

Monitoring and modelling spatio-temporal soil change in a semi-arid irrigated cotton- growing region of south-west NSW, Australia

– The impacts of land use and climatic fluctuations

Patrick Filippi

B. Sc. Agriculture (Hons)

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

2017

School of Life and Environmental Sciences

Faculty of Science

The University of Sydney

New South Wales

Australia



THE UNIVERSITY OF
SYDNEY

Statement of originality

This thesis is submitted to the University of Sydney in fulfillment of the requirements for the Doctor of Philosophy. The work presented in this thesis is, to the best of my knowledge and belief, original except as acknowledged in the text. I hereby declare that I have not submitted this material, either in full or in part, for a degree at this, or any other institution.

P. Filippi

Patrick Filippi

15/09/2017

Thesis Summary

Soil is an invaluable finite resource, and there is considerable interest in monitoring the status of soil, as well as the direction and degree of any changes in soil attributes. Land use change and agricultural management have the capacity to alter the properties of soil considerably over relatively short time scales, however, it is less clear how recent climatic shifts observed throughout the globe will influence changes in soil condition. In the semi-arid regions of eastern Australia, there has been an expansion of irrigated cotton production, which is considered to be a very intensive land use with vigorous management practices. These regions have also been exposed to significant fluctuations in rainfall patterns in the last decade or so. While these semi-arid areas are agriculturally important, they often possess distinct soil characteristics, such as high levels of alkalinity, salinity, sodicity, inorganic carbon, and low levels of organic carbon. This body of work focuses on monitoring the change in soil condition in the semi-arid irrigated cotton-growing district of Hillston in the lower Lachlan River valley catchment in south-west, New South Wales (NSW), Australia. Data from soil cores extracted to 1.5 m depth from two soil surveys performed in 2002 and 2015 were used to monitor the change in several important soil properties – pH, electrical conductivity, exchangeable sodium percentage, organic carbon, and inorganic carbon. It is anticipated that the significant shifts in land use and rainfall patterns could have altered the condition of soil during this period.

Rather than using traditional digital soil mapping techniques, such as regression kriging or machine learning approaches, this study focuses on using linear mixed models, which are particularly advantageous for monitoring changes in soil properties as they can account for correlation in space and time. In this work, the focus is on using bivariate linear mixed models (BLMMs) and multivariate linear mixed models (MLMMs) to create digital maps of the various soil properties. In the BLMM approach, one model is used to predict a soil property from both time points at a single depth, which results in improved soil maps that have a logical connection through time. The MLMM approach is similar to the BLMM approach, but multiple depths are also be modelled simultaneously in addition, which results in more coherent connections between the different sampling depths. Another strong advantage of using these approaches to monitor soil is that the correlation between the monitoring periods is used to improve the sensitivity of the model to detect statistically significant changes. Traditional laboratory methods of measuring certain soil properties can be expensive and laborious. This study used visible near infrared (VisNIR) spectroscopic techniques to rapidly predict soil exchangeable sodium percentage (ESP), soil organic carbon (SOC) content, and soil inorganic carbon (SIC) content to overcome this.

Chapter 2 is a review of the literature on monitoring and modelling soil change. This chapter describes the concept of spatio-temporal soil change and discusses the impact that anthropogenic activity and climatic shifts can have on a suite of soil properties. There is a particular focus on different methods of monitoring and modelling soil change, and many examples of studies are provided that range considerably in space (from field to continental scales) and time (from years to centuries). This review also discusses the many challenges in monitoring soil change, as well as the opportunities to detect changes in soil condition. To prevent repetition in the different research chapters of this thesis, Chapter 3 describes the study area, the soil datasets, soil property analysis, the covariates used for mapping and modelling, and the approaches used to model these soil properties in space and time.

Chapter 4 is the first research chapter of this thesis, where much of the materials and methods described in Chapter 3 are put into practice. This chapter looks at the status of soil pH and the change in pH from 2002 to 2015. An overall acidification trend through time was observed throughout the soil profile, with more widespread changes in the subsoil, however, only isolated parts of the study area were predicted to have experienced a statistically significant change in soil pH over time. This acidification was at least partly due to management practices associated with irrigated cotton production, but was primarily attributed to the leaching of basic cations and dissolution of carbonates, caused by increased water flow throughout the profile as a result of periods of very high rainfall. While soil acidification trends generally cause significant concern, this is less of an issue in the highly alkaline soils of this semi-arid region.

Chapter 5 concerns the changes in soil salinity and sodicity from 2002 to 2015 in the study area. While soil electrical conductivity (EC) is easily measured by traditional laboratory methods, analysing soil exchangeable sodium percentage (ESP) by traditional laboratory methods is both time and cost limiting. Because of this, soil spectroscopic techniques were used to predict soil ESP and fill the gaps in the dataset where necessary, with results showing that ESP could be predicted with high accuracy using this approach. Soils of the Hillston district were found to be generally non-saline, and some clear temporal changes in salinity were detected over the study period. While no considerable changes in soil salinity were detected under the natural land use, a mean desalination trend was observed in irrigated cotton production, and an increase was observed in soils under irrigated horticultural production. The contrasting directions of soil salinity was primarily attributed to the application of varying quantities and qualities of irrigation water, as well as the addition of salt-containing fertilisers in horticultural production. Much of the soil at Hillston was found to be sodic, and a trend of increasing soil sodicity through time was observed. The cause of this increase in

sodicity was not clear, although it was likely due to the continual addition of sodium to the soil system through irrigation and fertilisers. The observed trend of increasing ESP and decreasing EC in some parts of the study area is concerning for soils under irrigated agriculture, as this could have substantial negative implications for soil structural integrity.

Chapter 6 focuses on simultaneously predicting soil organic carbon (SOC) content in four dimensions (eastings, northings, depth and time) using multivariate linear mixed models. Rather than modelling each sampling depth at the two time points, as was done in Chapter 4 and 5, two time points (2002 and 2015) and two depths (0-0.1 and 0.3-0.5 m) are modelled in combination. Because analysing SOC content by traditional laboratory methods can be laborious when the soil sample contains inorganic carbon (SIC), as is often the case in arid regions, the SOC and SIC content of samples was predicted using soil spectroscopic techniques. In addition to using VisNIR spectra, easily measurable and readily available soil property data (pH and total carbon content) were also used as predictor variables in the Cubist models, and including this ancillary soil data as predictor variables gave very accurate predictions for both SOC and SIC content. The soils at Hillston were found to contain low levels of SOC, and topsoil maps of SOC content showed that some areas were predicted to have experienced an increase in SOC content during the study period. This increase in SOC content did not appear to be a function of land management, but rather a result of several years of high rainfall experienced in the study area.

Chapter 7 focuses on developing a methodology for modelling soil inorganic carbon (SIC) in space and time. Datasets of SIC are often zero-inflated (i.e. many zero values) and also highly skewed compared to datasets of other soil properties, and cannot be directly modelled using standard statistical modelling approaches or the BLMM/MLMM method used in the preceding research chapters. This chapter uses 'mixture models' to overcome these challenges, where SIC content maps for a subsoil layer (0.3-0.5 m) are created. This involves initially modelling the presence or absence of SIC in the study area, and then separately modelling SIC content using values from the dataset that are greater than zero only. The results from these two models are then combined, producing a mixture model. This approach resulted in high quality maps that showed a realistic distribution of SIC content in the study area. No clear changes in SIC content were detected over time, although both increases and decreases in SIC content were observed at a number of sites. These fluctuations in SIC content could be due to the dissolution and re-precipitation of carbonates to/from other layers of the soil profile as a result of increased water percolation through the soil. Overall, the mixture model approach was deemed successful, suggesting that this method could be adopted by other studies

that aim to model SIC content, or to model other environmental properties that have a similarly unusual distribution, such as soil contaminants.

Overall the use of BLMMs and MLMMs to model soil properties in space and time resulted in accurate predictions, logical soil maps, and greater ability to detect statistically significant changes in soil properties over time. The research presented here concludes that there have been changes in several aspects of the soil in the semi-arid irrigated cotton-growing district of Hillston. From an agricultural production viewpoint, there have been some positive changes in soil properties, such as the highly alkaline soils moving towards neutral, and the increase in SOC content, as well as undesirable shifts in soil condition, such as an increasing ESP. It was clear that irrigated cotton production impacted different soil properties from 2002 to 2015, although these changes were not necessarily negative. It was also apparent that the significant shifts in rainfall patterns during the study period also had an impact on many aspects of soil condition, suggesting that appreciable changes in soil condition can occur at the decadal scale as a result of changes in climatic conditions.

