soil quality indicators: pedological aspects. Pages 143-66 In *Soil quality for crop production and ecosystem health.* (Eds. Gregorich E and Carter MR) Elsevier: Amsterdam (Netherlands).
- MacEwan, R.J., 2014. Frameworks for soil matters. In *Soil Change Matters International Workshop.* Bendigo, Vic. 24-27 March 2014. Pp. 1-3. (Department of Environment and Primary Industries: East Melbourne).

- Marchant, B.P., Crawford, D.M. and Robinson, N.J., 2015. What can legacy datasets tell us about soil quality trends? Soil acidity in Victoria. *IOP Conf. Series: Earth and Environmental Science* 25 (2015) 012015. <http://dx.doi:10.1088/1755-1315/25/1/012015>.
- McBratney, A.B., 1992. On Variation, Uncertainty and Informatics in Environmental Soil Management. *Australian Journal of Soil Research* 30, 913–35.
- McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. *Geoderma*, 117, 3-52.
- Minasny, B., McBratney, A.B., Whelan, B.M., 2005. VESPER version 1.62. Australian Centre for Precision Agriculture, McMillan Building A05, The University of Sydney, NSW 2006. (<http://www.usyd.edu.au/su/agric/acpa>).
- Minasny, B., McBratney, A., Malone, B., Sulaeman, Y., 2010. Digital mapping of soil carbon. 19th World Congress of Soil Science, Soil Solutions for a Changing World 1 – 6 August 2010, Brisbane, Australia. Published on DVD.
- Nelson, M.A., Bishop, T.F.A., Odeh, I.O.A., Triantafyllis, J., 2011. An error budget for deferent sources of error in digital soil mapping. *European Journal of Soil Science* 62, 417–430.
- Oliver, M.A., Webster, R., 2014. A tutorial guide to geostatistics: Computing and modelling variograms and kriging. *Catena* 113, 56-69.
- Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences* 30, 683-691.
- Pelizaro, C., Benke, K., Sposito, V., 2011. A modelling framework for optimisation of commodity production by minimising the impact of climate change. *Applied Spatial Analysis and Policy*, 4, 201-222.
- Pettijohn, P.F., 1998. *Psychology: A ConnecText*, Dushkin Publishing Group, Inc, USA.
- Rayment, G.E., Lyons, D.J., 2011. *Soil chemical methods – Australasia*. CSIRO Publishing, Collingwood, Australia.

- Refsgaard, J.C., van der Sluijs, J.P., Højberg, A.L., Vanrolleghem, P.A., 2007. Uncertainty in the environmental modelling process – A framework and guidance. *Environmental Modelling & Software* 22, 1543-1556.
- Richter, D., Hofmockel, M., Callahan Jr., M.A., Powlson, D.S., Smith, P., 2007. Long-term soil experiments: Keys to managing earth's rapidly changing ecosystems. *Soil Science Society of America Journal* 71, 266–279.
- Richter, D., Tugel, A.J., 2012. Soil change in the Anthropocene: Bridging pedology, land use and soil management. In *Handbook of Soil Sciences, Properties and Processes*, 2nd Edition (Eds PM Huang, Y Li and ME Sumner). CRC Press, Florida.
- Richter, D., Yaalon, D.H., 2011. The changing model of soil revisited. *Soil Science Society of America Journal* 76, 766–778.
- Robinson, N.J., Benke, K.K., Hopley, J., MacEwan, R.J., Clark, R., Rees, D.B., Bardos, D., 2014. Multi-source data integration and identification of uncertainties affecting production of a digital soil map. p. 353. In Ed. Arrouays, D, et al., *GlobalSoilMap: Basis of the global spatial soil information system*, CRC Press, Boca Raton, Florida, USA.
- Saby, N.P.A., Bellamy, P.H., Morvan, X., Arrouays, D., Jones, R.J.A., Verheijen, F.G.A., Kibblewhite, M.G., Verdoodt, A., Uveges, J.B., Freudenschuss, A., Simota, C., 2008. Will European soil-monitoring networks be able to detect changes in topsoil organic carbon content? *Global Change Biology* 14, 2432–2442.
- Simonon, R.W., 1959. Outline of a generalised theory of soil formation. *Soil Science Society of America Proceedings* 23, 156–156.
- Slattery, W.J., Burnett, V.F., 1992. Changes in soil pH due to long term soil storage. *Australian Journal of Soil Research* 30, 169–175.
- Sposito, V., Benke, K., Pelizaro, C., Wyatt, R., 2010. Adaptation to Climate Change in Regional Australia: A Decision-Making Framework for Modelling Policy for Rural Production. *Geography Compass* 4, 335-354.
- Stein, M.L., 1999. *Interpolation of spatial data*. Springer, New York, USA.

- Steinitz, C., 1990. A framework for theory applicable to the education of landscape architects (and other environmental design professionals). *Landscape Journal* 9, 136–143.
- Steinitz, C., 2012. *A Framework for Geodesign: Changing Geography by Design*. ESRI Press.
- Tirez, K., Vanhoof, C., Hofman, S., Deproost, P., Swerts, M., Salomez, J., 2014. Estimating the contribution of sampling, sample pretreatment, and analysis in the total uncertainty budget of agricultural soil pH and organic carbon monitoring. *Communications in soil Science and Plant Analysis* 45, 984–1002.
- Tugel, A.J., Herrick, J.E., Brown, J.R., Mausbach, M.J., Puckett, W., Hipple, K., 2005. Soil change, soil survey, and natural resource decision making: A blueprint for action. *Soil Science Society of America Journal* 69, 738–747.
- Wagener, T., Gupta, H.V., 2005. Model identification for hydrological forecasting under uncertainty. *Stochastic Environmental Research and Risk Assessment* 19, 378–387.
- Walker, W.E., Harremoës, P., Rotmans, J., Van der Sluijs, J.P., Van Asselt, M.B.A., Janssen, P., Krayen Von Krauss, M.P., 2003. Defining Uncertainty, A Conceptual Basis for Uncertainty Management in Model-Based Decision Support. *Integrated Assessment* 4, 5–17.
- Webster, R., Oliver, M.A., 2007. *Geostatistics for Environmental Scientists*, Second Edition. John Wiley and Sons, Chichester, USA.
- Young, I.M., Crawford, J.W., 2004. Interactions and self-organisation in the soil-microbe complex. *Science* 304, 1634–1637.

Chapter 6 Improving information content in soil pH maps: a case study in south-western Victoria

The conceptual model of the Global Representation of Uncertainty in the Modelling Process (GRUMP) was presented in Chapter 5 as a systematic framework to integrate the various error sources that contribute to uncertainty. Quantification and enumeration of these error sources in mapping and modelling enable creators of this spatial soil information to convey to users where and what led to uncertainty in the delivered information. The aim of this research was to extend the conceptual logic developed in Chapter 5 as a more wholesome implementation of the GRUMP framework to a soil mapping application. This supports a primary research aim of this thesis to *develop an approach to accommodate, and illustrate to users of spatial soil information, the various error sources in modelling and mapping.*

The objectives of this chapter were to:

Define potential error sources that contribute to uncertainty in mapping of soil pH;

Implement the GRUMP framework to illustrate how the range of error sources contributes to uncertainty in the production of spatial soil information.

There is increasing attention directed to the identification and treatment of epistemic uncertainty (that is, lack of knowledge, context or information). A problem with epistemic uncertainty is that once a source is identified, how may it be incorporated into the total picture on uncertainty? The GRUMP framework has been proposed recently as one path to integrated uncertainty assessment. In order to combine epistemic uncertainty with statistical variability it is necessary to quantify epistemic uncertainty. In this chapter I provide examples of several important epistemic uncertainties and their quantitative

evaluation in the mapping of soil pH. Further progress towards integration of uncertainty from all sources will require development of global metrics, e.g. system total variance or system prediction interval.

6.1 Introduction

A primary threat to the quantity and quality of food production worldwide is soil acidification (FAO and ITPS, 2015). Globally, acid soils are estimated to affect 30% of the ice-free land mass (Uexküll and Mutert, 1995) with accelerated acidification due to the drainage of land, land use change and intensification (e.g. from native systems to productive agriculture), acid rain, and the application of acidifying nitrogen fertilizers. In Australia, 50 million hectares of agricultural land are estimated to have acid surface soils causing annual production losses of \$AUD 1585 million (NLWRA, 2002).

Realistic statements about soil properties, such as pH and its decrease over time (soil acidification), are a current priority for many countries including Australia (Department of Agriculture, 2014). To establish soil pH baselines and better target interventions to manage soil acidity, reliable spatial estimates are necessary. While some countries have invested in purpose built soil monitoring networks (Arrouays et al., 2012), many countries are using available soil information (often referred to as legacy data) to formulate a current picture of soil degradation and changes in soil condition (Marchant et al., 2015). Legacy data used for such purposes can be convenient but also present many potential issues that must be resolved. Inaccuracy in site location, bias in sample design, sparseness and clustering of sample sites (Marchant et al., 2013), imprecise and inaccurate field and laboratory measurements (Raupach, 1954; Raupach and Tucker, 1959; White, 1969; Laslett and McBratney, 1990), and out-dated analytical methodologies incongruent with current methods, all contribute errors to modelling and mapping.

So how do we make better maps to represent soil properties of interest? Ultimately, a better map should be new knowledge that can inform a user to make better decisions. The lack of a clear and concise representation of the soil property of interest and a reliance on

tacit knowledge can lead to incorrect assumptions on impacts of land use and management to soil pH. To best support land managers in their management of land, we need to identify the tacit understandings of land managers and present them with tailored spatial soil information that is congruent and timely for their decision making, easily interpreted and applied, and with certainty defined.

To make a map more 'certain' for users and thus reduce risk, approaches that can integrate data with associated errors, and reduce the effects of these errors, are required. Here systematic approaches can prove useful to accommodate and illustrate uncertainty in the development and delivery of a soil map. The error budget procedure of Nelson et al. (2011), as an example, combines the relative contribution from stochastic error sources in Digital Soil Mapping (DSM: McBratney et al., 2003) and has been adapted for soil salinity mapping purposes (Huang et al., 2015). Potential error sources considered in the error budget include environmental covariates (e.g. error in digital elevation models), soil property measurement error (e.g. accuracy and precision), model error and positional error with location of sites - all known as aleatory uncertainty (Walker et al., 2005; Benke et al., 2007). In DSM, a conventional focus of uncertainty assessment has been on statistical variability using error propagation or other stochastic methods. Epistemic uncertainty due to imperfect knowledge or assumptions can be significant in their contribution to uncertainty assessment. While error sources contributing to epistemic uncertainty can be significant (e.g. incorrect context or environment), it can generally be reduced through the attainment of new knowledge to reduce effects from misdiagnosis, misinterpretation or incorrect implementation. In mapping soil pH, epistemic uncertainties may include: measurement error for different pH analysis methods; temporal cycles in pH; model specification and assumptions; the incorrect environment and conditions in framing of models, assessments and legacy data used. Robinson et al.

(2015) describe a framework, the Global Representation of Uncertainty in the Modelling Process (GRUMP), to integrate epistemic and aleatory uncertainties. The GRUMP framework supports explicit definition, organisation and quantification of these error sources in a modelling process. The sources of aleatory and epistemic uncertainties can be quantified or enumerated through experiments, expert opinion or explicit knowledge. This enables comparison between uncertainty assessment techniques and the ability to modify these depending from user perceptions of uncertainty (McBratney, 1992; Robinson et al., 2015).

In this paper, we combine legacy data with model-based geostatistics to predict soil pH and associated error for south-western Victoria (Australia). We attempt to accommodate error sources contributing to epistemic uncertainty that have rarely been included in previous DSM applications, such as: the time of sampling and seasonal variability, differences in analytical methods , effects of land use change and variable soil sample depth in legacy data. In this example, these error sources are viewed as contributing to epistemic uncertainty. Spatial covariates representing soil forming factors are also used to improve our predictions. To transform spatial prediction and error estimates of soil pH into informative and usable products, a spatial simulation technique to approximate the likelihood (probability) of soil pH being less than critical agronomic thresholds and an explanation of how this process has led to improved information (better maps) is presented.

6.2 Methods

6.2.1 Study area

The study area of 14,000 km² in south-western Victoria comprises the catchments of the Hopkins River and Lake Corangamite (Figure 6.1). It is part of the Western Plains and Western Uplands geomorphological divisions of Victoria (Rees et al., 2010) with low-lying undulating plains of volcanic and sedimentary origin and Palaeozoic bedrock formations as upland residuals at various elevations. The volcanic plains comprise deposits from eruptions over the last 5 million years, including overlapping basalt flows with palaeosols and pyroclastic deposits from scoria cones and tuff. Soils of the Western Plains are of variable age and pedogenic development, the major soils being Sodosols, Chromosols and Vertosols (Isbell 2002, Robinson et al., 2003). Chromosols, Dermosols or profiles that can be strongly acidic (Kurosols) are found in the higher rainfall zones of the study area.

6.2.2 Land use

Historically, livestock production systems in south-western Victoria have dominated landscapes since European settlement in the 1830s. This includes sheep production systems (wool and meat), beef cattle production, dairy and mixed farming systems (Gibbons and Downes, 1964). Animal husbandry practices were supported by improved pastures and significant increases in livestock numbers. In 2000, the majority of the study area was either under improved pastures or native grassland (Figure 6.2a). This was determined from a supervised classification of Landsat scenes, air photo interpretation and field validation. When contrasted against the 2014 land use (Figure 6.2b) from the Victorian Land Use Information System (VLUIS; Morse-McNabb et al., 2015), over 300,000 ha of land (22% of the study area) converted from pasture to forestry or grain production (Figure 6.2c). The prediction accuracy of land use classes for 2014 ($R^2=0.66$)

was lower than the 2000 ($R^2=0.88$) land use map. This is attributed to the use of MODIS imagery in the 2014 assessment and a limited field calibration/validation program. The spatial trends of increasing cropping and where this expansion is occurring are consistent with where existing cropping enterprises were in 2000.

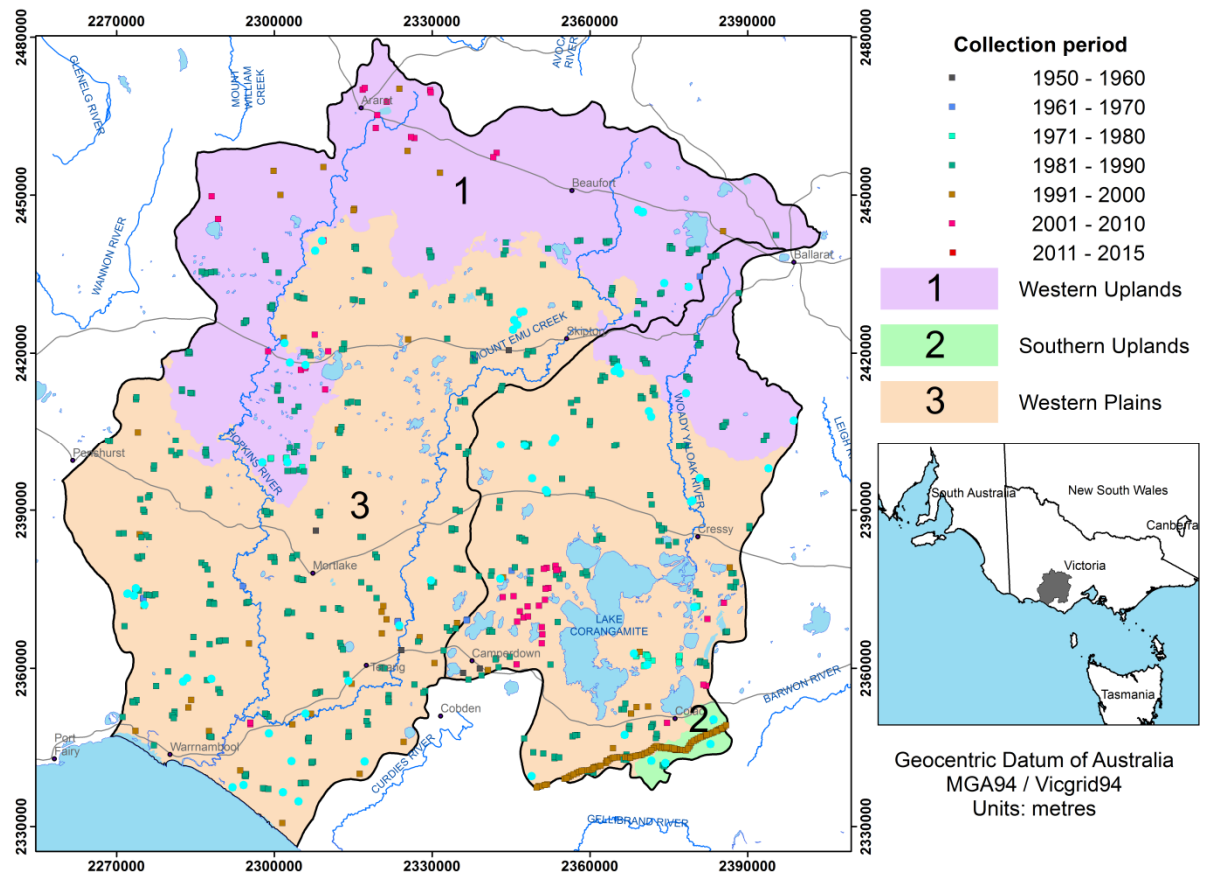


Figure 6.1. Soil sites and their space-time distribution in south-western Victoria.

6.2.3 Soil data

In the study area, 828 sites were sampled between 1957 and 2015 (Figure 6.1, Table 6.1) from 13 soil and land surveys. Sites include surface samples from profile descriptions, monitoring sites and soil fertility samples. Of the 828 sites, 174 (21%) are from paddocks where land use has changed since 2000 (Figure 6.2c) and 126 of these 174 sites were sampled prior to 2000.

Soil pH measurements from methods 4A1 (pH_w) and a modified 4B5 (pH_{wmir}) (Rayment and Lyons, 2011) used in this study are summarised in Table 6.2. Measurements since 2010 for Method 4A1 (equivalent to ISO 10390:2005) were obtained using a Radiometer Analytical SAS titration system comprising a PHM92 pH meter and CDM240 conductivity meter. Between 1992 and 2010 a comparable automated system was used with control samples and a test sample (Shelley, Personal communication) to account for instrument drift (Laslett and McBratney, 1990). Prior to this period, pH was determined using equipment from the same manufacturer with samples left to equilibrate to monitored room conditions prior to analysis. Error is reported as ± 0.1 pH units.

Table 6.1. Soil sites (N) for the collection periods.

Period	N	Period	N
1950-1960	5	1991-2000	144
1961-1970	15	2001-2010	74
1971-1980	22	2011-2015	124
1981-1990	444		

Where pH_w was not observed, Mid-Infrared (MIR) predictions (pH_{wmir}) were included. The MIR diffuse reflectance spectroscopy samples were finely ground to ensure a standardised fine grind particle size distribution ($>95\% < 100 \mu\text{m}$) and scanned using a PerkinElmer Spectrum One Fourier Transform MIR spectrometer at 8 cm^{-1} resolution, from 450 to 7800 cm^{-1} for one minute. Spectra were averaged and background readings collected every 10 samples. Predictions and error estimates were determined using Partial Least Squares Regression (PLSR) from a calibration model with over 11,000 samples.

Model statistics reported include an R^2 of 0.88, R^2 cross validation (CV) of 0.88, Root Mean Squared Error (RMSE) of 0.56 and RMSECV of 0.56.

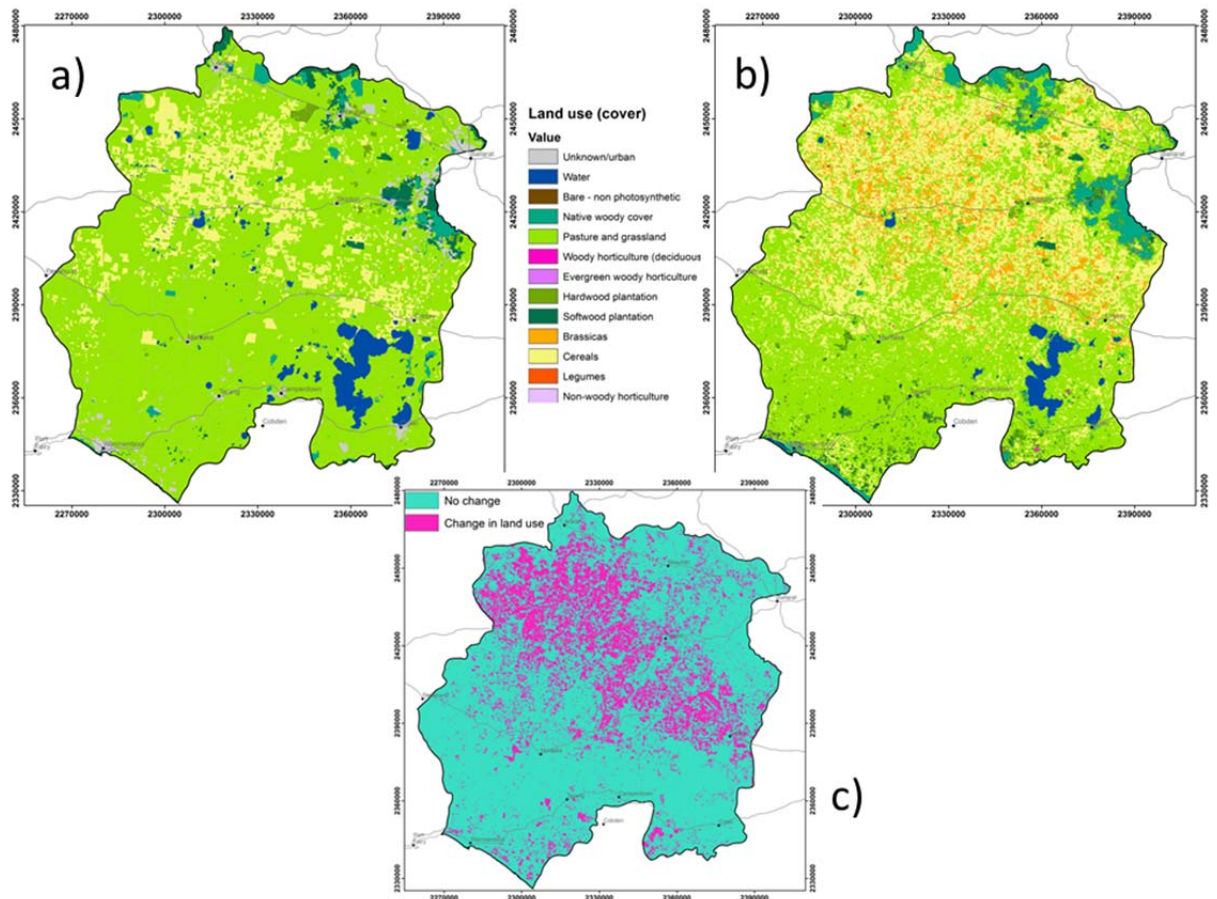


Figure 6.2. (a) Land use in 2000 and (b) 2014 and (c) changes between 2000 and 2014.

Table 6.2. Summary statistics for pH measurements.

Method	N	Minimum	1 st Quantile	Median	Mean	3 rd Quantile	Maximum	St. Dev	Skew
pHw	625	4.3	5.1	5.4	5.52	5.8	10.2	0.68	0.143
pHwmir	203	4.0	5.07	5.53	5.71	6.12	8.71	0.86	0.135

6.2.4 Digital Soil Mapping

McBratney et al. (2003) present a framework for predicting soil properties based on soil forming factors (Jenny 1941) across regions of interest. For each soil forming factor (soil, climate, organisms, relief/topography, parent material, time and spatial position, environmental variables were selected (Table 6.3) as potential fixed effects in the model-based geostatistics. Spatial covariates were projected to GDA94/Vicgrid94 and resampled to a 1000 m resolution using nearest neighbourhood for computational efficiency purposes in the model-based geostatistics.

Model-based geostatistics

To predict pH_W from available environmental spatial covariates and account for important factors in observed pH_W (e.g. time of sampling, sample depth, land use change), model estimation and prediction was performed using a Linear Mixed Model (LMM); see Lark et al. (2006) for further details.

The LMM separates the fixed effects ($\boldsymbol{\beta}$) as the linear model between pH_W and the important explanatory variables from the random effects (\mathbf{u}) which are modelled to identify spatial dependence as the error. The equation also has an error term ($\boldsymbol{\varepsilon}$):

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$$

where \mathbf{y} is the response variable (pH_W), $\boldsymbol{\beta}$ is a vector of unknown fixed effects and \mathbf{X} is a design matrix relating the response variable to those fixed effects; \mathbf{u} is a vector of unknown random effects and the design matrix \mathbf{Z} relates the observations (\mathbf{y}) to the random effects. The error term $\boldsymbol{\varepsilon}$ is a vector of the independent random errors from measurement imprecision or inaccuracy and variation from processes over short distances that are not represented in the sample set (the nugget variance).

Table 6.3. Environmental (spatial) covariates available for model-based geostatistics.

Factor	Variable name	Description	Agency/Source
S (soil)	Victoria land units	Victorian soil type mapping from harmonised legacy surveys with 3,300 land units	Department of Economic Development, Jobs, Transport and Resources (DEDJTR)
	Land use	Tertiary dominant land cover class for 2014.	Department of Economic Development, Jobs, Transport and Resources (DEDJTR); Morse-McNabb et al. (2015)
C (climate)	Mean rainfall (1960-1989)	Average annual rainfall (mm) between 1960 and 1989.	Australian Bureau of Meteorology
	Prescott Index	Prescott Index is an estimate of the water balance including leaching potential from evaporation and precipitation data	CSRIO; Gallant and Austin (2015)
O (organisms)	NDVI_2009	MODIS NDVI 2009 Timesat derivative (maximum amplitude) using a Savitzky-Golay filter.	DEDJTR; Eklundh and Jönsson (2015)
R (relief)	Elevation	Vicmap elevation DTM 20m is at a spatial resolution of 20m and is derived from data of various resolutions, accuracies and ages with increased details in local areas.	Department of Environment, Land, Water and Planning (DELWP)
	Slope	Slope gradient (%) – derived from the DTM 20m	Department of Environment, Land, Water and Planning (DELWP)
	MrVBF	Derived from Elevation – Multi-resolution Valley Bottom Flatness index	Gallant and Dowling (2003)
P (parent material)	Weathering intensity index	Weathering intensity index - degree that primary minerals are altered to secondary clay minerals and oxides.	Geoscience Australia; Wilford (2012)
	GRS-K	Gamma radiometric potassium concentration from natural gamma rays to a depth of approximately 40 cm.	DEDJTR

LMMs for the alternative models were fitted with spatial covariates as fixed effects and the spatial coordinates (covariance structure) as the random effects using the “likfit” function in the geoR package (Ribeiro and Diggle, 2001) which adopted the Maximum Likelihood (ML) procedure/algorithm. The null model was fitted in a similar fashion but only had the mean (no fixed effects) but with spatial coordinates as the random effects like the alternative models. To determine which spatial covariates (fixed effects) to use in the parsimonious model, we fitted each of the fixed term sequentially which was followed by a likelihood ratio tests between nested models to determine whether or not a fixed effect is included at 5% level of significance.

To illustrate the benefits of including spatial covariates and fixed effect factors (e.g. land use change, sample depth and sampling time – season), we fitted two models. Model 1 included no covariates/factors and is a spatial dependence model with a constant mean (the null model). Model 2 included the significant fixed effects found in previous step (parsimonious model) to demonstrate the reduction in prediction error from their inclusion in a model-based design. Spatial predictions on to a common grid (predicted locations) were performed from both sets of models (see below) and outputs were compared spatially.

Fixed effects used in modelling

Some of the major fixed effects used in the LMMs included Sample depth which was a two-level factor where surface samples either corresponded with the depth interval 0 – 10 cm, or not (e.g. were greater than this 10 cm interval; Figure 6.3). Sites were assigned to the four seasons in Victoria (Winter, Spring, Summer and Autumn) according to their sample date (Figure 6.3). Land use was determined from the 2014 land use spatial dataset with two classes assigned – pasture or crop. Land use change was spatially assigned as per the estimated change in land use between 2000 and 2014 (Figure 6.2c). Variance

estimates were also used to compare the two models. As Nelson et al. (2011) determined that uncertainty due to positional error was relatively small in the error budget using model-based approaches, we did not include this error source into the design of our investigation.

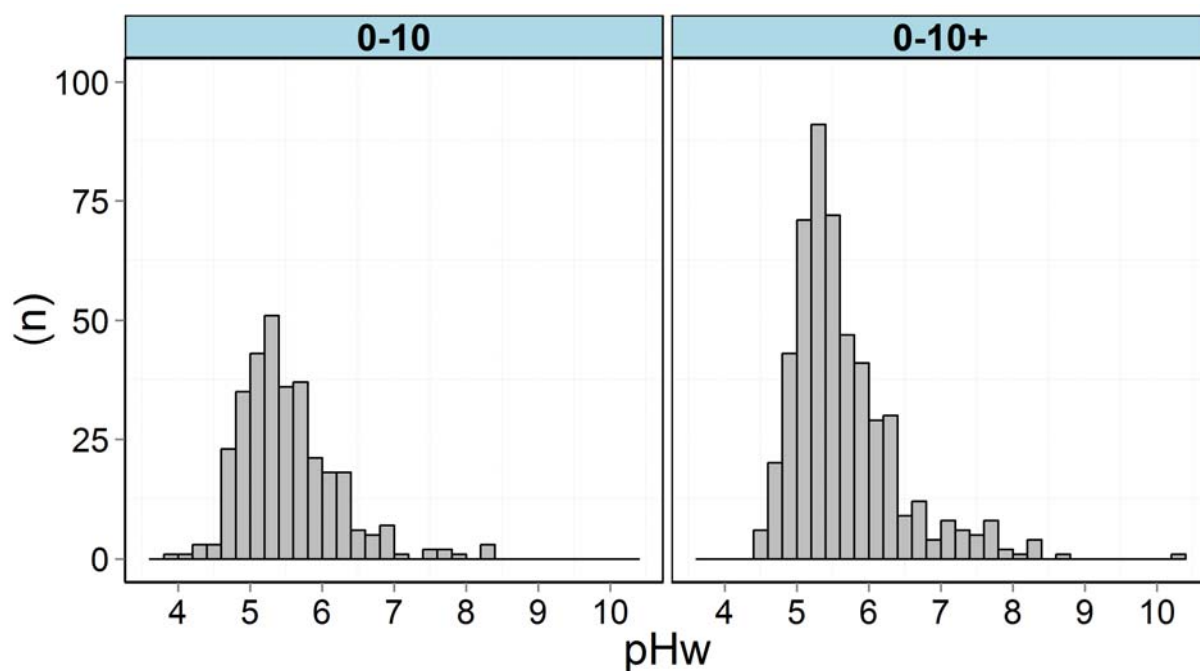


Figure 6.3. pH distribution for sample depth for the two levels (0 – 10 or 0 – 10+ cm).

Probability of pHw being below critical agronomic thresholds

A simulation of 5000 realisations for the conditional spatial distribution of pH was generated and stored at each of the predicted locations. These simulated pH data were then used to approximate the probability that pHw at a known location is above or below a pH value, e.g. 5.3 for perennial pastures and 6.0 for brassicas (see Slattery and Coventry, 1993; Slattery et al., 1995). The approach employed by Lark et al. (2014) adopted the likelihood scale of the Intergovernmental Panel on Climate Change (IPCC) to describe quantified uncertainty using verbal scales (Mastrandrea et al., 2010). These probabilities were then converted to verbal scales and mapped to support users with potential interventions on land use and management. The scheme used in this example

(Table 6.4) has not been translated into unambiguous management outcomes as recommended by Lark et al. (2014), rather the purpose is to highlight the use of probability methods to convey to users the relative uncertainty associated with the map estimates.

Table 6.4. Verbal scale for likelihood (probability) used for pHw scenarios (≤ 5.3 and ≤ 6.0)

Verbal descriptor	Likelihood (probability, %)
Exceptionally unlikely	0 - 1
Very unlikely	1 - 10
Unlikely	10 - 33
About as likely as unlikely	33 - 66
Likely	66 - 90
Very likely	90 - 99
Virtually certain	99 - 100

6.3 Results

6.3.1 Establishing fixed and random effects

The soil pHw LMM included legacy data factors and environmental covariate factors that were determined from the likelihood ratio test (Table 6.5). Legacy data factors included *Sample depth* and *Season* (Spring, Summer, Autumn and Winter). Factors including *pH measurement method* and *Land use change* were excluded as they added little improvement to the model. Environmental covariates included were *Elevation*, *Land use* (pasture or crop) and *NDVI_2009*. The seasonality of pHw differences is apparent in Figure 6.4 for the 0-10 cm sample depth where values are highest in winter and decrease across the following seasons (spring and summer) before increasing slightly in autumn.

This pH trend as a yearly cycle is consistent with findings of Slattery and Ronnfeldt (1992) and Conyers et al. (1997).

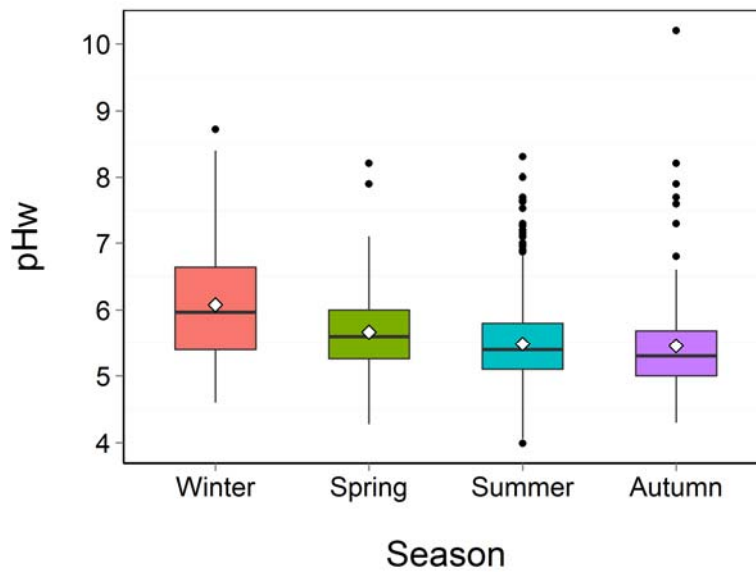


Figure 6.4. Seasonal differences in pHw.

Table 6.5. Summary statistics for environmental covariates and factors (e.g. Sample depth) used in the Likelihood ratio test for the LMM (Model 2).

Fixed effect	p value
Sample depth	0.015
Season	0.029
pH source	0.120
Land use	<0.01
Land use change	0.328
Victoria land units	0.096
Elevation	0.068
Mean rainfall	0.943
MrVBF	0.437
Prescott Index	0.433
Slope	0.225
Weathering Intensity Index	0.133
NDVI_2009	0.007

6.3.2 Spatial prediction of pHw

The prediction of pHw using the constant mean model (Model 1) was compared against the LMM with fixed effects (Model 2) to establish whether fixed effect factors were significant in their contribution to a reduced error. Figure 6.5 shows the two models for 0-10 cm and their associated variance estimates (season is Summer in Model 2). Where the land use was something other than pasture or grain production, no spatial predictions of pH were derived. These maps highlight that the addition of fixed effect factors led to an improvement in the AIC from 1595 to 1499, reduction in the pH prediction error and seasonal departures from a constant mean (Table 6.6). Associated model parameters and highly significant variables are provided as tables in the Appendix (section 6.6).

Table 6.6. Mean soil pH prediction and variance estimates for Model 2; and Model 1 (in brackets).

	Winter	Spring	Summer	Autumn
Prediction	5.48 (5.50)	5.63 (5.50)	5.34 (5.50)	5.57 (5.50)
Variance	0.36 (0.39)	0.36 (0.39)	0.36 (0.39)	0.36 (0.39)

Overall the mean predictions on a seasonal basis for Model 2 varied between -0.02 to 0.28 from the constant mean model (Model 1). The variance estimates for all seasons of Model 2 were smaller than Model 1, but only marginally. This is to be expected as the spatial covariates included in Model 2 result in the predictions being more accurate due to the additional information contained in these variables. The inclusion of fixed effect factors, *Season* and *Sample depth*, is significant and represents a departure from conventional model-based geostatistical approaches used in DSM.

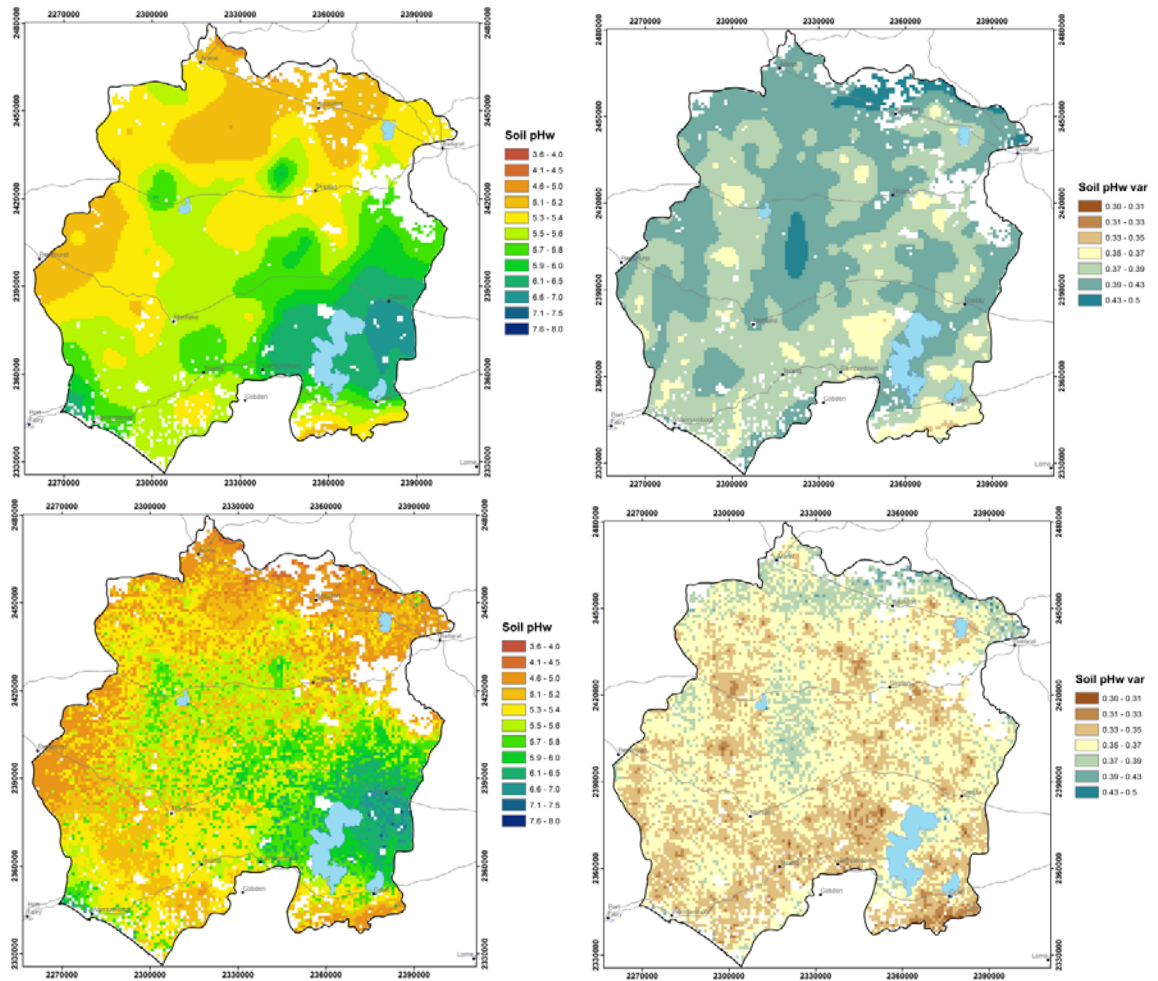


Figure 6.5. Spatial prediction (left) and variance estimates (right) for Model 1 (top) and Model 2 (bottom) for summer.

The seasonal differences in the map predictions for pHw and associated variance estimates defined in the LMM are evident for 0-10 cm (Figure 6.6 and 6.7). Winter and spring both display similar patterns with slight fluctuations in pHw for northern and western parts of the study area. There is a sharp contrast with Summer where there is considerably larger areas of lower pHw in south-western parts with values approaching 5.2 or lower. Autumn has similar patterns to Summer, but, it is evident that pHw values have increased slightly from a low in the summer months.

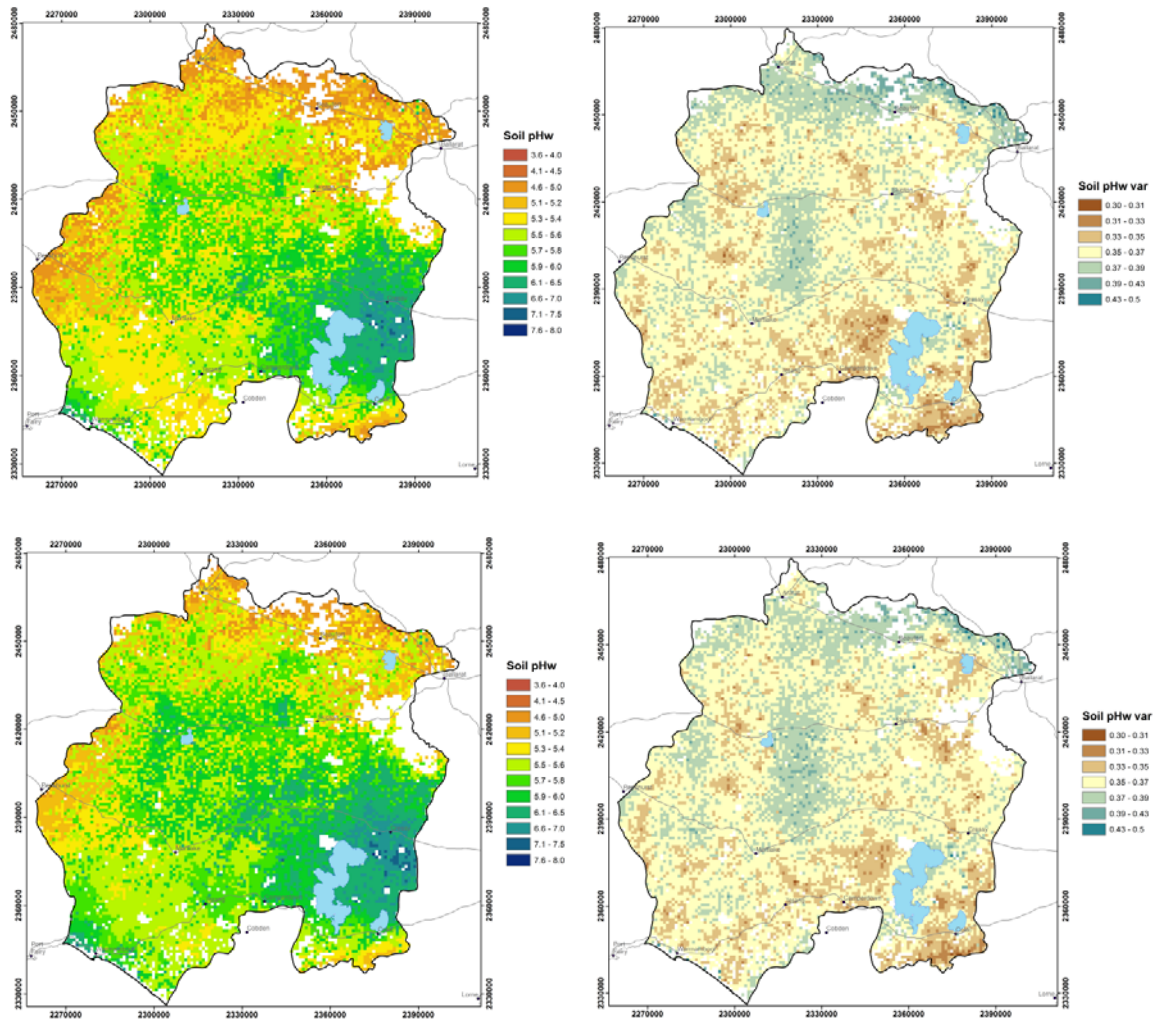


Figure 6.6. Spatial prediction (left) and variance estimates (right) for Winter (top) and Spring (bottom).

6.3.3 Probability of limiting soil pH conditions

The probability maps (Figures 6.8a and b) using the approach recommended by Lark et al. (2014) represent the uncertainty in spatial predictions of soil pHw. Maps have been presented using a red-yellow-blue colour scheme as this has been preferred by users of volcanic hazard maps (Thompson et al. 2015) to reflect a progression from hazard (red hues) to absence of hazard (blue hues). The probabilities have been translated into verbal scales (Table 6.4) to illustrate that a large proportion of the Hopkins and Corangamite

basins have *likely* limiting pHw conditions ($\text{pHw} \leq 6.0$) across seasons for cropping where brassicas and other acid sensitive species (e.g. lucerne) are included in management rotations - Figure 6.8a. Areas surrounding, and to the east of, Lake Corangamite as part of the Western Plains, are designated likelihood classes *unlikely* to *very unlikely* of having a topsoil $\text{pHw} \leq 6.0$. The map with probabilities of pHw being ≤ 5.3 ; Figure 6.8b, defines areas in the Western Uplands near Ararat where there is likely to be limiting agronomic conditions due to acidification and toxicity to plants from aluminium and manganese. Likewise, there are areas to the east of Penshurst that are likely to have limiting pHw conditions for pastures in livestock production enterprises. Acid tolerant varieties including perennial pastures are likely to be impacted at these pH values in these areas. In the south-east, it is unlikely that soil pHw will affect tolerant species except for land in the Heytesbury region, directly south of Lake Corangamite and Colac that abuts the Otway Range (the Southern Uplands). These landscapes have higher rainfall and known acid soils including Kurosols (Robinson et al., 2003).

6.4 Discussion

This study has shown that map users can have information with greater certainty by accounting for potential error sources using the GRUMP framework to make maps based on consideration of factors such as seasonal variability and land use effects. As legacy soil data issues are rarely, if ever addressed in soil maps, the research reported here shows that error sources reducing map certainty can be quantified and accounted for using systematic approaches.

Cyclic seasonal variability in soil pH is recognised (Slattery and Ronnfeldt, 1992; Conyers et al., 1997); however, to our knowledge, this is the first example where time of

year (translated into season) of sampling and temporal variability is accounted for in the creation of a soil pH map. The cyclic pattern of pH values rising from a low in summer and peaking in late autumn to mid-winter when the soil is at its wettest, and then decreasing in late spring to early summer when the soil is drying, is consistent with results from previous trials in north-eastern Victoria (Slattery and Ronnfeldt, 1992). For gross changes in soil acidification to be quantified, variations in soil pH will need to exceed temporal flux and other error sources (e.g. spatial variability) to enable detection.

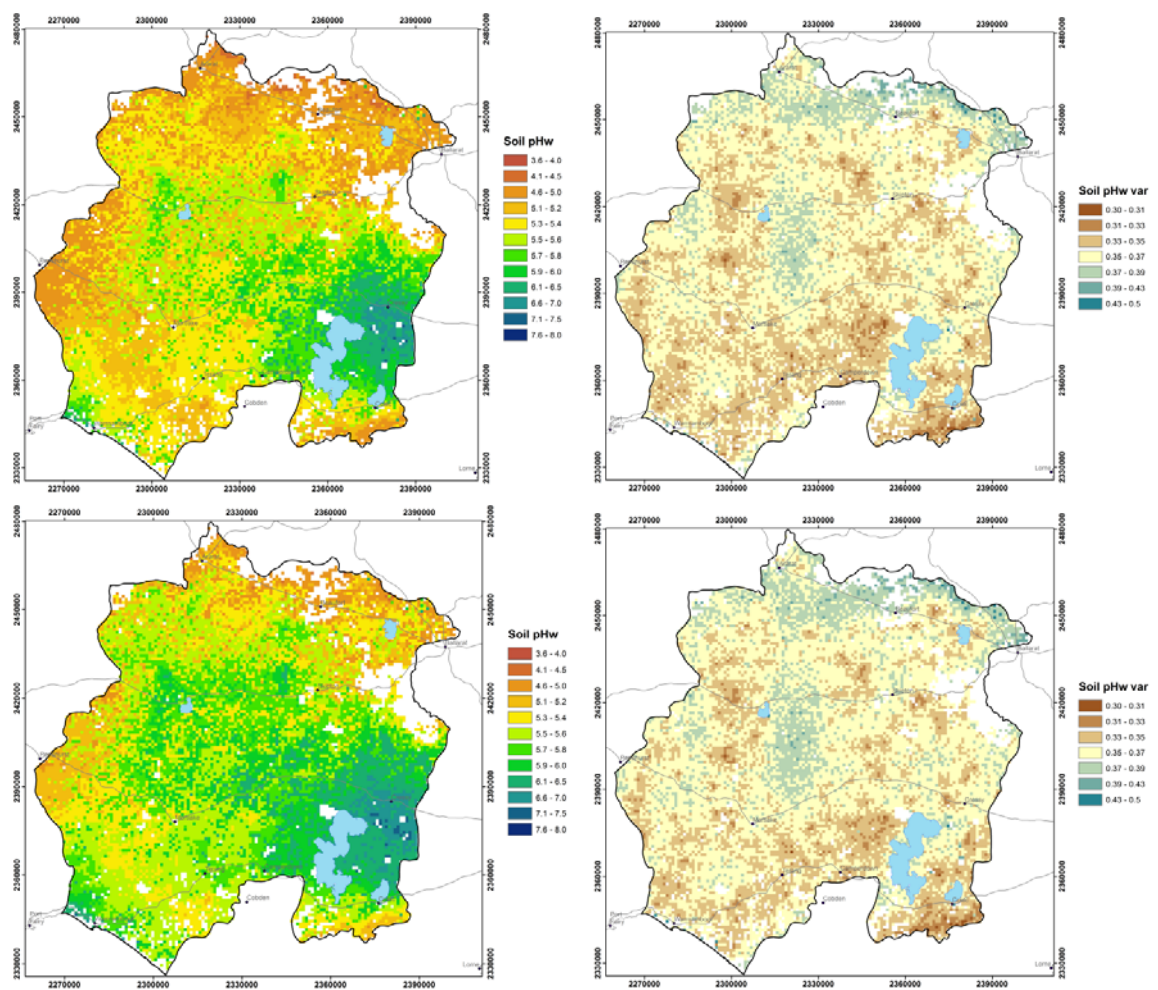


Figure 6.7. Spatial prediction (left) and variance estimates (right) for Summer (top) and Autumn (bottom).

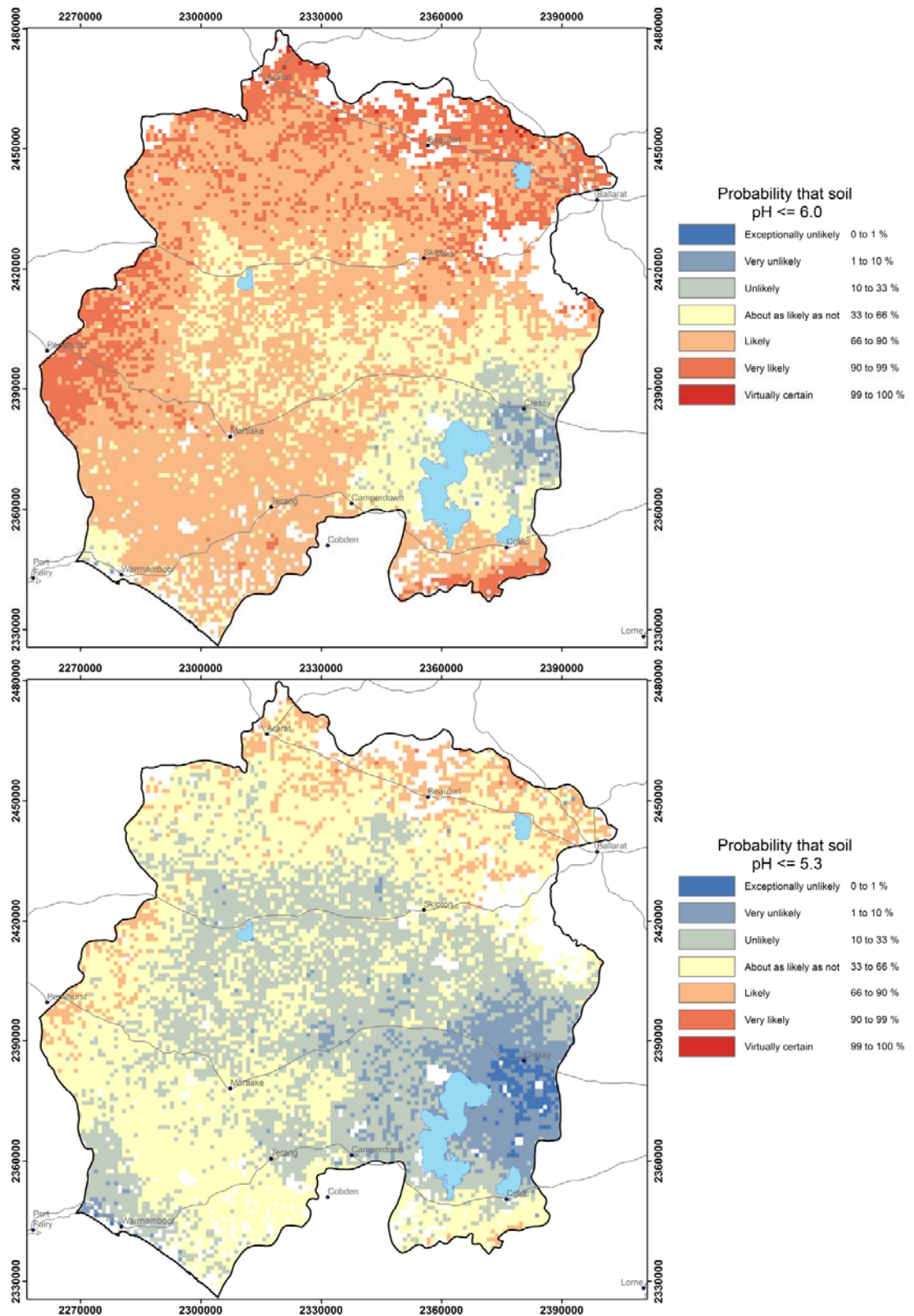


Figure 6.8. (a) Mapped probabilities that pH_w was less than, or equal to 6.0 (top), and (b) 5.3 (bottom) for autumn.

Land use change, while marginally significant, represents a potential factor relevant to changes in soil pH. Changes in pasture composition and a reliance on legume based pasture (e.g. subterranean clover) have been recognised as causing more land to become acid in Australia (Coventry, 1985). In south-western Victoria, transformations in farming systems from those focused on pastures to those of cropping, or forestry, are occurring. Soil acidification has been observed in pastures for this region (Crawford et al., 1994) and conversion to cropping may further accelerate the Net Acid Addition Rate (NAAR). Slattery et al. (1998) collated data from previous research and found that acidification rates for cereal-legume rotations (1.0 to 7.5 kmol (H⁺)/ha per year) were considerably higher than pastures (0.16 and 3.6 kmol (H⁺)/ha per year).

Changes in pH with depth are well known, and from this study different sample depths were found to contribute to differences in soil pH. Variability in soil pH with depth is recognised and is strongly aligned to soil type. Variations within 0 to 10 cm can also be considerable (McLaughlin et al., 1990). From this investigation, dealing with samples derived from multiple depths, there were significant differences in pH observed. Our finding that there was no significant discrepancy between the different pH methods with associated uncertainties in the LMM is consistent with Nelson et al. (2011) where analytical error was found to be relatively minor in the total uncertainty assessment.

The inclusion of environmental covariates as fixed effects representing soil forming factors has improved predictions of soil pH. This was identified by Nelson et al. (2011) as a way of reducing model error. The study area in south-western Victoria is large in comparison to error budget and uncertainty studies (e.g. Nelson et al., 2011; Huang et al., 2015) and includes legacy data with real artefacts and errors that compound to affect spatial prediction of soil pH. By harmonising some of these epistemic error sources, under

guidance from the GRUMP framework, model-based geostatistics have enabled more certain soil pH maps to be created.

The application of probability based schemes with verbal scales to convey uncertainties are recent developments in communication of soil information to users (see Lark et al., 2014; Marchant et al., 2015). This technique has been implemented in this study to convey to land users that areas with limiting soil pH conditions are likely and represent a potential limitation to plant production. By accounting for factors that contribute to uncertainty in model-based approaches, and use of verbal scales to convey uncertainty, we can make maps of greater utility than conventional soil pH maps.

Our argument for maps presented in this paper being better than previous maps is that: (i) conventional maps produced from previous survey programs did not include a measure of uncertainty or error (i.e. we do not know how good they truly are); (ii) national attempts to predict soil pH have been based on exhaustive soil fertility test datasets (e.g. NLWRA 2002) with sites georeferenced to localities and samples purposively collected to inform management; (iii) no analyses to our knowledge have considered seasonal variability and other potential error sources in the cyclic behaviour of soil pH; (iv) the implementation of conditional simulations, together with critical agronomic thresholds and use of verbal uncertainty scales, provides land users with map information of direct applicability.

There are further opportunities to improve the approach reported here. Firstly, model specification could be refined to integrate further epistemic error sources including expert opinion and additional data (e.g. field pH determinations) using probabilistic methods such as Monte Carlo simulation. This can be informed by the error budget approach. Secondly, the implementation of a model ensemble approach that incorporates pre-existing maps and new maps from other techniques (e.g. data mining) should be

considered. Thirdly, we recognise that there remain large deficiencies in adequate soil site representation across temporal and spatial domains for this region. This issue requires further attention if we are to refine and improve soil pH maps, especially as legacy data that is often used in DSM may represent past soil conditions that have been modified by new and different farming systems with different acidification rates. Development of future soil maps should consider these factors through repeated analysis of suitable sites in a purpose-built soil monitoring network with a spatio-temporal statistical design that meets the desired certainty negotiated between the soil scientist, geo-statistician and decision maker.

6.5 Conclusion

There is a continuing need for vigilance in monitoring changes in soil acidification that can have harmful impacts on primary production and the environment. Prediction of soil properties and their spatial distribution, however, is subject to uncertainties related to the accuracy of legacy data. In this paper, a legacy dataset from south western Victoria was used with model-based geostatistics to produce maps of soil pH that addressed a variety of error sources, such as the time of sampling, seasonal variability, analytical method differences, effects of land use change and variability in soil samples. Using a linear mixed model (LMM), significant factors contributing to uncertainty in results were identified and used to produce more informative maps with improved information content for prediction of soil pH. The resulting soil maps displayed uncertainty in soil properties and the probability of being below agronomic critical production thresholds. Using probabilities from conditional simulations in combination with critical thresholds for

production of acid sensitive species, it was possible to define different areas in south-western Victoria that are likely to be below these thresholds.

6.6 Appendix

Parameter estimates and model statistics for the constant mean (Model 1) and LMM (Model 2).

	Model 1	Model 2
Variogram function	Matérn	Matérn
κ (shape parameter)	0.5	0.1
$\hat{\mu}^2$ (constant mean)	5.452	7.364
$\hat{\sigma}^2$ (partial sill)	0.220	0.208
$\hat{\tau}^2$ (nugget variance)	0.325	0.210
$\hat{\phi}$ (range)	0.158	0.391
asympt range	0.474	0.545
AIC	1595	1499
$\log L$	-794	739

Highly significant terms (variables) added sequentially for the LMM.

	df	SS	Wald statistic	Pr (Chi sq)
Intercept	1	1134.24	16110.8	$< 2.2e^{-16}$
Sample depth	1	1.30	18.4	$1.744e^{-05}$
Season	3	1.83	26.0	$9.410e^{-06}$
Land use	1	2.51	35.7	$2.315e^{-09}$
Elevation	1	3.16	45.0	$2.016e^{-11}$
NDVI_2009	1	1.03	14.6	0.0001299
residual		0.07		

References

- Arrouays, D., Marchant, B.P., Saby, N.P.A., Meersmans, J., Orton, T.G., Martin, M.P., Bellamy, P.H., Lark, R.M., Kibblewhite, M., 2012. Generic Issues on Broad-Scale Soil Monitoring Schemes: A Review. *Pedosphere* 22(4), 456-469.
- Benke, K.K., Hamilton, A.J., Lowell, K.E., 2007. Uncertainty analysis and risk assessment in the management of environmental resources. *Australasian Journal of Environmental Management* 14(4), 243-249.
- Conyers, M.K., Uren, N.C., Helyar, K.R., Poile, G.J., Cullis, B.R., 1997. Temporal variation in soil acidity. *Soil Research* 35(5), 1115-1130.
- Coventry, D.R., 1985. Changes in agricultural systems on acid soils in southern Australia, Crop and pasture production - science and practice. Australian Society of Agronomy, Hobart.
- Crawford, D.M., Baker, T.G., Maheswaran, J., 1994. Soil pH changes under Victorian pastures. *Soil Research* 32(1), 105-115.
- Department of Agriculture, 2014. The National Soil Research, Development and Extension Strategy, Securing Australia's Soil, For profitable industries and healthy landscapes. Department of Agriculture: Canberra.
- Diggle, P.J., Ribeiro Jnr, P.J., 2007. Model-based geostatistics. Springer.
- Eklundh, L., Jonsson, P., 2015. TIMESAT 3.2 with parallel processing. Software Manual, Department of Earth and Ecosystem Sciences, Lund University, Sweden, www.nateko.lu.se/TIMESAT.
- FAO, ITPS, 2015. Status of the World's Soil Resources (SWSR) - Technical Summary, Food and Agriculture Organisation of the United Nations and Intergovernmental Technical Panel on Soils, Rome, Italy.
- Gallant, J.C., Austin, J.M., 2015. Derivation of terrain covariates for digital soil mapping in Australia. *Soil Research* 53(8), 895-906.
- Gallant, J.C., Dowling, T.I., 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resources Research* 39(12), 1347.

- Gibbons, F.R., Downes, R.G., 1964. A study of the land in south-western Victoria, Soil Conservation Authority, Victoria.
- Huang, J., Zare, E., Malik, R.S., Triantafilis, J., 2015. An error budget for soil salinity mapping using different ancillary data. *Soil Research* 53(5), 561-575.
- Isbell, R.F., 2002. The Australian Soil Classification. Revised Edition. Australian Soil and Land Survey Handbooks Series 4. CSIRO Publishing, Melbourne.
- Jenny, H., 1941. Factors of soil formation. McGraw-Hill Book company.
- Lark, R.M., Ander, E.L., Cave, M.R., Knights, K.V., Glennon, M.M., Scanlon, R.P., 2014. Mapping trace element deficiency by cokriging from regional geochemical soil data: A case study on cobalt for grazing sheep in Ireland. *Geoderma* 226–227(0), 64-78.
- Lark, R.M., Cullis, B.R., 2004. Model-based analysis using REML for inference from systematically sampled data on soil. *European Journal of Soil Science* 55(4), 799-813.
- Laslett, G.M., McBratney, A.B., 1990. Estimation and implications of instrumental drift, random measurement error and nugget variance of soil attributes—a case study for soil pH. *Journal of Soil Science* 41(3), 451-471.
- Marchant, B.P., Crawford, D.M., Robinson, N.J., 2015. What can legacy datasets tell us about soil quality trends? Soil acidity in Victoria. *IOP Conference Series: Earth and Environmental Science* 25(1), 012015.
- Marchant, B.P., Lark, R.M., 2007. The Matérn variogram model: Implications for uncertainty propagation and sampling in geostatistical surveys. *Geoderma* 140, 337-345.
- Marchant, B.P., Viscarra Rossel, R.A., Webster, R., 2013. Fluctuations in method-of-moments variograms caused by clustered sampling and their elimination by declustering and residual maximum likelihood estimation. *European Journal of Soil Science* 64(4), 401-409.
- Mastrandrea, M.D., Field, C.B., Stocker, T.F., Edenhofer, O., Ebi, K.L., Frame, D.J., Held, H., Kriegler, E., Mach, K.J., Matschoss, P.R., Plattner, G., Yohe, G.W.,

- Zwiers, F.W., 2010. Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties, Intergovernmental Panel on Climate Change, <http://www.ipcc.ch>.
- McBratney, A.B., 1992. On variation, uncertainty and informatics in environmental soil management. *Soil Research* 30(6), 913-935.
- McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. *Geoderma*, 117, 3-52.
- Mclaughlin, M., Baker, T., James, T., Rundle, J., 1990. Distribution and forms of phosphorus and aluminium in acidic topsoils under pastures in south-eastern Australia. *Soil Research* 28(3), 371-385.
- Morse-McNabb, E., Sheffield, K., Clark, R., Lewis, H., Robson, S., Cherry, D., Williams, S., 2015. VLUIS, a land use data product for Victoria, Australia, covering 2006 to 2013. *Scientific Data* 2, 150070.
- Nelson, M.A., Bishop, T.F.A., Triantafilis, J., Odeh, I.O.A., 2011. An error budget for different sources of error in digital soil mapping. *European Journal of Soil Science* 62(3), 417-430.
- NLWRA, 2002. *Australians and Natural Resource Management*. Commonwealth of Australia.
- Raupach, M., 1954. The errors involved in pH determination in soils. *Australian Journal of Agricultural Research* 5(4), 716-729.
- Raupach, M., Tucker, B.M., 1959. The field determination of soil reaction. *Journal of the Australian Institute of Agricultural Science* 25(2), 129-133.
- Rayment, G.E., Lyons, D.J., 2011. *Soil chemical methods: Australasia*, 3. CSIRO publishing, Melbourne.
- Rees, D., Robinson, N., Imhof, M., Joyce, E.B., MacEwan, R., 2010. Spatial frameworks to support digital soil mapping, 2010 19th World Congress of Soil Science, Soil Solutions for a Changing World, Brisbane, Australia, pp. 50-53.

- Ribeiro Jnr, P.J., Diggle, P.J., 2001. geoR: A package for geostatistical analysis. R-NEWS 1(2).
- Robinson, N., Rees, D., Reynard, K., MacEwan, R., Dahlhaus, P., Imhof, M., Boyle, G., Baxter, N., 2003. A land resource assessment of the Corangamite region, Department of Primary Industries, Epsom Vic.
- Robinson, N.J., Benke, K.K., Norng, S., 2015. Identification and interpretation of sources of uncertainty in soils change in a global systems-based modelling process. *Soil Research* 53(6), 592-604.
- Slattery, W., Coventry, D., 1993. Response of wheat, triticale, barley, and canola to lime on four soil types in north-eastern Victoria. *Australian Journal of Experimental Agriculture* 33(5), 609-618.
- Slattery, W., Morrison, G., Coventry, D., 1995. Liming effects on soil exchangeable and soil solution cations of 4 soil types in north-eastern Victoria. *Soil Research* 33(2), 277-295.
- Slattery, W.J., Edwards, D.G., Bell, L.C., Coventry, D.R., Helyar, K.R., 1998. Soil acidification and the carbon cycle in a cropping soil of north-eastern Victoria. *Soil Research* 36(2), 273-290.
- Slattery, W.J., Ronnfeldt, G.R., 1992. Seasonal variation of pH, aluminium, and manganese in acid soils from north-eastern Victoria. *Australian Journal of Experimental Agriculture* 32(8), 1105-1112.
- Uexküll, H.R., Mutert, E., 1995. Global extent, development and economic impact of acid soils. *Plant and Soil* 171(1), 1-15.
- Walker, W.E., Harremoës, P., Rotmans, J., Van der Sluijs, J.P., Van Asselt, M.B.A., Janssen, P., Kreyer Von Krauss, M.P., 2003. Defining Uncertainty, A Conceptual Basis for Uncertainty Management in Model-Based Decision Support. *Integrated Assessment* 4, 5–17.
- White, R.E., 1969. On the measurement of soil pH. *The Journal of the Australasian Institute of Agricultural Science*, 3-14.

Wilford, J., 2012. A weathering intensity index for the Australian continent using airborne gamma-ray spectrometry and digital terrain analysis. *Geoderma* 183-184, 124-142.

Chapter 7 Assessment of error sources in measurements of field pH: effect of operator experience, test kit differences and time-of-day

As highlighted in the literature review (Chapter 3) and Chapter 6, there are significant deficiencies in the spatial and temporal coverage of soil sites to adequately fit spatial inference models (McBratney et al., 2003). However there are large collections of legacy measurements of soil properties from quantitative and qualitative methods that may hold practical value in supplementing disparate soil site datasets. For example field pH as a measurement is open to many uncertainties identified in Chapters 3 and 5, but there are thousands of records available in state, territory and national databases that may hold value in assessments of soil pH for calibration or validation purposes. Research presented in this chapter aims to establish a model for field pH against laboratory pH and how operator experience and field pH test kit influence the quality of predictions. In the absence of laboratory measurements for conventional soil mapping there has been a high dependence on field pH measurements for screening and soil classification purposes. This would suggest little reason to suspect that in the absence of laboratory observations that field pH won't serve as a useful replacement. The research contributes to thesis objectives to:

Understand and account for potential error sources in input soil data to spatial inference systems;

Support the prediction of soil properties linked to soil health, e.g. pH.

Various methods exist to measure soil pH, and while there is general agreement between existing published laboratory and field based methods, the latter are subject to uncertainties including test kit reliability, accuracy, precision and environmental factors. The contribution of this study is to quantify three uncertainties that affect the conversion between field pH and laboratory pH measurements, namely operator experience, choice of test kit and the time-of-day for measurement. Soil samples from western Victoria, representing the pH range 4.5 to 10.0, were used in a randomised complete block design with ten assessors split into two groups representing experienced and inexperienced users. Statistical analysis of laboratory and field pH was based on using the Maximum Likelihood Functional Relationship (MLFR) to determine whether or not there was any bias between the two methods. Significant differences were found between experienced and inexperienced users, and between test kits. The findings of this chapter provide the potential to screen samples and reduce error in input data to Digital Soil Mapping assessments, and apply the confidence and prediction intervals for uncertain data which can inform error propagation analysis in mapping and modelling.

7.1 Introduction

Soil pH is the most frequently measured soil chemical property and provides invaluable background context to understanding chemical, physical and biological interactions and behaviours of soil and regolith with the biosphere and hydrosphere. Not only does pH have a critical role, as the expression of acidity or alkalinity and its impact on the availability and solubility of nutrients, it is also used for soil classification purposes, land use and land capability assessment and for modelling and understanding of agro-ecosystems.

Internationally there are numerous methods used to measure soil pH in field and laboratory environments. In the laboratory, different ratios of soil and water or saline solutions are used. Historically, in Australia, laboratories have measured pH in suspensions of soil and water by shaking one part soil with 5 parts water for one hour (ISO 10390:2005; Method 4A1 in Rayment and Lyons, 2011). To better account for seasonal variability in insoluble salts due to rainfall or management interventions, such as fertilizer addition (White 1969), water was supplemented with a weak salt solution, i.e. 0.01 M CaCl_2 (Method 4B1 in Rayment and Lyons, 2011). Arguably, laboratory pH methods are the most reliable in comparison to field pH procedures; however, field based pH assessment is rapid, inexpensive and results are instantly available to users, such as soil scientists, extension and advisory providers.

Field measurement of soil pH (hereon referred to as field pH) has been in use for 100 years, with methods required to be rapid, accurate, cheap and easily ascertained (Wherry, 1920; Mason and Obenshain, 1939). The sequential development of pH measurement includes methods that added salt solution to the soil (e.g. CaCl_2 or KCl) and those that added water to the soil and observed colour changes of indicators as related to concentrations (Wherry, 1920). As field methods evolved, further comparison studies

were undertaken to assess the usefulness of indicator methods in comparison with standard electrometric laboratory methods (Mason and Obenshain 1939). In Australia, enhancements to the makeup of indicator solution and methodology to apply barium sulphate onto a soil-indicator paste (Raupach, 1950; Raupach and Tucker, 1959) led to the establishment of the colorimetric procedure (Method 4G1 in Rayment and Lyons, (2011) that is still widely used today.

Field pH provides a simple, expedient and reliable approach to measuring pH for soil survey and advisory services at various scales (Raupach and Tucker, 1959; National Committee on Soil and Terrain, 2009). Measurement of field pH using the colorimetric method of Raupach and Tucker (1959) has been undertaken as standard practice in soil and land surveys across Australia for over 60 years. Extensive collections of field pH measurements exist in state, territory and national databases, such as the Victorian Soil Information System (VSIS, Hunter et al., 2010) and Australian Soil Resource Information System (ASRIS, www.asris.csiro.au). Also contained within these databases are less frequent companion sets of laboratory pH observations for pH in 1:5 soil-to-water suspension (hereon referred to as pH_w or lab pH), and with 0.01M $CaCl_2$ extract.

Complementary field and laboratory measurements of soil pH on samples enable comparison of these methods and evaluation of method performance. Comparative studies of various pH measurement modalities have been carried out in the past (Mason and Obenshain, 1939; Steinhardt and Mengel, 1982; Slattery and Ronnfeldt, 1992). It has been demonstrated that there is reasonable agreement between lab pH and field pH, measured from the same soil sample where a single operator was responsible for field measurements (Baker et al., 1983). Steinhardt and Mengel (1982) specifically evaluated the performance of a colorimetric indicator field method against the laboratory method for determining the accuracy of predicting soil pH. However, while the authors identified

some of the potential error sources that result in variation between field and laboratory pH methods, the scope of this and early studies failed to investigate factors affecting the strength of agreement between different methods of measuring pH for extremely acid to alkaline soils.