Acknowledgements

The last few years have been by far the most challenging and rewarding of my academic life. I have learnt more in these years than I could have ever imagined, and the knowledge that I have gained is far different to what I envisaged at the beginning of the project. Firstly, I would like to thank the Cotton Research Development Corporation (CRDC) and the Australian Postgraduate Awards (APA) for my scholarship, research funding, and the opportunities to travel across Australia and the world to learn, and to share this research.

To my main supervisor, Stephen Cattle, I greatly appreciate the opportunity to do this PhD under your guidance. You have put up with my countless questions and have taught me some valuable lessons about research. You have trained me to constantly make sure that the research that we carry out is relevant to the end outcome, and that connecting the research to farmers and grower groups simply and effectively is an absolute necessity. While I complained about your particular and nit-picking editing at the time, I admire the effort you put in to improve my scientific writing technique. I have never met anyone that has such a genuine passion for soil, and some of that has unquestionably been instilled in me.

Tom Bishop was not an associate supervisor at the start of the project, but roping Tom in as a supervisor has been one of the best decisions I have made (thanks to Floris and Willem). Tom has taught me about spatio-temporal modelling, mapping, coding, statistics, and heaps of other useful (and not so useful) things. I greatly appreciate his patience and ability to explain complex things (and his never ending enthusiasm for a 5 o'clock beer). I not only consider Tom a great mentor, but also a great friend. To my other associate supervisor, Inakwu Odeh, I thank you for your advice at the commencement of the project to get it underway. Your life advice over the years has been insightful and it has been a joy to receive it. I will never utter or hear the words "I don't believe it!" without hearing your voice. I also appreciate the contributions from Matt Pringle, who was of great assistance with the complicated spatio-temporal modelling techniques used in this thesis.

The field sampling campaigns were not an easy task, and I couldn't have done it without the help of Lori Watson, Ali Barlow and Ed Jones. I also thank Ed for teaching me the fundamentals of R, being there to bounce ideas off, and keeping me sane when times were the exact opposite. Thanks to Budiman Minasny for his guidance in the creation of the extensive literature review. Thanks to all the current and former academic staff in the soils group for your advice over the years; Alex, Balwant, Brett, Damien, Uta, Brendan, Nathan, and Kanika. Thanks to all the other students that have helped me in some way or simply been a friend over the last few years (Liana, Mario, Phil, Vanessa, Niranjana, Jose, Maryam, Kai, Sanji, Stacey, Wartini, Yumi, Yuxin, Sonam, Senani). A special thanks to my Korean superstar; Jay, I thank you for helping me in the lab and for being a great friend. We all miss you and hope that you will return again to permanently annoy us with your crazy ways.

While many people have helped and inspired me throughout the last few years, there are a string of people that have encouraged my passion for agriculture, the environment, and science before I started University. A special thanks to my Uncle Scott, who since I can remember has shared my passion for nature and science and has encouraged me to pursue my dreams to work in this field. Special thanks to two of my high school teachers; Mrs Amore and Mr Kobus. The fundamental understanding of science Mr Kobus passed on to me before I was thrown into university life was instrumental, and he was a big part of the reason I decided to pursue a career in science. I would also like to thank my boys – Mitchell, Gelato, Joel, Rhys, Fadi, and Sultan, for their support and love over the last decade or so.

Lastly, I would like to thank my family for their constant support and inspiration. Thanks to my mum Sharon and my dad Claudio for the never ending encouragement, meals and clothes-washing (a huge bonus of living at home when you're an adult). Also, to my brother Tim, for being an inspiration and showing me an example of the great things that you can achieve through hard work and determination. Thanks to my girlfriend, Julie, who has been by far my biggest supporter along the way, and is always proud of everything that I do. Your love and care has helped me immensely. To my grandparents – Nonna, Nonno, Nanna, and Pa – I appreciate everything that you have done for me.

Now it is time to end my 20-year campaign as a student, start living in the real world, and attempt to positively contribute to agricultural and environmental knowledge.

Publications and presentations made from this thesis

Chapters of this thesis that have been published in peer reviewed scientific journals:

Chapter 2

Filippi, P, Minasny, B, Cattle, SR, Bishop, TFA (2016) Monitoring and Modeling Soil Change: The Influence of Human Activity and Climatic Shifts on Aspects of Soil Spatiotemporally. *Advances in Agronomy* **139**, 153-214.

Other publications and presentations made from this thesis:

Abstracts and presentations:

Filippi P, Cattle S, Bishop T, Odeh I (2015) Monitoring the impact of dramatic landuse shifts on soil in a fragile semi-arid landscape. Pedometrics 2015, University of Córdoba, Spain, 14th - 18th September, 2015

Filippi P, Cattle S, Bishop T, Odeh I (2016) Monitoring the impact of irrigated cotton production on soil condition in a semi-arid landscape in Australia. World Cotton Research Conference - 6, Goiânia, Brazil, 2nd - 6th May, 2016

Filippi P, Cattle S, Bishop T, Odeh I (2016) Monitoring and modelling soil change in a semi-arid landscape with distinctly contrasting landuses. New Zealand Society of Soil Science and Soil Science Australia conference - Soil, a balancing act downunder, Queenstown, New Zealand, 12th - 16th December, 2016

Posters:

Filippi P, Cattle S, Odeh I, Bishop T (2014) Monitoring soil change at Hillston, NSW – utilising those old surveys. *Soil Science Australia Conference 2014*, Melbourne Cricket Ground, Australia, November 2014.

Filippi P, Cattle S, Bishop T, Odeh I (2015) Soil change under intensifying landuse and shifting climates. *Research Symposium 2015*, Faculty of Agriculture and Environment, The University of Sydney, Australia, July 2015.

Publications in preparation from this thesis:

Chapter 4

Filippi P, Cattle SR, Bishop TFA, Pringle MJ, Odeh IOA (2017) Monitoring and modelling spatio-temporal changes in topsoil and subsoil pH at decadal scales in an irrigated semi-arid landscape using bivariate linear mixed models

Chapter 5

Filippi P, Cattle SR, Bishop TFA, Pringle MJ, Odeh IOA (2017) Monitoring changes in topsoil and subsoil salinity and sodicity in an irrigated semi-arid landscape at decadal scales

Chapter 6

Filippi P, Cattle SR, Bishop TFA, Pringle MJ, Odeh IOA (2017) Simultaneous 4D space-time predictions of soil organic carbon by multivariate linear mixed models

Chapter 7

Filippi P, Cattle SR, Bishop TFA, Pringle MJ, Odeh IOA (2017) Modelling subsoil inorganic carbon in space and time with mixture models

In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Patrick Filippi

15/09/2017

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Associate Professor Stephen R. Cattle

15/09/2017

Contents

1. General Introduction	1
2. Review of the literature – Monitoring & modeling soil change - the influence of human activity & climatic shifts on aspects of soil spatio-temporally	6
Abstract.....	7
2.1. Introduction	8
2.2. Soil properties most susceptible to anthropogenic activity and climate change	9
2.2.1. Anthropogenic activity.....	10
2.2.2. Climatic shifts.....	13
2.3. Methods of measuring and monitoring soil change	17
2.3.1. Space-for-time-substitution/chronosequences/paired-sites	19
2.3.2 Long-term soil experiments	21
2.3.3 Soil monitoring networks.....	23
2.3.3.1. <i>Model-based sampling strategies</i>	25
2.3.3.2. <i>Design-based sampling strategies</i>	29
2.3.3.3. <i>Non-statistical monitoring programs</i>	30
2.3.4. Legacy data	32
2.3.4.1. <i>Utilising legacy data without resampling</i>	33
2.3.4.2. <i>Utilising legacy data with resampling</i>	34
3.4.2.1. <i>New site</i>	35
3.4.2.2. <i>Same site</i>	36
2.3.4.3. <i>Utilising archived soil samples</i>	37
2.3.4.4. <i>Laboratory test database</i>	38
2.3.5. Summary of methods to monitor/measure soil change.....	42
2.4. Challenges in monitoring soil change	43
2.4.1. Detecting a change in soil properties	43
2.4.1.1. <i>Soil spatial variation</i>	44
2.4.1.2. <i>Magnitude of change</i>	45
2.4.1.3. <i>Covariance between surveys and returning to sampling sites</i>	46
2.4.2. Operational issues and sample support	47
2.4.3. Disentangling effects of climate & landuse	48
2.5. Modelling soil change into the future.....	48
2.6. Conclusions and future research needs.....	52
References	54