Globally, there is a current focus on the delivery of digital soil maps (McBratney et al., 2003) exploiting available legacy soil data (Carré et al., 2007) for initiatives such as the GlobalSoilMap project (www.globalsoilmap.net). For many states, territories and nations, significant deficiencies may exist in measured, accessible and available laboratory pH data. As a consequence, there is a potential role for legacy pH observations over geographically widespread areas to complement available laboratory pH data for digital soil mapping purposes (de Caritat et al., 2011; Hopley et al., 2014). The extensive collections of field pH observations in state, territory and national government organisation databases may also be valuable in establishing a baseline of soil condition where design-based monitoring systems are absent.

At present, the documented pH datasets for field pH and lab pH measurements are large, but limited by the numerous confounding error sources that contribute to measurement uncertainty. Some of these unaccounted sources of uncertainty in field pH measurement include:

- assessor (experience level);
- pH test kits (different brands);
- soil characteristics (pH range and value);
- time-of-day (light quality), and
- age of test kit.

From practical field experience in conducting field pH measurements, there are many effects that could potentially bias the relationship between lab pH and field pH. For example, it has been reported that although Australian measurements of the spectral content of daylight have been similar to northern hemisphere measurements, there is a higher level of irradiance in the ultraviolet spectral region (Dixon, 1978). The effect on colour card matching and pH assessment over the course of the day is unknown but there may be bias towards a higher pH reading.

Print quality of colour cards provided by different commercial field pH kits may introduce inaccuracy and uncertainty in pH test kits. The performance of indicator test kits can deteriorate over time due to solvents with aged dyes or impurities (Mason and Obenshain, 1939). Also, batch-to-batch variations in the kit indicators and solvents may introduce perceptible shifts in performance. Very little research has been reported on these effects or on the potential impact of colour interpretation in the field.

The aim of this study is to address this gap in knowledge on sources of uncertainty affecting soil pH determination by investigating how those factors may affect the relationship between field and lab pH and quantifying the potential bias introduced by each factor. Two experiments to account for error sources in both field and laboratory pH using Linear Models and the Maximum Likelihood Functional Relationship (MLFR) as proposed by Ripley and Thomson (1987) were designed to test the following hypotheses:

1. there is a significant assessor effect on the analytical bias between field and lab pH;
2. there is a significant pH level effect, and
3. there is a significant test kit effect.

The effect of light quality was considered as time-of-day and has been used as a blocking factor in this study. The findings from these experiments will provide support for recommendation of a more "robust" measurement methodology of field pH in future applications such as soil surveys and contribute to the harmonization of existing legacy field pH datasets with laboratory pH data used in digital soil mapping and monitoring applications.

7.2 Materials and methods

7.2.1 Materials

Soil samples and laboratory analyses

Samples were selected from over 1800 soil monitoring and reference site samples that were analysed for pH_w between 2011 and 2014. These samples were selected as they correspond with various pH levels represented in commercially available field pH test kits (4.5, 5.0, 5.5 (x2), 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, 9.0, 9.5 and 10.0). The prepared <2 mm samples (Figure 7.1) from sites across western Victoria obtained initial laboratory pH values ± 0.02 of the field kit pH levels. Samples were included from various Soil Orders of the Australian Soil Classification (Isbell 2002) including Chromosols, Dermosols, Sodosols, Calcarosols and Vertosols. Key soil properties, including depth of sample, are presented in Table 7.1.

Laboratory analysis for the experiments was undertaken in triplicate to estimate error in laboratory measurement. Measurements were determined using a Radiometer Analytical (Lyon, France) titration system comprising PHM92 pH meter, CDM240 conductivity meter and SAC950 sample changer. The instrument was calibrated according to the manufacturer's specifications with a reported laboratory precision of $<\pm 0.1$ pH units.

Initial pH_w results were from numerous batches, and as a consequence there is greater batch-to-batch variability in these results in comparison with the second and third measurement that were obtained in the single batch. All batches included two control samples, as recommended, to account for instrument drift (Laslett and McBratney, 1990).

Table 7.1. Site, sampled depth, ASC order and soil properties.

Site	ASC (Isbell 2002)	Depth (cm)	Clay % ¹	E.C. (dS/m) ²	Org. C % ³
USFS_VP100	CH	10-20	33	0.06	1.65
USFS_VP11	SO	0-10	21	0.10	2.30
USFS_VP32	SO	0-10	23	0.15	2.97
USFS_VP36	CH	80-90	51	0.18	0.55
USFS_VP38	SO	40-50	49	0.17	0.74
USFS_VP5	CH	90-100	86	0.13	0.37
USFS_VP66	CH	0-10	22	0.12	2.78
USFS_VP71	CH	90-100	34	0.09	0.38
USFS_VW150	CA	60-70	53	2.43	0.28
USFS_VW55	CA	80-90	28	0.72	0.53
CSMP_89_C1	DE	38-75	41	0.26	10.05
CSMP_100_C1	VE	69-92	34	0.87	3.07
SW22	DE	5-25	10	0.14	9.60

¹ From laboratory or Mid-Infra-Red (MIR) prediction

² Method 3A1 from Rayment and Lyons (2011)

³ Method 6B3 or 6B4 from Rayment and Lyons (2011)



Figure 7.1. Samples used in experiments with pH level.

Psychophysical assessment of field pH

Experimental assessment of field pH using the pH test kit followed the standard protocol for a psychophysical experiment involving human perceptual judgements recorded on a psychometric scale (Benke et al., 1988). Psychophysical measures were in the form of colour assessments using a colour card with 16-step scale for matching colour against treated soil samples for field pH determination. To compare and contrast regular users of the field pH kit such as trained pedologists involved in soil and land survey (Experienced group) against those that may have used a kit irregularly or not at all (Inexperienced group), two groups of assessors (Assessor Type) were selected based on their test results from an online colour-blind test called the Farnsworth-Munsell 100 Hue Colour Vision Test (Farnsworth, 1943). The ten subjects were male and female technical and scientific staff volunteers. All subjects had 20/20 vision wearing their normal correction. Ages of subjects ranged from 35 to 60 years. Each subject carried out three colour assessment sessions (two on the first day at Early and Late afternoon and one on the second day at noon).

Field pH test kits

Two commercially available soil pH test kits were used in this study (referred to as Kit 1 and Kit 2) and were based on the Raupach and Tucker (1959) field pH determination procedure. Both kits used the same assessment procedure where a soil sample (<1 teaspoon) was mixed with the indicator solution until a thick paste was established. The paste is then dusted with BaSO₄ (barium sulphate) powder (used as an optical enhancing agent) and the colour assessed against the colour card after 1 to 2 minutes to find a nearest match.

Time-of-day (light quality)

One of the major influencing factors in colour differentiation is light quality (or lack of it) which reflects the background environmental lighting, glare from the light source and veiling reflection. This is directly influenced by the time-of-day for measurement of field pH. Time-of-day, in the remainder of this paper, will be used interchangeably with light quality. Since time could not be randomised, it was fixed as a blocking factor with two classes: 1PM and 5PM.

It was decided that early and late afternoon (Period) would be good surrogates for good and poor quality light respectively. Both experiments were conducted outside in April 2015 on sunny days with clear blue skies.

7.2.2 Experimental design

Two experiments were conducted to test hypotheses 1 and 2 (Experiment 1) and hypothesis 3 (Experiment 2).

Experiment 1

At each time period, samples were randomly allocated to the 10 assessors for field pH assessment. Each participant was randomly allocated samples of the 13 pH levels to detect if any significant difference in colour differentiation between assessors exists and if differentiation is consistent across the full spectrum of colours (or pH levels). This was phase A of Experiment 1.

Phase B of this experiment involved each assessor completing pH assessments in triplicate on at least 3 pH levels (for example, Assessor 1 might be allocated pH levels 4.0, 6.5 and 9.5 and Assessor 2 might receive pH levels, 4.5, 6.0 and 10, etc.). One assessor in each group (Experienced or Inexperienced) assessed pH levels on four samples to complete the set of measurement errors for each of the pH levels*Assessor

Type. This data was combined with the triplicate lab pH data to provide measurement error estimates on both field pH and lab pH enabling an assessment of potential bias by fitting models that accommodate for errors in both field and lab pH.

The above two phases (A and B) were combined into one single experiment in a full factorial of Assessor*pH level in a randomised complete block design (RCBD), where time-of-day were used to group pH assessments as the blocking factor. Phase B was incorporated using the same design but with an extra randomisation of Assessor to pH level for conducting triplicate field pH measurements. The same randomisation was fixed for 1PM and 5PM for practical reasons, that is, Assessor and pH level pairing were consistent and an extra replication for a better estimate of the Assessor consistency.

Experiment 2

In this experiment, Kit Type, Assessor (and Assessor Type) and pH level were included in a split-plot design where Assessor was used as a blocking factor, pH level was the whole-plot factor and Kit Type was the sub-plot factor respectively.

7.2.3 Statistical analyses

Exploratory analyses

Trellis plots were used to plot data from both experiments in order to explore any potential relationship between variables as a basis to inform further formal statistical modelling. In Experiment 1 (phase A), field pH was plotted against lab pH in panels (Figure 7.2), where each trellis/panel represented each Assessor (A-J). In the same experiment, field pH was again plotted against lab pH in panels, but this time each trellis/panel represented Assessor Type (Experienced and Inexperienced).

In Experiment 2, field pH was again plotted against lab pH in panels, but this time the panels were extended to include a double layer of Assessor and Kit Type where each panel represented a combination of those two factors.

All plots were constructed using the lattice 0.20-31 package (Sarkar 2008) in R and implementing modified codes to accommodate our data structure and visual display requirements. All plots were performed in the R statistical software (R Development Core Team 2015).

Formal analyses

To compare the performance of Assessors in Experiment 1 (phase A) and Kit Type in Experiment 2, a relevant measure was necessary to compare how well an assessor managed to measure the field pH of their allocated samples. The closer the field pH values are to the lab pH values, the higher the precision of the Assessor or Kit Type in determining pH value. An absolute difference between field pH and lab pH was used as the variable of interest.

In Experiment 1 (phase A), the absolute difference was analysed using Analysis of Variance (ANOVA). The treatment structure was specified with fully factorial effects for Assessor Types in full factorial combination with Level (pH levels). The treatment structure was set as Assessor Type* pH level, the blocking structure was specified as Samples nested within Assessor and nested within Period (Period/Assessor/Sample).

To detect potential bias between lab pH and field pH, an estimated measurement error for both methods (field method and lab method) was produced in Experiment 1, phase B. Given that triplicate samples were allocated to both the Experienced and Inexperienced groups on both experimental periods (Early and Late afternoon), it is possible to look at

the potential bias for all combinations and of Assessor Type*Period as well as a combined data (ignoring the groups).

Given that data were available for all combinations of time-of-day and Assessor Type, four scenarios were tested: 1. Experienced and 1PM; 2. Experienced and 5PM; 3. Inexperienced and 1PM; and 4. Inexperienced and 5PM. For each combination and the combined data, two models for field pH and lab pH were fitted: Linear Model (LM) and the Maximum Likelihood Functional Relationship (MLFR). Both models were adapted to test (1) if the intercepts were significantly different from 0, and (2) if the slopes were significantly different from 1, both of which formed the basis for our bias detection.

In Experiment 2, the absolute difference between test kits was analysed using an ANOVA appropriate for a split-plot design. The treatment structure was specified with fully factorial effects for Kit Type in full factorial combination with pH level. This was coded in GenStat as Kit Type*pH level. Assessor was specified as the blocking structure. Residual diagnostics performed in the analysis of Experiment 1 (phase A) were similarly performed here.

In all the ANOVA analyses (for Experiment 1 and 2), residual values were examined graphically to check for distributional normality and constant variance assumptions. Observations with standardised residuals greater than 3.0 were excluded from the analyses. The absolute difference data was square root transformed during analysis to establish normal distribution and constant variance. Least significant differences (5% level) were used to separate the means, subject to significant F-tests.

ANOVA analyses in Experiment 1 (phase A) and Experiment 2 were performed using the GenStat® statistical package (GenStat® Release 16.1, Copyright 2013, VSN International Ltd). The LM model was fitted using modified code based on a built-in LM function. The

MLFR function was written based on the methodology described in Ripley and Thomson (1987). Both functions were implemented using the R statistical package (R Development Core Team 2015).

7.3 Results

7.3.1 Experiment 1

A trellis plot of lab pH versus field pH is shown below in Figure 7.2. Each panel from A-J represents the information for each Assessor. In each panel, lab pH (x-axis) is plotted against field pH (y-axis) with least-squares lines fitted to the data. The fitted model is plotted against the 1:1 line (in red) with slope=1 and intercept=0 for comparison. The estimates for intercept and slope of the LM are printed in each panel, along with the estimated R^2 . Each Assessor produced a different fit for the least-squares model with different intercept and slope estimates. This implies that there were different abilities between assessors to determine pH measurements in the field using a specific field pH kit.

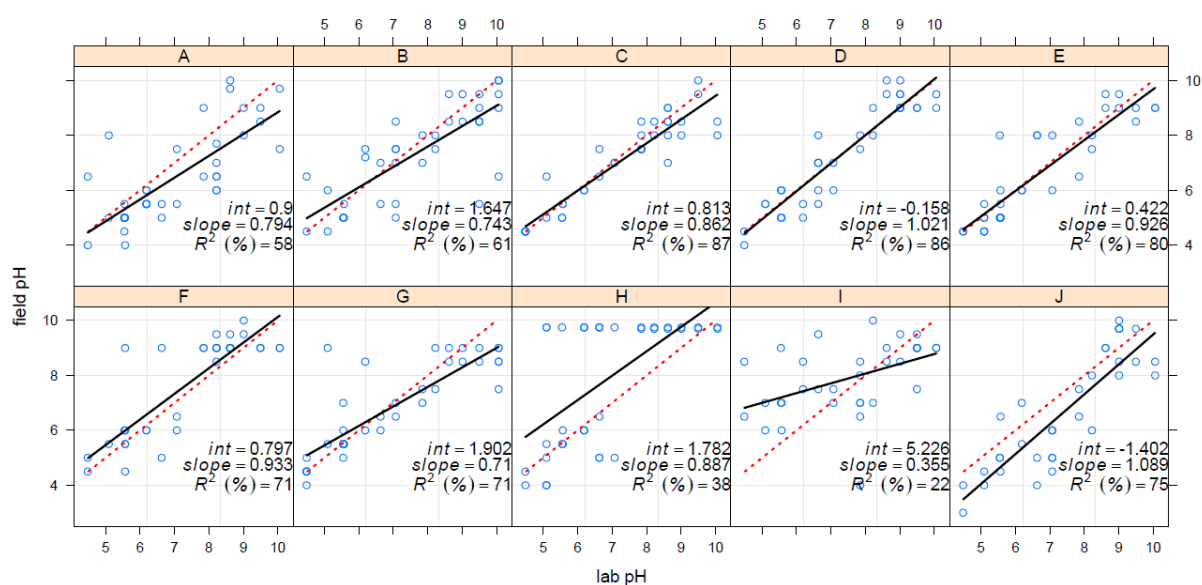


Figure 7.2. Trellis plot of field pH versus lab pH by Assessor.

A trellis plot of lab pH versus field pH (Figure 7.3), where the trellis is either Experienced (Yes) or Inexperienced (No), suggests that the two groups are different. The slope, intercept parameter and estimated R^2 were all different. This implies that a significant difference exists between the experienced and inexperienced Assessors in their ability to conduct soil pH measurements.

The ANOVA results showed that the main effects of Type ($P < 0.05$) and pH Level ($P < 0.001$) were significant but the interaction was not (Table 7.2). This implies that experienced Assessors were able to more accurately determine pH than inexperienced Assessors. The magnitude of error (getting the pH wrong) varied with pH level. It appeared that the degree of difficulties varies from one pH level to the next and this was consistent for all Assessors.

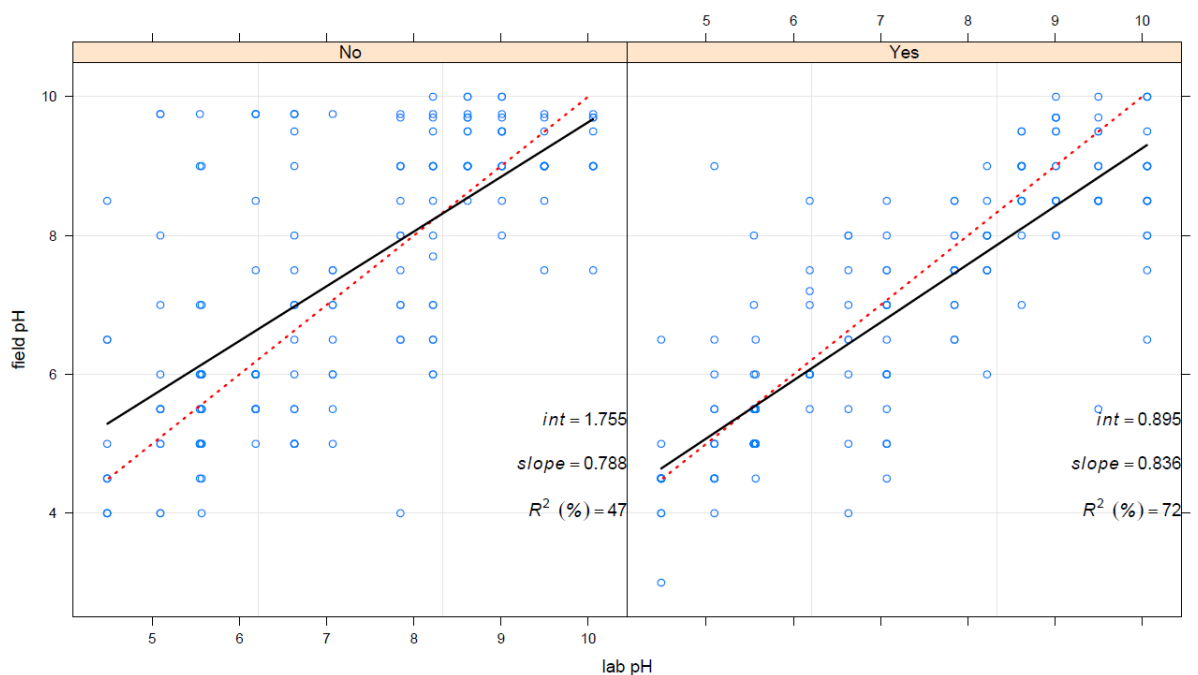


Figure 7.3. Trellis plot of field pH versus lab pH by Experience.

Table 7.2. ANOVA for the absolute difference between lab pH and field pH with two types of assessors and thirteen levels of pH. Mean values (back-transformed mean) are presented.

Factor	Absolute difference
Type of Assessors	
Experienced	0.75 (0.56)
Inexperienced	0.90 (0.81)
LSD(5%)	0.14
pH Levels	
4.5	0.61
5.0	0.83
5.5	0.73
5.5	0.85
6.0	0.85
6.5	1.08
7.0	0.80
7.5	0.88
8.0	0.86
8.5	0.77
9.0	0.70
9.5	0.66
10.0	1.12
LSD(5%)	0.24
F-test probabilities	
Types of Assessor (T)	0.03
pH Levels (L)	<0.001

In phase B of Experiment 1, mean pH and corresponding measurement errors for all the samples using a laboratory pH meter and standard field technique (pH kit) were averaged over all four groups then modelled using the MLFR and LM functions in R. Summary statistics are provided in Table 7.3. For the LM, the intercept or α (1.214) is significantly ($P < 0.05$) different from 0; the slope or β (0.8064) is also significantly ($P < 0.01$) different from 1, signifying that there was a bias between lab pH and field pH in both the intercept

and the slope. For MLFR, the intercept ($\alpha = 0.342$) was marginally ($P < 0.1$) different from 0 and slope ($\beta = 0.9341$) was significantly ($P < 0.05$) different from 1. The MLFR result is much more conservative than the LM as the standard error of the parameter ($\alpha = 0.182$ and standard error for β (0.0269) are tighter (better estimated). However, both methods (MLFR and LM) showed that there was a bias between lab pH and field pH. Figure 7.4 below shows that the LM (red line) and MLFR (green line) deviates from the 1:1 line (black). Both reveal bias with the LM biased at both extremes, whereas MLFR is biased at the high end only (indicating pH is more alkaline).

Table 7.3. Summary of LM and MLFR model parameters between pH data measured in the field (from our experiment) as the response variable (y) and pH data measured in the laboratory as the fixed variable (x).

	LM	MLFR
Intercept (α)	1.2137	0.3424
Standard error of Intercept (s.e.(α))	0.5156	0.1820
Probability $\alpha \neq 0$	0.0186	0.0599
Slope (β)	0.8064	0.9341
Standard error of Intercept (s.e.(β))	0.0695	0.0269
Probability $\beta \neq 1$	0.0053	0.0142

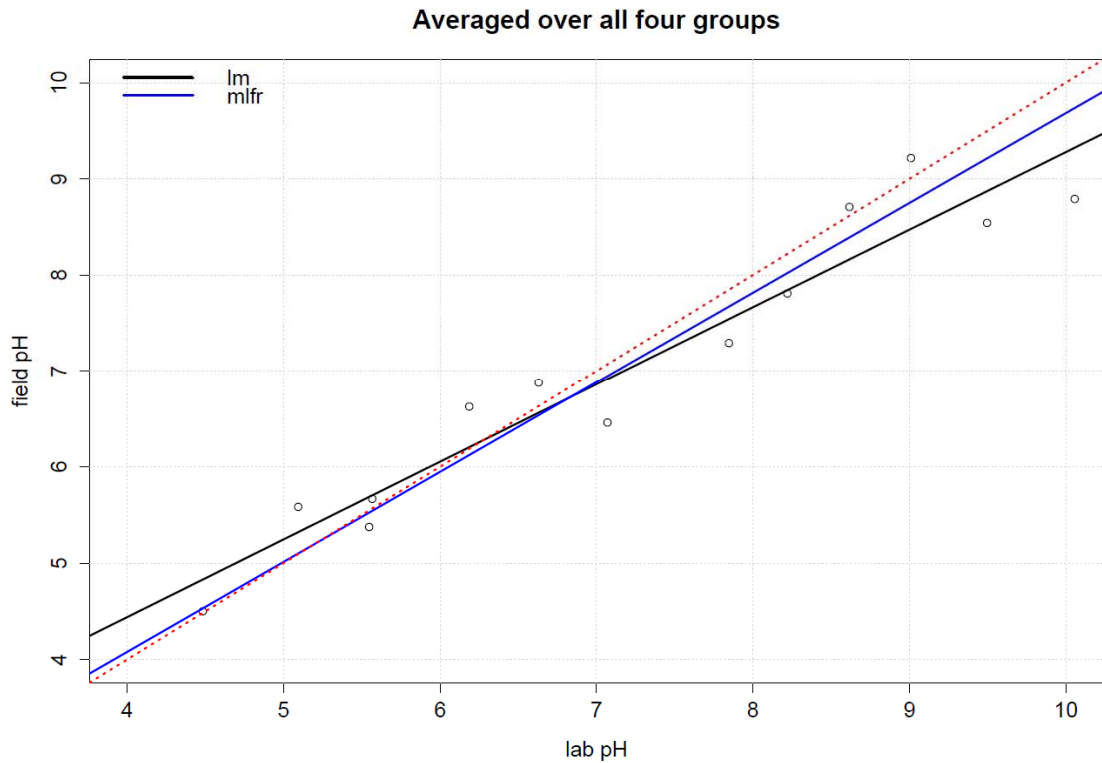


Figure 7.4. Field pH versus lab pH with fitted models using LM (black line) and MLFR (blue line). The red dashed line is the 1:1 line.

In the following analysis, all samples were split based on the combination of Assessor Type by time-of-day. The results of each combination are summarised in Table 7.4 and Figure 7.5. In Table 7.4, only statistics for the MLFR fit are presented as they are more robust. The MLFR parameter fits for the four scenarios of Assessor Type and Light quality showed that the slope and intercept parameters were biased. The LM fits are still shown in Figure 7.5 for comparison. Experienced assessors were positively biased at 1PM ($\alpha = 0.7572$) and 5PM ($\alpha = 0.3477$) with intercepts significantly different from 0 at $P < 0.01$. Inexperienced assessors were also biased at 1PM ($\beta = 1.5476$) and 5 PM ($\beta = 1.4273$) with slopes also significantly different from 1 at $P < 0.01$.

Table 7.4. Summary of MLFR model parameters for field pH (y) and lab pH data (x) for four scenarios: (Case 1) Experienced and 1PM, (Case 2) Experienced and 5PM, (3) Inexperienced and 1PM and (4) Inexperienced and 5PM.

	Case 1	Case 2	Case 3	Case 4
Intercept (α)	0.7572	0.3477	1.5476	1.4273
Standard error of Intercept (s.e.(α))	0.3393	0.2381	0.5273	0.4933
Probability $\alpha \neq 0$	0.0257	0.1443	0.0033	0.0038
Slope (β)	0.8705	0.9178	0.8256	0.8315
Standard error of Intercept (s.e.(β))	0.0430	0.0347	0.0604	0.0580
Probability $\beta \neq 1$	0.0026	0.0177	0.0039	0.0037

7.3.2 Experiment 2

Two trellis plots are presented including field pH versus lab pH by Kit Type and Assessors (Figure 7.6), and field pH versus lab pH by Kit Type and Assessor Type (Figure 7.7). Each panel represents the information for each Assessor by Kit. The parameter (α and β) estimates for each panel are different, indicating a significant effect of Kit and Assessor (experience) for the bias between field pH and lab pH. The Assessor is the same as the panels in Figure 7.2.

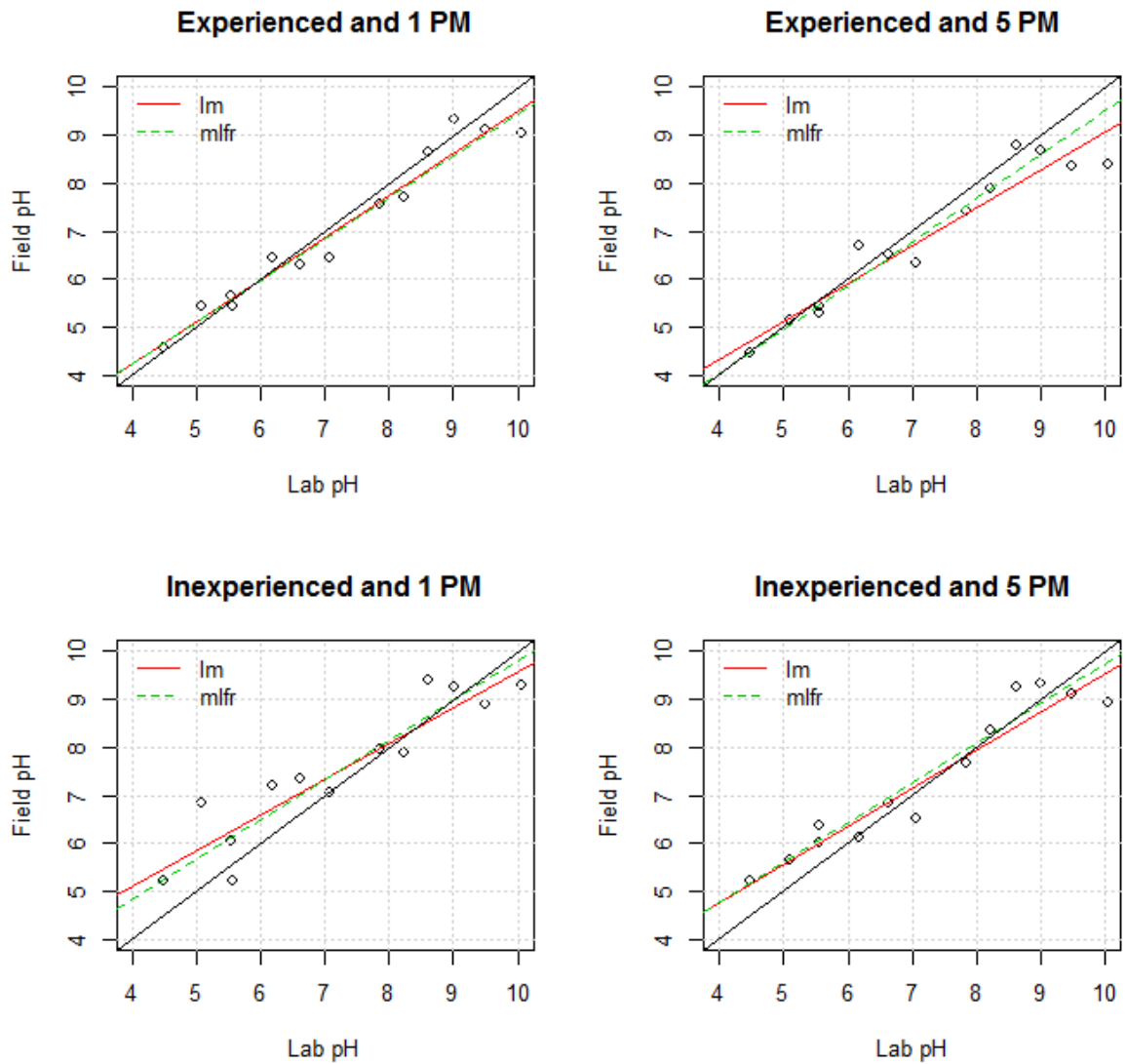


Figure 7.5. Field pH versus lab pH for the four different scenarios - (Case 1) Experienced and 1PM (top left-hand corner), (Case 2) Experienced and 5PM (top right-hand corner), (3) Inexperienced and 1PM (bottom left-hand corner) and (4) Inexperienced and 5PM (bottom right-hand corner). The black line is the 1:1 line.

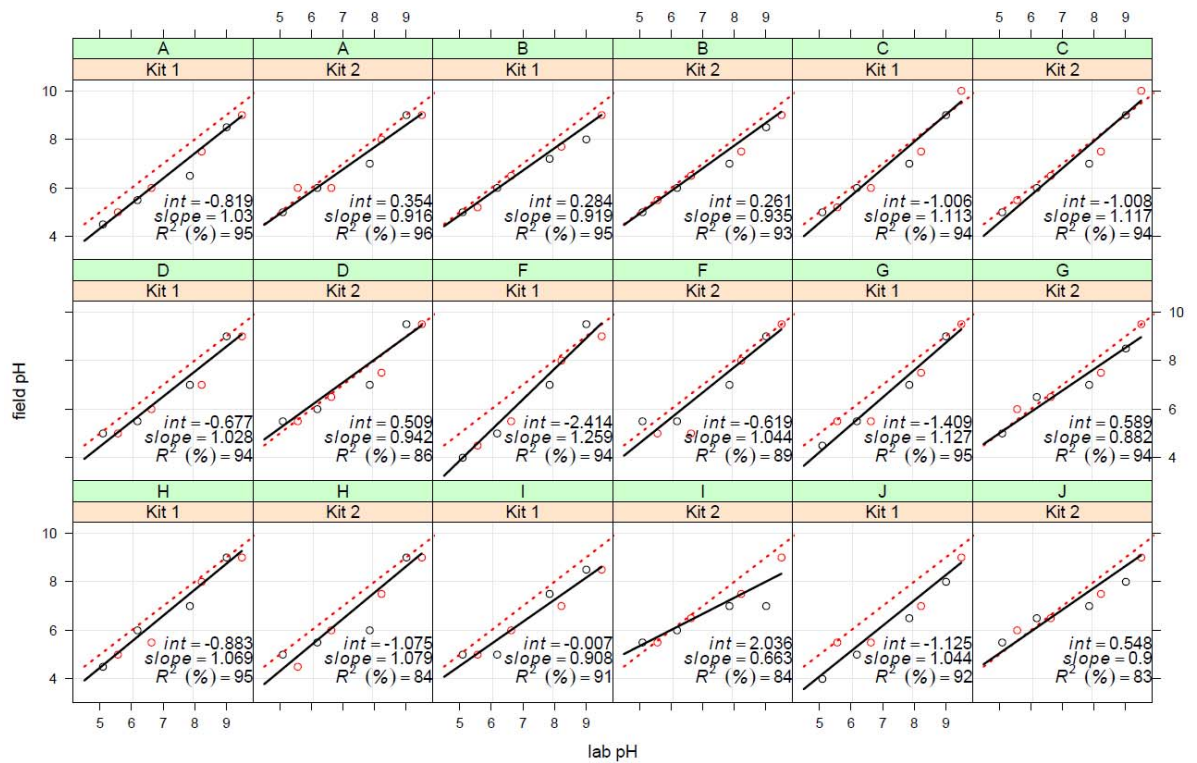


Figure 7.6. Trellis plots of field pH versus lab pH by Kit by Assessor.

Figure 7.6 showed that almost all fitted linear models were below the 1:1 line in the panels, indicating that field pH (measured by assessors) were almost always underestimating the lab pH (assumed to be the true pH) and this was consistent for all assessors regardless of experience. For each Assessor, it is possible to compare between Kits using the paired panels. For example, panel 1 (Assessor A using Kit 1) can be compared with panel 2 (Assessor A using Kit 2). Similarly, panel 3 and 4 can be used to compare Assessor B using Kit 1 and 2 and so on. For each Assessor, a comparison between the fitted linear models to the 1:1 lines by Kit 1 and Kit 2 can be used to determine any potential difference between Assessors, Kits and their interactions. From this, we obtain the following summary:

- Assessor A: performed better using Kit 2;
- Assessor B: no difference;
- Assessor C: no difference;
- Assessor D: performed better using Kit 2;
- Assessor F: performed better using Kit 2;
- Assessor G: performed better using Kit 2;
- Assessor H: no difference;
- Assessor I: no difference, and
- Assessor J: performed better using Kit 2.

In Figure 7.7 we can compare Kit 1 versus Kit 2 as well as Experienced (Yes) versus Inexperienced (No). Looking at all four panels, there was a difference between Kit 1 and Kit 2 where Kit 2 produced results that were closer to the lab results. This was consistent, regardless of the assessors' experience. Both Figures 7.6 and 7.7 indicate that there might be a significant difference between Kit and Assessor but no significant interaction between Assessor and Kit.

The ANOVA results (Table 7.5) identified that the main effects of Kit type ($P < 0.001$) and pH Level ($P < 0.01$) were significant but the interaction was not. The Kit type effect implied that using Kit 2, the assessors were able to obtain more accurate pH measurements than using Kit 1. The magnitude of error (getting the pH wrong) varied with pH levels as in Experiment 1. The degree of difficulties varied from one pH level to the next and this was consistent for both kits (on the whole).

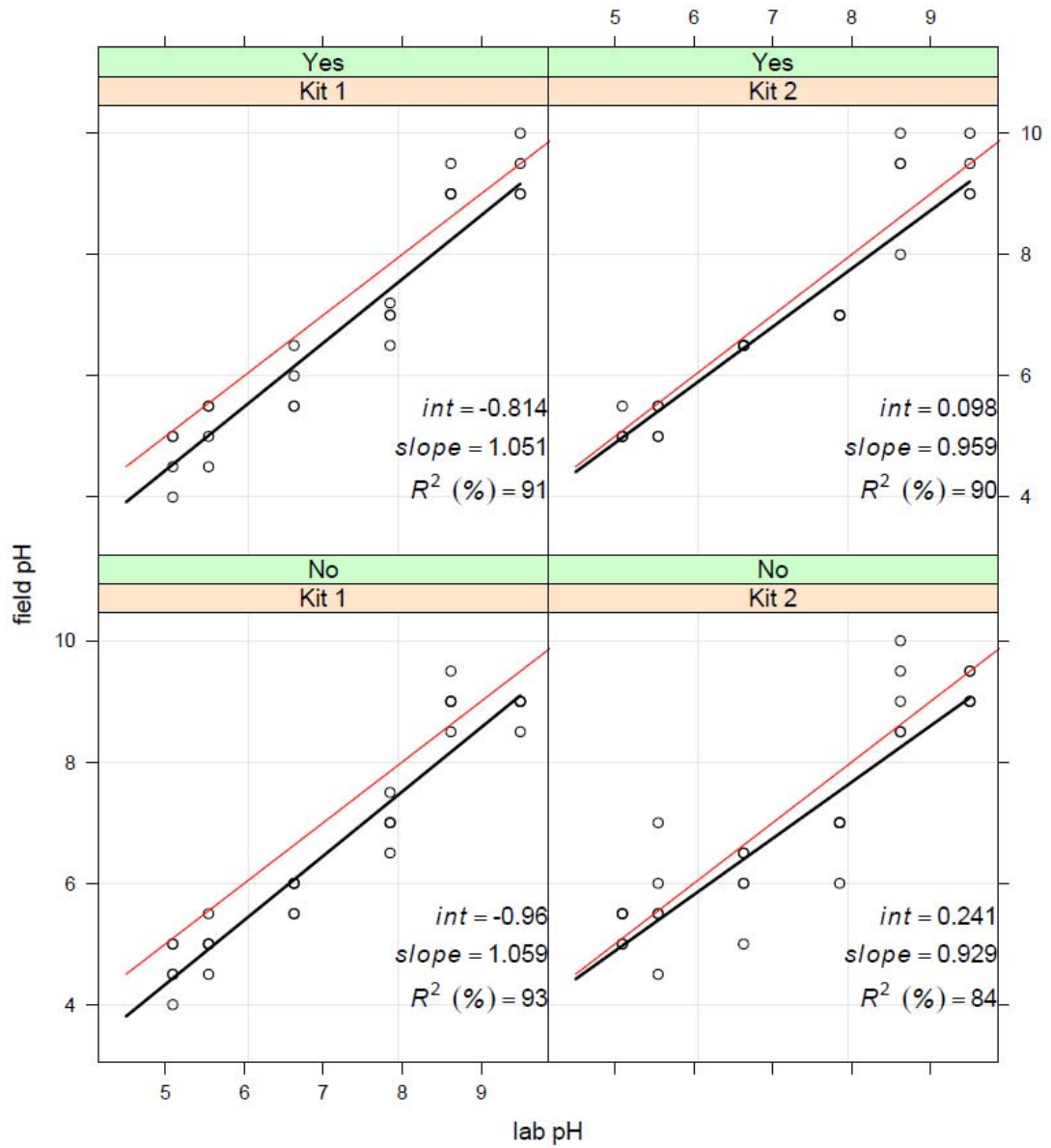


Figure 7.7. Trellis plot of field pH versus lab pH for Kit by Assessor Type.

Table 7.5. ANOVA for the absolute difference between lab pH and field pH with two kits and twelve levels of pH. Mean values are presented.

Factor	Absolute difference
Kit Type	
Kit 1	0.603
Kit 2	0.473
LSD(5%)	0.0924
pH Levels	
4.5	0.372
5.0	0.357
5.5	0.411
5.5	0.481
6.0	0.506
6.5	0.602
7.0	0.653
7.5	0.919
8.0	0.681
8.5	0.602
9.0	0.452
9.5	0.473
LSD(5%)	0.2532
F-test probabilities	
Types of Kits (K)	0.006
pH Levels (L)	<0.001

7.4 Discussion

7.4.1 Assessor experience

This study confirms that pH measured in the field has many potential sources of error, one of which is the experience of the user (Assessor). The experiments highlight that inexperienced field pH assessors under-perform against experienced assessors, and

therefore a greater uncertainty, bias and error with field pH assessments can be expected from inexperienced assessors. This would suggest that for users with limited, or no previous experience using a field pH kit, there is likely to be greater error in the pH determination and therefore greater caution required when using these measurements for decision making, e.g. lime application. This does not account for spatial or temporal variability which are additional sources of uncertainty besides measurement error and epistemic error sources explored in the experimental design. From Experiment 1 to Experiment 2, there is the potential for those with limited to no experience to learn from others that participated in earlier assessments. This is akin to on-the-job training where junior or 'inexperienced' surveyors learn the field determination method under the guidance of an experienced operator with field pH determination. These findings suggest that introductory training and guidance from experienced users in the application of a field pH kit can be extremely beneficial to achieve accuracy and precision in pH determinations. Ongoing quality assurance and control should also be considered as part of regular testing regimes for persons measuring field pH.

While differences between the two assessor groups were evident, there was no clear relationship between the pH level of assessment and the assessor group across the pH range of this study. There were pH levels that were more difficult to assess than others such as pH levels 6.5 and 10. The experiment reveals that the differences in performance are most likely due to the interpretation of the colour card at these pH values rather than the quality or age of the indicator solution or barium sulphate.

There are difficulties in interpretation of the colour graduation on the cards, especially for males which have a deficiency in the red/green region (as evident in results from the Ishihara colour chart). In collated soil site information from soil and land surveys in Victoria contained in the VSIS, there are 51 reported surveyors that have participated in

studies where field pH observations have been collected for 3398 sites. Of the 51 surveyors, only 20% are female. It is unclear how many of the surveyors were properly assessed for vision impairment or were adequately trained for field pH determination, although, often in the field surveyors would cross-reference with one another especially if uncertain on the pH assignment class.

7.4.2 Model and bias

All participants in this study demonstrated different abilities to predict pH using the field determination method. This was reflected in the different bias, error and model fit for every assessor. As there is bias represented in the LM and MLFR models between field pH and lab pH, there is a need for users of such assessments to be prudent as field pH results in this study do not agree perfectly with pH data measured in the laboratory. While both the LM and MLFR display a bias between lab pH and field pH, the MLFR provides a much improved fit than the LM which is biased at both high and low pH values, whereas the MLFR is biased only for high pH values. Further improvements of the MLFR over the LM are evident where the standard error of the model parameters is considerably less than those of the LM.

Baker et al. (1983) and Steinhardt and Mengel (1981) have established quite different results for bias in the relationship between field and laboratory pH measurements. This study also achieved systematic differences (bias) for the different assessors and assessor groups. Strong agreement between field pH and lab pH has been found where one experienced assessor completes all field pH assessments (Baker et al., 1983). From our study the samples were specifically chosen to represent the spread of pH levels represented in the colour cards, but also variations in soil properties such as colour, depth, clay % and organic carbon content that may contribute to error in measurement. This provides a degree of confidence in the agreement between methods being maintained for

a significantly larger sample size. It is unclear, and beyond the scope of this evaluation, if soil colour made a difference to perception of pH level.

7.4.3 Time-of-day (light quality)

Another potential source of uncertainty in field determination of pH is the quality of light. Although this study did not formally test light quality, as we were unable to randomise time in our experimental design, we did assess, using a LM and MLFR, if there were differences between two different times of day that were intended to represent good quality and poor quality light. The results demonstrated that with measurement errors for both the lab pH and field pH, we were able to detect bias in the slopes and intercepts for the four scenarios of time-of-day and assessor group. The MLFR model for the experienced assessor group was better than the inexperienced for 1PM and 5PM. Observation time was not a significant factor for both experienced and inexperienced assessor groups. This is not surprising given that light quality (brightness and glare) for the two times of day of the experiment (1PM and 5PM) was relatively similar. Glare as a light quality factor was noted as an issue in pH assessment by assessors of both groups. On the first day (Experiment 1), the light quality at 5PM was considered as good, if not better than the light quality at 1PM. In our pre-experiment design, we had expected that light quality later in the day would be poorer, but this effect was not observed. It is expected that light quality will be a significant factor for both experienced and inexperienced assessors in future experiments, if we can replicate and quantify true “good” and “poor” light quality in our design.

7.4.4 Kit type differences

A final source of uncertainty considered in this study was difference between commercially available field colorimetric indicator pH kits. Both kits used in this study resulted in underestimated lab pH for all assessors. There were consistently better results

achieved for all assessors regardless of experience for Kit 2 in comparison with Kit 1. The pH levels differences between the kits were inconsistent although there were some pH levels (e.g. 7.5) that had an absolute difference between the two kits close to 1. This could potentially be due to a number of error sources including indicator and BaSO₄ impurities or slight differences that were apparent in the colour cards for the two kits. This requires further investigation as users of field pH kits need confidence in the ability to easily contrast the treated sample with colours represented on the indicator card.

The comparison of kits has highlighted that it is prudent to remove kit type error as a potential source and use one kit type only. Batch to batch variation in kits is potentially a substantial source of error, especially where impurities exist in solvents and reagents, but this was not able to be factored into the experimental design for this study.

7.4.5 General comments

Field pH measurements have been used for soil survey and agricultural advisory work for over 60 years, highlighting the robustness, simplicity and reliability of the procedure. Field determination of soil pH using the Raupach and Tucker (1959) procedure can produce reliable results in comparison to laboratory pH. In particular, field pH determination has provided a role in the screening of samples for potential laboratory analysis, should it be required. A benefit of the current field pH method is that there has been no change to the methodology and chemical constituents since its conception. In contrast, modifications to laboratory techniques over the last 60-years including stirring effects and operator differences are likely to represent sources of uncertainty in legacy pH data greater than currently reported values, e.g. ± 0.1 pH unit. This suggests that as a method for determining soil reaction, it has been an adequate servant for many soil mapping activities over this period.

In the absence of representative laboratory measurements, there is little evidence to suspect that field determinations with greater uncertainty cannot serve as useful replacements for laboratory measurements in spatial and temporal assessments for mapping and monitoring purposes. The findings from this study support the wider use of legacy field pH data for soil mapping purposes at regional to national scales. A mapping technique that could utilise legacy field pH observations in partnership with lab pH is a linear model of coregionalization (LMCR; Webster and Oliver, 2001) using a model-then-calculate, or, calculate-then-model approach described by Orton et al. (2014).

The two experiments reported in this study provide an account of error sources that add to field pH uncertainty. By understanding the nature and magnitude of these errors, we can determine and understand the error bounds represented by the confidence and prediction intervals and provide information on error propagation in mapping and modelling applications. Further investigation to understand the errors in soil survey should be considered to screen legacy soil pH observations prior to use in regional monitoring or mapping applications. The differences found between experienced and inexperienced operators of field pH kits can also be used to guide cleansing of field pH from various sources, such as data from citizen science and crowd sourcing (Rossiter et al., 2015).

Other factors unaccounted for in explaining differences between field and laboratory pH include oxidation effects due to soil storage conditions (Slattery and Burnett, 1992) and incorporation of pedogenic segregations (e.g. calcareous) into the <2mm fraction through differences in sample preparation procedures.

While dealing with legacy data can be problematic due to insufficient metadata to isolate effects due to operator experience, test kit differences and light quality characteristics at time of observation, this should not preclude the capture of new error sources in future. Practical suggestions to increase the certainty in field pH data include: a level of training

to provide assessor certification across the soil pH range; field kits should be regularly tested against known standards, and the test kit should be identified in metadata associated with field measurements. Also to be noted are date and time of observation recorded, and assessor and other factors that may contribute to potential significant differences between field and laboratory measurements (e.g. soil moisture status, observed segregations, depth).

An important consideration when assessing pH in the field or laboratory, or producing maps for planning and land use decision-making, is what is the intended use or purpose of the data. While high precision and accuracy is generally useful, it is often the critical pH ranges relevant to management (e.g. effect plant production, nitrate leaching into groundwater and waterways or corrosion of infrastructure) that are sought. Using the diagnostic pH ranges described by Slattery et al. (1999) as a guide, the critical range of 5.3 to 5.8 is where accurate measurements are most valuable due to the sensitivity of grain and pasture cultivars from the effects of exchangeable manganese and aluminium at these levels. Below this threshold there are implicit and known significant impacts to plant production where remediation actions are necessary. But is high accuracy and precision required here? Likewise, above a pH of 5.8, there are few limitations except where trace elements such as zinc and molybdenum are less available to plants at pH values of 8 and above. Unpublished investigations by the authors identify interquartile range (IQR) values for field pH values 5, 5.5 and 6 against laboratory measurement as 4.9-5.5, 5.1-5.6 and 5.4-6.1. These IQRs suggest that field pH determinations around this diagnostic range are more than just useful indicators especially given that the amount of agricultural land to have pH values in this pH range or below was expected to double to 43-64 million hectares in the coming decade (Dolling et al., 2001).

7.5 Conclusion

Field pH is a useful indicator of soil condition and has practical value for soil pre-screening and rapid classification. Field observations may have additional utility in soil mapping where there is insufficient data available from laboratory pH determinations. While field pH determinations are not as accurate as laboratory measurements, they do provide valuable support for laboratory measurements that are spatially and temporally sparse or biased. This evaluation study of field pH test kits has demonstrated that user experience with a pH test kit will have an impact on the prediction accuracy and uncertainty. This study also confirmed that sources of uncertainty in field pH assessments, such as choice of kit, will affect the accuracy and bias of pH determination in comparison to laboratory measurements.

Using the field colorimetric method, some pH levels at the extreme range were more difficult to determine than others, regardless of assessor experience. There is likely to be bias between field and laboratory measurements and there are distinct benefits from using a kit free from impurities and with a colour card that is consistent with colours expressed in treated samples. Mixing of commercial kits when attempting to harmonise legacy measurements because of differences between the kits may introduce additional uncertainty. The experimental methodology implemented for this study could be modified to accommodate further test subjects and potential error sources, such as within-kit and between-kit variability, or to consider spatial and temporal variability as additional factors. It is recommended that further investigation is pursued on the possible effects of sample size and gender on test kit performance and reliability.

References

- Baker, D.E., Rayment, G.E., Reid, R.E., 1983. Predictive relationships between pH and sodicity in soils of tropical Queensland. *Communications in Soil Science & Plant Analysis* **14**(11), 1063–1073.
- Benke, K.K., Skinner, D.R., Woodruff, C.J., 1988. Convolution operators as a basis for objective correlates of texture perception. *IEEE Transactions on Systems, Man, and Cybernetics* **18**(1), 158–163.
- Carré, F., McBratney, A.B., Mayr, T., Montanarella, L., 2007. Digital Soil assessments: beyond DSM. *Geoderma* **142**, 69–79.
- Dixon, E.R., 1978. Spectral distribution of Australian daylight. *Journal of the Optical Society of America* **68**, 437–450. <http://dx.doi.org/10.1364/JOSA.68.000437>.
- Dolling, P.J., Moody, P., Noble, A., Helyar, K., Hughes, B., Reuter, D., Sparrow, L., 2001. Soil Acidity and Acidification, final report to National Land and Water Resources Audit for Project 5.4C.
- de Caritat, P., Cooper, M., Wilford, J., 2011. The pH of Australian soils: field results from a national survey. *Soil Research* **49**(2), 173–182. doi.org/10.1071/SR10121.
- Farnsworth, D., 1943. The Farnsworth–Munsell 100–Hue and Dichotomous Tests for Color Vision. *Journal of the Optical Society of America* **33**, 568–574. [doi:10.1364/josa.33.000568](http://doi.org/10.1364/josa.33.000568).
- Hopley, J., Rees, D.B., MacEwan, R.J., Clark, R., Benke, K., Imhof, M.P., Bardos, D.C., 2014. A digital soil map of Victoria—VicDSMv1. In ‘GlobalSoilMap: Basis of the global spatial soil information system’ (Eds D Arrouays, N McKenzie, J Hempel, AR de Forges, AB McBratney) pp. 185–190. (CRC Press: London).
- Hunter, D., Williams, S., Robinson, N., 2010. VSIS a new system for Victorian soil data. 19th World Congress of Soil Science, Soil Solutions for a Changing World, 1 – 6 August 2010, Brisbane, Australia.
- Isbell, R.F., 2002. The Australian Soil Classification. Revised Edition. (CSIRO Publishing: Melbourne).

- Laslett, G., McBratney, A.B., 1990. Estimation and implications of instrument drift, random measurement error and nugget variance of soil attributes—a case study for soil pH. *Journal of Soil Science* **41**, 451–471.
- Mason, D., Obenshain, S., 1939. A Comparison of Methods for the Determination of Soil Reaction. *Soil Science Society of America Journal* **3**, 129–137. doi:10.2136/sssaj1939.036159950003000c0027x.
- McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. *Geoderma* **117**(1-2), 3-52.
- R Development Core Team, 2015. R: A language and environment for statistical computing, R version 3.1.3. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3–900051–07–0, www.R–project.org/.
- Raupach, M., 1950. An indicator outfit for soil reaction tests in the field. *Journal of the Australian Institute of Agricultural Science* **16**, 108.
- Raupach, M., Tucker, B.M., 1959. The field determination of soil reaction. *Journal of the Australian Institute of Agricultural Science* **25**(2), 129–133.
- Rayment, G.E., Lyons, D.J., 2011. Soil chemical methods: Australasia (Vol. 3). (CSIRO Publishing: Melbourne).
- Ripley, B.D., Thompson, M., 1987. Regression techniques for the detection of analytical bias. *Analyst* **112**, 377–383.
- Sarkar, D., 2008. Lattice: Multivariate Data Visualization with R. Springer.
- Slattery, W.J., Conyers, M.K., Aitken, R.L., 1999. Soil pH, aluminium, manganese and lime requirement. In ‘Soil analysis, an interpretation manual’. (Eds KI Peverill, LA Sparrow, DJ Reuter) pp. 103-128. (CSIRO Publishing: Melbourne).
- Slattery, W., Ronnfeldt, G., 1992. Seasonal variation of pH, aluminum, and manganese in acid soils from north–eastern Victoria. *Australian Journal of Experimental Agriculture* **32**(8), 1105–1112. doi:10.1071/ea9921105.
- Steinhardt, G.C., Mengel, D.B., 1981. Comparison of pH measurement by field and laboratory methods. *Communications in Soil Science & Plant Analysis* **12**(1), 71–78.

- Wherry, E.T., 1920. Soil Acidity and a Field Method for Its Measurement. *Ecology* **1**(3), 160–173.
- White, R.E., 1969. On the measurement of soil pH. *Journal of the Australian Institute of Agricultural Science* **35**, 3–14.

Chapter 8 The 3D distribution of phyllosilicate clay minerals in western Victoria

Soil properties can serve multiple functions and contribute towards the delivery of many ecological services. Conventional soil survey has focused on properties that are either easy to observe, and interpret for an intended use, e.g. agronomic decision, land evaluation assessment and hydraulic modelling. Although some properties are important to understand the services delivered by soil, they are rarely measured or observed, e.g. hydraulic conductivity. General reasons given for not observing these properties include their expense, they are often time consuming and difficult to obtain (as specified in the Chapter 3). One such property is clay mineralogy which is recognised for its role in carbon turnover and storage, buffering of soil pH and ultimately the chemical behaviour of soils.

This chapter presents a novel approach using legacy clay mineral determinations from X-Ray Diffraction (XRD) in combination with new spectral techniques (Mid Infra-Red Spectroscopy, MIR) and spatial inference systems (Digital Soil Mapping) to map soil clay minerals. These topics have all been discussed in the literature review (Chapter 3) and separately in following chapters (e.g. use of MIR in prediction of pH as a measurement method in Chapter 6). The key objective of this chapter was to:

Produce spatial predictions of soil properties (e.g. clay mineralogy) connected to soil functions supporting agriculture.

The mineralogy of the clay fraction of soils is a major determinant of the behaviour of soil. Conventionally these clay minerals have been determined using techniques such as X-ray Diffraction (XRD), but new complementary methods such as infrared spectroscopy can be used to rapidly and economically predict these minerals. This paper presents a

methodology to predict these clay minerals at high-resolution that adhere to GlobalSoilMap (GSM) standards. Mid-infrared (MIR) spectroscopic models were formulated for clay minerals kaolinite, illite and smectite using partial least squares regression (PLSR) and legacy quantitative XRD determinations. Very strong models were achieved for kaolinite, illite and smectite and the root mean square error of cross validation (RMSECV) were all ≤ 12 wt.%. Spectroscopic models were applied to 11,500 samples from western Victoria and harmonized to the GSM specified depth intervals using equal area splines. Clay minerals were then mapped using data mining model trees with a 10-fold cross validation to derive a mean prediction estimate and 90% prediction interval. Spatial models accounted for 26 to 77% of the total variation with kaolinite predictions for all 6 GSM depths $\geq 65\%$. Kaolinite is the dominant soil clay mineral of western Victoria for uplands and weathered volcanic terrains. Illite concentrations are higher where associated with weathered granitic plutons and in aeolian deposits of semi-arid environments. Smectite tends to occur associated with depressions of plains (volcanic and sedimentary). Further supplementation of additional sites and samples for landscapes with relatively sparse observations should contribute to refined and improved maps of these clay minerals.

The delivery of spatial soil information for clay minerals should support future assessments to quantify and understand the role and distribution of soil functions and services. In combination with examples presented in Chapters 5 and 6, there is the potential to further our knowledge on the resilience and the buffering of soils to changes caused by climatic or management factors linked to acidification and primary production.

8.1 Introduction

The mineralogical composition of the clay fraction ($<2\mu\text{m}$) is a key determinant of soil physical and chemical properties and the regulation of biogeochemical processes. Clay is a generic term for the fine particle size, less than $2\mu\text{m}$ in soil, but the mineralogy of clay and the variability of particles less than $2\mu\text{m}$ is highly diverse depending on the source material of primary minerals, the physical and chemical weathering environment and time (Gilkes, 1990). Primary minerals (generally $>2\mu\text{m}$), and more so secondary minerals that are reactive with their environment (Churchman and Lowe, 2012) support key functions of ecosystem services including the filtering and storage of water, adsorption of soil organic carbon and supply of available nutrients to plants (e.g. potassium), retention of heavy metals as contaminants and providing a physical medium for infrastructure. The clay minerals (phyllosilicates, otherwise known as the layer silicates) comprise a single octahedral alumina sheet linked to either; a tetrahedral sheet of silica (1:1 layer silicate), or sandwiched between two tetrahedral silica sheets as a 2:1 layer silicate. Due to the clay's dominant specific surface area characteristics for interactions with plants, nutrients, metals and organic compounds, the clay fraction is largely responsible for the chemical behaviour of soils (Gilkes, 1990). The importance of phyllosilicate group minerals to organic matter storage and turnover is recognized (Torn et al., 1997; Fontaine et al., 2007; Yuan and Theng, 2012) and emphasized in global efforts to reduce greenhouse gas emissions through soil carbon sequestration (Amundson et al., 2015). In contrast, there has been a noted decline in mineralogical research (Hartemink et al., 2001) and failure to include mineralogy information with spatial modelling and mapping of soil properties (Grunwald, 2009).

8.1.1 Measurement of clay minerals

Crystalline clay minerals have conventionally been characterized and quantified from monochromatic x-rays using X-ray Diffraction (XRD) techniques. Quantification and identification of mineral phases in soils derived from alteration and formation processes (e.g. transformation or neoformation) such as kaolinite, illite, halloysite, smectite and vermiculite has been the mainstay of clay mineralogy determination for over 80 years (Churchman and Lowe, 2012). As a method, XRD determination has improved significantly due to increased sensitivity and reliability of equipment (Gilkes, 1990) and advances in assessment techniques. Four common XRD analytical methods are described by Kahle et al. (2002) including full-pattern methods that trace the entire diffractogram with mean or calculated diffraction patterns (Hughes et al., 1994) and quantify phases using the Rietveld Method (Rietveld, 1969).

A complementary method is infrared spectroscopy (IR), that requires relatively little sample preparation, in contrast to XRD, uses assessment techniques that are quantitative and precise, and analysis is rapid and thus economic (Madejova' and Komadel, 2001; McBratney et al., 2006; Viscarra Rossel 2011; Mulder et al., 2013). An additional benefit of IR is that minerals with poorly crystalline structures (e.g. iron and manganese oxides) can be easily identified from their prominent absorption features, enabling their quantities to be better predicted (Viscarra Rossel et al., 2009). IR spectroscopy is a non-destructive technique that interrogates characteristic molecular bond vibrations that occur in the infrared region of the electromagnetic spectrum. The implementation of Diffuse Reflectance Infrared Fourier Transform (DRIFT) spectroscopy for soils analysis in the visible (VIS), near-infrared (NIR) and mid-infrared (MIR), as summarized by Soriano-Disla et al. (2014), has grown rapidly leading to national (Hicks et al., 2015) and global spectral libraries being developed (Viscarra Rossel, 2009). Authors including Bellon-

Maurel and McBratney (2011), Chang et al. (2001), Janik et al. (1995, 1998, 2007), Janik and Skjemstad (1995), McBratney et al. (2006), Reeves III (2010), Viscarra Rossel et al. (2006), Viscarra Rossel and McBratney (2008), and have published on the application of spectroscopy for soil property determination.

8.1.2 Prediction of clay minerals using IR (VIS, NIR and MIR)

Few IR studies have quantitatively predicted the mineral phases of soil. Janik et al. (1995) found general correspondence between MIR and qualitative XRD mineral estimates from surface samples that were ground to <200 μm . Viscarra Rossel et al. (2009) using VIS-NIR reflectance spectra processed using continuum removal techniques, and Clark and Roush (1984) also achieved good agreement with XRD phase estimates from samples ground to less than 50 μm . Yitagesu et al. (2011) using continuum removed spectra for the 3–5 and 8–14 μm wavelength region for <2 μm achieved useful results for quantifying clay minerals from spectrally distinct bands. Malone et al. (2014a) applied a shape-fitting algorithm to estimate clay mineral abundance using mineral reference spectra and diagnostic wavelengths prior to digital soil mapping. Both Janik et al. (1995) and Viscarra Rossel et al. (2009) used whole soil samples in qualitative XRD analysis; but overall, there has been little research on the prediction of mineral composition for whole soil or separated fractions (e.g. clay) using DRIFT spectroscopy. Furthermore, there is little published information on the application of MIR spectroscopy to quantitatively predict major phyllosilicate minerals including kaolinite, illite and smectite.

8.1.3 Mapping of clay minerals

Viscarra Rossel (2011) highlights a global absence of soil mineralogy maps that would benefit assessments of soil functions supporting ecosystems services. Mineralogical maps based upon soil association mapping for England and Wales from samples characterized for soil clay mineralogy at Rothamsted (now Rothamsted Research) have been derived for

Great Britain (Loveland, et al., 1999), and recent application of VIS and NIR spectroscopy using Digital Soil Mapping (DSM; McBratney et al., 2003) approaches have delivered the first digital maps of soil mineral distribution at national scales (Viscarra Rossel et al., 2010; Viscarra Rossel 2011) and regional scales (Mulder et al., 2013).

The occurrence of clay minerals and their relative abundance are attributed to the five genetic factors of soil formation defined by Jenny (1941): climate, relief, parent material, living organisms and time. These soil forming factors are primary influences on soil and the association of clay mineralogy with other properties, e.g. structure, cation exchange and water characteristics. For clay mineralogy: climate (current and past) affects weathering rate, erosion and deposition of soil; relief often produces localized leaching and weathering effect through interaction with hydrological regimes; parent material provides the host lithology from which primary minerals are inherited, with the weathering sequence acting on it to produce secondary clay minerals; living organisms contribute to dissolution of primary and secondary silicates (Jackson, 1957), the production of biomass and ground cover that shelters soil from erosional events; and time which influenced all the aforementioned soil forming processes. These factors and processes form a pedogenic framework that can be applied to prediction of the occurrence of clay minerals based on environmental correlation principles (McKenzie and Ryan, 1999).

This paper presents an approach to quantify clay mineral abundance using quantitative XRD analysis with MIR spectroscopy to formulate predictive models. This was implemented using an MIR spectral library linked to georeferenced soil sites to map the spatial occurrence and quantity of clay minerals (kaolinite, illite and smectite) in western Victoria, Australia. Spatial covariates used to derive maps according to GlobalSoilMap

specifications (Arrouays et al., 2014) are appraised for their connections with clay mineral distribution and relationship to soil forming factors.

8.2 Methods

8.2.1 Study area

The study area of 135,000 km² (western Victoria, Australia) is characterized by a Mediterranean climate with mean annual rainfall varying between <300 mm in the north to over 2000 mm in the south. Landscapes are diverse, reflecting their geomorphic origins, from marine shoreline deposition, structural faulting and uplift, lacustrine and alluvial deposition, widespread aeolian accession of calcareous loess, periodic volcanic eruption and drainage displacement. The geomorphology has been mapped using a hierarchical classification of landforms and landscapes, known as the Victorian Geomorphology Framework (VGF) (Rees et al., 2010) with five tier-one (Figure 8.1) and twenty tier-two units. The tier-one divisions (North Western Dunefields and Plains, Northern Riverine Plains, Western Uplands, Western Plains and Southern Uplands) serve as a spatial system to classify areas with common processes and land types while simplifying the immense range of geological, landform, climate, soils and vegetation variation encountered (Rees, 2000). Comprising a range of sedimentary, igneous and metamorphic source lithologies, soil types are dominated by calcareous uniform to gradational profiles (Northcote, 1979) or Calcarosols, (Isbell, 2002) in the north to texture contrast (Chromosols, Sodosols and Kurosols) and uniform clays (Vertosols) and sands (Podosols and Tenosols) in the south. Primary agricultural industries include wool, red meat (lamb and beef) and extensive cereal and pulse production across the northern plains

that continue to extend further south into traditional pasture-based farming systems in response to drying conditions of the past two decades.

8.2.2 Soil sites

Soil samples used for MIR spectroscopy were sourced from the Victorian Soil Archive (VSA) (Johnstone et al., 2010) and are georeferenced to sites in the Victorian Soil Information System (VSIS). In total, 2795 sites (11,532 samples) from soil and land surveys of different scales undertaken during the last 80 years by various state and federal government agencies were used (Figure 8.1). Samples with associated clay mineralogy predictions from MIR calibration models for kaolinite, illite and smectite (described in section 8.2.5) were harmonized to the specified depth intervals (0-5, 5-15, 15-30, 30-60, 60-100 and 100-200 cm) of the GlobalSoilMap project (Arrouays et al., 2014) using an equal area spline (Bishop et al., 1999; Malone et al., 2009).

8.2.3 MIR spectra acquisition

All samples had been air dried and sieved to ≤ 2 mm prior to storage in the VSA. A subsample of approximately 20 g was finely ground in a 10 cm steel ring and puck bowl for 60 seconds using a Rocklabs ring mill (Rocklabs, Auckland, NZ) to provide a standardised particle distribution ($>95\%$ <100 μm). Grinding of samples using this procedure to a homogenised finely ground specimen enables more pronounced spectral absorptions of clay minerals to be acquired (Le Guillou et al., 2015) and is in accordance with the current practices used in constructing state and national MIR calibration datasets (Janik et al., 1998; Janik and Skjemstad, 1995).

MIR spectra were acquired using a PerkinElmer Spectrum One Fourier Transform MIR spectrometer equipped with a diffuse reflectance accessory to collect MIR spectra at 8 cm^{-1} resolution, from 450 to 7800 cm^{-1} . Sixty scans were co-added to reduce signal-to-

noise, with a resulting collection time of one minute. A background reading was collected every 10 samples, or every 30 minutes, whichever occurred first. The spectra were transformed from reflectance (R) to apparent absorbance ($A = \log 1/R$) with spectra converted into the GRAMS .spc format. All spectra were collected in a purpose fitted laboratory to minimise effects from temperature, humidity and carbon dioxide.

8.2.4 XRD calibration samples

The source samples selected for development of MIR calibrations for kaolinite, illite and smectite are from 16 soil and land surveys undertaken between 1970 and 2013. 63 samples contained in the VSA with quantitative XRD determinations between 1995 and 2013 were used to formulate preliminary calibration models. An additional 17 samples were sourced from the study by Sultan (2006) and 11 from The Clay Minerals Society (www.clays.org) with accompanying quantitative XRD determinations (Chipera and Bish, 2001). These samples were included to improve the spatial coverage of calibration samples, and also to provide clay mineral extreme values or endmembers (e.g. N95% or b5%) for model development. A further 28 samples were selected from the VSA and quantitative XRD undertaken. This significantly improved the MIR calibration models described in the following section by accounting for samples with erroneous predictions (e.g. high uncertainty or prediction value exceeding 100%) and reducing the overall prediction uncertainty. This provided 117 potential samples for calibration purposes (Figure 8.2). Summary statistics for soil properties associated with these samples are provided in Table 8.1

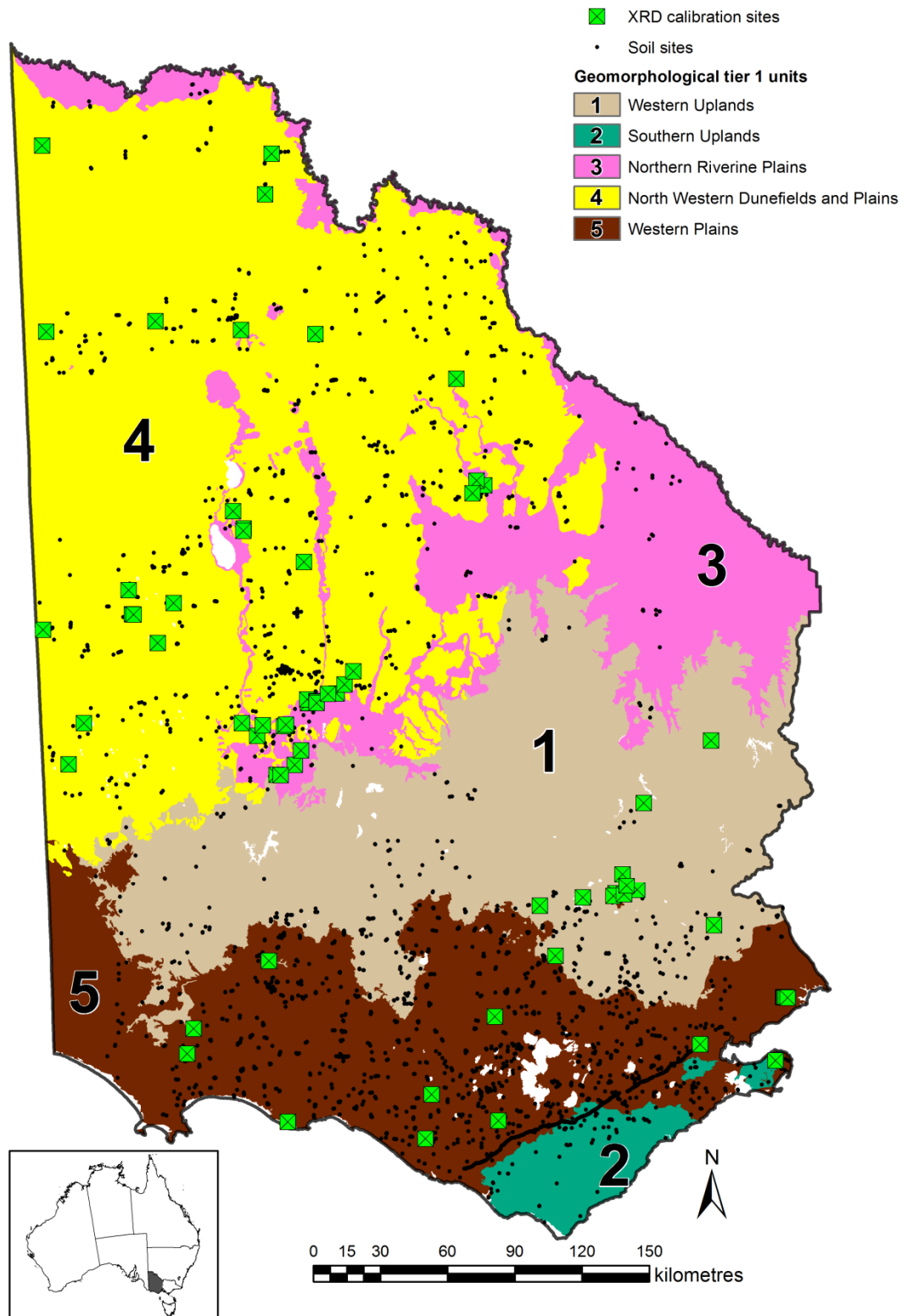


Figure 8.1. Soil sites and quantitative XRD calibration sites against 1st tier geomorphological divisions of the VGF for western Victoria, Australia.

All quantitative XRD analysis was undertaken on the fine earth fraction ($\leq 2 \mu\text{m}$). Initially soil samples that were $\leq 2 \text{ mm}$ were pre-treated either with sodium hexametaphosphate ($(\text{NaPO}_3)_6$) or sodium chloride (NaCl) and deionized water. Mechanically separated samples were made up to a volume (0.5 or 1.0 L) and allowed to settle for 6 to 16 hours before a subsample of the suspension was siphoned off and dried at 105°C . Where ultrasonic dispersion was used (in preference to mechanical dispersion and sedimentation processes), deionized water was added and centrifuged possibly several times until the supernatant was clear. The retained $\leq 2 \mu\text{m}$ fraction was flocculated with NaCl then centrifuged. Iterative treatment using 1M acetic acid (CH_3COOH) and 0.25M CaCl_2 then washing with water and ethanol was undertaken to create Ca-saturated clays for drying.

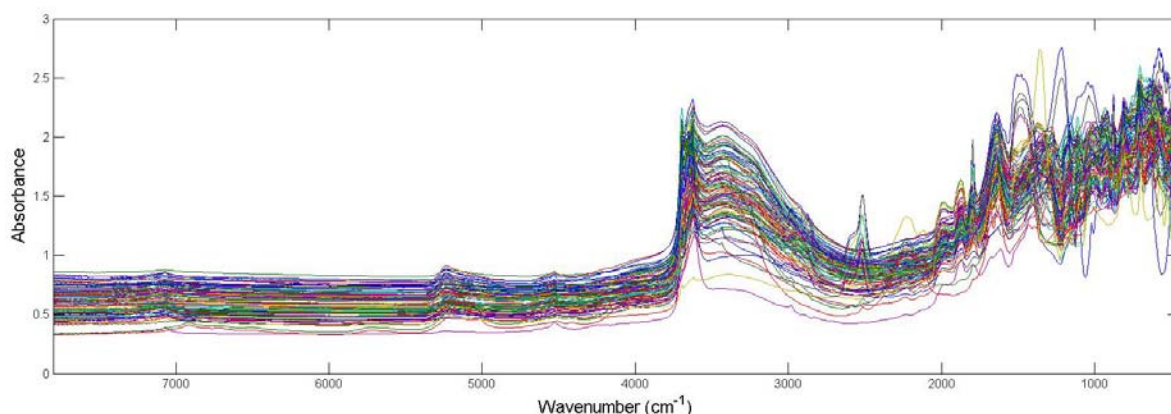


Figure 8.2. MIR spectra of clay mineral calibration samples.