3. Materials and methods – study area, soil datasets, soil property analysis and modelling approaches 63

3.1. Introduction	64
3.2. Study area	64
3.2.1. Physiography	64
3.2.2. Soil types and geological units	67
3.2.3. Climate and climatic fluctuations	68
3.2.4. Land use and land use change	69
3.3. Soil datasets	72
3.3.1. 2002 dataset	72
3.3.2. 2015 dataset	72
3.4. Analysis of soil properties	72
3.4.1. Traditional laboratory methods	73
3.4.1.1. Soil pH	73
3.4.1.2. Soil electrical conductivity	73
3.4.1.3. Soil exchangeable sodium percentage	73
3.4.1.4. Soil carbon – total, organic and inorganic	73
3.4.2. Visible near infrared (VisNIR) predictions	74
3.4.2.1. Soil exchangeable sodium percentage	74
3.4.2.2. Soil organic and inorganic carbon	75
3.5. Harmonising sampling depths	76
3.6. Collation and processing of covariates for modelling and mapping	76
3.6.1. Covariates for modelling and mapping	76
3.6.2. Principal component analysis	78
3.7. Linear mixed models (LMMs)	78
3.7.1. Model selection and map production	80
3.7.2. Assessment of model quality	80
3.7.3. Change calculation and statistical significance of change	81
3.8. Concluding remarks	82
References	82

4. Monitoring and modelling spatio-temporal changes in topsoil and subsoil pH at decadal scales in an irrigated semi-arid landscape, using bivariate linear mixed models 84

Abstract	85
4.1. Introduction	86
4.2. Materials and methods	88
4.2.1. Study area and changes in climate and land use	88
4.2.2. Soil datasets	88
4.2.3. Covariates and principal component analysis	89
4.2.4. Modelling and mapping	89
4.2.4.1. Bivariate linear mixed models (BLMMs)	89
4.2.4.2. Model selection and map production	90

4.2.4.3. Assessment of model quality	90
4.2.4.4. Change calculation and statistical significance of change	91
4.3. Results	92
4.3.1. Soil pH summary statistics	92
4.3.2. Principal component analysis (PCA) of numerical predictor variables	95
4.3.3. Bivariate linear mixed models	96
4.3.4. Soil pH maps and modelled data	99
4.3.5. Statistical significance of change in soil pH.....	104
4.4. Discussion.....	106
4.4.1. Soil pH, acidification, and driving factors.....	106
4.4.2. Spatial patterns of soil pH maps	109
4.4.3. Statistical significance of pH change.....	109
4.4.4. Bivariate linear mixed models – benefits and opportunities.....	110
4.5. Conclusions & future directions.....	112
References	113

5. Monitoring changes in topsoil and subsoil salinity and sodicity in an irrigated semi-arid landscape over a 13-year period..... 115

Abstract.....	116
5.1. Introduction	117
5.2. Materials and methods	119
5.2.1 Soil datasets	119
5.2.2. Laboratory analyses	119
5.2.3. Visible near infrared (VisNIR) spectroscopy analyses.....	119
5.2.4. Modelling and mapping	121
5.2.4.1. Predictor variables	121
5.2.4.2. Bivariate linear mixed models (BLMMs) and statistical significance of change.....	121
5.2.4.3. Model selection and map production	122
5.2.5. Determination of salinity and sodicity ratings.....	123
5.3. Results	124
5.3.1. Soil EC summary statistics.....	124
5.3.2. Soil ESP prediction by VisNIR	126
5.3.3. Soil ESP summary statistics.....	127
5.3.4. Bivariate linear mixed models	130
5.3.5. Modelled maps	131
5.3.5.1. Soil EC modelled maps	131
5.3.5.2. Soil ESP modelled maps.....	136
5.3.6. Statistical significance of predicted change.....	138
5.3.6.1. Soil EC.....	138
5.3.6.2. Soil ESP.....	140
5.4. Discussion.....	140
5.4.1. Changes in soil properties and possible driving factors.....	140
5.4.1.1. Salinity.....	140
5.4.1.2. Sodicity.....	144
5.4.2. Prediction of ESP by VisNIR.....	146

5.4.3. Bivariate linear mixed models	147
5.4.3.1. <i>Model quality and future opportunities</i>	147
5.4.3.2. <i>Comparison of univariate and bivariate linear mixed models</i>	148
5.5. Conclusions & future directions.....	149
References	150

6. Simultaneous 4D space-time predictions of soil organic carbon by multivariate linear mixed models 153

Abstract.....	154
6.1. Introduction	155
6.2. Materials and methods.....	156
6.2.1. Soil datasets	156
6.2.3. Soil analyses	157
6.2.3.1. <i>Traditional laboratory methods</i>	157
6.2.3.2. <i>Visible near infrared (VisNIR) spectroscopy predictions</i>	157
6.2.4. Models and mapping	159
6.2.4.1. <i>Multivariate linear mixed model</i>	159
6.2.4.2. <i>Covariates, model selection, map production, and assessment of model quality</i>	159
6.3. Results.....	160
6.3.1. Visible near infrared (VisNIR) spectroscopy predictions	160
6.3.2. Soil organic carbon summary statistics.....	165
6.3.3. Multivariate linear mixed model	168
6.3.3.1. <i>Final model and model quality</i>	168
6.3.3.2. <i>Soil organic carbon content maps</i>	169
6.3.4. Statistical significance of soil organic carbon change	171
6.4. Discussion.....	171
6.4.1. Soil organic and inorganic carbon predictions with visible near infrared (VisNIR).....	171
6.4.2. Spatio-temporal patterns of soil organic carbon.....	174
6.4.3. Multivariate linear mixed models.....	176
6.5. Conclusions and future directions	178
References	178

7. Modelling subsoil inorganic carbon in space and time with mixture models..... 181

Abstract.....	182
7.1. Introduction	183
7.2. Materials and methods.....	185
7.2.1. Soil datasets and soil property analyses	185
7.2.2. Soil inorganic carbon prediction by mixture models	185
7.2.2.1. <i>The probability model for the presence/absence of SIC</i>	186
7.2.2.2. <i>SIC content model with above zero values only</i>	186
7.2.2.3. <i>Mixture model</i>	187
7.2.3. Prediction of SIC by difference of STC and SOC maps produced with MLMs	187

7.3. Results.....	187
7.3.1. Summary statistics of inorganic carbon.....	187
7.3.2. Inorganic carbon presence/absence predictions with Random Forest.....	191
7.3.3. Soil inorganic carbon content predictions with bivariate linear mixed models.....	194
7.3.4. Inorganic carbon maps from mixture models.....	195
7.3.5. Statistical significance of change in inorganic carbon.....	196
7.3.6. Inorganic carbon maps from difference of total carbon and organic carbon.....	197
7.4. Discussion.....	198
7.4.1. The dynamics of soil inorganic carbon in the study area.....	198
7.4.2. Modelling approaches.....	201
7.4.2.1. <i>Mixture models</i>	201
7.4.2.2. <i>SIC difference from STC and SOC maps</i>	203
7.5. Conclusions and future directions.....	203
References.....	204

8. Concluding remarks..... 206

8.1. Overview.....	207
8.2. Future research questions.....	215
8.2.1. Monitoring temporal changes in soil.....	215
8.2.2. Spatio-temporal modelling.....	217
8.2.3. VisNIR predictions of soil properties.....	219
8.3. Closing statements.....	219