Where samples were mechanically dispersed, samples were saturated with magnesium (1M MgCl_2) and glycerol to orientate clays for XRD analysis.

XRD patterns were collected on instruments including a Phillips PW1800 microprocessor-controlled diffractometer, Siemens D500 or D501 diffractometer, or PANalytical X'Pert Pro diffractometer, all operating at 40 kV with $\text{Co-K}\alpha$ radiation. Scans were collected between 3° and up to 80° in steps of 0.017 to 0.05° (instrument

dependent). Quantitative assessment of the diffraction patterns was usually performed from full-pattern-fitting based on reference standard patterns using software such as SIROQUANT (www.siroquant.com). Quantified mineral phases were normalized to 100% without amorphous components so reported values may be overestimated.

8.2.5 Spectroscopic calibration models

Clay mineral calibrations for kaolinite, illite and smectite were built using partial least squares regression (PLSR) and the SIMPLS algorithm (de Jong, 1993) on a single y-block datum (quantitative XRD data) at a time using the calibration datasets. Leave-one-out cross validation (LOOCV) was used solely to assist identification of outliers in the initial calibration dataset. “Venetian blind” cross validation with ten data splits was used to determine the root mean square error of cross validation (RMSECV). Further outlier assessment procedures included using Mahalanobis distance (Hicks et al., 2015) and reported XRD uncertainties to identify these outlier samples.

Table 8.1. Soil properties for MIR calibration samples.

Property ^a	n	Minimum	1 st Quantile	Median	Mean	3 rd Quantile	Maximum
Clay %	42	2.00	34.25	47.25	45.58	58.00	83.00
EC dS/m ⁻¹	89	0.01	0.07	0.16	0.42	0.31	4.35
pH _w	89	4.5	6.3	7.4	7.49	8.6	10.1
pH _C	73	4.0	5.5	6.6	6.79	8.1	9.1
Exch Ca cmol _C /kg	75	0.10	2.85	5.90	8.41	9.70	41.00
Exch Na cmol _C /kg	75	0.00	0.35	1.00	2.76	3.05	33.00
Exch K cmol _C /kg	75	0.00	0.36	0.80	1.01	1.54	3.50
Exch Mg cmol _C /kg	75	0.10	3.50	5.20	7.10	7.80	42.00
TOC _M	89	0.06	0.46	1.03	1.58	2.05	8.95
Total P	71	0.10	0.20	0.40	0.54	0.70	2.30
Avg. depth	89	0.025	0.100	0.375	0.443	0.650	1.360

n = number of observations

^a Clay % (Mikhail and Briner, 1978); EC (3A1), pH_w (4A1), pH_C (4B2), Exch Ca/Na/K/Mg (15A1/15D3), TOC_M (6B4), Total P (9A) in Rayment and Lyons (2011)

Spectra were scale centred with a zero mean and converted to unit variance, normalized using extended multiplicative scatter correction (Gallagher, 2005) and then transformed using a first derivative polynomial Savitzky Golay smoothing function (Savitzky and Golay, 1964) that was fitted using 15 points surrounding the transformed spectra wavelength.

Calibration models were initially assessed against the number of latent variables (LVs) estimated using the plots of root mean square error of calibration (RMSEC) and RMSECV. A randomization function to determine the optimal number of LVs based on the statistical confidence limit for each individual loading variable (Wiklund et al., 2007), rather than an overall estimate of parsimony as indicated by F tests and predicted residual error sum of squares (PRESS) plots was used. Here the confidence level was set at 95%.

Uncertainty estimates for individual MIR predictions were made with a sample specific standard error of prediction technique for multivariate analysis using PLSR. The constructed model using training samples, where the predictor and predictand are known, was then used to compare against reference samples not used in model development. This provided an estimate of the average prediction uncertainty, otherwise known as the root mean squared error of prediction (RMSEP). Faber and Bro (2002) suggest the accommodation of heteroscedastic errors in the predictors to calculate the variance of the prediction error by:

$$\sigma_{PE} \approx [h(\|\beta\|^2 V_{\Delta x} + V_e + V_{\Delta y}) + \|\beta\|^2 V_{\Delta x} + V_e]^{1/2}$$

where h is the scalar of the unknown sample leverage, $\|\cdot\|$ are the Euclidean norm, $V_{\Delta y}$ is the variance of the measurement error (uncertainty) for the calibration method and $V_{\Delta x}$ is the independently and identically distributed error assumption for $V_{\Delta x}$. The method provides a sample specific RMSEP rather than a standard error of prediction and

compromises by not accounting for bias and selecting less factors in the model. Where bias is negligible, the MSEC can be used to calculate the predictor error by:

$$\sigma_{PE} \approx [(1 + h)MSEC - V_{\Delta y}]^{1/2}$$

Methods including PLS weightings (Wong et al., 2005) and regression coefficients (Haaland and Thomas 1988; Viscarra Rossel et al., 2008) were used to assess the most influential MIR frequencies in the regression model. Variables were assessed for this study using variable importance for projection (VIP, Kvalheim et al., 1994) and the selectivity ratio (SR, Rajalahti et al., 2009). The SR and VIP are determined by:

$$VIP_j = \sqrt{p \sum_{k=1}^h (SS(b_k t_k / (\omega_{jk} / \|\omega_k\|)^2)) / \sum_{k=1}^h (SS(b_k t_k))}$$

where $SS(b_k t_k) = b_k^2 t_k^t t_k$ (1), and:

$$SR_j = \sigma_{ej}^2 / \sigma_{rj}^2$$

where σ_{ej}^2 = explained for each variable, σ_{rj}^2 = residual for each variable.

VIP scores >1 are considered important in the given model (Chong and Jun, 2005) and larger SR scores indicate more useful variables in the model. Lower scoring variables can be excluded without model performance degradation. Both the VIP and SR were used to assess the important frequencies in the PLS models and their relationship with diagnostic absorption features of the MIR spectrum.

The final calibration models were applied to the spectra from the 2795 sites to predict kaolinite, illite and smectite abundance (wt.%).

8.2.6 Mapping clay minerals

Mapping was undertaken with guidance from the SCORPAN framework (McBratney et al., 2003) to predict the relative abundance of clay minerals across the study region with primary factors:

$$S_x = f(s, c, o, r, p, a, n)$$

where S_x is the predicted soil property (e.g. clay mineral), s is soil information from a prior map, remote or proximal sensing, or from expert knowledge, c represents the climate at a point, o is the organisms, r is the topography/landscape attributes, p is parent material, a equals time and n is the spatial position. For this study, the first five factors (s , c , o , r and p) were used in mapping.

Spatial covariates used as environmental predictors were chosen to represent these spatial soil forming factors (Table 8.2). All covariates were resampled to a 50m resolution in the GDA94 Vicgrid94 projection using nearest neighbour interpolation. This was undertaken for computational efficiency and proportional upscaling and downscaling of input variables. Spatial models for the clay minerals were undertaken using the M5 system that constructs tree-based models that can handle high data dimensionality in production of multivariate linear models (Quinlan, 1992). The Cubist package, as implemented in the R environment (Kuhn et al., 2014), implements a tree structure via if/then conditions resulting in regression models at the ‘leaves’ that can be further pruned (reducing input parameters) or combined to reduce the estimated error. These smaller and simpler models enable better prediction and reduced error (Quinlan, 1992).

Map predictions and uncertainty estimates were derived using a 10-fold cross validation with no constraints on the number of rules to be derived from Cubist. Predictions from the 10 models were averaged to produce a mean value with prediction intervals (as per Kidd

et al., 2015) at the 5th and 95th quantiles determined from held-back data for each cross validation (Malone et al., 2014). Diagnostic statistical measures including R^2 , root-mean-square error (RMSE), bias and concordance (ρ_c) using the Lins concordance correlation coefficient (Lin 1989) and percentage of values within the 90% prediction interval were used to assess map outputs.

Table 8.2. Environmental covariates used in Digital Soil Mapping.

Factor	Variable name	Description	Agency/Source
S (soil)	Victoria land units	Victorian Soil type mapping from harmonised legacy surveys with 3,300 land units	Department of Economic Development, Jobs, Transport and Resources (DEDJTR)
C (climate)	Evapotranspiration	Average annual areal actual evapotranspiration.	Australian Bureau of Meteorology
	Pan evaporation	Average annual pan evaporation.	Australian Bureau of Meteorology
	Mean rainfall (1960-1989)	Average annual rainfall (mm) between 1960 and 1989.	Australian Bureau of Meteorology
	Prescott Index	Prescott Index is an estimate of the water balance including leaching potential from evaporation and precipitation data.	CSRIO; Gallant and Austin (2015)
O (organisms)	NDVI 2009	MODIS NDVI 2009 Timesat derivatives (amplitude, base, beginning, large integer, end, left derivative, max, length, right derivative, middle, small integer) using a Savitzky–Golay filter.	DEDJTR; Eklundh and Jönsson (2015)
	NDVI 2011	MODIS NDVI 2011 Timesat derivatives (amplitude, base, beginning, large integer, end, left derivative, max, length, right derivative, middle, small integer) using a Savitzky–Golay filter.	DEDJTR; Eklundh and Jönsson (2015)
	Maxlen205k23k	Maximum growing season length from 16 day time periods (MODIS time-series imagery) over the period 2001-2009.	DEDJTR; Eklundh and Jönsson (2015)
	Smaint201k35k	Small integer (area under the EVI curve, i.e. green vegetation produced during the growing season) of the maximum growth for any growing season between 2001 and 2009.	DEDJTR; Eklundh and Jönsson (2015)
	TM_2004	Landsat TM National Mosaic for 2004 (Bands 1, 2 and 3)	Geoscience Australia
	TM50	Landsat TM National Mosaic (no fixed date) – 4 bands	Geoscience Australia
	Veg_fpar	MODIS vegetation indices for photosynthetic vegetation and non-photosynthetic vegetation (max, min, mean, median, standard deviation)	Geoscience Australia
	Veg_landcover_EVI_trend	Dynamic Land Cover indices for Australia from MODIS (class, max, min, mean)	Geoscience Australia; Lymburner et al. (2011)

Table 8.2 (continued)

Factor	Variable name	Description	Agency/Source
O (organisms)	Veg_persistent_green_veg	Persistent Green Vegetation Fraction from Landsat (2000-2010)	Geoscience Australia
	Lu2005	Land use class in 2005	Department of Environment, Land, Water and Planning (DELWP)
	Luhist2005v2	Land use history classes (1800-2005)	DELWP
	PPL	Broad land use classes (primary production landscapes)	DEDJTR
	SDLC_2014	Secondary Dominant Land Cover (SDLC) from classified MODIS imagery for 2014	DEDJTR
	TDLC_2014	Tertiary Dominant Land Cover (SDLC) from classified MODIS imagery for 2014	DEDJTR
R (relief)	DTM20	Vicmap elevation DTM 20 m is at a spatial resolution of 20 m and is derived from data of various resolutions, accuracies and ages with increased details in local areas.	Department of Environment, Land, Water and Planning (DELWP)
	Aspect	Derived from DTM20 – Orientation for a cell measured in degrees.	DEDJTR
	Flow accumulation	Derived from DTM20 - is the accumulated weight of all cells flowing into each downslope cell.	DEDJTR
	Flow direction	Derived from DTM20 – defines the direction from the cell to its steepest downslope neighbour.	DEDJTR
	MrVBF	Derived from DTM20 – Multi-resolution Valley Bottom Flatness index	Gallant and Dowling (2003)
	Curvature (plan and profile)	Derived from DTM20 – These are curvatures of the land surface. Plan curvature is the curvature of the surface perpendicular to slope and profile curvature is the curvature relative to the slope direction.	DEDJTR
	Slope	Derived from DTM20 – is the rate of maximum change from a cell to its lowest neighbour.	DEDJTR
	Topographic index	Derived from DTM20 – a function calculated from slope and upstream contributing area.	Beven and Kirkby (1979)
P (parent material)	Geology (1:250k)	1:250,000 geological map units	DEDJTR
	Geomorphology (1:250k)	Geomorphology of Victoria (1:250,000)	DEDJTR
	Weathering intensity index	Weathering intensity index - degree that primary minerals are altered to secondary clay minerals and oxides.	Geoscience Australia; Wilford (2012)
	GRS – TDose, K, Th, U and ratios	Gamma radiometric concentrations from natural gamma rays to a depth of approximately 40 cm. Ratios are from combinations of the four principal bands.	DEDJTR

8.3 Results

8.3.1 Clay mineral MIR calibration models

Abundance of the clay minerals kaolinite, illite and smectite, for the calibration samples used in the PLSR derived MIR models, are illustrated in Figure 8.3 and summarized in Table 8.3. Not all samples possess quantitative XRD values for all three layer silicates, therefore the number of samples used in model development were less than the 117 available. This may be due to the dominance of crystalline phases identified and exclusion of interstratified phases, or that these layer silicates were absent or below detection limits in the quantitative XRD determination.

Table 8.3. Summary statistics for samples used for calibration purposes (* number of samples used in PLSR model).

	N	nM*	Min	Med.	Mean	Max	St. dev	Skewness
Kaolinite	107	102	7	36	42.6	99	24.5	0.5
Illite	90	87	2	23.0	26.8	95	18.0	1.6
Smectite	98	94	0.6	30.5	33.8	100	24.9	0.9

Kaolinite was observed in 107 of the calibration samples and ranged from 7 to 99%. Moderate to strong correlation coefficients were found using the Pearson's product-moment statistic with pH_w and pH_c ($r=0.67$), clay % ($r=0.47$) and Total P ($r=0.62$) the most strongly related to kaolinite. Illite calibration samples ($n=90$) were positively skewed and were square root transformed prior to formulating a PLSR model. Samples ranging between 2 and 95% exhibited a weak correlation with TOC_M ($r=0.3$). All other correlations for illite were very weak to negligible. Smectite samples ranged from 0.6 to 100% and were moderately to strongly correlated with pH_w ($r=0.5$) and pH_c ($r=0.64$), exchangeable sodium ($r=0.47$) and Total P ($r=0.45$).

The multivariate PLSR model for kaolinite incorporated 102 samples with 5 samples excluded when assessed against the reference method and spectral residuals criteria (Figure 8.4). The RMSE for the calibration set was 5.41% and a R^2 of 0.95. A small negative bias (-0.06) was found and the cross validation statistics reveal a RMSE of 11.06% and R^2 of 0.77. Seven latent variables (factors) were included in the model with a ratio of percentage deviation (RPD) of 2.2 and a ρ_c of 0.962. There was no noticeable benefit from implementing a Support Vector Machine (SVM) for kaolinite (results not included) in preference to PLSR. For the illite model, 87 samples were used (3 excluded) with values square root transformed to account for skewness in the calibration data. Illite abundances for samples from Victoria in the calibration set were generally <45% with only 4 samples occurring between 45 and 75%. The two Clay Mineral Society samples of 89 and 95% provided samples at the high end for calibration purposes. A RMSE of 3.45% and R^2 of 0.96 was achieved for the calibration model with a RMSE and R^2 for the cross validation of 10.21% and 0.69. The bias was again small (-0.26) with a modest calibration RPD of 1.8 and a ρ_c of 0.969 from seven latent variables applied in the PLSR model. Again the SVM improvements were marginal and are not reported.

The smectite model included 94 samples, with 4 samples removed using the outlier criteria. The majority of smectite abundances in the calibration set were below 50% (n=75) with four samples above 70% included. The calibration model results include a RMSE of 5.77%, R^2 of 0.94, cross validation RMSE of 11.87% and R^2 of 0.76 for cross validation. Like kaolinite and illite, the negative bias was small (-0.28), with seven latent variables included in the model, a RPD of 2.1 and ρ_c of 0.961 achieved.

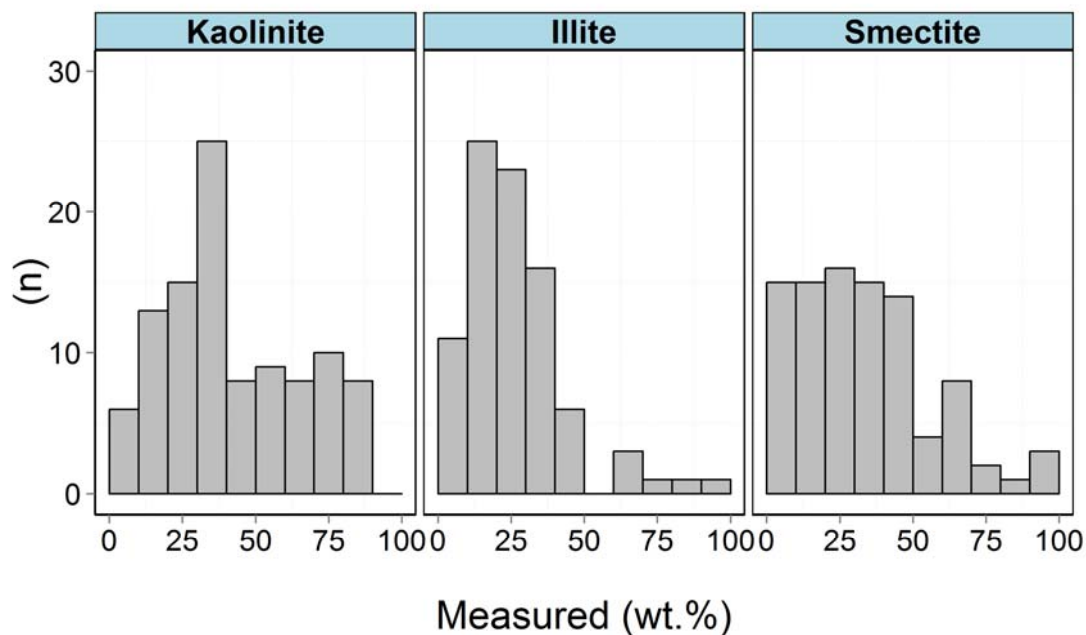


Figure 8.3. Histogram of clay mineral (kaolinite, illite and smectite) content of samples used in PLSR calibration models.

8.3.2 Key absorption features of clay minerals (kaolinite, illite, smectite)

The key variables to the layer silicate calibration models varied among the three clay minerals (Figure 8.5). For kaolinite, there was general correspondence between high VIP and SR scores for ranges between 930-1120, 3060-3400, 3600-3950, and 5230-5440 and 7000-7200 cm^{-1} . Alumino-silicate vibrations at 1020 cm^{-1} (Nguyen et al., 1991) and hydroxyl stretching at 3620 and 7100 cm^{-1} (well-ordered double or triplet OH feature depending upon crystallinity) are attributed to kaolinite. The hydroxyl stretching vibrations are considered to be diagnostic for kaolinite and a better diagnostic than using XRD (Clark 1999).

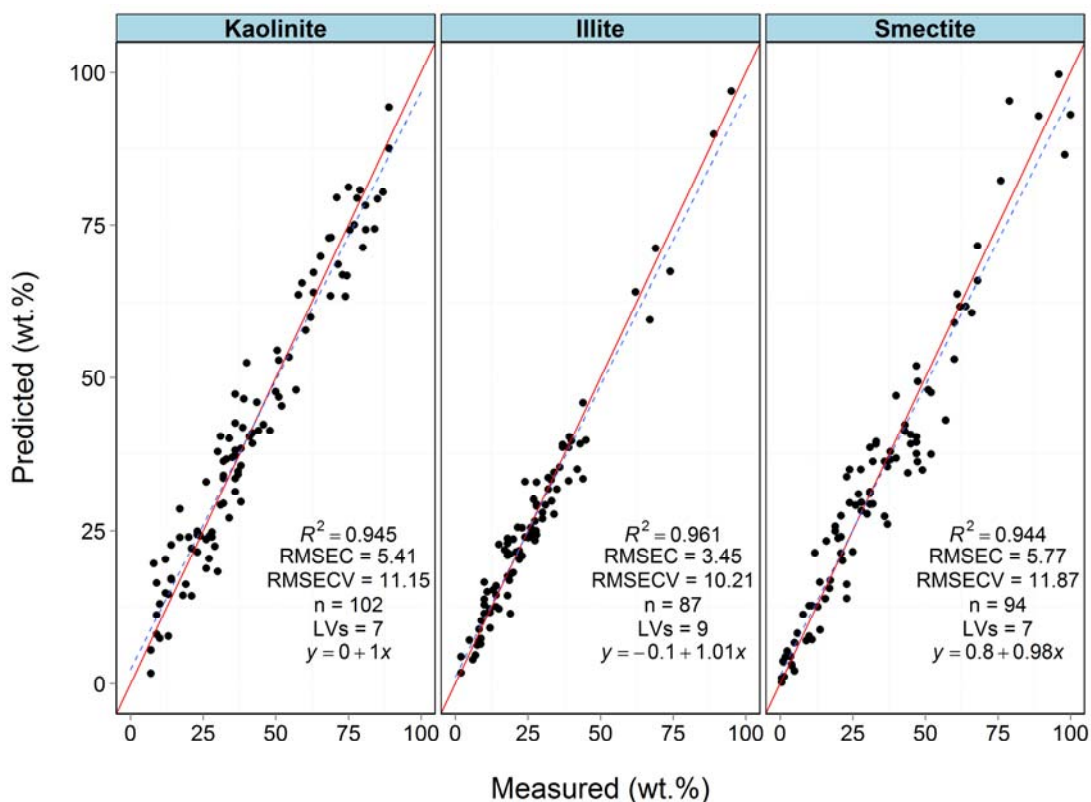


Figure 8.4. Clay mineral calibration models for kaolinite, illite and smectite.

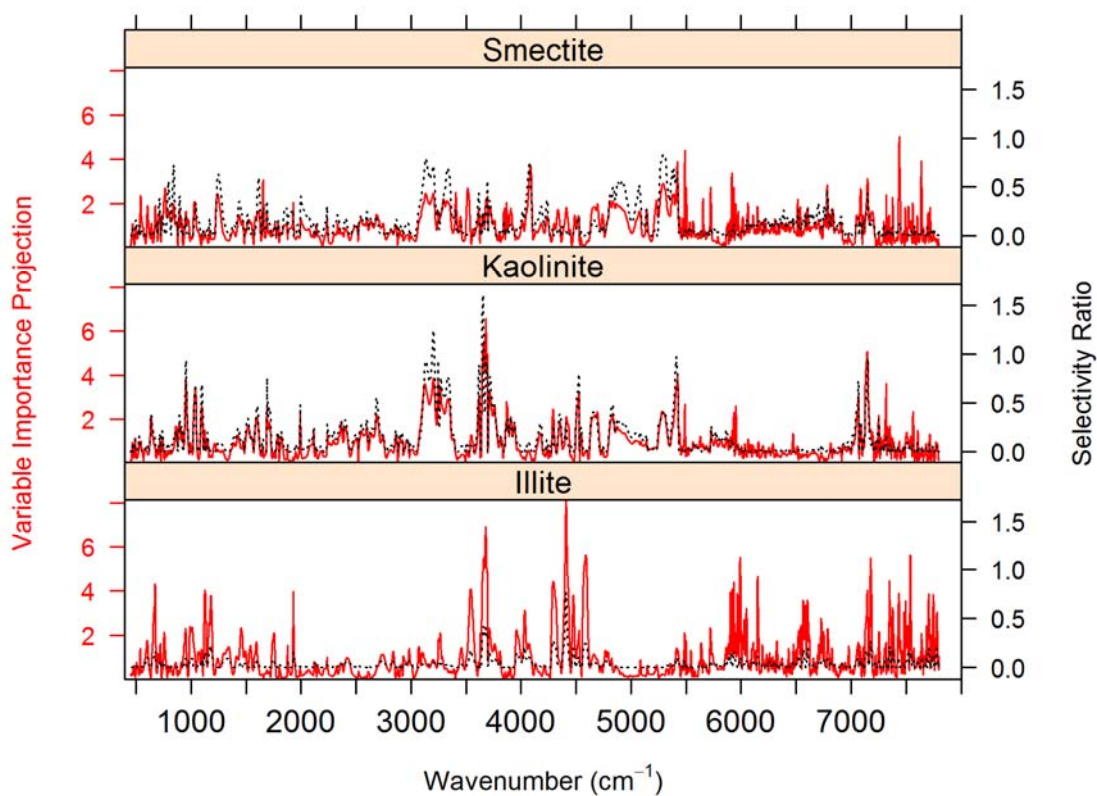


Figure 8.5. VIP and SR scores for the PLSR clay mineral models.

The 2:1 layer silicates can vary greatly due to their ionic substitution properties. The model for smectite has observable distinguishing frequencies in the 880-910, 1200-1360, 3050-3440, 3950-4120 and 5240-5470 cm^{-1} regions. The strong absorbance of the MIR spectra by smectite, and overlap with kaolinite including a poorly defined Al-OH bond at 3620 and 4533 cm^{-1} was observed. Alumino-silicate substitutions for smectite below 920 cm^{-1} are also likely (Nguyen et al., 1991).

There were considerably fewer corresponding variables in the illite model with two notable regions at 3600-3730 and 4200-4460 cm^{-1} . Isolated water vapour as the OH stretch at 3652 cm^{-1} and a short band at 4270 cm^{-1} are potentially related to these distinguishing features in the VIP and SR scores for illite.

8.3.3 Application of clay mineral spectroscopic models to MIR spectra with sites

Clay mineral prediction models were applied to the 2795 sites (11,532 samples) with MIR spectra (Figure 8.6). The kaolinite model is characterised by predictions from 5 to 100%. The distinctive 'hull' shape to this model reflect the absence of kaolinite calibration samples at low (<5%) and high (>95%) values. Over 99% of errors from the model are <12%. Very low kaolinite predictions were found to correspond with pH_w values > 8.4, field textures are \geq light clay (LC) and electrical conductivity (EC) is > 0.3. High kaolinite predictions were associated with exchangeable calcium <6, EC between 0.05 and 0.5, and pH_w > 9.5 or < 5.5.

Illite predictions were generally <50% (note one sample was excluded from view as this recorded a prediction of >150%). Like the kaolinite model, prediction errors were nearly always < 12%. Low illite predictions are all from south-western Victoria with an average depth of > 0.5m and EC is < 1. Samples where high illite predictions were determined have pH_w values > 7.3, high salinity (EC > 2) and as a result there is little difference

between pH_w and pH_c measurements. These samples are predominantly from sites in the North Western Dunefields and Plains (Mallee soils) where illite is recognised as the dominant clay mineral of aeolian depositional layers with carbonates (Stace et al., 1968; Wetherby and Oades, 1975).

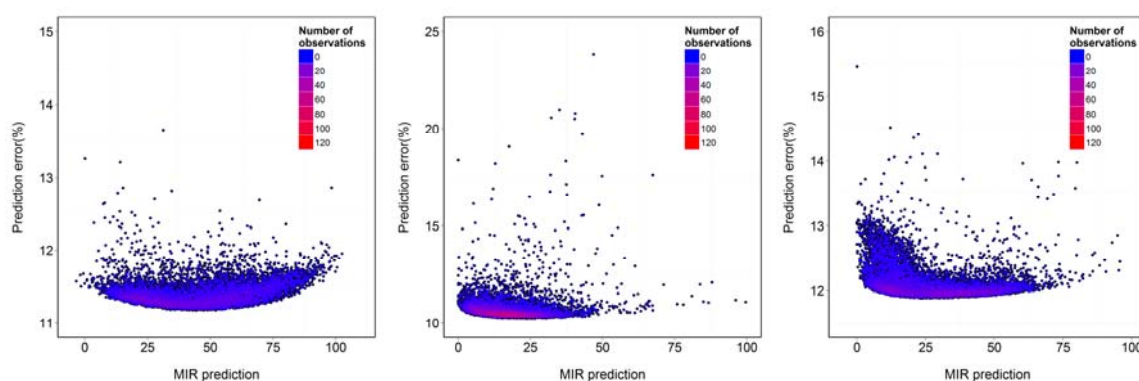


Figure 8.6. Clay mineral predictions for kaolinite (a), illite (b), smectite (c) and associated errors for samples.

The majority of samples in the smectite model have observed values of $< 70\%$ with nearly all errors $< 13\%$. A noticeable increase in prediction error occurs where smectite predictions are $< 20\%$. Very low predictions ($< 10\%$) are often from sites in south-western Victoria with $pH_w < 6.2$, exchangeable calcium < 6 and textures lighter than clay loam (CL). High smectite predictions for samples also have high exchangeable calcium, pH_w is above 6.5, clay % is $> 30\%$ and EC is high (> 0.5). These samples are generally found in north-western Victoria or as local depressions (e.g. swamps) in the volcanic plains for example.

The three applied models were then aggregated to form a ternary plot (Figure 8.7) that emphasized high kaolinite, moderate smectite and low illite predictions for samples from western Victoria. Clay mineral predictions were harmonised according to the GSM depth

ranges (Table 8.4) using equal area splines as inputs to the formation of model trees in Cubist.

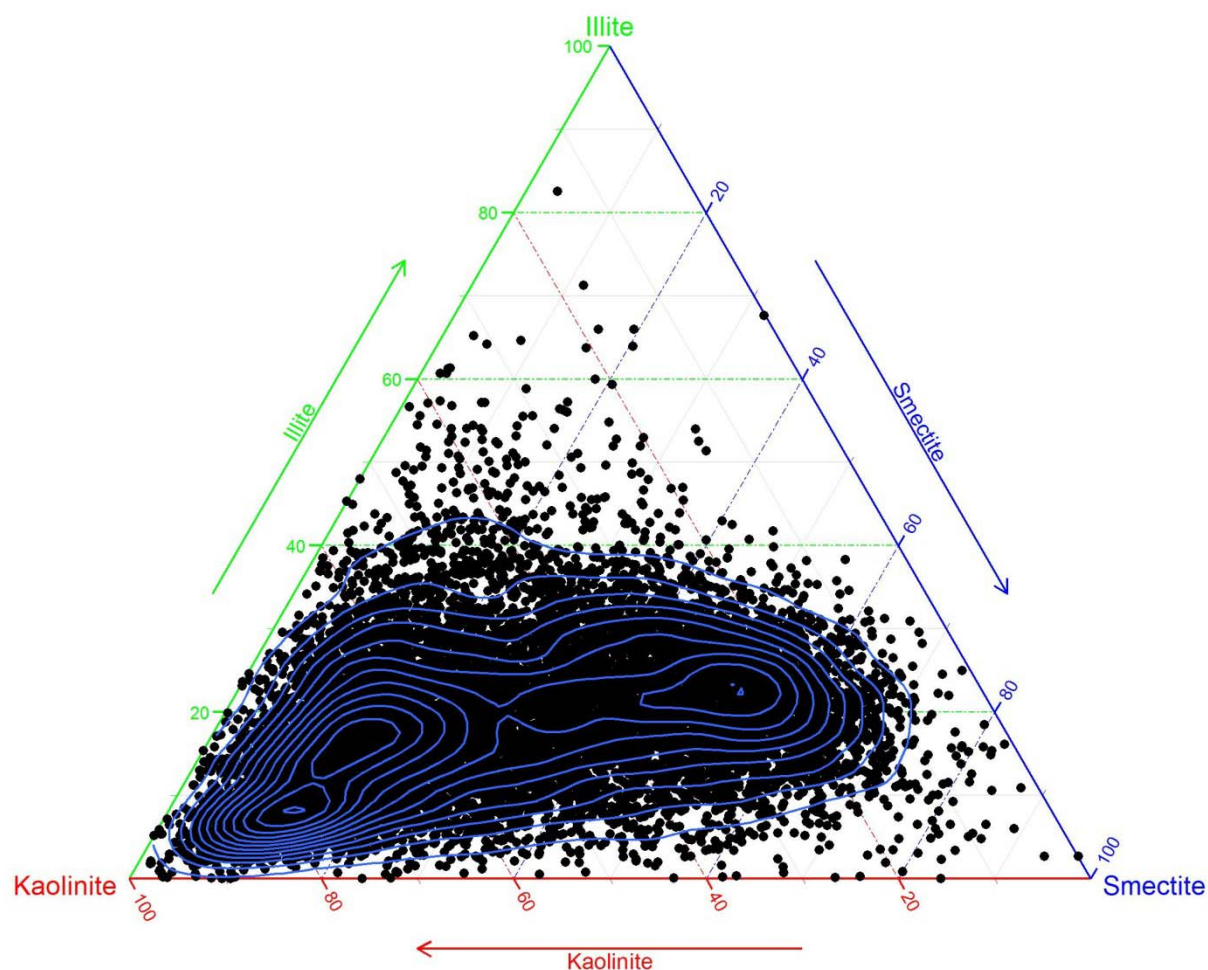


Figure 8.7. Ternary diagram of kaolinite, illite and smectite predictions from the spectroscopic models.

8.3.4 Mapping of clay mineral abundances

Output mapping statistics for kaolinite, illite and smectite are presented in Table 8.5 and mean prediction with upper and lower limit maps presented in Fig. 8.8 to 8.13. The 10-fold cross validation statistics of the six GSM depths for kaolinite yielded a RMSE between 8.2 and 12.3% and a R^2 between 0.65 and 0.77. These results compare favourably with those achieved by Viscarra Rossel (2011) of 0.52 and 0.46 for 0-20 and

60-80 cm. Strong concordance statistics were achieved for all depths with the 60-100 cm depth interval achieving the best results. Validation statistics were generally poorer than calibration results although this difference was consistent and relatively small. Between 87 to 89 % of values occur within the 90% prediction interval.

Table 8.4. Summary statistics for clay mineral data used in the Cubist model trees.

Depth range (cm)	N	Minimum	1 st Quantile	Median	Mean	3 rd Quantile	Maximum
Kaolinite							
0-5	2795	9.03	39.26	51.12	49.97	60.84	92.90
5-15	2795	6.73	37.94	49.65	48.90	59.85	90.67
15-30	2785	6.79	34.39	46.39	46.69	58.72	97.14
30-60	2727	2.18	29.39	44.13	45.42	59.93	97.60
60-100	2207	2.25	23.55	40.24	43.76	63.15	99.92
100-200	522	3.36	20.51	30.48	38.26	53.99	99.30
Illite							
0-5	2783	0.59	12.53	17.00	17.97	22.05	70.29
5-15	2788	0.11	13.09	17.3	18.20	22.05	67.12
15-30	2779	0.17	12.55	16.68	17.78	21.70	65.19
30-60	2722	0.38	9.75	14.69	15.76	20.06	66.02
60-100	2198	0.25	8.12	13.26	14.43	19.42	67.02
100-200	528	0	9.08	16.23	16.29	22.67	63.02
Smectite							
0-5	2793	0.90	12.26	17.61	20.56	27.49	70.68
5-15	2794	1.24	12.28	18.08	21.13	28.81	74.99
15-30	2785	0	12.2	20.45	23.09	32.92	87.83
30-60	2727	0.49	13.45	25.26	26.95	39.31	82.05
60-100	2207	0.39	16.04	30.52	31.48	45.90	100.51
100-200	525	0.84	20.11	38.40	36.03	49.89	110.55

Model statistics for illite were mixed. The prediction error was considerably less than those for kaolinite ranging from 5.2 to 6.9% although the coefficient of determination was markedly less varying between 0.26 at 0-5 cm to 0.6 at 60-100 cm. Prediction of illite was worst at the surface (0-5 cm) and best at 60-100 cm. The poorer model statistics for illite

in contrast to kaolinite and smectite was also observed by Viscarra Rossel (2011) with R^2 values of 0.41 and 0.4 for the two depth intervals (as noted earlier). Concordance values also improved with depth due to strong correlations below 30 cm. Validation statistics were similar for RMSE and ρ_c , however the R^2 was notably lower (0.18 to 0.37) than the calibration results. Once again there was good agreement with the 90% PI with 87 to 90% of predictions occurring in this interval.

Table 8.5. Map diagnostic results for kaolinite, illite and smectite. Standard deviation for map diagnostics is provided in ().

Depth (cm)	range	Calibration				Validation				
		RMSE	R^2	Bias	ρ_c	RMSE	R^2	Bias	ρ_c	% within 90% PI
Kaolinite										
0-5		8.6 (0.3)	0.65 (0.02)	0.05 (0.22)	0.78 (0.02)	10.1 (0.5)	0.52 (0.05)	-0.01 (0.6)	0.70 (0.03)	0.89 (0.02)
5-15		8.2 (0.4)	0.68 (0.03)	0.04 (0.1)	0.81 (0.02)	9.9 (0.3)	0.54 (0.03)	0.03 (0.73)	0.72 (0.02)	0.89 (0.01)
15-30		8.9 (0.4)	0.68 (0.03)	0.08 (0.21)	0.80 (0.02)	10.8 (0.4)	0.53 (0.04)	0.06 (0.79)	0.71 (0.03)	0.89 (0.02)
30-60		10.5 (0.4)	0.69 (0.02)	-0.11 (0.18)	0.82 (0.01)	12.8 (0.7)	0.56 (0.04)	-0.16 (0.74)	0.73 (0.02)	0.88 (0.02)
60-100		10.9 (0.3)	0.77 (0.01)	-0.13 (0.21)	0.87 (0.01)	14.2 (1.2)	0.62 (0.06)	0.33 (1.17)	0.78 (0.04)	0.88 (0.02)
100-200		12.3 (1.2)	0.71 (0.05)	-0.32 (0.36)	0.83 (0.03)	16.5 (2.5)	0.50 (0.09)	-0.48 (2.91)	0.68 (0.06)	0.87 (0.05)
Illite										
0-5		6.9 (0.3)	0.26 (0.06)	-0.75 (0.11)	0.40 (0.07)	7.3 (0.5)	0.18 (0.05)	-0.74 (0.36)	0.33 (0.05)	0.90 (0.02)
5-15		6.0 (0.2)	0.36 (0.04)	-0.55 (0.13)	0.52 (0.04)	6.8 (0.3)	0.22 (0.06)	-0.56 (0.39)	0.39 (0.05)	0.89 (0.02)
15-30		6.1 (0.2)	0.36 (0.04)	-0.67 (0.09)	0.52 (0.04)	6.9 (0.5)	0.21 (0.06)	-0.67 (0.44)	0.38 (0.05)	0.89 (0.02)
30-60		6.0 (0.2)	0.44 (0.03)	-0.67 (0.09)	0.61 (0.03)	7.1 (0.9)	0.25 (0.08)	-0.68 (0.55)	0.46 (0.08)	0.90 (0.02)
60-100		5.2 (0.2)	0.60 (0.03)	-0.42 (0.11)	0.74 (0.02)	6.7 (0.5)	0.36 (0.09)	-0.49 (0.40)	0.58 (0.07)	0.88 (0.02)
100-200		6.1 (0.3)	0.55 (0.04)	-0.51 (0.15)	0.71 (0.03)	7.3 (1.2)	0.37 (0.10)	-0.56 (1.22)	0.57 (0.08)	0.88 (0.05)
Smectite										
0-5		7.4 (0.2)	0.55 (0.02)	-0.76 (0.12)	0.71 (0.02)	8.8 (0.4)	0.37 (0.05)	-0.76 (0.67)	0.58 (0.03)	0.88 (0.02)
5-15		6.9 (0.2)	0.63 (0.02)	-0.68 (0.1)	0.77 (0.02)	8.9 (0.6)	0.42 (0.05)	-0.69 (0.51)	0.63 (0.04)	0.89 (0.01)
15-30		7.7 (0.3)	0.66 (0.03)	-0.68 (0.12)	0.79 (0.02)	9.7 (0.8)	0.46 (0.07)	-0.65 (0.62)	0.66 (0.05)	0.88 (0.02)
30-60		9.0 (0.5)	0.67 (0.04)	-0.47 (0.15)	0.80 (0.03)	11.0 (0.9)	0.51 (0.06)	-0.35 (0.6)	0.70 (0.04)	0.88 (0.03)
60-100		10.0 (0.4)	0.66 (0.03)	-0.52 (0.15)	0.80 (0.02)	12.1 (0.8)	0.53 (0.03)	-0.42 (0.78)	0.71 (0.02)	0.90 (0.02)
100-200		12.1 (0.9)	0.58 (0.06)	-0.51 (0.42)	0.74 (0.05)	16.4 (4.3)	0.32 (0.17)	-0.26 (2.97)	0.51 (0.21)	0.88 (0.06)

Similar to kaolinite, smectite mapping model statistics were robust across all depth intervals. The RMSE increased with depth to a maximum of 12.1 ± 0.9 at 100-200 cm and R^2 between 0.55 (0-5 cm) and 0.67 (60-100 cm). Smectite map predictions of the 0-5 cm depth range are comparable to those of Viscarra Rossel (2011) for the surface (R^2 of 0.61 for 0-20 cm). Considerable improvements were also observed in the prediction of smectite with depth. Like illite, models were negatively biased however the concordance statistics were considerably better (between 0.71 and 0.8). Validation results generally were slightly poorer while similar trends of model improvement with depth to 60-100 cm were noted. Again, there was good agreement with the 90% PI with 88 to 90% corresponding with this interval.

8.3.5 Interpretation of prediction variables and relationship to western Victoria (kaolinite, illite, smectite)

Major predictor variables used in condition statements and regression models from the Cubist spatial models to predict the three clay minerals for two sample depths (0-5 and 60-100 cm) are illustrated in Fig. 8.14 to 8.16. Explanatory variables are grouped according to the *s*, *c*, *o*, *r* and *p* factors of the SCORPAN model and usage is illustrated according to mean application across the 10 models. Effectively, the condition statements are the coarse spatial stratification of landscapes while the models reflect local processes and variations (Viscarra Rossel 2011).

Conditional statements of the 0-5 cm kaolinite model include *p* predictors gamma radiometric potassium (pot) and tier 2 geomorphological mapping (gmu250t2), *s* factor land units and *o* factor (veg_par_median) for all implementations. In condition statements for 60-100 cm the *r* factor elevation (mASL) and *o* factor (veg_par_mean) were used extensively with the *s* factor land units.

Kaolinite models for 0-5 cm are represented by variables of *s*, *c*, *o*, *r* and *p* factors. The *o* variables include 2009 NDVI Timesat derivatives (begin, end, length, smallint, largeint, median and base) and veg_fpar (mean and median) as important predictors with *c* variables (etaaan, rainann, meanrain6089, prescott and prescott_lg). The more frequently used variables in models for 60-100 cm also include *o* variables from 2009 and 2011 NDVI Timesat derivatives (e.g. begin, end, length and base), veg_fpar_mean and Landsat imagery (TM50_2 and TM50_3). Again, the same *c* variables (etaaan, rainann, meanrain6089, prescott and prescott_lg) were also important as they were for 0-5 cm. Prominent *p* factors include gamma radiometrics (thor, pot, tc and th_pot) and gmu250t2. The *r* factor elevation (mASL) was also frequently used in these models.

Variables used in condition terms for prediction of illite (0-5 cm depth) include *p* factors pot, geology mapping (geol250), the *o* factor veg_fpar_std (standard deviation) and *c* variable etaaan. The *c* factors (rainann and meanrain6089) were important in the 60-100 cm depth range with *p* factors pot, geol250 and thorium to potassium ratio (th_pot). Models for 0-5 and 60-100 shared important *o* variables from 2009 NDVI (begin, end and length). More frequently used *o* factors for 0-5 cm also include 2011 NDVI (begin, end and length) and for 60-100 cm veg_fpar (mean, median and mean_pv). The *c* variables etaaan, rainann and prescott were prominent in models across both depths while prescott_lg and meanrain6089 were also widely applied for 60-100 cm. The *p* factor pot was important for both depth intervals while thor and tc were also important. The *r* factor mrvbf was used in all 10 models.

The smectite spatial models for 0-5 cm include the *s* factor land units and *o* factor veg_par_median as the two standout variables of conditional statements. Likewise for the 60-100 cm, land units were prominent, with additional *o*, *p* and *r* factors including veg_par_mean, gmu250t2, tc and elevation. Models for both depths were also strongly

influenced by *o* factors: 2009 NDVI (begin, end and length) and veg_fpar (mean and median). Other important *o* factors for 0-5 cm were 2009 NDVI variables base, largeint and smallint. *p* factors tc and gmu250t2 were used by all 10 models for 60-100 cm with pot used in 8 models for 0-5 cm. Additional *c* variables etaaan, rainann, meanrain6089, prescott and prescott_lg were also prominent in the 60-100 cm models.

Across the spatial prediction models *s*, *c*, *o*, *r* and *p* factors have been used for conditional statements and models. Variables that were more frequently used include radiometric potassium that has been found associated with potassium rich dust of illite and K-feldspar (Cattle et al., 2003). Variables such as geol250 and gmu250t2 have proven useful for coarse scale stratification where used in conditional statements for illite and kaolinite, and land units when used in prediction of smectite for 0-5 cm. *o* factors were the most used variables in models to predict fine scale processes and variation, and were rarely used in conditional statements. *c* factors were also used extensively in models like the *o* factors. Even though these *o* and *c* variables are represented as coarse resolution datasets, there is adequate differentiation to represent fine scale processes reflected in the distribution of kaolinite, illite and smectite for the study region of western Victoria. Surprisingly, the higher spatial resolution *r* variables were rarely used in condition statements or models and suggests that local variations were not significantly associated with terrain or that sites weren't adequately located to represent the local range variation.

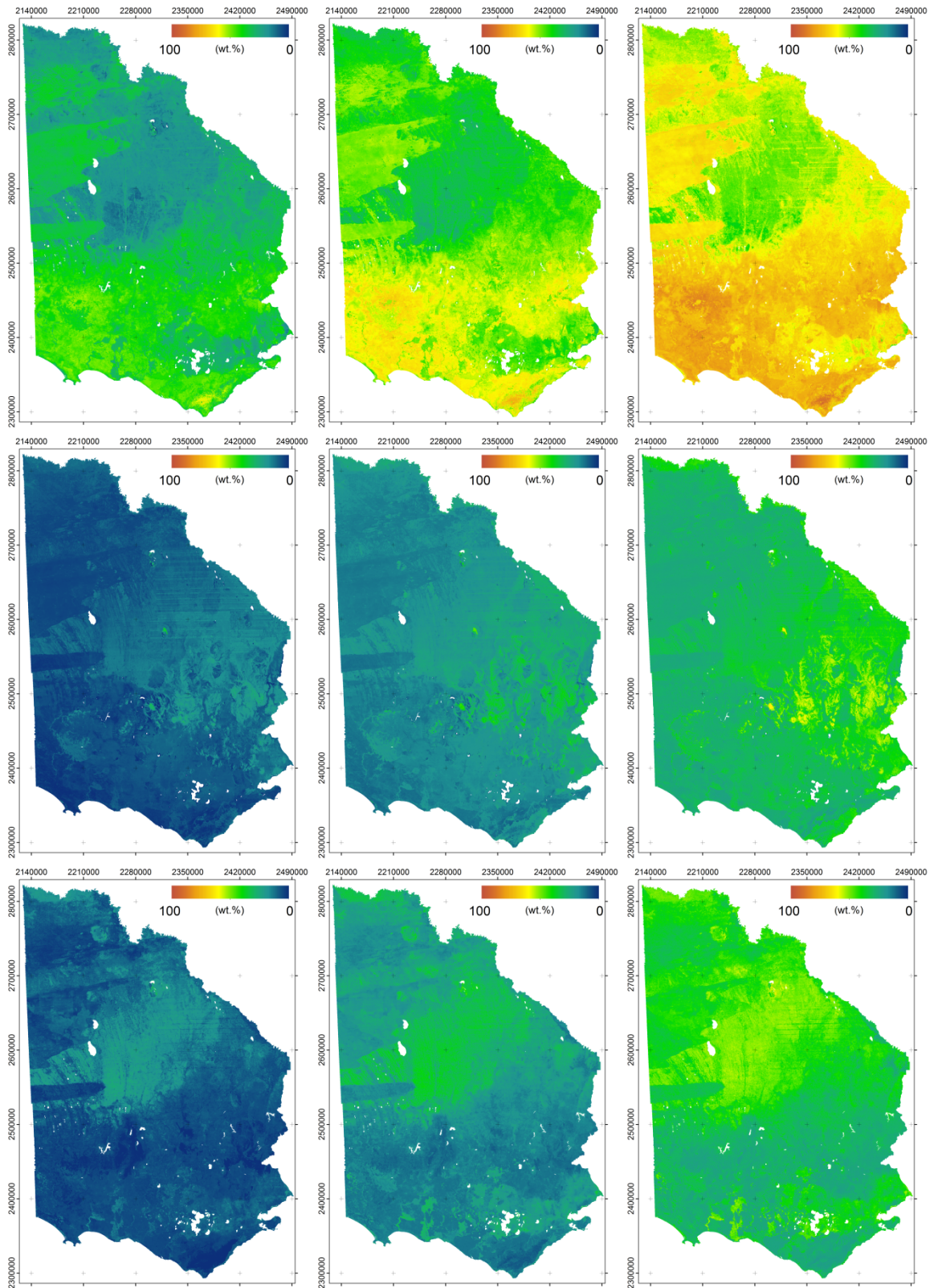


Figure 8.8. Map predictions (lower, mean and upper) from left to right for kaolinite (top), illite (middle) and smectite (lower) for 0-5 cm.

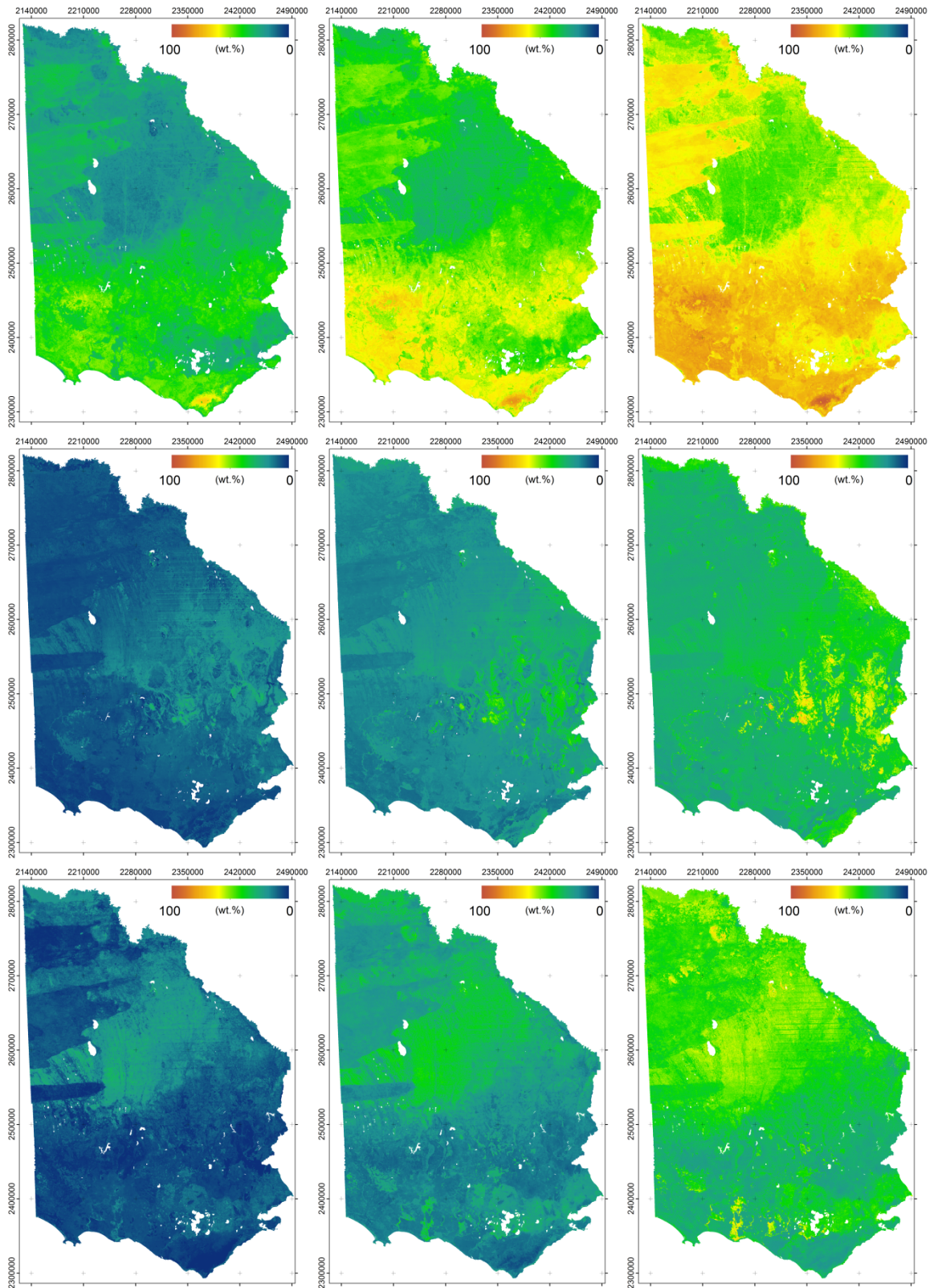


Figure 8.9. Map predictions (lower, mean and upper) from left to right for kaolinite (top), illite (middle) and smectite (lower) for 5-15 cm.

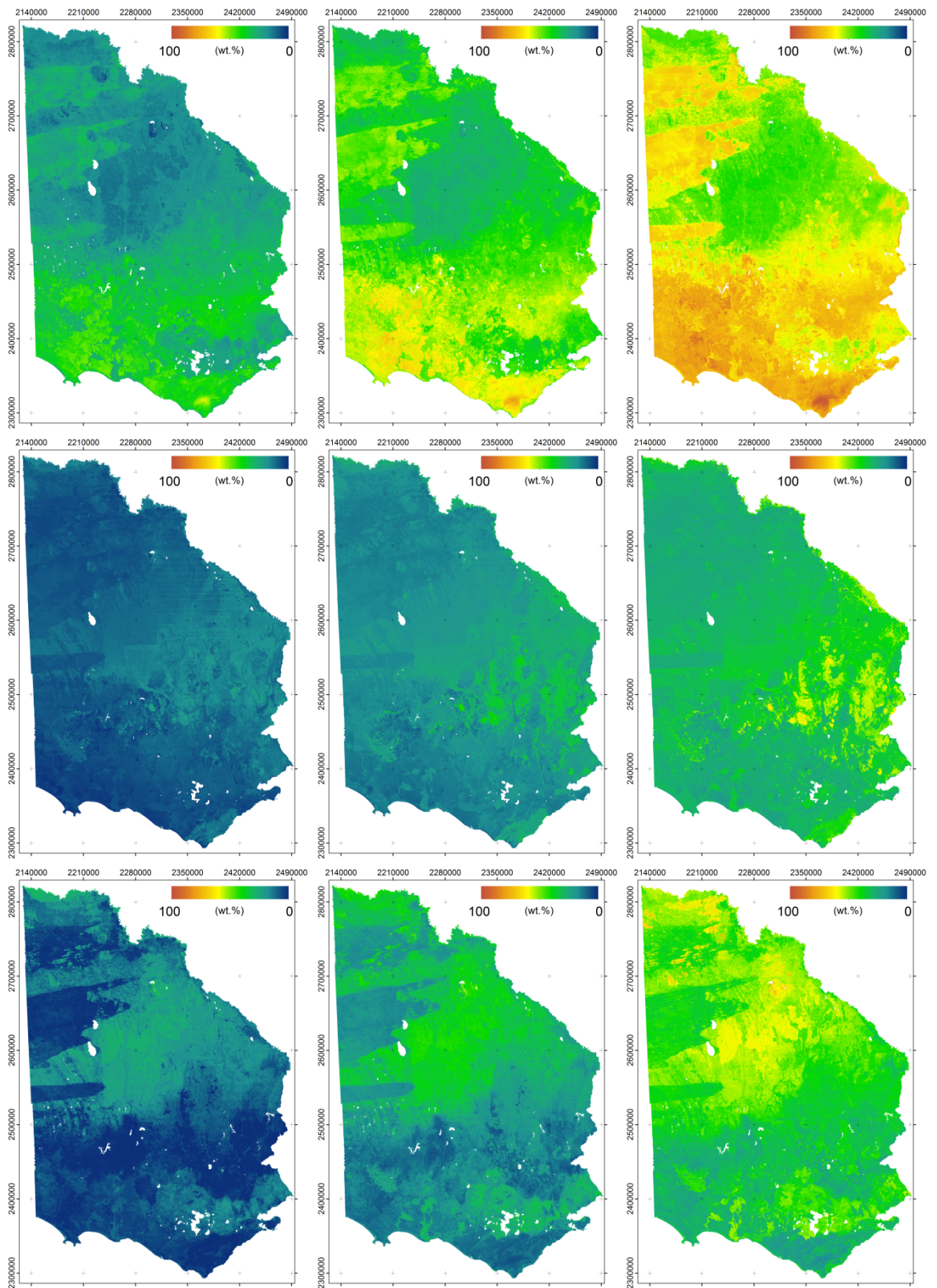


Figure 8.10. Map predictions (lower, mean and upper) from left to right for kaolinite (top), illite (middle) and smectite (lower) for 15-30 cm.

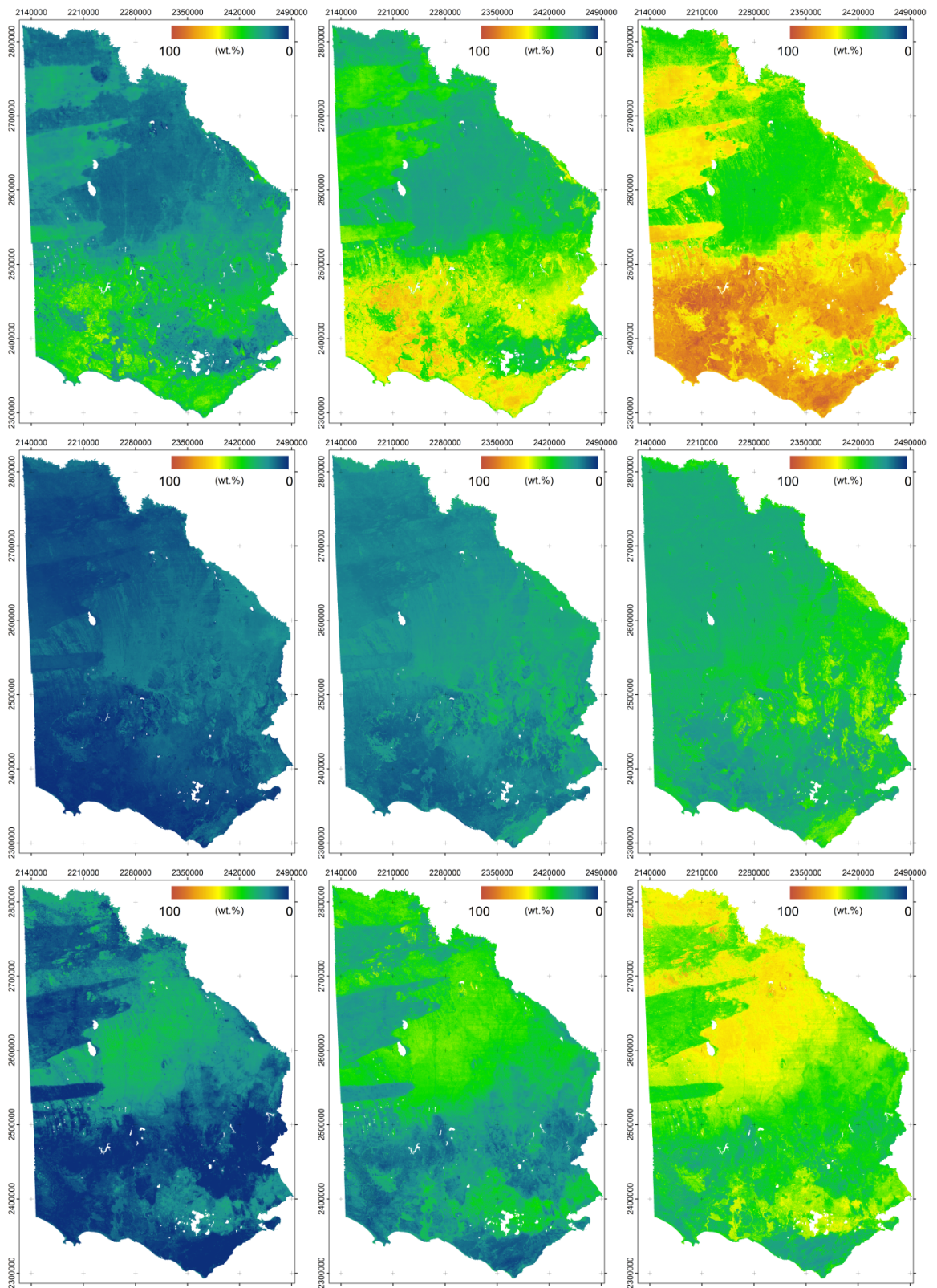


Figure 8.11. Map predictions (lower, mean and upper) from left to right for kaolinite (top), illite (middle) and smectite (lower) for 30-60 cm.

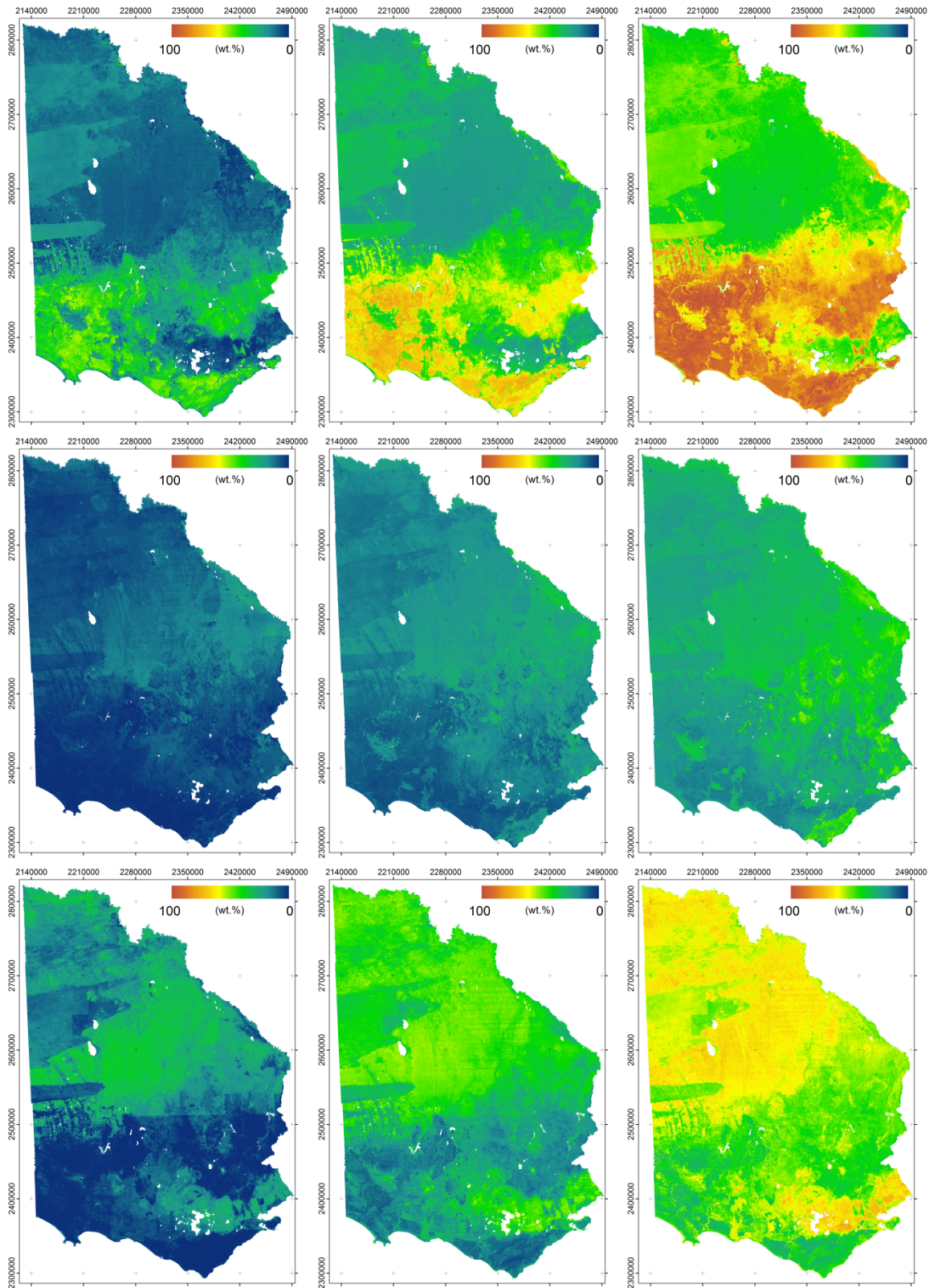


Figure 8.12. Map predictions (lower, mean and upper) from left to right for kaolinite (top), illite (middle) and smectite (lower) for 60-100 cm.

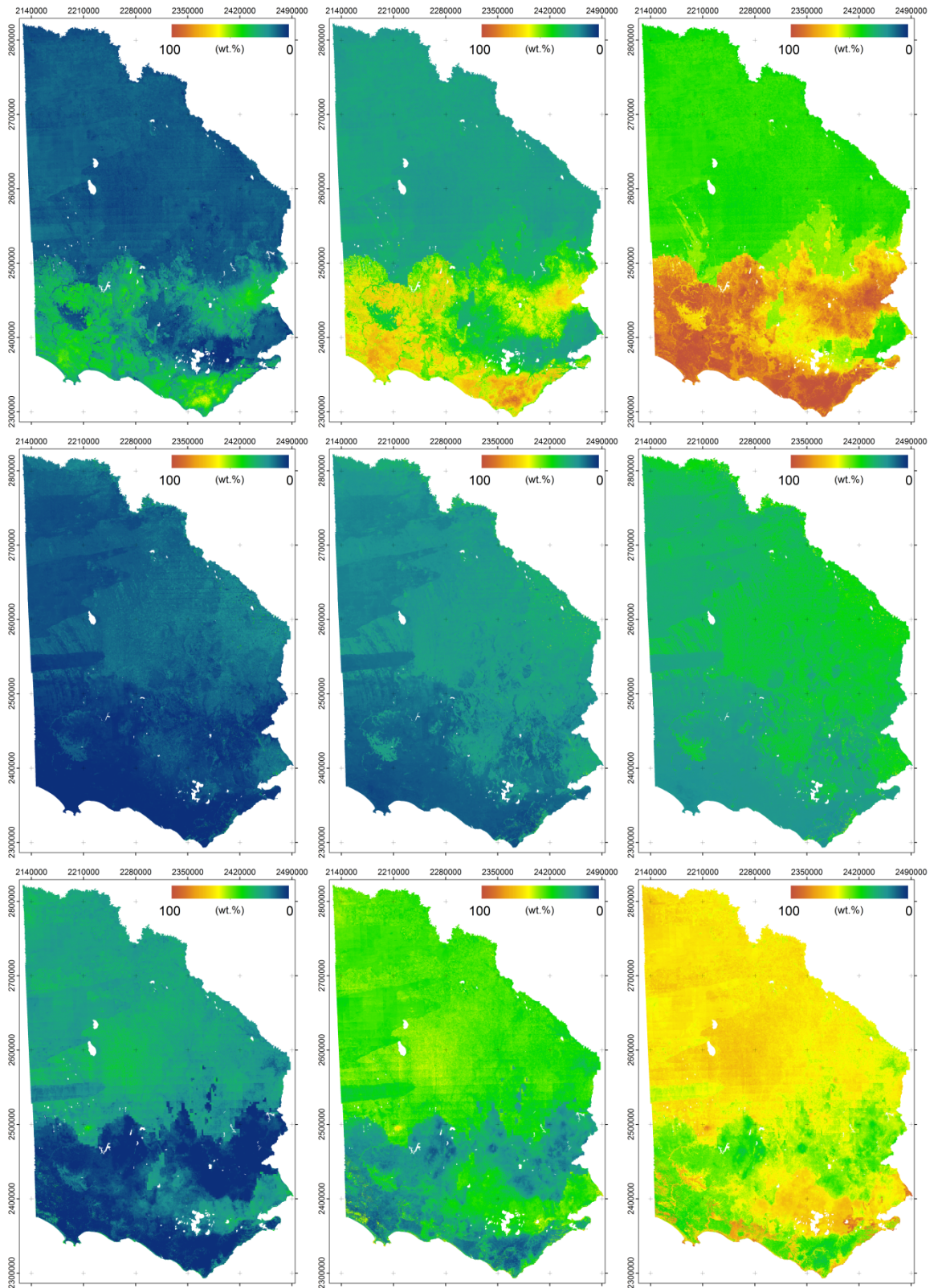


Figure 8.13. Map predictions (lower, mean and upper) from left to right for kaolinite (top), illite (middle) and smectite (lower) for 100-200 cm.

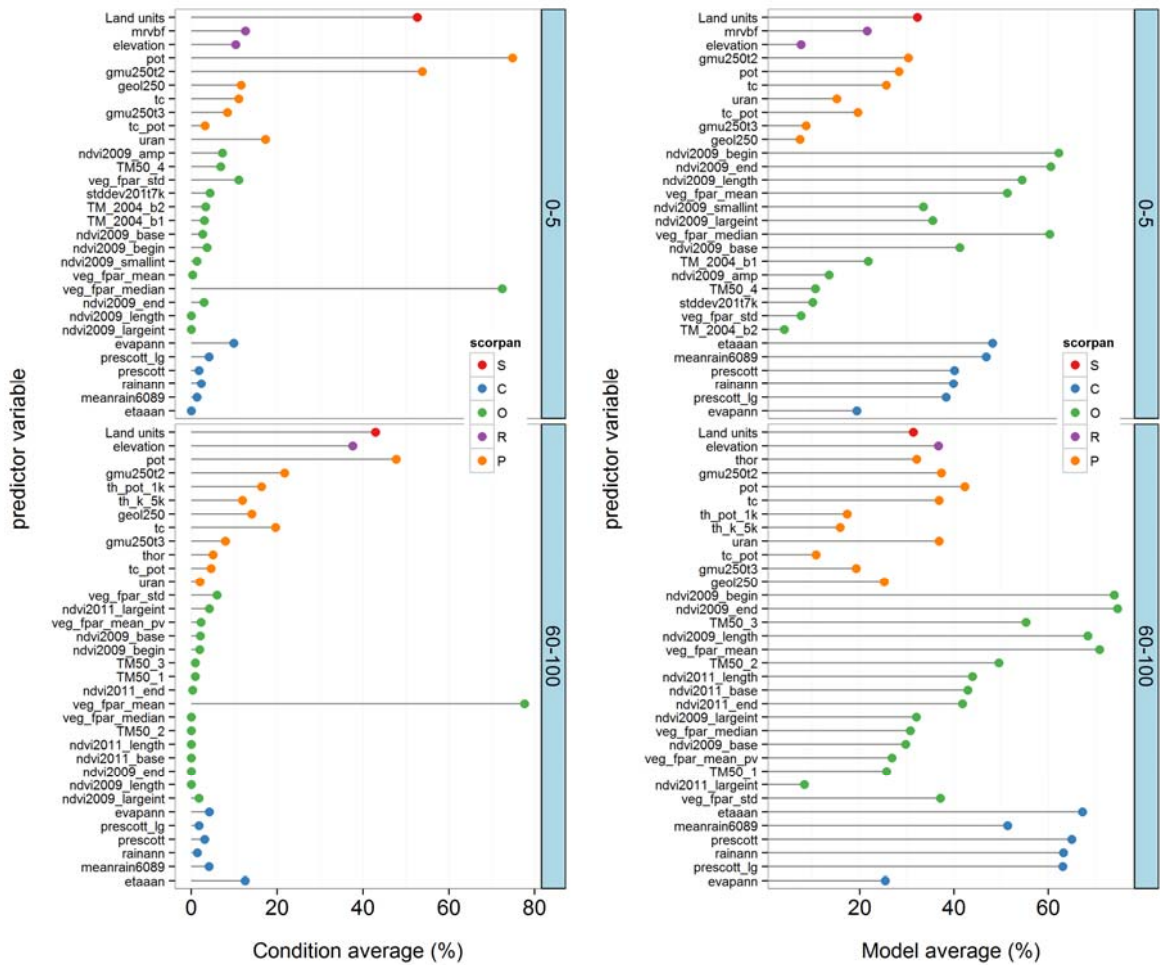


Figure 8.14. Major predictor variables used for kaolinite maps.

8.3.6 Comparison of clay mineral predictions against existing studies and investigations

Map predictions have been compared with published findings on soil clay mineral occurrence in Victoria.

Kaolinite is considered the dominant clay mineral (Norrish and Pickering 1983) for Australian soils. Smectite is of minor occurrence; however, interstratified kaolin-smectites (Norrish and Pickering, 1983; Churchman et al., 1994) are believed to be common to Australian soils. Clay mineral maps for western Victoria reflect this dominance of kaolinite, especially for southern regions with erosional landscapes, high rainfall and deep weathering profiles such as the Dundas Tableland. For the volcanic

plains, kaolinite content increases with age of the basalt flow and depth of profile (Mokma et al., 1973). This assessment is consistent with derived kaolinite prediction maps. There can be considerable halloysite interspersed with kaolinite and pure samples that can be easily mistaken for kaolinite, although actual concentrations of halloysite are possibly underestimated due to handling effects from drying that reduce its likelihood of detection (Norrish and Pickering, 1983). Krasnozems of the newer and older volcanics provinces occur with kaolinite either being exclusive (Hosking et al., 1957), or the dominant clay mineral phase (Sargeant and Skene, 1970) with minor illite or intermixed vermiculite-chlorite.