Appendix..... 220

List of Figures

Fig. 2.1 – Examples of a random stratified design-based sampling scheme (left) and a systematic grid model-based sampling scheme (right) of a district in western NSW, Australia	25
Fig. 2.2 – Changes in topsoil (0-15 cm) SOC contents across England and Wales between 1978 and 2003 – (a) SOC contents in original samples (1978-83), (b) predicted annual rates of change calculated from the changes over the different sampling intervals (Bellamy <i>et al.</i> 2005).....	26
Fig. 2.3 – Changes in topsoil (0-15 cm) pH across England and Wales between 1978 and 2003 – (a) pH of original samples (1978-83), (b) predicted annual rates of change calculated from the changes over the different sampling intervals (Kirk <i>et al.</i> 2010)	27
Fig. 2.4 – Measured SOC content (g C kg^{-1}) at survey points in the LUCAS topsoil survey (0-30 cm) (de Brogniez <i>et al.</i> 2015)	29
Fig. 2.5 – Maps of changes in topsoil (0-15 cm) C stocks in South Korean rice paddy fields from 1999-2003, and 2003-2007 (Minasny <i>et al.</i> 2012).....	32
Fig. 2.6 – Changes in C stocks (Tg) at various intervals from 1920-2010 in Java, Indonesia, grouped into kabupaten (regions) (Minasny <i>et al.</i> 2011)	34
Fig. 2.7 – Map of predicted SOC for 2000 (a), 2010 (b), and predicted temporal SOC change from 2000-2010 (c) for the Cox’s Creek Catchment, NSW, Australia (Karunaratne <i>et al.</i> 2014) Journal compilation © CSIRO 2014. http://www.publish.csiro.au/nid/84/paper/SR13081.htm	36
Fig. 2.8 – Spatial distribution of cantonal medians of SOC in topsoil (g/kg) for the three periods 1990–1994, 1995–1999 and 2000–2004 in the Franche-Comté region, France (white polygons lacked sufficient quantities of data) (Saby <i>et al.</i> 2008a).....	39
Fig. 2.9 – Changes in mean SOC contents of grassland (1984–2000), maize land (1984–2004) and arable land (1984–2004) in the Netherlands. The mean annual change in SOC is indicated as $\Delta\text{C} / \Delta\text{t}$, in g/kg/yr (Reijneveld <i>et al.</i> 2009).....	39
Fig. 2.10 – Change in topsoil pH of pasture and cropping soils across Victoria, Australia from 1973-1994 (Marchant <i>et al.</i> 2015)	40
Fig. 2.11 – Change in topsoil available potassium in parts per million of pasture and cropping soils across Victoria, Australia from 1973-1994 (DPI 2012).....	40
Fig. 2.12 – Predicted SOC stocks for an area in the Hunter Valley, NSW, Australia, in the year 2010 (a), and 2020 (b) (Minasny <i>et al.</i> 2013).....	50
Fig. 2.13 – Predicted change in SOC in t C ha^{-1} in cropland (a) and grassland (b) under climate-only changes across Europe from 1990-2080, and predicted change in SOC in t C ha^{-1} in cropland (c) and grassland (d) under climate change, technology change, and net primary production (NPP) changes across Europe from 1990-2080 (Smith <i>et al.</i> 2005)	51
Fig. 3.1 – The location of the study area (Hillston) within Australia	65
Fig. 3.2 – The locations of soil cores extracted in the 2002 and 2015 soil surveys overlaid on a satellite image with notable landmarks labelled	66
Fig. 3.3 – The broad soil types of the study area, derived from geological units of the Hillston 1:100,000 mapsheet (Cameron <i>et al.</i> 2000)	67
Fig. 3.4 – Annual rainfall received at Hillston from 2002 to 2015 (BOM 2017).....	68
Fig. 3.5 – Soil core locations, and land use classes for the study area in the year 2002 (left), and 2015 (right) (Office of Environment and Heritage 2017).....	71
Fig. 4.1 – Mean change in soil pH (1:5 soil:water) from 2002 to 2015 of co-located sites based on different soil types (left) and land use (right) to the depth of 1.2 metres (with standard error bars).94	

Fig. 4.2 – Correlation of 2002 with 2015 soil pH at co-located samples in the six sampling depths, with the line of best fit and <i>r</i> values (Pearson correlation) included	95
Fig. 4.3 – Maps of soil pH values (1:5 soil: water) for the 0-0.1 and 0.1-0.3 m depths in 2002 and 2015 overlain with sample locations, and the change in soil pH in this time period (2015-2002)	101
Fig. 4.4 – Maps of soil pH values (1:5 soil: water) for the 0.3-0.5 and 0.5-0.8 m depths in 2002 and 2015 overlain with sample locations, and the change in soil pH in this time period (2015-2002)	102
Fig. 4.5 – Maps of soil pH values (1:5 soil: water) for the 0.8-1.2 and 1.2-1.5 m depths in 2002 and 2015 overlain with sample locations, and the change in soil pH in this time period (2015-2002)	103
Fig. 4.6 – Maps of z-scores showing the statistical significance of the change in soil pH for each sampling depth	105
Fig. 5.1 – Locations of soil cores extracted in the 2002 and 2015 soil surveys, and sites analysed by traditional laboratory methods, and predicted from VisNIR spectra	121
Fig. 5.2 – Mean change in soil EC (dS/m) from 2002 to 2015 of co-located sites by land use (left) and soil type (right) to 1.2 metres with standard error bars (Ve = Vertosol)	126
Fig. 5.3 – Observed and predicted ESP (%) of calibration (left) and validation (right) datasets predicted with VisNIR spectra and Cubist models (1:1 line overlaid).....	127
Fig. 5.4 – Change in soil exchangeable sodium percentage (%) of laboratory measured samples only at co-located sites between 2002 and 2015 for the six sampling depths	129
Fig. 5.5 – Mean soil exchangeable sodium percentage (%) of laboratory measured samples from 2002 by soil type to 1.5 metres with standard error bars (Ve = Vertosol)	130
Fig. 5.6 – Electrical conductivity (EC) for 2002 and 2015 for the 0-0.1 m and 0.1-0.3 m depths and the EC change (dS/m 1:5 soil: water)	133
Fig. 5.7 – Electrical conductivity (EC) for 2002 and 2015 for the 0.3-0.5 m and 0.5-0.8 m depths and the EC change (dS/m 1:5 soil: water).....	134
Fig. 5.8 – Electrical conductivity (EC) for 2002 and 2015 for the 0.8-1.2 m and 1.2-1.5 m depths and the EC change (dS/m 1:5 soil: water).....	135
Fig. 5.9 – Exchangeable sodium percentages (ESP) for 2002 and 2015 for the 0-0.1 m and 1.2-1.5 m depths and change in ESP (%)	137
Fig. 5.10 – Z-scores showing the statistical significance of the change in EC (dS/m 1:5 soil: water) for the six different sampling depths	139
Fig. 5.11 – Z-scores showing the statistical significance of the change in ESP for the 0-0.1 m and 0.8-1.2 m sampling depths.....	140
Fig. 5.12 – Soil electrical conductivity (EC) values for (1:5 soil: water dS/m) for 0-0.1 m sampling depths in 2002 (left) and 2015 (right) produced from separate univariate linear mixed models (ULMMs)	149
Fig. 6.1 – Observed vs. predicted soil organic carbon (SOC) content (%) values of calibration (left) and validation (right) datasets predicted with VisNIR spectra and mid-depth (top); and VisNIR spectra, mid-depth, soil pH and soil total carbon contents (bottom) as predictor variables in Cubist models	163
Fig. 6.2 – Observed vs. predicted soil inorganic carbon (SIC) content (%) values of calibration (left) and validation (right) datasets predicted with VisNIR spectra and mid-depth (top); and VisNIR spectra, mid-depth, soil pH and soil total carbon contents (bottom) as predictor variables in Cubist models.....	164

Fig. 6.3 – Temporal change in soil organic carbon (SOC) contents at sites sampled in both 2002 and 2015 surveys under the cotton (left) and natural (right) land use at the 0-0.1 m (top) and 0.3-0.5 m sampling depth (bottom).....	167
Fig. 6.4 – Observed vs. predicted SOC content (%) using LOSOCV and predicted with the MLMM ..	169
Fig. 6.5 – Soil organic carbon (SOC) content (%) of the 0-0.1 (top) and 0.3-0.5 m (bottom) sampling depths in 2002 and 2015, modelled with a single multivariate linear mixed model, and the change in SOC during this time period (2015-2002)	170
Fig. 6.6 – Z-scores showing the statistical significance of the change from 2002 to 2015 in soil organic carbon content for the 0-0.1 and 0.3-0.5 m sampling depths.....	171
Fig. 7.1 – The distribution of soil inorganic carbon content (%) at the 0-0.1, 0.1-0.3 and 0.3-0.5 m sampling intervals at all sites	188
Fig. 7.2 – Change in the soil inorganic carbon content (%) at the 0-0.1, 0.1-0.3 and 0.3-0.5 m sampling intervals at co-located sites that included a presence of carbonates in both 2002 and 2015	190
Fig. 7.3 – Mean change in soil inorganic carbon contents (%) from 2002 to 2015 of co-located sites under the cotton and natural land use to the depth of 0.5 metres (with standard error bars)	191
Fig. 7.4 – ROC (receiver operator characteristic) curve for out-of-bag predictions of the presence or absence of soil inorganic carbon with a Random Forest model.....	192
Fig. 7.5 – The importance of each predictor variable in the Random Forest model of the probability SIC with the contribution to mean decrease in accuracy (left) and mean decrease in node impurity (right)	193
Fig. 7.6 – The predicted distribution of the presence or absence of soil inorganic carbon for year 2002 (left) and 2015 (right) for the 0.3-0.5 m sampling depth.....	194
Fig. 7.7 – Soil inorganic carbon content (%) in 2002 (left) and 2015 (right) for the 0.3-0.5 m sampling depth produced from a bivariate linear mixed model using above-zero values only	195
Fig. 7.8 – Soil inorganic carbon content (%) for the 0.3-0.5 m depth produced from mixture models	196
Fig. 7.9 – The change in soil inorganic carbon content (%) from 2002 to 2015 for the 0.3-0.5 m depth	196
Fig. 7.10 – Z-scores showing the statistical significance of the change in soil inorganic carbon from 2002 to 2015 for the 0.3-0.5 m depth	197
Fig. 7.11 – Soil inorganic carbon content (%) for the 0.3-0.5 m depth produced from the difference between maps of soil total carbon and maps of soil organic carbon.....	198
Fig. A1 – Visible near infrared (VisNIR) spectra of all archived soil samples ($n = 1291$) from the 2002 ($n = 385$) and 2015 ($n = 906$) soil surveys	221

List of Tables

Table 2.1 – Major global soil degradation issues resulting from human activity/agriculture.....	12
Table 2.2 – Summary of soil properties/characteristics expected to be significantly impacted by climate change, and the mechanisms driving that change	16
Table 2.3 – Summary of methods to monitor/measure soil change - (Expense = low (*) – high (****)).....	42
Table 3.1 – The typical amount of water percolation that occurs on an annual basis, and throughout the study period (2002-15) for characteristic land uses in the Hillston district	70
Table 3.2 – List and specifications of environmental covariates used in model development.....	77
Table 4.1 – Mean soil pH for all sites at different sampling depths, time points and land uses.....	93
Table 4.2 – The proportion of variance that each principal component explains, the cumulative variance, and the standard deviation from the principal component analysis (PCA) of the 19 numerical covariates.....	96
Table 4.3 – Results from the principal component analysis (PCA) showing the proportion (loading) that each input variable contributes to the first seven principal components	96
Table 4.4 – Covariates (predictor variables) included in the models and the coefficient matrix (estimated fixed effects) for the six soil layers	97
Table 4.5 – Summary of the random effects from the pH BLMMs at different sampling depths – the auto-correlation nugget ($c_{0,1,1}$ and $c_{0,2,2}$), the cross-correlation nugget ($c_{0,1,2}$), the auto-correlation structural semivariance ($c_{1,1,1}$ and $c_{1,2,2}$), the cross-correlation structural semivariance ($c_{1,1,2}$), and the range (ϕ , or ϕ). The range is presented in metres, and is multiplied by 3. The subscript 1 refers to the 2002 survey, and the subscript of 2 refers to the 2015 survey.....	98
Table 4.6 – Quality statistics of the models for the six soil layers based on leave one out cross-validation	98
Table 4.7 – Rate of change on a per year basis by land use from modelled maps of whole study area during the 2002-2015 period (pH unit yr ⁻¹).....	100
Table 4.8 – Area extent of predicted statistically significant decreases in soil pH by landuse across the study area at different levels of significance	106
Table 5.1 – Soil salinity ratings for soils of varied proportions of clay, and the legend values used in the salinity maps in this study – adapted from Hazelton and Murphy (2007)	124
Table 5.2 – Soil sodicity ratings and the legend values used in the sodicity maps in this study – adapted from Hazelton and Murphy (2007).....	124
Table 5.3 – Electrical conductivity (dS/m 1:5 soil: water) statistics for all sites at different sampling depths, time points and land uses	125
Table 5.4 – Soil exchangeable sodium percentage (%) statistics for 2002 and 2015 laboratory-measured (Lab) samples, and 2015 VisNIR-predicted (Pred) samples at different soil depths, time points and land uses	128
Table 5.5 – Model quality statistics for EC and ESP BLMMs based on leave-one-out cross-validation	131
Table 6.1 – Pearson’s correlation (r) between response variables (SOC and SIC content) and predictor variables (STC content and pH).....	161
Table 6.2 – Model quality statistics for SOC and SIC content predictions from Cubist models with varying combinations of input variables.....	162

Table 6.3 – Combined laboratory-measured and VisNIR-predicted soil organic carbon content (%) statistics of all samples for different depth increments, time points and land uses.....	165
Table 6.4 – Mean change in SOC content (%) from 2002 to 2015 for co-located sites.....	166
Table 6.5 – Included covariates in the final MLMM of SOC content	168
Table 6.6 – Quality statistics of the soil organic carbon content MLMM using LOSOCV	168
Table 7.1 – Soil inorganic carbon content (%) statistics of sites that possessed values above zero, and the number of samples that contained a zero value or a value above zero, by land use, depth and time point.....	189
Table 7.2 – Confusion matrix – out-of-bag estimate of error rate = 20.76% – based on threshold probability of 0.5.....	192
Table 7.3 – Included covariates in the final BLMM of SIC content.....	195
Table 7.4 – Quality statistics of the SIC content BLMM modelled with above-zero values only	195
Table 7.5 – Quality statistics of the STC content MLMM using LOSOCV	198
Table 7.6 – Included covariates in the final STC content MLMM.....	198

Chapter 1

1. General Introduction

Soil is an extremely valuable finite natural resource that provides many irreplaceable services to the earth system. These include food and fibre production, water cycling and filtration, waste decomposition, nutrient cycling, and carbon sequestration. The world is currently faced with the challenge of international food security, and global soil resources will play a pivotal role in meeting this challenge, as well as a suite of other global environmental sustainability challenges (McBratney *et al.* 2014). It is known that anthropogenic activity can change the properties of soil considerably over relatively short time scales, however, it is less clear how the recent climatic shifts observed throughout the globe will influence changes in soil condition.

The agriculturally-important semi-arid regions of eastern Australia generally possess arable soils that have valuable attributes for cropping, but they are also characterised by several inherent soil issues and constraints, such as low levels of soil organic carbon, and high alkalinity, salinity, and sodicity (Onus *et al.* 2003). There have been some significant shifts in land use and land management in these semi-arid areas, as well as considerable fluctuations in rainfall patterns over the last decade or so. During the early 2000s, much of eastern Australia experienced the ‘Millennium Drought’, which was considered to be the worst drought ever recorded in many districts. In 2010, the emergence of La Niña ended this extended dry period, and several consecutive years of very high rainfall were received. The semi-arid region surrounding the township of Hillston in south-western NSW is a particularly good example of an area that has experienced both of these land use and climatic changes. At Hillston, irrigated cotton production is the primary agricultural enterprise, and irrigated cotton is considered to be one of the most intensive agricultural crops grown in Australia, which can place significant pressure on the soil. Irrigated cotton receives large amounts of applied water, substantial fertiliser inputs, requires constant cultivation, and heavy machinery is regularly driven over the soil. While land management had to be adjusted at Hillston to adapt to the fluctuating rainfall patterns, the area under irrigated agriculture (cotton and horticulture) has continued to expand in the last decade or so, with this primarily due to the increased utilisation of groundwater from aquifers.

These changes in land use and climate have the potential to alter a suite of soil properties. Despite this, most of the attention in soil monitoring research has been on monitoring the response of soil organic carbon to these driving factors of change, with other important soil attributes typically disregarded. In addition to this, very few studies have considered monitoring beyond the topsoil, but subsoil constraints such as acidity, alkalinity, salinity and sodicity, are often the most limiting factors to agricultural production in many areas. Studies that disregard subsoil properties usually do so due to the laborious and cost-intensive nature of collecting these samples. Similarly, soil properties that

are difficult to measure by traditional laboratory methods, such as exchangeable sodium percentage (ESP), are often ignored due to resource limitations. Recently, there has been significant focus on the use of spectroscopic tools, such as visible near infrared (VisNIR), to predict soil properties due to the cost and labour savings that come with these approaches (Viscarra Rossel *et al.* 2006). There is considerable promise in using these techniques to predict difficult-to-measure soil properties, and this may encourage the monitoring of a suite of important soil properties, rather than just those soil properties that can be easily measured by traditional laboratory methods.

There are various approaches to monitoring spatio-temporal soil change, and each method possesses individual benefits and limitations. Recently, there has been much focus on using soil data that already exists, otherwise known as 'legacy data' (Marchant *et al.* 2015). While legacy data has its limitations, it is both cost- and time-efficient, and can be invaluable in detecting changes in soil condition if handled correctly. Legacy soil data can encompass many different types of data, but the information usually represents soil at a single time point, and there is rarely a time-series of data available. To gain a time-series of data, sites that were sampled in legacy soil surveys can be resampled at a different time point. This soil information from the same site, extracted at different time points, is often referred to as 'co-located' data, and this can be particularly beneficial in evaluating temporal changes in soil condition.

Many soil monitoring studies involve the modelling of soil properties to create digital soil maps (DSM), and there is an increasing amount of free and readily-available spatial information that can be used as covariates to do this. There are various methods of modelling soil properties to create DSMs, but certain approaches are advantageous for soil monitoring purposes, particularly when a time-series of co-located soil data is available. There is often a strong correlation between the values of a soil property extracted at different time points, and it is possible to utilise this correlation to improve the accuracy of soil property predictions. One modelling approach that can take advantage of this is multivariate linear mixed models (MLMMs) (Lark and Papritz 2003). In this approach, a soil property can be simultaneously modelled at multiple time points, rather than independently modelling soil at each time point. Using MLMMs creates soil maps that have a coherent relationship through time, and is particularly beneficial when the dataset from one time point possesses more sampling locations than another. Another advantage of using MLMMs to model/monitor soil properties is that the covariance at sites can be included in the calculation of the contrast variance. This contrast variance is then used in the calculation of a test statistic, which subsequently allows greater sensitivity in determining whether a statistically significant change in a soil property has occurred over time. Most monitoring studies ignore the statistical significance of

any observed changes, and simply show the difference of soil maps at different time points, which can lead to considerable misinterpretation about actual changes in soil properties.