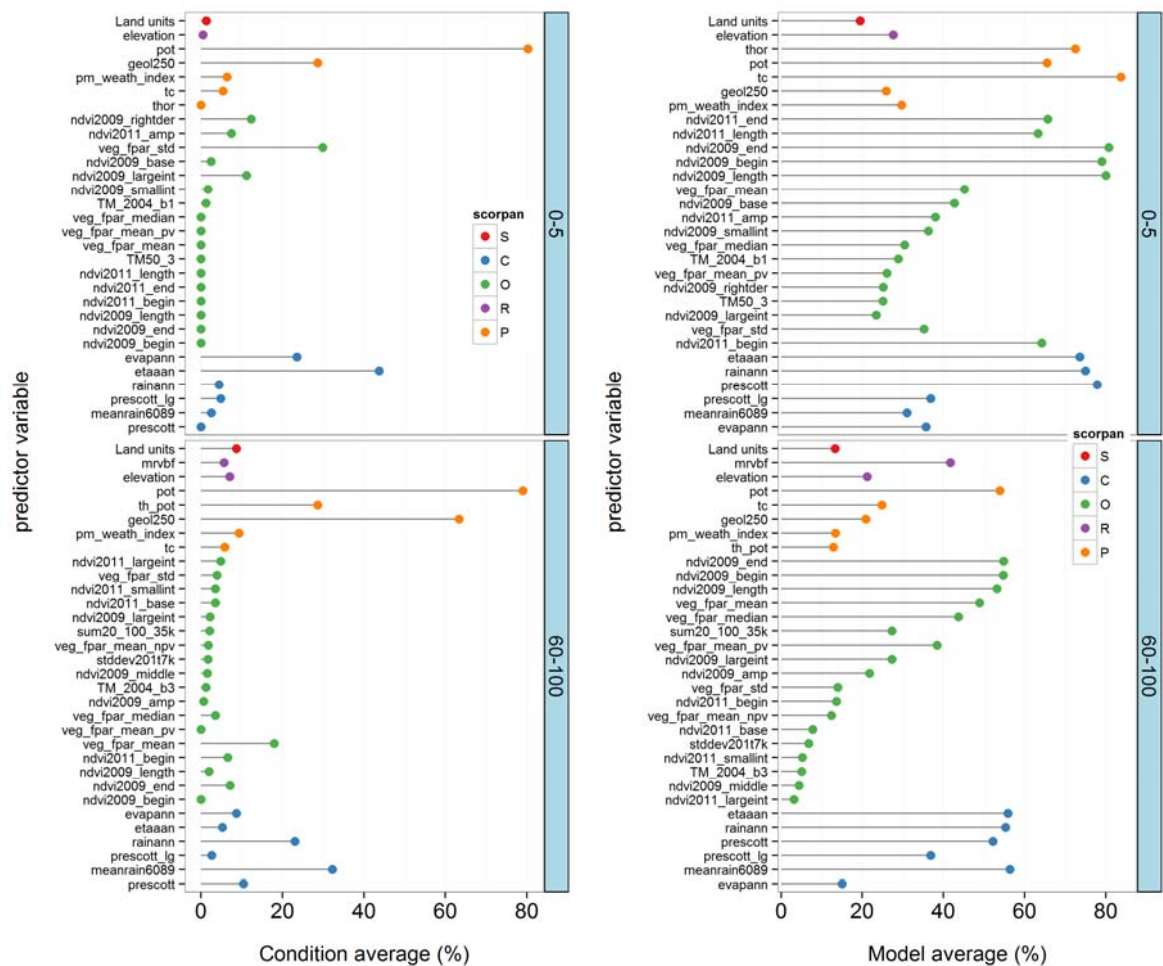


Figure 8.15. Major predictor variables used for illite maps.

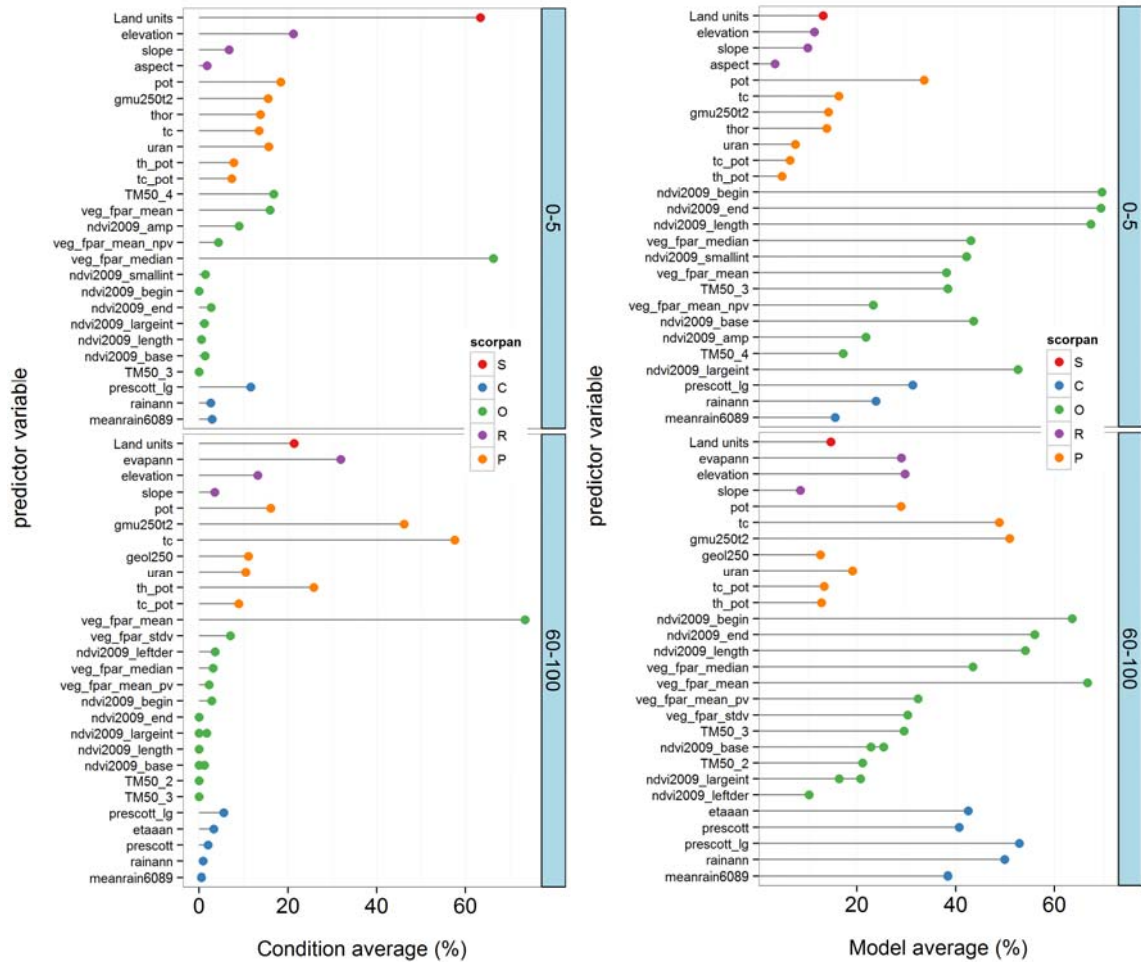


Figure 8.16. Major predictor variables used for smectite maps.

Granitic plutons and their weathering profiles contain variable quantities of kaolinite in the clay fraction with some illite and accessory minerals haematite, goethite and gibbsite (Hosking et al., 1957). Soils of Palaeozoic sedimentary rocks in central western Victoria are dominated by kaolinite with variable quantities of illite and smectite present (Sultan, 2006).

Illite occurrence in soils is greatest in central areas of western Victoria that are characterized by granitic and sedimentary hills with tertiary and quaternary alluvial and colluvial aprons. Granitic plutons at Mafeking, the Mirranatwa Granite in the Grampians

Ranges and Stawell Granite, for example, are clearly delineated in mapping with higher concentrations of illite than surrounding landscapes. This is likely due to residual weathering of muscovite and biotite leaving illite in K-rich sands. Spatial predictions suggest that illite concentration decreases with depth which is consistent with findings by Norrish and Pickering (1983). This may be attributed to the aeolian source material dominated by illite and kaolin blanketing upland landscapes of western Victoria, akin to profiles of the alpine regions with elevated illite at the surface (Johnston, 2001). Aeolian deposition on to the basalt plains has left quantifiable volumes of illite in clay and silt fractions (Jackson et al., 1972). A concentration of illite near the surface, especially for Mallee soils, may be due to aeolian deposition although the poor model prediction for 0-5 cm may be due to surface activities (e.g. cultivation, wind erosion), leading to high variability of illite, that are not accommodated in the applied spatial covariates.

Extremely weathered to fresh bedrock samples of the Otway Group Sandstone from southern slopes of the Otway Ranges include illite (16-18%) and minor kaolinite of 1-3% (Hall, 2004). Map predictions for illite for all depths have potentially underestimated illite occurrence in the Otway Ranges due to an absence of available sites and samples for modelling, or that further alteration of illite may occur where bedrock has further degraded to higher quantities of kaolinite. In the Wimmera plains, illite concentrations are comparable to kaolinite in low-lying areas associated with the Wimmera River. Here soils are derived from alluvial processes with potential bedrock sources from the western uplands including granites, sedimentary strike ridges and aeolian illite and kaolinite.

Smectite occurrence tends to reflect localized processes and weathering decay sequences as an early stage of this process. As the age of volcanic deposits and associated weathering in southern Victoria can be quite variable, smectite occurrence, while relatively low at the surface, increases with depth especially for black clay soils derived

from basalt. For the black clay soils, smectite phases are often the dominant clay mineral (Briner and Jackson, 1970). Average smectite concentrations for these soils have been observed at approximately 40%. Interstratified smectite-kaolinite has been found in soils associated with volcanic lake and lunette complexes of the western plains at Kariah, although the kaolinite layers in these clays can be poorly crystalline (Churchman et al., 1994). The plains of northern Victoria including the Wimmera have abundant smectite in the clay fraction of soils found in swales between, and blanketing NNW trending tertiary beach ridges. These uniform self-mulching Vertosols have smectite concentrations varying between 20 and 50%. Maps illustrate that smectite % increases with depth and this may reflect more advanced weathering at the soil surface (Churchman et al., 1994; Viscarra Rossel, 2011).

8.4 Discussion

Spectroscopic models for kaolinite, illite and smectite using MIR spectroscopy, combined with spatial prediction models using model-tree algorithms, has produced reliable maps of clay mineral distribution that adhere to GSM standards. By mapping these layer silicates, this provides an opportunity to further quantify and understand processes linked to fertility and resultant primary production as a key ecosystem service delivered by soil.

8.4.1 MIR calibration models

The spectroscopic models for kaolinite, illite and smectite in the $<2 \mu\text{m}$ fraction from whole soil samples were highly useful and relatively robust given the number of available samples with quantitative XRD assessments for calibration purposes. This may be expected given the diagnostic absorption, water and hydroxyl features that are characteristic of clay minerals represented in this spectral range (Clark et al., 1990).

Initial calibration models (not presented) were supplemented by additional quantitative XRD analysis for atypical samples (e.g. high clay mineral prediction or high relative error) to improve the overall ‘robustness’ of models for Victorian soils. There was no decline in model diagnostics from addition of the atypical samples to models, rather a noticeable contraction in prediction error and a reduction in the range of predicted values to theoretically possible values, e.g. 0 to 100 %. Cross-validation was applied in model development due to the small number of samples available for calibration purposes. 10% of samples for each k-fold were retained for validation purposes, reducing potential for model overfitting and impacts from outliers.

Quantitative XRD data used for calibration purposes collected over two decades has proven invaluable in development of quantified clay mineral models. Uncertainty estimates for these XRD determinations have been provided but were not accommodated in spectral models at this stage. Ideally, these XRD determinations can be compared with other approaches to determine mineral abundance such as relative spectral determinations using continuum removal to normalize reflectance spectra (Clark and Roush, 1984; Viscarra Rossel, 2011). Clay mineral abundance could be more adequately represented by addressing deficiencies in the distribution of samples for calibration purposes, e.g. above 50% for illite and for low and high kaolinite concentrations (e.g. <5% and >90%). Overall, the models provide estimates of clay mineral abundance that, as a minimum, are comparable to qualitative XRD with uncertainty estimates of $\pm 10\%$ (Briner and Palmer, 1985). Ideally, the coupling of quantitative XRD with IR spectroscopy offers operational advantages in quantitative prediction of phyllosilicates, oxides and carbonates (Janik et al., 1995).

Spectroscopic models were developed using PLSR for this study. This multivariate technique is relatively simple to apply and interpret and can handle colinearity in input

data (Wold et al., 2001; Soriano-Disla et al., 2014). A limitation of PLSR is the sensitivity to asymmetrical (skewed) input distributions in contrast to SVM, artificial neural networks or boosted regression trees (Viscarra Rossel and Behrens, 2010). For this study, only illite quantitative XRD data was transformed (square root) with no improvements in smectite and kaolinite prediction from various transformation algorithms. Assessment of diagnostic absorptions and spectral ranges using the VIP and SR for these clay minerals models proved useful in qualifying model development. Implementation and comparison of these two methods for assessment of critical spectral features has proven successful (Farfés et al., 2015) and provides an improvement on the use of regression vectors (Kvalheim, 2010). Implementation of SVMs in preference to PLSR or other multivariate assessment techniques has resulted in improved predictions for IR spectroscopy (Viscarra Rossel and Behrens, 2010). For this study, improvements were marginal, and therefore not pursued, but should be considered in spectral model ensembles using the range of multivariate techniques available.

8.4.2 Maps of soil clay minerals

The DSM procedure used for this study has been successfully implemented in Tasmania, Australia for state soil attribute mapping (Kidd et al., 2015) and adapted to model the distribution of soil organic carbon for mainland France (Mulder et al., 2016). The implementation of model trees with a 10-fold cross validation has reduced the modelling bias, averaged outputs and enabled systematic prediction of the 90% prediction interval based on the approach of Malone et al. (2014b). Residual errors from the spatial model trees were not assessed for spatial dependence but could be considered for further refinements to clay mineral maps.

Initial selection of covariates was based on their availability and spatial continuity for the study area. These covariates were used in a ‘first run’ of the mapping procedure using

clay mineral predictions from a reduced calibration set. Important covariates used in development of condition statements and regression models across the three clay minerals were retained and superfluous covariates removed from further model iterations.

Sites and samples used in spatial models have been applied, with no formal sample design, for their distribution in space or time. MIR predictions for these samples are consistent with trends observed in map outputs and published findings that:

- kaolinite content decreases with depth, potentially indicating greater weathering at the soil surface than with depth;
- mean and median values for kaolinite are at least twice that of illite;
- illite also decreases with depth and over 50% of predictions occur between 8 and 22%;
- smectite concentration increases with depth and mean/median predictions are slightly higher than illite.

This suggests a sequence of abundance whereby kaolinite > smectite > illite for western Victoria, with both illite and kaolinite decreasing with depth and smectite increasing. This sequence will vary though depending upon the landscape and history of soil formation.

There are some limitations in the soil site distribution for this study. There are areas of the Southern Uplands (e.g. Otway Ranges), Western Uplands, Northern Riverine Plains, and public lands of the North Western Dunefields and Plains (linear and parabolic dunefields) which are inadequately represented by sampled sites. An additional finding is that there is a deficiency of deep soil samples (>1 m) with fewer than 530 observations (<19% of total sites) available for use in the 10 model applications. This has most likely led to the deterioration of spatial model performance for the 100-200 cm depth range.

Uncertainties represented by the 90% PI for this study are from the averaged predictions of the 10 models which are consistent with approaches applied by Kidd et al. (2015) and

Viscarra Rossel et al. (2015). Upper and lower prediction intervals are spatially consistent with patterns observed for mean predictions. The majority of prediction intervals occurred with a relative uncertainty of around 40%, although there are areas with higher and lower uncertainties depending on site density, covariate quality and representativeness of soil landscape features. Uncertainties were found to increase with depth, reflecting the fewer sites available for model development (Viscarra Rossel et al., 2015). Temporal uncertainty was not considered in this study due to the relative stability and inadequate sensitivity of clay formation to time (MacEwan, 1997). Further enhancements leading to reduced uncertainty are to be undertaken using the GRUMP framework (Robinson et al., 2015) to assess and accommodate uncertainties due to input data error sources, model assumptions and qualitative understandings not used in production of clay mineralogy maps. This should reduce the residual error from the spatial models and this can be explored further for spatial dependence across landscapes.

Map predictions for kaolinite and smectite were reliable and illite results slightly less so. Model diagnostics improved with depth and the best results were achieved for the 60-100 cm depth range, indicating a strong alignment with environmental predictors and indicating less spatial variability. This is reflected in the strong association of map predictions, with pronounced topographic and lithological discontinuities in the study region, such as the deserts along the western border with South Australia. Model diagnostics compare favourably with those of Viscarra Rossel (2011) with continuous clay mineral predictions derived for the six GSM depth intervals. While it is reassuring that both clay mineral mapping attempts have achieved similar results for some landscapes using different data sources and IR spectroscopy techniques, there remain landscapes that have been under or over classified for some clay minerals (e.g. kaolinite in the Southern Uplands) that require further improvement.

The distribution of the layer silicates in soil reflects the principal factors of soil formation (Jackson, 1957). Climate, relief (topography), parent material, organisms and vegetation all influence the distribution and quantity of clay minerals in soil. All *s*, *c*, *o*, *r* and *p* factors were used in models for the broad scale stratification of the study region for clay mineral abundance. The variables most frequently represented are gamma radiometric potassium, geomorphological and soil-landform polygonal mapping. These variables are strongly aligned to parent material and the changes observed in soil mineral distribution of the parent lithology when other factors such as climate are held constant (Jackson, 1957). Not only are the primary minerals affected by their parent material, secondary clay minerals that are reactive are also influenced by the inherent nature of their host rock. Weathering processes of transformation and neoformation impacting the parent material result in the occurrence and accumulation of soil clay minerals (Wilson, 1999). Vegetation (MODIS vegetation indices) was also an important predictor. This may be due to the preferential vegetation clearance of land in Victoria where soils of higher agricultural capability were cleared first with inherent soil properties (e.g. clay mineralogy) that were favoured for farming. This trend continued leaving land of lower agricultural potential uncleared (Sheffield and Morse-McNabb, 2015). It is likely that vegetation, rainfall and evaporation are linked in their contribution to the models, however, further investigation is required to define linkages between vegetation and clay mineral distribution.

Regional or local scale variations represented in regression models were dominantly *c*, *o*, and *p* factors. Rainfall, estimated evapotranspiration and the Prescott index were all extensively used in predicting clay mineralogy. These climatic variables affect weathering rates of parent material and the transformation of primary to secondary clay minerals. Gamma radiometrics and geomorphological mapping was also used frequently in models

to delineate local differences in clay mineral occurrence associated with parent material, weathering state and erosional process of alluvial, colluvial and aeolian deposits. NDVI vegetation derivatives from Timesat parameters were also used across the depth intervals for the three clay minerals and potentially reflect local soil differences and variations due to pedological indicators, e.g. waterlogging and drainage, salinity, rock occurrence, cementation and soil strength.

8.4.3 Looking forwards

To our knowledge, the maps produced for this study are the first attempts to predict clay mineral abundance according to GSM specified depth intervals using MIR spectroscopy. Refinement and significant enhancement of maps can be achieved by using uncertainty estimates to guide future soil site collection with samples collected representative of the depth range (0 to 2 m) for the GSM project. This can be supported through implementation of sampling designs such as latin hypercube sampling approaches (Minasny and McBratney 2006; Clifford et al., 2014). There also remain many potential samples in the archive that could be used to fill some of these voids for the major geomorphological divisions of the Southern Uplands, Western Uplands and Northern Riverine Plains. Potential improvements through the combining of IR spectral libraries should also elicit more detail and better estimates of clay mineral occurrence for these landscapes.

A limitation for this study was the availability of accessible quantitative XRD measurements. Re-analysis of earlier determinations using the latest pattern-fitting software may yield higher precision in diffraction results and less uncertainty associated with these values. Available legacy qualitative XRD data could also be used to contrast, compare and integrate with quantitative data either to augment DSM models, or use as an independent validation set. Further improvements may occur in response to available new

environmental predictors such as passive and active satellite sensors such as Sentinel (1 and 2), combined with existing platforms such as MODIS and Landsat that offer better signal-to-noise ratios and high temporal frequency of image collection.

8.5 Conclusion

This study presents the first example where MIR spectroscopy and spatial models have been implemented to predict clay mineral distribution according to GSM specified depth intervals. This has been successfully implemented for the three prominent clay minerals: kaolinite, illite and smectite. Not only do the maps provide new predictions of these layer silicates to 2m in depth, they also enable users to elucidate new understandings linked to ecosystems services such as food production and the identification of potential engineering and infrastructure hazards due to expansive clay soils. The methods deployed for this study can be easily implemented elsewhere by organisations where MIR spectral libraries exist.

Data derived from this study will also provide reference spectra and mineralogy assessments that can be used in derivation of future spectroscopy calibrations. The outputs of the study (the mineral prediction maps) will be used as inputs to understand yield variability in pasture and cereal systems of Victoria and to support research on chemical stabilization processes of SOM and the potential for carbon sequestration. Further development of regional and state calibrations using MIR spectroscopy to predict other soil minerals (e.g. oxides and carbonates) is also envisaged.

References

- Amundson, R., Berhe, A.A., Hopmans, J.W., Olson, C., Sztein, A.E., Sparks, D.L., 2015. Soil and human security in the 21st century. *Science* 348(6235).
- Arrouays, D., McBratney, A.B., Minasny, B., Hempel, J.W., Heuvelink, G.B.M., MacMillan, R.A., Hartemink, A.E., Lagacherie, P., McKenzie, N.J., 2014. The GlobalSoilMap project specifications, GlobalSoilMap. CRC Press, pp. 9-12.
- 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 & Biochemistry* 43(7), 1398-1410.
- Beven, K.J., Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology / Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological Sciences Bulletin* 24(1), 43-69.
- Bishop, T.F.A., McBratney, A.B., Laslett, G.M., 1999. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. *Geoderma* 91(1–2), 27-45.
- Briner, G.P., Jackson, M.L., 1970. Mineralogical Analysis of Clays in Soils Developed from Basalts in Australia. *Israel Journal of Chemistry* 8(3), 487-500.
- Briner, P., Palmer, B.P., 1985. Determination of Clay Minerals in Soil and Sediments by X-Ray Diffraction. State Chemistry Laboratory, Melbourne, pp. 5.
- Cattle, S.R., Meakin, S.N., Ruzkowsky, P., Cameron, R.G., 2003. Using radiometric data to identify aeolian dust additions to topsoil of the Hillston district, western NSW. *Soil Research* 41(8), 1439-1456.
- Chang, C.-W., Laird, D.A., Mausbach, M.J., Hurburgh, C.R., 2001. Near-Infrared Reflectance Spectroscopy–Principal Components Regression. Analyses of Soil Properties. *Soil Sci. Soc. Am. J.* 65(2), 480-490.
- Chipera, S.J., Bish, D.L., 2001. Baseline studies of the Clay Minerals Society Source Clays: Powder X-Ray Diffraction Analyses. *Clays and Clay Minerals* 49(5), 398-409.

- Chong, I.-G., Jun, C.-H., 2005. Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems* 78(1–2), 103-112.
- Churchman, G.J., Lowe, D.J., 2012. Alteration, Formation and Occurrence of Minerals in Soils. In: P.M. Huang, Y. Li, M.E. Sumner (Eds.), *Handbook of soil sciences: resource management and environmental impacts*. CRC Press, Taylor & Francis, Florida, U.S.A., pp. 72.
- Churchman, G.J., Slade, P.G., Self, P.G., Janik, L.J., 1994. Nature of interstratified kaolin-smectites in some Australian soils. *Soil Research* 32(4), 805-822.
- Clark, R.N., 1999. Spectroscopy of Rocks and Minerals, and Principles of Spectroscopy. In: A.N. Renez (Ed.), *Manual of Remote Sensing. Remote Sensing of the Earth Sciences*. John Wiley and Sons, pp. 3-58.
- Clark, R.N., King, T.V.V., Klejwa, M., Swayze, G.A., Vergo, N., 1990. High spectral resolution reflectance spectroscopy of minerals. *Journal of Geophysical Research: Solid Earth* 95(B8), 12653-12680.
- Clark, R.N., Roush, T.L., 1984. Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *Journal of Geophysical Research: Solid Earth* 89(B7), 6329-6340.
- Clifford, D., Payne, J.E., Pringle, M.J., Searle, R., Butler, N., 2014. Pragmatic soil survey design using flexible Latin hypercube sampling. *Computers & Geosciences* 67, 62-68.
- de Jong, S., 1993. SIMPLS: an alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems* 18, 251-263.
- Eklundh, L., Jonsson, P., 2015. TIMESAT 3.2 with parallel processing. Software Manual, Department of Earth and Ecosystem Sciences, Lund University, Sweden, www.nateko.lu.se/TIMESAT.
- Faber, N.M., Bro, R., 2002. Standard error of prediction for multiway PLS: 1. Background and a simulation study. *Chemometrics and Intelligent Laboratory Systems* 61(1–2), 133-149.

- Farrés, M., Platikanov, S., Tsakovski, S., Tauler, R., 2015. Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation. *Journal of Chemometrics* 29(10), 528-536.
- Fontaine, S., Barot, S., Barre, P., Bdioui, N., Mary, B., Rumpel, C., 2007. Stability of organic carbon in deep soil layers controlled by fresh carbon supply. *Nature* 450(7167), 277-280.
- Gallagher, N.B., 2005. *Extended Multiplicative Scatter Correction Applied to Mid-Infrared Reflectance Measurements of Soil*. Eigenvector Research Incorporated.
- Gallant, J.C., Austin, J.M., 2015. Derivation of terrain covariates for digital soil mapping in Australia. *Soil Research* 53(8), 895-906.
- Gallant, J.C., Dowling, T.I., 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resources Research* 39(12), 1347.
- Gilkes, R.J., 1990. Mineralogical insights into soil productivity: an anatomical perspective. *Transactions of the international congress of soil science*, 63-73.
- Grunwald, S., 2009. Multi-criteria characterization of recent digital soil mapping and modeling approaches. *Geoderma* 152(3-4), 195-207.
- Guillou, F.L., Wetterlind, W., Viscarra Rossel, R.A., Hicks, W., Grundy, M., Tuomi, S., 2015. How does grinding affect the mid-infrared spectra of soil and their multivariate calibrations to texture and organic carbon? *Soil Research* 53(8), 913-921.
- Haaland, D.M., Thomas, E.V., 1988. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Analytical Chemistry* 60(11), 1193-1202.
- Hall, C.J., 2004. *Weathering and strength parameters of Otway Group rocks*, University of Ballarat, Ballarat, Victoria., 170 pp.
- Hartemink, A.E., McBratney, A.B., Cattle, J.A., 2001. Developments and trends in soil science: 100 volumes of *Geoderma* (1967-2001). *Geoderma* 100(3-4), 217-268.

- Hicks, W., Viscarra Rossel, R.A., Tuomi, S., 2015. Developing the Australian mid-infrared spectroscopic database using data from the Australian Soil Resource Information System. *Soil Research* 53(8), 922-931.
- Hosking, J.S., Neilson, M.E., Carthew, A.R., 1957. A study of clay mineralogy and particle size. *Australian Journal of Agricultural Research* 8(1), 45-74.
- Hughes, R.E., Moore, D.M., Glass, H.D., 1994. Qualitative and Quantitative Analysis of Clay Minerals in Soils. In: J.E. Amonette, L.W. Zelazny (Eds.), *Quantitative Methods in Soil Mineralogy*. Soil Science Society of America, Madison, WI.
- Isbell, R.F., 2002. *The Australian Soil Classification. Revised Edition*. Australian Soil and Land Survey Handbooks Series 4. CSIRO Publishing, Melbourne.
- Jackson, M.L., 1957. Frequency Distribution of Clay Minerals in Major Great Soil Groups as Related to the Factors of Soil Formation. *Clays and Clay Minerals* 6(1), 133-143.
- Jackson, M.L., Gibbons, F.R., Syers, J.K., Mokma, D.L., 1972. Eolian influence on soils developed in a chronosequence of basalts of Victoria, Australia. *Geoderma* 8(2-3), 147-163.
- Janik, L.J., Merry, R.H., Forrester, S.T., Lanyon, D.M., Rawson, A., 2007. Rapid Prediction Of Soil Water Retention Using Mid Infrared Spectroscopy. *Soil Sci. Soc. Am. J.* 71(2), 507-514.
- Janik, L.J., Merry, R.H., Skjemstad, J.O., 1998. Can mid infrared diffuse reflectance analysis replace soil extractions? *Australian Journal of Experimental Agriculture* 38(7), 681-696.
- Janik, L.J., Skjemstad, J.O., 1995. Characterization and analysis of soils using mid-infrared partial least-squares .2. Correlations with some laboratory data. *Soil Research* 33(4), 637-650.
- Janik, L.J., Skjemstad, J.O., Raven, M.D., 1995. Characterization and analysis of soils using mid-infrared partial least-squares .1. Correlations with XRF-determined major-element composition. *Soil Research* 33(4), 621-636.
- Jenny, H., 1941. *Factors of soil formation*. McGraw-Hill, New York, NY.

- Johnston, S.W., 2001. The influence of aeolian dust deposits on alpine soils in south-eastern Australia. *Soil Research* 39(1), 81-88.
- Johnstone, P., Robinson, N., Imhof, M., Rees, D., Hunter, D., Crawford, D., Wells, S., Walsh, R., Faderson, A., Williams, S., Kitching, M., 2010. Systems integration: enhancing soil information delivery to Victoria, 2010 19th World Congress of Soil Science, *Soil Solutions for a Changing World*, Brisbane, Australia, pp. 11-14.
- Kahle, M., Kleber, M., Jahn, R., 2002. Review of XRD-based quantitative analyses of clay minerals in soils: the suitability of mineral intensity factors. *Geoderma* 109(3-4), 191-205.
- Kidd, D., Webb, M., Malone, B., Minasny, B., McBratney, A., 2015. Digital soil assessment of agricultural suitability, versatility and capital in Tasmania, Australia. *Geoderma Regional* 6, 7-21.
- Kuhn, M., Weston, S., Keefer, C., Coulter, N., Quinlan, J.R., 2014. Cubist: Rule- and Instance-Based Regression Modeling, R package version 0.0.18, Available at: <https://cran.r-project.org/package=Cubist>.
- Kvalheim, O.M., 2010. Interpretation of partial least squares regression models by means of target projection and selectivity ratio plots. *Journal of Chemometrics* 24(7-8), 496-504.
- Kvalheim, O.M., Brakstad, F., Liang, Y., 1994. Preprocessing of analytical profiles in the presence of homoscedastic or heteroscedastic noise. *Analytical Chemistry* 66(1), 43-51.
- Lin, L.I., 1989. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 45(1), 255-268.
- Loveland, P.J., Wood, I.G., Weir, A.H., 1999. Clay mineralogy at Rothamsted: 1934-1988. *Clay Minerals* 34, 165-183.
- Lymburner, L., Tan, P., Mueller, N., Thackway, R., Lewis, A., Thankappan, M., Randall, L., Islam, A., Senarath, U., 2011. The National Dynamic Land Cover Dataset. National Earth Observation Group, Geoscience Australia, Canberra, Australia.

- MacEwan, R.J., 1997. Chapter 6 Soil quality indicators: Pedological aspects. In: E.G. Gregorich, M.R. Carter (Eds.), *Developments in Soil Science*. Elsevier, pp. 143-166.
- Madejova', J., Komadel, P., 2001. Baseline studies of the clay minerals society source clays: infrared methods. *Clays and Clay Minerals* 49(5), 410-432.
- Malone, B.P., McBratney, A.B., Minasny, B., Laslett, G.M., 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. *Geoderma* 154(1-2), 138-152.
- Malone, B.P., Hughes, P., McBratney, A.B., Minasny, B., 2014a. A model for the identification of terrons in the Lower Hunter Valley, Australia. *Geoderma Reg.* 1, 31-47.
- Malone, B.P., Minasny, B., Odgers, N.P., McBratney, A.B., 2014b. Using model averaging to combine soil property rasters from legacy soil maps and from point data. *Geoderma* 232-234, 34-44.
- McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. *Geoderma* 117(1-2), 3-52.
- McBratney, A.B., Minasny, B., Viscarra Rossel, R., 2006. Spectral soil analysis and inference systems: A powerful combination for solving the soil data crisis. *Geoderma* 136(1-2), 272-278.
- McKenzie, N.J., Ryan, P.J., 1999. Spatial prediction of soil properties using environmental correlation. *Geoderma* 89(1-2), 67-94.
- Mikhail, E.H., Briner, G.P., 1978. Routine particle size analysis of soils using sodium hypochlorite and ultrasonic dispersion. *Soil Research* 16 (2), 241-244.
- Minasny, B., McBratney, A.B., 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers & Geosciences* 32(9), 1378-1388.
- Mokma, D.L., Jackson, M.L., Syers, J.K., Gibbons, F.R., 1973. Mineralogy and radioisotope retention properties of a chronosequence of soils developed in basalts of Victoria, Australia. *Journal of Soil Science* 24(2), 199-214.

- Mulder, V.L., de Bruin, S., Weyermann, J., Kokaly, R.F., Schaepman, M.E., 2013. Characterizing regional soil mineral composition using spectroscopy and geostatistics. *Remote Sensing of Environment* 139, 415-429.
- Mulder, V.L., Lacoste, M., Richer-de-Forges, A.C., Martin, M.P., Arrouays, D., 2016. National versus global modelling the 3D distribution of soil organic carbon in mainland France. *Geoderma* 263, 16-34.
- Nguyen, T.T., Janik, L.J., Raupach, M., 1991. Diffuse reflectance infrared fourier transform (DRIFT) spectroscopy in soil studies. *Soil Research* 29(1), 49-67.
- Norrish, K., Pickering, G., 1983. Clay minerals, *Soils: an Australian viewpoint*. Division of Soils, CSIRO, Melbourne, pp. 281-308.
- Northcote, K.H., 1979. A factual key for the recognition of Australian soils. Rellim Technical Publications, Coffs Harbour.
- Quinlan, J.R., 1992. Learning with continuous classes. In: A. Sterling (Ed.), *AI '92*. World Scientific, Singapore, pp. 343-348.
- Rajalahti, T., Arneberg, R., Kroksveen, A.C., Berle, M., Myhr, K.-M., Kvalheim, O.M., 2009. Discriminating Variable Test and Selectivity Ratio Plot: Quantitative Tools for Interpretation and Variable (Biomarker) Selection in Complex Spectral or Chromatographic Profiles. *Analytical Chemistry* 81(7), 2581-2590.
- Rees, D., Robinson, N., Imhof, M., Joyce, E.B., MacEwan, R., 2010. Spatial frameworks to support digital soil mapping, 2010 19th World Congress of Soil Science, *Soil Solutions for a Changing World*, Brisbane, Australia, pp. 50-53.
- Rees, D.B., 2000. Land systems of Victoria, Edition 3. Technical report (Centre for Land Protection Research, Vic.); no. 56. Dept. of Natural Resources & Environment, Epsom, Vic.
- Reeves III, J.B., 2010. Near- versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: Where are we and what needs to be done? *Geoderma* 158(1–2), 3-14.
- Rietveld, H., 1969. A profile refinement method for nuclear and magnetic structures. *Journal of Applied Crystallography* 2(2), 65-71.