The literature review in this thesis focuses on the various methods of monitoring and modelling spatio-temporal changes in soil condition, and the benefits and disadvantages of the different approaches are discussed. A review of current knowledge surrounding the impact that climatic fluctuations and agricultural management practices have on a suite of important soil properties is also undertaken. In this thesis, soil data from a baseline legacy soil survey conducted in 2002 is used, along with soil data from a repeat soil survey undertaken in 2015 in a semi-arid irrigated cotton-growing landscape surrounding the township of Hillston in south-west NSW, Australia. There is a common theme of fully-utilising and maximally-benefiting from the data and techniques that are available, including the use of already-existing soil information, the use of modelling approaches that takes advantage of available data, and the use of spectroscopic analytical techniques that are both cost- and labour-efficient. More specifically, aims of this thesis with associated research questions are:

- 1.** To explore the status and change of multiple soil properties over a 13-year period in a semi-arid irrigated cotton-growing region using legacy and newly-collected soil survey data
 - a.** What is the status of soil condition at Hillston?
 - b.** Has soil condition changed at the various sampling depths to 1.5 m since the original 2002 survey?
 - c.** How have different land uses and land management impacted the change in soil properties?
 - d.** How has fluctuating rainfall patterns influenced the change in soil properties?
 - e.** Is any detected change in soil properties desirable or undesirable? Has soil condition improved or declined?

- 2.** To use bivariate linear mixed models (BLMMs) and multivariate linear mixed models (MLMMs) to monitor, model and map soil properties in space and time using co-located soil data
 - a.** Can these models be used to accurately predict different soil properties in space and time?
 - b.** How does the inclusion of temporal correlation improve soil maps?
 - c.** How does the inclusion of both temporal and vertical spatial correlation improve soil maps?
 - d.** Are these approaches advantageous in detecting statistically significant temporal changes in soil properties?

3. To use visible near infrared (VisNIR) spectroscopy in combination with traditional laboratory methods to predict properties
 - a. How accurately can soil exchangeable sodium percentage be predicted using these spectroscopic techniques?
 - b. How does the inclusion of readily-available, ancillary, laboratory-measured soil data such as soil total carbon content and pH improve the prediction of soil organic and inorganic carbon content?

4. To determine a spatio-temporal modelling/mapping technique that can sufficiently deal with the zero-inflated and highly-skewed datasets of soil inorganic carbon (SIC) content
 - a. Can the 'mixture model' approach be used to do this?
 - b. Can maps of SIC content accurately be created by taking the difference of soil total carbon and soil organic carbon content maps?

References

- Lark RM, Papritz A (2003) Fitting a linear model of coregionalization for soil properties using simulated annealing. *Geoderma* **115**, 245-260.
- Marchant BP, Crawford DM, Robinson NJ (2015) What can legacy datasets tell us about soil quality trends? Soil acidity in Victoria. IOP Conference Series: Earth and Environmental Science. **25** No. 1. IOP Publishing.
- McBratney A, Field DJ, Koch A (2014) The dimensions of soil security. *Geoderma* **213**, 203-213.
- Onus A, Cattle S, Odeh I (2003) How do Lachlan Valley cotton soils compare to northern NSW? *Australian Cottongrower* **24**, 28–30.
- Viscarra Rossel, RA, Walvoort, DJJ, McBratney, AB, Janik, LJ, Skjemstad, JO (2006) Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* **131**, 59-75.

Chapter 2

2. Review of the literature – Monitoring & modeling soil change - the influence of human activity & climatic shifts on aspects of soil spatio-temporally

Published as:

Filippi, P, Minasny, B, Cattle, SR, Bishop, TFA (2016) Monitoring and Modeling Soil Change: The Influence of Human Activity and Climatic Shifts on Aspects of Soil Spatiotemporally. *Advances in Agronomy* **139**, 153-214.

Abstract

Soils naturally change through time, but anthropogenic activity has significantly altered the rate and direction of soil change. As well as further impacts of human activity on soil into the future, it is also expected that recent climatic shifts will have an important effect. Certain soil properties are more susceptible to change than others, and it is essential that we focus on monitoring a suite of these important soil properties. There are a variety of methods of measuring and monitoring these changes in soils throughout the world, each having their own advantages and disadvantages. Historically, long-term soil experiments and space-for-time-substitution studies were relied upon, however, there has been a shift in focus to assessments of soil change over larger areas. For this reason there has been increasing encouragement for nations and regions to implement structured soil monitoring networks. These networks are effective, but time and resource consuming to implement and maintain, and are simply not an option for many parts of the world, which has led to recent focus on utilising existent soil data, or 'legacy data' to detect changes in soil over space and time. Whether legacy data is in the form of soil surveys or laboratory databases, it does come with limitations, but also presents many opportunities in detecting historical and present changes in soil condition. There are several challenges in soil monitoring studies, such as detecting a significant change in soil properties, the issue of soil spatial variation, operational issues, and the difficulty in disentangling the effects of climate and management. While it is important to know of historical changes in soil condition, it is imperative to predict how climate and management will influence soil change into the future, and this can be done by using temporal soil models. There are two broad types of models available; process-based and empirical models, and there have been many recent increases in the number and quality of these models. As we strive to move away from laborious and expensive soil surveys, these models become more invaluable, however they should not necessarily be seen as a replacement for monitoring studies, but as an addition. This review reinforces the cruciality of soil monitoring, and suggests that we should focus on the wealth of soil legacy data available to us from previous soil measurement exercises. Rather than a sole emphasis on a select few topsoil properties, we should focus on a suite of important soil attributes at various vertical depths. There seems to be comprehensive understanding of the impact different agricultural activities have on soil change, but the same cannot be said for the impact of climatic shifts, and this must improve. Furthermore, it is essential to utilise available statistical analysis methods to detect change, and to provide figures of model quality in these studies. With all this in place, the accurate representation of past and future changes in soil condition is possible, providing a guide for future landuse adaptation.

2.1. Introduction

Soil is an extremely valuable finite natural resource, although its importance has often been overlooked. Humans are concerned with change in the environment, but as visual creatures this concern has typically been focused on visible changes on top of the landscape rather than changes in what lies beneath. The world is currently faced with the global environmental sustainability challenges of food security, water security, energy sustainability, climate stability, biodiversity, and ecosystem service delivery, leading to concern on a global scale (McBratney *et al.* 2014). When these environmental challenges are analysed, it is clear that soil has a fundamental part to play in all of them (McBratney *et al.* 2014). The recognition that soil is fundamental in overcoming all of these global environmental challenges, and not solely food security, has led to increased efforts into maintaining and improving global soil resources (McBratney *et al.* 2014).

It is well-known that soils inherently change through time, with this change propelled by a variety of climatic and anthropogenic factors and their combined impacts (Tugel *et al.* 2006). The growing human population and their strain on soil resources has significantly transformed the nature, intensity, and speed of change in soil properties and is perhaps the greatest driving factor of soil change over short time scales (Tugel *et al.* 2006; Hartemink *et al.* 2008). Landuse and land management impacts on soil are particularly important in understanding soil change; however, providing information about climatic changes that occur on the human time scale must not be neglected (Tugel *et al.* 2006). The role that climate change plays in altering soil condition, and how soils respond to changes in temperature, rainfall and atmospheric composition has recently received much attention on the global scale (Brevik 2012).

In order to ensure the security of our soil resources and to better understand climatic and anthropogenic impacts on soils, it is essential that we assess and monitor soil condition, and the direction and nature of soil change on a regional, national and global scale (Wilson *et al.* 2010). Soil monitoring is essential for the early detection of changes in soil condition (Morvan *et al.* 2008), as soil degradation typically occurs slowly and is often irreversible (Arrouays *et al.* 2012). Many nations have realised this importance for soil monitoring, and it has recently been strongly promoted worldwide to assess soil condition to prioritise investment in natural resource management. Although the development of soil monitoring networks (SMNs) is particularly useful, it is not the sole method of assessing how soil has changed over time. It can be extremely expensive to implement SMNs and information on the extent of soil change is not immediately obtained, as rates of change can only be predicted after soil properties have been measured twice (Marchant *et al.* 2012). Because of these limitations, many have turned to more cost efficient methods of assessing soil

change that produces immediate information, with a strong recent trend of using soil data that already exists, otherwise known as soil legacy data.

While monitoring soils is an important method of providing information for natural resource management, it is essential that SMNs and soil change studies consider the beneficial supplementary activities of soil mapping and modelling (McKenzie *et al.* 2002; Arrouays *et al.* 2014). The integration of soil monitoring, mapping and modelling provides an opportunity to fully utilise gathered information, and to achieve maximum benefit from this information (McKenzie *et al.* 2002). Digital soil mapping (DSM) allows huge datasets to become easily interpretable, and is very effective at displaying and understanding the spatial information and variability of soil (Bui 2006). There have been recent advances in DSM techniques and the availability of environmental covariate information, with increasingly accurate predictions of soil properties at unvisited locations able to be made. Much of the attention of DSM has been solely focused on modelling the spatial variation and distribution of soil properties; however, DSM techniques are now increasingly being used to display temporal shifts in soil properties concurrently. The use of DSM techniques to show changes in soil properties is growing, particularly as soil change and soil security become more pressing issues in the scientific and political community.

While monitoring provides information regarding past and present soil conditions, the inclusion of future temporal modelling in soil change studies delivers insight into rates of soil change over relatively long time scales into the future, providing crucial information for decisions on natural resource management funding and policy (McKenzie *et al.* 2002). Unlike soil monitoring, predicting soil change into the future is not plagued with massive financial costs, and thus provides an extremely useful, cost-efficient method for predicting how changes in future climates and landuses will affect the soil resource. Not all methods and approaches to assessing and monitoring soil change provide the appropriate information to adequately perform mapping and modelling, and unfortunately, many studies that do have the relevant information still fail to combine these activities, eliminating the opportunity to benefit from the collected soil data in the greatest way (McKenzie *et al.* 2002).

2.2. Soil properties most susceptible to anthropogenic activity and climate change

We know that soil changes over time, but certain properties of soil are more susceptible to change, while others are more resistant to driving factors such as climatic shifts and anthropogenic activity (Tugel *et al.* 2006). For example, clay content of a soil is unlikely to change significantly as a result of increasing temperatures or applications of organic amendments on a decadal scale. In contrast, it could be expected that soil organic carbon (SOC) levels would likely be affected as a result of the same shifts in climate and management over the same time scale.

The susceptibility of certain soil attributes to change, and the perceived importance of these attributes, is a major factor in determining the soil properties of focus in soil monitoring and change studies. Soil properties that are considered to be of both agricultural and environmental importance are often of principal interest. Soil degradation indicators such as erosion, salinity, pH, fertility indicators and SOC are common aspects monitored in soil change studies, although this varies according to individual, regional and national interests and concerns. In particular, SOC has recently taken much of the spotlight, principally due to the growing idea that SOC sequestration could potentially counteract greenhouse gas emissions (Wilson *et al.* 2010), and its worth as an indicator of overall soil condition due to its relations with chemical, physical and biological properties of soil (McKenzie *et al.* 2002; Jones *et al.* 2008).

An important concept to consider in this discussion is how easily a degraded soil property can be replenished, or the ‘degree of reversibility’ of a soil property (Hartemink 2006). It is easier to reverse degradation of certain soil properties, for example, depleted soil phosphorus reserves can be replenished by the application of inorganic fertilisers, but restocking SOC levels to their natural state would prove much more difficult (Hartemink 2006). The depth that change has occurred also has significant implications, as it is generally simpler to alter topsoil properties than those deeper down the soil profile (Hartemink 2006). The inherent resilience of the soil also plays an important role in soil change. Soils have an intrinsic capability to resist and recover from external disturbance, but this varies with soil type (Lal 1997).

2.2.1. Anthropogenic activity

The effect of human activities on the soil resource has long been a topic of interest, with much literature discussing this subject (Tugel *et al.* 2006; Richter *et al.* 2007; Richter *et al.* 2011). Almost no characteristic of soil is immune to alteration from human activities, and it is well known that humans can have a significant effect on most soil properties over relatively short timescales. Whether this activity is deforestation, mining, landuse change, fertiliser application, tillage, etc., there can be significant alterations of soil chemical, physical and biological properties, with the potential to greatly accelerate soil degradation (Stockmann *et al.* 2013).

Humans can drive soil change through a multitude of actions, but the most globally significant is undoubtedly agriculture. While landuses such as mining and landfill/waste-sites may have a severe and intense impact on the soil resource at a particular location, there is no landuse that mantles the earth as extensively as agriculture. There is an assumption that human activity and agriculture inherently leads to degradation of the soil resource, but this is not always the case, with several positive impacts resulting from human activity. For example, human activities are known to

have lowered SOC levels in some areas (Saby *et al.* 2008a; Karunaratne *et al.* 2014), but increased SOC levels in other locations (Minasny *et al.* 2012; Ross *et al.* 2013).

As aforementioned, the various roles human activity and agriculture play in altering different soil properties are extensive, but there are certain human activities that have a strong impact on very important soil attributes globally that should be highlighted (Table 1). These activities largely correlate to the major soil degradation issues that the world faces; soil acidification, erosion, salinisation, compaction, and loss of fertility (Oldeman 1994). Soil acidification is largely driven by the removal/displacement of basic cations in the soil, which is typically caused through excessive leaching in areas continuously irrigated or through their removal when plant biomass is continuously harvested. Soil erosion is often triggered by the conversion of land from its natural state into arable land, where groundcover is removed and the soil exposed, making it vulnerable to displacement from wind and water. Soil salinisation can be brought on by continuous irrigation as salts are accumulated in the soil (because the water is used by the plants, but the salt naturally found in the irrigation water remains in the soil), or through dryland salinity, where salt is brought to the surface by capillary action in areas with salty shallow water tables, and is particularly accentuated by the removal of trees and perennial plants. The physical deterioration of soils is also a major degradation issue, with activities such as repeated trafficking of heavy machinery and pressure from livestock causing significant soil compaction. The loss of soil fertility is commonplace in agricultural soils worldwide and is caused by constantly 'mining' the soil for nutrients without replenishment from some external source, and eventually results in nutrient depletion. The impact humans have on SOC stocks have also become a major concern in recent decades, and activities such as deforestation and continuous cultivation have dramatically reduced SOC stocks in many areas of the world. These major, human-induced soil degradation issues and the activities driving them are summarised in Table 1.

Table 2.1 – Major global soil degradation issues resulting from human activity/agriculture

Soil degradation issue	Anthropogenic driving factor and nature of soil change	Reference	Example rates of change from literature (topsoil)	Reference	Time before sig. difference	Relative sensitivity to change
<i>Acidification</i>	- Continuous irrigation leaches basic cations out of soil	(Helyar <i>et al.</i> 1990)	-0.002 to -0.1 pH unit yr ⁻¹	(Singh <i>et al.</i> 2003)	Decades	*
	- Removal of basic cations in plant biomass harvesting - Application of N fertilisers leaves excess H ⁺ ions in soil		-0.015 to -0.04 pH unit yr ⁻¹	(Guo <i>et al.</i> 2010)		
<i>Erosion</i>	- Landuse change and soil exposure increases incidence of wind/water erosion	(Bennett 1929)	-5 to 20 t ha ⁻¹ yr ⁻¹ -10.2 to 11.1 t ha ⁻¹ yr ⁻¹ -17 t ha ⁻¹ yr ⁻¹	(Singh <i>et al.</i> 1992) (Yang <i>et al.</i> 2003) (Pimentel <i>et al.</i> 1995)	Days – decades	***
	<i>Salinisation</i>	- Continuous irrigation accumulates salt in soil - Removal of deep-rooted plants causes water table rise	(Ghassemi <i>et al.</i> 1995)	-0.03 ECe dS m ⁻¹ yr ⁻¹ -0.5 to +3.5 ECe dS m ⁻¹ yr ⁻¹	(Herrero and Pérez-Coveta 2005) (Shirokova <i>et al.</i> 2000)	Decades
<i>Physical deterioration</i>	- Soil compaction and bulk density increases under heavy machinery and livestock	(Hamza and Anderson 2005)	BD: +0.004 g/cm ³ yr ⁻¹ Po: -0.15% yr ⁻¹	(Islam and Weil 2000) "	Years – decades	**
	<i>Soil fertility</i>	- Continuous mining of soil nutrients leads to depletion - Soil chemical degradation affects nutrient availability	(Hartemink 2006)	N: -14 to -20 kg ha ⁻¹ yr ⁻¹ P: -2 to -3 kg ha ⁻¹ yr ⁻¹ K: -12 to -18 kg ha ⁻¹ yr ⁻¹	(Stoorvogel <i>et al.</i> 1993) " "	Years – decades
<i>Soil organic carbon</i>	- Decreases as a result of soil exposure, removal of native vegetation etc.	(Guo and Gifford 2002)	-0.26 g kg ⁻¹ yr ⁻¹ -0.18 g kg ⁻¹ yr ⁻¹	(Saby <i>et al.</i> 2008b) (Sleutel <i>et al.</i> 2003)	Years – decades	***
	- Increases through management techniques, such as green manure cropping and composting		+0.12 g kg ⁻¹ yr ⁻¹	(Goedts and van Wesemael 2007)		