- Robinson, N.J., Benke, K.K., Norng, S., 2015. Identification and interpretation of sources of uncertainty in soils change in a global systems-based modelling process. *Soil Research* 53(6), 592-604.
- Sargeant, I.J., Skene, J.K.M., 1970. Some properties of the krasnozems of Southern Victoria, Australia. *Soil Research* 8(3), 281-295.
- Savitzky, A., Golay, M.J.E., 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry* 36(8), 1627-1639.
- Sheffield, K., Morse-McNabb, E., 2015. Using satellite imagery to assess trends in soil and crop productivity across landscapes. *IOP Conference Series: Earth and Environmental Science* 25(1), 012013.
- Soriano-Disla, J.M., Janik, L.J., Viscarra Rossel, R.A., MacDonald, L.M., McLaughlin, M.J., 2014. The performance of visible, near-, mid-infrared reflectance spectroscopy for prediction of soil physical, chemical and biological properties. *Applied Spectroscopy Reviews* 49, 136-186.
- Stace, H.C.T., Hubble, G.D., Brewer, R., Northcote, K.H., Sleeman, J.R., Mulcahy, M.J., Hallsworth, E.G., 1968. *A handbook of Australian Soils*. Rellim Technical Publications, Glenside, South Australia.
- Sultan, K., 2006. Clay Mineralogy of Central Victorian (Creswick) Soils: Clay Mineral Contents as a Possible Tool of Environmental Indicator. *Soil and Sediment Contamination: An International Journal* 15(4), 339-356.
- Torn, M.S., Trumbore, S.E., Chadwick, O.A., Vitousek, P.M., Hendricks, D.M., 1997. Mineral control of soil organic carbon storage and turnover. *Nature* 389(6647), 170-173.
- Viscarra Rossel, R.A., 2009. The Soil Spectroscopy Group and development of a global soil spectral library. *NIR news* 20(4), 17-18.
- Viscarra Rossel, R.A., 2011. Fine-resolution multiscale mapping of clay minerals in Australian soils measured with near infrared spectra. *J. Geophys. Res.* 116(F4), F04023.

- Viscarra Rossel, R.A., Behrens, T., 2010. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma* 158(1–2), 46-54.
- Viscarra Rossel, R.A., Bui, E.N., de Caritat, P., McKenzie, N.J., 2010. Mapping iron oxides and the color of Australian soil using visible–near–infrared reflectance spectra. *Journal of Geophysical Research: Earth Surface* 115(F4).
- Viscarra Rossel, R.A., Cattle, S.R., Ortega, A., Fouad, Y., 2009. In situ measurements of soil colour, mineral composition and clay content by vis–NIR spectroscopy. *Geoderma* 150(3–4), 253-266.
- Viscarra Rossel, R.A., Chen, C., Grundy, M.J., Searle, R., Clifford, D., Campbell, P.H., 2015. The Australian three-dimensional soil grid: Australia’s contribution to the GlobalSoilMap project. *Soil Research* 53(8), 845-864.
- Viscarra Rossel, R.A., Jeon, Y.S., Odeh, I.O.A., McBratney, A.B., 2008. Using a legacy soil sample to develop a mid-IR spectral library. *Soil Research* 46(1), 1-16.
- Viscarra Rossel, R.A., McBratney, A.B., 2008. Diffuse Reflectance Spectroscopy as a Tool for Digital Soil Mapping. In: Hartemink, A.E., McBratney, A.B., Mendonça-Santos, M.L. (Eds.), *Digital Soil Mapping with Limited Data*. Springer, pp. 165-172.
- 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.
- Wetherby, K., Oades, J., 1975. Classification of carbonate layers in highland soils of the northern Murray Mallee, S.A., and their use in stratigraphic and land use studies. *Soil Research* 13(2), 119-132.
- Wiklund, S., Nilsson, D., Eriksson, L., Sjöström, M., Wold, S., Faber, K., 2007. A randomization test for PLS component selection. *Journal of Chemometrics* 21, 427-439.

- Wilford, J., 2012. A weathering intensity index for the Australian continent using airborne gamma-ray spectrometry and digital terrain analysis. *Geoderma* 183–184, 124-142.
- Wilson, M.J., 1999. The origin and formation of clay minerals in soils: past, present and future perspectives. *Clay Minerals* 34, 7-25.
- Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* 58(2), 109-130.
- Wong, J.W.H., Durante, C., Cartwright, H.M., 2005. Application of Fast Fourier Transform Cross-Correlation for the Alignment of Large Chromatographic and Spectral Datasets. *Analytical Chemistry* 77(17), 5655-5661.
- Yitagesu, F.A., van der Meer, F., van derWerff, H., Hecker, C., 2011. Spectral characteristics of clay minerals in the 2.5–14 μm wavelength region. *Appl. Clay Sci.* 53 (4), 581–591.
- Yuan, G., Theng, B.K.G., 2012. Clay-Organic Interactions in Soil Environments. In: P.M. Huang, Y. Li, M.E. Sumner (Eds.), *Handbook of soil sciences: resource management and environmental impacts*. CRC Press, Taylor & Francis, Florida, U.S.A., pp. 20.

Chapter 9 Conclusions

The primary objective of this thesis was to advance knowledge on Digital Soil Mapping (DSM) and how practitioners and users of spatial soil information can benefit from the presented research. This chapter summarises the key findings from the literature review (Chapter 3) and five chapters in response to the three research objectives posed in Chapter 1. Research conclusions are summarised from these investigations that can be used to support future developments in DSM.

Discussion

This thesis presents examples of research to:

better understand the spatial soil information needs of users;

identify error sources contributing to aleatory and epistemic uncertainty that should be accommodated in modelling and mapping applications, and

harmonise legacy methods with new analytical techniques to map soil properties linked to supporting functions delivered by soil.

Chapter 4 presents a concise summary of the spatial soil information needs of biophysical modellers. A repeat survey of this user group was timely, given the global focus on soil and land related issues of primary production to meet global food requirements, poverty and concerns on the impact of climate change. The survey identified key soil properties that affect model sensitivity and align with GlobalSoilMap priorities, e.g. plant available water capacity, clay%, and organic carbon. Given the paucity of published literature on user needs for spatial soil information, the survey provides a valuable benchmark to understand how needs of modellers continue to evolve in response to environmental and agricultural issues.

A focus of chapters 5, 6 and 7 was to accommodate various error sources in spatial soil information (aleatory and epistemic) using a systematic framework for mapping and modelling purposes. Traditional thinking on uncertainty has been largely reductionist, and there is a need to think holistically to accommodate the many potential error sources using systems-based models. Conventional uncertainty assessment has focused on statistical uncertainty and overlooked other possible error sources due to lack of information, e.g. measurement error, expert opinion, incorrect context or environment. Chapter 5 presents the Global Representation of Uncertainty in the Modelling Process (GRUMP) as a framework to embody, and to illustrate to users these error sources and how they were accommodated in the map or model.

Implementations of the GRUMP framework using examples are provided in chapters 5, 6 and 7. Chapter 6 demonstrates different elements of uncertainty by focusing on one soil property (pH) and how expert opinion can be an important source of measurement error. The implementation of the GRUMP in Chapter 7 demonstrates how factors that are often overlooked in mapping and modelling, such as accounting for temporal variability in a soil property and changes in land use (context), can contribute to soil change and error in maps and models.

A novel example using legacy data, archive soil samples and Mid-Infrared (MIR) spectroscopy with a spatial inference system to map clay minerals is presented in chapter 8. This example exploited the wealth of stored archive samples to map a property that is expensive to acquire and are poorly connected to existing model approaches. The method was applied successfully to map soil clay mineral distribution for kaolinite, illite and smectite according to GlobalSoilMap specifications for western Victoria.

The thesis makes a valuable contribution to support future assessments of soil functions and processes. Land degradation issues such as acidification are a global issue and examples presented in chapters 5, 6 and 7 provide better understandings on this its distribution and delivery of information that is more certain and tailored towards meeting land manager's needs. Novel research to map the clay mineral distribution in soil (Chapter 8) could be used to support spatial quantification of soil services and functions such as plant health and production, carbon sequestration and the resilience of soil in response to climatic and human induced impacts.

Summaries of research findings

Research objective 1: Identify what are users' needs for spatial soil information and how this has changed in Victoria over the last century

Key messages

- Eight key properties are of immediate focus to support biophysical models. These properties are: critical lower limit/permanent wilting point (CLL/PWP); drained upper limit/field capacity (DUL/FC); hydraulic conductivity (Ksat); clay proportion (clay%); bulk density; organic carbon; soil depth; and effective rooting depth. Soil pH and clay mineralogy became the properties of interest for this thesis as they are used regularly in land evaluation and are linked to soil functions and processes, e.g. buffering capacity.
- The number of biophysical models being used for landscape analysis has decreased, but the frequency with which those models that are still being used has increased (especially for point/site based models).

- A common response from modellers is that high resolution soils data is often not required for their purposes – rather their preference is for accessible, available and contemporary soil data.
- Data harmonisation and uncertainty are important to ensure future relevance and accuracy of the models from the influx of new soil data sources, e.g. in-field sensors.
- The needs of users through time have reflected the major challenges or issues to government, e.g. settlement in the 1880's to 1950's, agricultural production from the 1950's to 1980's and 2010's, conservation from the 1930's through to the 1990's, and urban development from the 1980's to 2000's. Currently there is a global-national-state focus on primary production and meeting global food security needs.

Throughout the last century for many countries across the globe, market forces (e.g. government policy intervention or industry development such as adoption of precision agriculture) are the principle causes for ebbs and flows in the supply-driven or demand-driven provision of soil information. The connection between supply and demand ultimately defines the value of a service, and this is no different for spatial soil information.

This disequilibrium due to either excess supply or demand tends to correspond with soil survey phases throughout history. For example, general purpose surveys reflect periods of over supply with the impetus for survey largely driven by the pedological community. The need for spatial soil information requires a clear mandate supported by the citizens rather than purely supply driven scientific considerations (Bouma and Drooge, 2007).

Conversely, surveys undertaken for specific purposes correspond with demand driven periods in history where there was a perceived need by communities, industries or government for soil information to assist their decision making. Here the connection with real-world situations rather than just research studies with little relevance to current agro-ecological conditions are required.

What makes this relationship between supply and demand volatile is the time factor. The provision of new soil information relies upon time in the supply relationship. Here the market must be responsive to identify these future demands and that adequate resourcing is provided to respond to this demand. Meeting the needs of biophysical modellers as primary users of soil information is one such example. Modellers demand for soil information has fluctuated depending upon the environmental and agricultural production questions posed and scale (both space and time) of information sought. In Australia, the agriculture versus environment paradigm for soil mapping (Bouma, 1989) has endured for nearly a century and currently oscillates in favour of agriculture.

What is clearly evident from the literature on spatial soil information is that benefits to users far outweigh the costs of investment. So why does the perception of poor investment persist for soil survey? Is it purely that government now sees soil survey as a user-pays argument and therefore an option to invest elsewhere with higher perceived benefit-to-cost ratios, e.g. human health? Both Manderson and Palmer (2007) and Martin (1980) advocate that new soil survey and the provision of spatial soil information should not be borne by farmers, rather well-resourced organisations such as government that should be committing to long-term programs to support custodians of the land and producers of food and fibre for global communities.

As any good business operator will attest to, in dynamic markets there is always a need to be responsive and to identify what are the current, and the future needs of consumers. In this thesis, Chapter 3 describes the evolving spatial soil information demands in Victoria, Australia. The initial demand for soil survey due to irrigation failure along the Murray River and the transition of the market with advances in technology, changes in societal structures and demographics, and understandings of the agro-ecological environment are defined.

The survey of a biophysical modelling community (Chapter 4) established that spatial soil information sought by these users should be contemporary and available. While a decrease in the number of biophysical models being used (and possibly developed?) was observed, there has been a corresponding increase in the application of models for agricultural or environmental purposes. This is especially true for point based models such as APSIM (Keating and McCown 2001) and its commercially accessible equivalent Yield Prophet (www.yieldprophet.com.au/yp/wfLogin.aspx).

So how sensitive are these models to spatial soil information? A common response from modellers was that high resolution soil information is often not required for their purposes, rather they favour easy access to spatial soil information online. This is consistent with responses from modellers as part of a user needs analysis for the Australian Soil Resource Information System (ASRIS). Wood and Auricht (2011) found that *'modellers will continue to require access to raw (primary site) data and continue to develop their own derived information and surfaces'*.

The use of biophysical models for agricultural industries was greatest where industries were larger users of land area. For intensive industries that operate on a small areal footprint (e.g. horticulture), there is a focus on technological innovations rather than the

implementation of biophysical models to support their operations. This may be due to high resolution demands of these industries that are beyond the scale of most currently available spatial soil information. For agricultural industries where spatial soil information are used, soil properties related to water availability (especially for plants) were identified as the most critical to the implementation of these models. Information providers need to be responsive to market demand by providing spatial soil information that are focused on these soil water properties of interest.

A contrary argument is: why are properties such as pH and EC that are recognised in land evaluation techniques as key properties to land capability and suitability assessments, of little or no interest, to the modelling community? The possible decline in biophysical model development corresponds with sentiments expressed by Bouma (2001) that there is a tendency to make plant growth models more sophisticated while soil and land properties remain static leading to unbalanced models. The relative importance of soil as a medium for supporting plant growth and production systems in models appear to be downplayed. There is a definite need for model developers and soil information providers to work more closely through participatory action (Bouma 2001) to formulate comprehensive solutions to agronomic and environmental problems that are defined.

The development of big data sources such as new sensors and sensor networks will produce vast quantities of data on the environment including soil. These new sources of data will need to be evaluated, harmonised, and included in modelling solutions that incorporate uncertainty and risk. The advent of community sourced soil data, access to tacit knowledge, precision agriculture systems and physical collection of new soil samples to calibrate spatial soil information products (Rossiter et al., 2015) are exciting prospects to the delivery of contemporary and high-resolution spatial soil information.

Research objective 2: Develop an approach to accommodate, and illustrate to users of spatial soil information, the various error sources in modelling and mapping.

Key messages

- The Global Representation of Uncertainty in the Modelling Process (GRUMP) framework provides a systematic design to incorporate error sources in a comprehensive uncertainty approach.
- Epistemic uncertainties can be significant for users of maps and models if not embodied in the map making or model implementation process.
- Legacy data sources of soil information can be valuable to understand agro-ecological condition and changes to soil, however, users must be cautious and aware of potential limitations and constraints associated with the data.
- Understanding fitness-for-use and the risk perception for users of spatial soil information through participatory action should be encouraged to support ongoing and valued services delivered to users.

Uncertainty analysis in modelling and mapping has focused on statistical variability and error as the uncertainty metrics reported to users of these products. For global sensitivity and uncertainty assessments to be successful, the process should include a model selection phase (epistemic uncertainty) coupled with the statistical variability via error propagation or other approaches (aleatory uncertainty).

A new framework formulated to support uncertainty assessment (the GRUMP) has been defined in Chapter 5 that builds upon foundations of uncertainty assessment from Refsgaard et al. (2007) and Walker et al. (2003). The approach includes model inputs,

model structure and implementation with additional error sources such as expert opinion, legacy data issues and software operations. The explicit illustration and enumeration of the error sources and uncertainty to users is also of benefit to support the modelling and mapping communities towards iterative improvement of maps and models. The GRUMPs global applicability and benefit to users is that it enables modellers and users to be conscious of uncertainties that may be forgotten or assumed as trivial, where as they may be detrimental to the running of a model or delivery of a useful map.

By considering epistemic uncertainties that are often taken-for-granted (or assumed) in mapping and modelling procedures, the framework provides a global system-based approach that considers error sources more comprehensively than previous approaches. Once epistemic uncertainty can be quantified then it can be integrated with statistical variability and uncertainty classified by global metrics, such as total system variance (assuming normally distributed errors), or system prediction interval (assuming skewed distributions).

In DSM, no map is perfect and all contain some level of error due to the array of integrated factors used in its creation. Unfortunately, uncertainty analysis is often considered at the end of an analytical process rather than its coordination with the spatial prediction technique from the planning phase to map production. All too often, there are successes and failures in the implementation of DSM without adequate examination of the environmental characteristics, the pedogenic hypothesis and justification, the spatial soil information available for calibration and validation, the adequacy of the spatial prediction method, and above all, the assumptions and epistemic uncertainties not considered in the initiation phase of a project.

Practical examples of factors used in the implementation of the GRUMP framework are presented in this thesis (chapters 5, 6 and 7). Error sources embedded in legacy data can be relatively important when considering factors such as land use change and resultant impacts to soil. The field pH kit example (Chapter 7) demonstrate that there are numerous factors (e.g. assessor experience and test kit variety) contributing to measurement error as an epistemic uncertainty. This has practical implications to those using field pH kits for mapping or modelling, or to manage pH levels through addition of ameliorants such as lime. Guidelines are suggested including training and certification of users, regular testing of field kits and collection of associated metadata (e.g. time of observation) towards more certain predictions. There is now the capability to screen these legacy records more closely as the magnitude of errors associated with field pH data has been identified. Field pH data that is cleansed can be used as a covariate to support modelling approaches for more precise and accurate measurements such as the mapping of laboratory pH.

The spatio-temporal assessment of change in soil properties using legacy data and model-based designs is gaining considerable interest (refer to Appendix A) as results are expediently derived and it is generally cheaper to implement than formal design based monitoring systems. An example presented in Chapter 5 demonstrates that simple statistical analysis that don't account for epistemic uncertainties or statistical assumptions can result in misleading conclusions. This can be the case when sample size and statistical power is insufficient for assessment of change in hypothesis testing resulting in type I (reject H_0 when true) or type II (accept H_0 when false) errors. For users such as government or regional authorities that make decisions on misleading information, there may be long-term serious consequences to the environment as a result.

By accounting for potential error sources using the GRUMP framework to make maps based on consideration of factors such as seasonal variability and land use affects, map

users can have information with greater certainty and utility to make decisions. As legacy soil data issues such as temporal variability of a soil property such as soil pH are rarely, if ever addressed in soil maps, there are tremendous opportunities to improve the quality and utility of soil maps for users.

There remains a greater need for soil mapping and monitoring to come together. As there are considerable changes occurring in farming systems, management practices and the continual motivation to produce more from less, these impacts on land need to be more closely monitored and understood. The case study presented in chapter 6 highlights that there is a real need to engagement between land users, soil scientists and spatial scientists to achieve goals and outcomes that are shared, e.g. reduction in the primary production losses due to soil acidification.

Potential enhancements that should be encouraged in mapping and modelling are multiple Monte Carlo simulation experiments as a special case of a GRUMP implementation. Here the inputs are replaced by probability distributions and then each distribution would be progressively replaced with the mean value, and a corresponding new Monte Carlo simulation executed in a manner analogous to step-wise multivariate regression involving forward selection and backward elimination. The output distribution for each separate simulation can be represented by an uncertainty metric, such as the variance or confidence interval, and then compared with the corresponding input distributions in the form of a stochastic sensitivity plot. The inputs responsible for the greatest output uncertainty were consequently identified and ranked. This approach has similarities to the error budget approach suggested by Nelson et al. (2011) to identify the primary error sources in the production of a digital soil map.

Importantly, the notions of uncertainty and fit-for-purpose need to be combined using risk concepts to highlight the importance, and risk acceptable to a decision maker. The assessment of fitness-for-use is valuable as it contrasts a risk-in-a-decision with the risk acceptable to the decision maker, is simple and cheap to implement and easily understood by users (Agumya and Hunter, 1999). In chapter 7, a scenario is highlighted where it is the critical pH ranges linked to primary production that are of greatest interest to a land manager. Therefore, while high precision in pH determination is useful, in many respects the higher precision is potentially only warranted in the pH range of 5.3 to 5.8 where cultivars become sensitive to toxic effects from macro and micro nutrients. This pH range represents a high risk zone to the farmer, whereas above this range the impacts of decisions are likely to be of less consequence, and below the impacts is implicit, e.g. reduced plant growth.

Ultimately the goal of all mapping and modelling should be to provide greater certainty to users. Model ensemble techniques as suggested by Finke (2012) and Malone et al. (2014) are worthy of greater consideration in DSM exercises to benefit from the wealth of spatial inference techniques and environmental covariates that are now freely accessible.

Research objective 3: Investigate the potential use of legacy data supplemented with new spectroscopic predictions to predict the regional distribution of two key soil properties - pH and clay mineralogy – for areas of western Victoria.

Key messages:

- Accurate quantitative spectroscopic models for clay mineral composition and pH using whole soil samples (<2 mm) can be achieved using MIR spectroscopy and analysis methods such as Partial Least Squares Regression (PLSR).

- Relatively few calibration samples are required to develop a robust spectroscopic model for clay mineralogy.
- By incorporating legacy data with new spectral methods and tailored probability based schemes, maps of direct applicability can be used for on-ground management, e.g. pH.
- There remains a deficiency of soil samples at depth (>1m). As a result potential predictive models deteriorate due to the low number of available sites. Integration with other sources or samples, e.g. bore logs and regolith interpretations may improve model predictions and support prediction of the clay mineral continuum to greater depth.
- Vegetation indices were useful predictors of clay mineral occurrence. This requires further exploration but hints at possible links between soil processes attributed to clay mineralogy (e.g. fertility, soil-water regime) and primary production.

Prediction of soil properties such as pH is commonly undertaken in DSM as it is seen as a key indicator of threatening processes such as acidification and impacts to agriculture (e.g. reduced plant production) and environment (e.g. acid sulphate soils). While clay mineralogy is a fundamental contributor to the chemical and physical behaviour of soil, it is unclear why mineralogy has not been of greater focus in spatial modelling and mapping efforts to date (Grunwald, 2009). With current and existing legacy survey programs across the globe, there are opportunities using new spectroscopic techniques and archived soil samples, and legacy quantitative measurements to develop inexpensive and rapid assessments of properties such as clay mineralogy. The example presented in this thesis

for western Victoria (Chapter 8) suggests relatively few calibration samples are required to produce reliable estimates of difficult to measure (expensive or time consuming) properties such as clay mineralogy.

The application of model-trees for DSM purposes has proven reliable and effective at various scales across the globe (e.g. Kidd et al., 2015; Viscarra Rossel et al., 2015; Mulder et al., 2016). Further enhancements to the approach implemented in this thesis warrant investigation, e.g. accounting for spatial dependence in soil clay mineral predictions. The implementation of spatial inference model ensembles that can harness the strengths and mitigate the weaknesses of the various combined approaches should also be considered. This has been highlighted by Finke (2012) and Malone et al. (2014) to reduce associated uncertainties in maps delivered to users. The implementation of linear mixed models using REML-EBLUP as advocated by Lark et al. (2006) could be implemented more widely. This technique has proven prohibitive for large data sets, but enhancements through the inclusion of filtering (Cressie and Kang, 2008), clustering and ranking algorithms to address this issue are being pursued (Chia and Robinson, Personal Communication). The implementation of a Linear Mixed Model (LMM) incorporating additional factors contributing to model uncertainty was successfully implemented to map soil pH using maximum likelihood and conditional simulation. This approach with use of critical thresholds for agronomic production purposes enables spatial soil information to be presented to users in a form that can be more directly linked and used for land management decisions.

Samples and independent validation

In most DSM applications, there is a dependency on legacy sites and samples for calibration and validation purposes. These soil sites may be clustered in space and time with no overall coordinated formal sample design across these domains. Ideally,

independent validation using a random sample design is advocated, although cost and time of implementation often prohibit this from occurring for DSM at all scales. The cross-validation approach used in the 3D prediction of clay mineralogy in this thesis is based upon the method advocated by Malone et al. (2014) with a 10-fold cross-validation procedure that averaged the model tree predictions from the 10 model runs using 90% of the data for calibration and 10% for validation purposes. While not optimal and vulnerable to bias due to use of a non-probability derived sample set, the cross validation is considered better than no validation at all. Furthermore, this cross-validation procedure enabled production of lower and upper prediction interval maps to give users guidance on the map qualities including accuracy and uncertainty.

A high dependence upon legacy data, archived samples and new spectroscopic prediction also mean that it is either by chance or good luck, if the sample data set adequately represents the pedo-geomorphic diversity for the region of interest. In the clay mineralogy example, an inadequate spatial coverage for physiographic regions including the Southern Uplands and Northern Riverine Plains of Victoria were identified. From expert opinion (Dahlhaus, Personal Communication) and other existing evidence on mineral occurrence for these landscapes, the clay mineral composition of soils in these landscapes may be inadequately represented in the spatial predictions, e.g. low illite predictions for the Southern Uplands. This is further complicated by a poor representation of sites for these landscapes with samples at depth > 1 m (as discussed earlier). For the pH example (Chapter 6), a combination of laboratory, MIR and field pH values were used to provide the most comprehensive set of pH observations to complement land use and management system changes over time. Supplementing these infrequent samples with new sites and samples, or other sources of clay mineral or pH data from bore lithological samples (e.g.

piezometer nests) or crowd sourced data could be considered where these samples are preserved.

Spectroscopy

Analytical methods such as X-Ray Diffraction (XRD) continue to be refined and improved with higher sensitivity and reliability. When coupled with complimentary techniques such as scanning and transmission electron microscope and particle accelerator techniques available through a synchrotron, the prediction of clay minerals with high precision is now achievable. Spectroscopic analysis using samples stored in soil archives provides a rapid approach to populating spatial soil information systems with missing or not observed properties such as clay mineralogy. The implementation of spectroscopy for predicting clay minerals is not new (Soriano-Disla et al., 2014) although few studies have quantitatively assessed these. Predominantly, these soil mineralogy studies have focused on the visible and near-infrared frequency regions on the electromagnetic spectrum. The prediction of pH from MIR spectroscopy has proven favourable although sample diversity can impact model calibration (Soriano-Disla et al., 2014). Results presented in Chapter 6 for Victoria include a R^2 of 0.88 and RMSE of 0.56 (roughly equivalent to the error associated with field pH determination). The analysis presented in chapter 8 is novel for a number of reasons: 1. the use of MIR to predict the dominant layer phyllosilicates for extensive sample sets (>10,000 samples) has not been undertaken previously for Australian soils; 2. quantitative XRD measurements were used for calibration purposes rather than relying on particular diagnostic absorption features and analytical techniques such as continuum removal that are 'relative' determination procedures; 3. spectroscopic mineral models were based on whole soil samples with predictions for the <2 μm fraction only.

Various spectroscopic modelling techniques such as continuum removal have been successfully implemented to predict various soil properties. PLSR which is used widely for quantitative prediction of soil properties was successfully applied to predict clay mineral composition using quantitative XRD data for calibration purposes. The advantage of PLSR models is their simplicity to implement and interpret by examining significant wavelengths through cross-comparison techniques of the variable importance of projection (VIP) and selectivity ratio (SR) scores. By reducing the spectral data set to the wavelengths of interest as defined from diagnostic absorption features in the literature, or the VIP and SR scores, this will remove much of the spectra with relatively little deterioration of the spectroscopic model. Future options to improve the spectroscopic models would benefit from the inclusion of Support Vector Machines (SVM) which can account for non-linearity as part of spectroscopic model ensembles.

Where improvements may be possible to clay mineral models is by accounting for uncertainty in quantitative XRD measurements that have associated measurement errors defined. This is often the case for legacy quantitative measurements where errors are significantly larger than contemporary assessments using latest advances in quantitative XRD analytical procedures and instrumentation. Potential re-examination of these legacy x-ray spectra may enable revised quantitative measurements for clay minerals of interest and improvement of spectroscopic models.

Sometimes in mapping and modelling exercises we are unsure how well the initial sample set represents the true population. In the development of spectroscopic models for kaolinite, illite and smectite, calibration samples and predictions for illite were non-Gaussian with few values above 60%, and for smectite there are few above 80%. Generally for soils of western Victoria the quantity of kaolinite and illite decreased with depth while smectite increased. Both illite and smectite exhibited positively skewed

distributions while kaolinite was normally distributed as the dominant clay mineral of soils in western Victoria. In general terms the relative order of abundance is kaolinite > smectite > illite.

Overall research conclusions

Spatial soil information and its provision have been beneficial to all branches of society. Today there are many challenges and threats to using soil in a sustainable manner. With advances in technology and increasing knowledge on landscape process, providers of soil information must be responsive and pro-active in the formulation of innovative solutions for primary production and ecological purposes.

The research presented in this thesis aims to improve the current understanding on DSM and how information that soil mapping practitioners deliver can be improved by: understanding the specific needs of users; provide greater certainty in the spatial soil information delivered to users, and produce spatial information on soil that is linked to services and functions of soil that benefits from legacy soil data and information.

The key findings established through research investigations include:

- Users of spatial soil information are seeking accessible and contemporary soil properties that will support models and assessments on land for agricultural and environmental purposes. Soil moisture characteristics, carbon content and clay% are sought by modellers for application at global to local scales. By focussing on these, we concentrate efforts on key provisioning roles of soil and factors linked to degradation of soil resources.

- A new uncertainty framework was presented that supports practitioners and users of spatial soil information to consider the diversity of error sources and implications to uncertainty in maps and models. The GRUMP provides a systematic approach where error sources of significance to users of maps can be embodied, illustrated and quantified in the uncertainty assessment.
- Sources of uncertainty in legacy data were defined for application in soil mapping examples for pH. By considering these factors in more holistic approaches to DSM, there are opportunities to deliver information that are more certain, and therefore of greater utility and of less risk to decision-makers. This is particularly useful given the importance of soil acidification as a global issue and losses to production that are known to occur in western Victoria.
- Spectroscopic models for clay minerals and spatial inference techniques were developed to predict the 3D spatial distribution of phyllosilicates. This integration of spatial models and spectral models provide a valuable example to exploit available archive samples to predict properties never before assessed due to their expense or difficulty to acquire for large geographical areas.

Future work

The research presented has established that there are still many questions that remain to be addressed, and consequently were not covered in this thesis. These questions are focused on DSM and the provision of usable spatial soil information by a community of current and future users.

- The example of using spectroscopic models with spatial inference techniques for clay mineralogy highlight the potential to expediently predict and validate the

occurrence, distribution and impact of properties linked to functions and processes delivered by soil, e.g. primary production. Supporting industries and communities with relevant information can lead to shared contributions to lift production and protect the environment, e.g. spatial delineation of soil limiting nutrition to grains production, degraded soils, threatened biomes in soil. This can be achieved through the collections of archived soil samples, new analytical techniques and synthesizing these together in spatial prediction models. Model ensemble techniques can be tested as part of real applications (Digital Soil Assessments) focus on land use and management, e.g. soil acidification and yield response to amelioration.

- Spatial inference model ensembles are advocated and offer the potential of reduced uncertainty and more accurate spatial soil information for users. Better spatial information should be the number one priority for all DSM practitioners.
- The expanded uncertainty logic and framework presented in this research can be refined to include Monte Carlo simulation methods and other stochastic and epistemic error sources in DSM applications. Deployment of uncertainty approaches guided by the GRUMP framework would be beneficial to support isolation and identification of error sources using different spatial inference methods, e.g. data mining techniques as compared to a Linear Mixed Model using geostatistical techniques to model error contributions. This should enable users and practitioners to focus and investment their efforts to reduce significant error and provide greater certainty in spatial soil information.
- The GRUMP framework provides an ‘illustrative’ guide to consider uncertainty in mapping and modelling. The concept can be deployed as a fully functional and

operational system akin to the Data Uncertainty Engine (Heuvelink and Brown, 2007).

- The focus on research investigations has principally been on soil pH, but there are many soil properties with real and potential error sources that may be overlooked in existing DSM examples. A priority should be to further knowledge on key soil properties and their errors linked to soil services and functions such as soil-moisture characteristics, organic carbon and clay%.
- Results presented for the prediction of kaolinite, illite and smectite are encouraging, and suggest that other minerals could be expediently predicted using 3D spatial prediction techniques. Investigations on the links of these minerals with other properties of global interest such as carbon sequestration potential, engineering characteristics and natural capital concepts are encouraged.
- Ongoing engagement with users is advocated. There is a continuing need to test and evaluate needs of users for spatial soil information to guide the delivery of maps that are pertinent, trustworthy and reliable. This will support a clear mandate for spatial soil information that is driven by users rather than suppliers. This must include modellers in participatory solutions to real-world solutions for agricultural and environmental issues.
- Currently there is a diverse set of DSM techniques to map and predict soil properties or classes, but it is uncertain which technique will return the best results given available soil observation and environmental variables from one area to another. DSM techniques and approaches require further evaluation to understand their robustness and utility across different landscapes and the contributing factors that can be resolved leading to higher certainty information delivered.

- New sources of soil data and information on earth are becoming available including citizen science, crowd sourcing, new sensor networks and satellite platforms. Integration of these data sources potentially represent ways of delivering information that is of a higher resolution, is cheap and inexpensive to acquire, can provide greater certainty and usability for decision making. Federating data from disparate databases using web services (WFS) and then dynamically modelling using web processing services (WPS) should enable refinement and higher certainty in delivery of soil coverages via Web Coverage Service (WCS).

Digital soil mapping has evolved as a discipline over the last 40 years that now integrates field, laboratory and proximal soil observations with quantitative methods to infer soil properties and classes of various spatial and temporal scales (Grunwald, 2010). While DSM has progressed to an operational phase to support needs across various spatial and temporal scales, there are tremendous opportunities to enhance the science and the development of DSM through the participation and interaction of communities including modellers and information users. This thesis has gleaned some useful findings from DSM assessments in western Victoria that can be used, but more importantly, should be considered in the provision of usable spatial soil information.

References

Agumya, A., Hunter, G.J., 2000. Assessing "Fitness for Use" of Geographic Information: What Risk Are We Prepared to Accept in Our Decisions? In: K. Lowell, A. Jaton

- (Eds.), *Spatial Accuracy Assessment: Land Information Uncertainty in Natural Resources*. CRC Press pp. 35-43.
- Bouma, J., 1989. Using Soil Survey Data for Quantitative Land Evaluation. In: Stewart, B.A., (Ed.), *Advances in Soil Science* 9, Springer, pp. 177-213.
- Bouma, J., 2001. The role of soil science in the land use negotiation process*. *Soil Use and Management* 17(1), 1-6.
- Bouma, J., Droogers, P., 2007. Translating soil science into environmental policy: A case study on implementing the EU soil protection strategy in The Netherlands. *Environmental Science & Policy* 10(5), 454-463.
- Cressie, N., Kang, E.L., 2010. High-Resolution Digital Soil Mapping: Kriging for Very Large Datasets. In: Viscarra Rossel, R.A., et al. (Eds.), *Proximal Soil Sensing, Progress in Soil Science*. Springer, The Netherlands, pp. 49-63.
- Finke, P.A., 2012. On digital soil assessment with models and the Pedometrics agenda. *Geoderma* 171–172, 3-15.
- Grunwald, S., 2009. Multi-criteria characterization of recent digital soil mapping and modeling approaches. *Geoderma* 152(3-4), 195-207.
- Grunwald, S., 2010. Current State of Digital Soil Mapping and What is Next. In: Boettinger, J.L., et al. (Eds.), *Digital Soil Mapping: Bridging Research, Environmental Application, and Operation*. Springer, London, pp. 3-12.
- Heuvelink, G.B.M., Brown, J.D., 2007. Towards a soil information system for uncertain soil data. *Developments in Soil Science*, pp. 97-106.
- Keating, B.A., McCown, R.L., 2001. Advances in farming systems analysis and intervention. *Agricultural Systems* 70(2–3), 555-579.
- Kidd, D., Webb, M., Malone, B., Minasny, B., McBratney, A., 2015. Eighty-metre resolution 3D soil-attribute maps for Tasmania, Australia. *Soil Research* 53(8), 932-955.

- Lark, R.M., Cullis, B.R., Welham, S.J., 2006. On spatial prediction of soil properties in the presence of a spatial trend: the empirical best linear unbiased predictor (E-BLUP) with REML. *European Journal of Soil Science* 57(6), 787-799.
- Malone, B.P., Minasny, B., Odgers, N.P., McBratney, A.B., 2014. Using model averaging to combine soil property rasters from legacy soil maps and from point data. *Geoderma* 232–234, 34-44.
- Manderson, A., Palmer, A., 2006. Soil information for agricultural decision making: a New Zealand perspective. *Soil Use and Management* 22(4), 393-400.
- Martin, J.J., 1980. Justification and financial responsibility for soil survey. Department of Agriculture, Melbourne, pp. 3.
- Mulder, V.L., Lacoste, M., Richer-de-Forges, A.C., Martin, M.P., Arrouays, D., 2016. National versus global modelling the 3D distribution of soil organic carbon in mainland France. *Geoderma* 263, 16-34.
- Nelson, M.A., Bishop, T.F.A., Triantafilis, J., Odeh, I.O.A., 2011. An error budget for different sources of error in digital soil mapping. *European Journal of Soil Science* 62(3), 417-430.
- Refsgaard, J.C., van der Sluijs, J.P., Højberg, A.L., Vanrolleghem, P.A., 2007. Uncertainty in the environmental modelling process – A framework and guidance. *Environmental Modelling & Software* 22, 1543-1556.
- Rossiter, D.G., Liu, J., Carlisle, S., Zhu, A.X., 2015. Can citizen science assist digital soil mapping? *Geoderma* 259–260, 71-80.
- Walker, W.E., Harremoës, P., Rotmans, J., Van der Sluijs, J.P., Van Asselt, M.B.A., Janssen, P., Kreyer Von Krauss, M.P., 2003. Defining Uncertainty, A Conceptual Basis for Uncertainty Management in Model-Based Decision Support. *Integrated Assessment* 4, 5–17.
- Wood, B., Auricht, C., 2011. ASRIS / ACLEP user needs analysis. Australian Collaborative Land Evaluation Program.

Appendix A

Appendix B

Appendix C