NB: It is important to highlight that these rates of change are often not linear. (BD = bulk density, Po = total porosity, N = nitrogen, P = phosphorus, K = potassium, ECe = electrical conductivity of a saturated soil extract) (Sensitivity = low (*) – high (***))

2.2.2. Climatic shifts

In contrast to the effects of anthropogenic activities on soil change, the effects of climate change in the anthropocene on soil properties are not yet well understood, and there is much debate regarding its potential impact on soil condition. Globally, there have been recent observations of changes in atmospheric temperature and composition, and precipitation amounts and patterns, with further changes anticipated. We do not know the severity of these future changes, but we can expect that this will influence the nature and extent of shifts in soil condition (Brevik 2012). While there is a multitude of soil properties understood to be significantly affected by anthropogenic activity and management, there are believed to be a smaller number of attributes susceptible to changes in climate over the human time scale (Tugel *et al.* 2006). However, it has been suggested that climate change will impact soil chemical, physical and biological properties and hence ecosystem functionality and agricultural productivity (Allen *et al.* 2011; Brevik 2012).

It is commonly accepted that one of the soil properties most vulnerable to climate change is SOC content, and this is partially the reason for the recent emphasis on SOC research and monitoring. Climate change is expected to significantly alter SOC dynamics, and consequently alter many of the other chemical, physical and biological properties associated with SOC (Brevik 2012). Increasing atmospheric and soil temperatures are expected to significantly decrease carbon storage in soils through increased decomposition of labile pools of carbon (Kirschbaum 1995; Paré *et al.* 2006). This is partially a result of the alteration of soil microbial community structure and function observed with increasing temperatures, which hastens the degradation of SOC, and thus releases more CO₂ into the atmosphere (Frazer 2009). In general, SOC stocks increase as mean annual temperature decreases worldwide (Post *et al.* 1982; Stockmann *et al.* 2013) and the sensitivity of SOC decomposition rates to temperature increases is believed to be much greater at lower temperatures than at warmer temperatures. Kirschbaum (1995) proposed that a 1°C increase in temperature could cause a SOC loss of 10% in areas with an annual mean temperature of 5°C, whilst regions with an annual mean temperature of 30°C would only lose 3% (Kirschbaum 1995). Temperature increases are also expected to significantly alter nitrification and nitrogen mineralisation (Rustad *et al.* 2000).

While we have some knowledge that SOC is affected by increasing atmospheric and soil temperatures, we are still unsure of the significance this will have on global SOC levels. Several regional and national studies at decadal scales have shown that despite increasing air temperatures, SOC levels have remained constant or increased in some areas of the world (Fantappiè *et al.* 2011; Yan *et al.* 2011; Minasny *et al.* 2012). These increases were attributed to improved land management practices and indicate that the effect of management supercedes that of a warming

climate, however, this theory can only be true for land areas that are actively managed, and have little significance in un-managed and ‘natural’ areas (Minasny *et al.* 2012). Overall, we have some preliminary understanding of the impact increasingly fluctuating temperatures will have on SOC, soil nitrogen and soil microbial communities, but the effects of this on other important soil attributes is poorly understood.

Precipitation patterns and amounts are anticipated to shift significantly, and it is generally expected that dry areas will become drier, wet areas will become wetter, and that rainfall events will become more severe (Trenberth 2011). It is expected that these shifts will have a significant impact on the soil resource, as water influences many soil properties. The predicted changes are both positive and negative, and often contradictory. For example, increased rainfall amounts in some areas could increase leaching, which may further accentuate soil acidification, but may also reduce salinity issues. It is known that SOC is susceptible to increased decomposition under heightened moisture availability, but it is difficult to anticipate how this will influence global SOC stocks (Craine and Gelderman 2011). Reduced rainfall and drought in some areas may lead to salt accumulation in soils, and decrease groundcover, resulting in a greater potential for wind and water erosion. It could also be expected that increased intensity of rainfall events will increase the severity of soil water erosion. In addition to the direct impact of shifting rainfall patterns and amounts on soil, the indirect impacts of this shift will also play a strong role, as this will influence plant growth and the availability of irrigation water, which will be a major contributor to land management decisions made by landholders (Raison and Khanna 2011).

In regard to changes in atmospheric composition, the combination of rising temperatures and CO₂ concentrations may alter net primary productivity (NPP) and the amount and type of organic inputs to soil, potentially increasing SOC levels (Stockmann *et al.* 2013), and causing further complications in predicting future global SOC stocks. Elevated atmospheric CO₂ concentrations are also expected to have a significant impact on nitrogen mineralisation, immobilisation, availability and nitrification (Müller *et al.* 2009; Norby *et al.* 2010).

The degree of change in soil properties and characteristics as a result of climate change will be significantly influenced by soil type and the specific location. As soil is such an interactive and complex system, the alteration of certain properties will in turn affect the whole soil environment (Reth *et al.* 2005). Furthermore, the indirect effects that climate change will have on how humans interact with the soil will also play a significant role. Overall, it is evident that we have a rudimentary understanding of the impact that climate change may have on some important soil characteristics; however, due to the complex nature of this issue, we are uncertain of the impact this will have on a

global scale. A summary of the properties and characteristics expected to be significantly impacted by climate change is provided in Table 2.

Table 2.2 – Summary of soil properties/characteristics expected to be significantly impacted by climate change, and the mechanisms driving that change

Soil properties /characteristics	Expected change and climatic driving factors	Time before sig. change	Reference
<i>Organic carbon</i>	- Increased temperatures – increased SOM decomposition – lower SOC stocks	Decades	(Kirschbaum 1995) (Craine and Gelderman 2011) (Hyvönen <i>et al.</i> 2007)
	- Heightened soil moisture – increased SOM decomposition – lower SOC stocks - Expected increased plant production – increased SOC stocks		
<i>Salinity</i>	- Reduced rainfall could lead to salt accumulation – increased salinity	Decades	(Chhabra 2002) (Chhabra 2002) (Raison & Khanna 2011)
	- Increased rainfall will enhance leaching of salts – decreased salinity - Increased rainfall – change water table depth – increased dryland salinity		
<i>Nitrogen</i>	- Temperature increases – increased N mineralisation and nitrification – mixed results about direction of change	Years - decades	(Rustad <i>et al.</i> 2000) (Groffman <i>et al.</i> 2009) (Müller <i>et al.</i> 2009) (Norby <i>et al.</i> 2010)
	- Soil moisture – significantly affects N cycling processes		
	- Elevated CO ₂ – increased N mineralisation and immobilisation & decreased nitrification		
	- Elevated CO ₂ – decrease in soil N availability		
<i>pH</i>	- Increased rainfall will enhance leaching of basic cations, exacerbating soil acidification	Decades	(Tang & Rengel 2003) (Rengel 2011)
	- Potential greater biomass production could increase acidification through product removal		
<i>Bulk density/ porosity</i>	- Largely linked to the impacts that shifting SOM levels will have on bulk density – decreased SOC – increase BD, increased SOC – decreased BD	Decades	(Allen <i>et al.</i> 2011)
<i>Soil erosion</i>	- Increased occurrence of severe rainfall events – potential to increase water erosion	Hours – years – decades	(Eldridge <i>et al.</i> 2011) (Murphy <i>et al.</i> 2011)
	- Increased occurrence of drought – potential to increase wind erosion		
<i>Soil biota</i>	- Increased temperatures – shifts in composition and activity of microbial communities	Years – decades	(Haynes 2008) (Thonon and Klok 2007)
	- Shifting rainfall patterns – impact earthworm and termite populations		
<i>CEC</i>	- Decreased SOM levels – decreased CEC	Decades	(Magdoff and Weil 2004) (Blake <i>et al.</i> 1999) (Allen <i>et al.</i> 2011)
	- Anticipated acidification – decreased CEC levels - Low CEC of soils – increased leaching of base cations – increased acidification		
<i>Aggregate stability</i>	- Repeated wetting and drying results in reduction in aggregate stability	Decades	(Kay <i>et al.</i> 1994)
<i>Inorganic carbon</i>	- Increased rainfall could increase dissolution of soil carbonates	Decades	(Suarez 2006)

2.3. Methods of measuring and monitoring soil change

The study of soil change is far from one-dimensional, and there are a variety of approaches and methods that can, and have been, undertaken (Richter and Markewitz 2001). The particular approach chosen to assess soil change depends on available funding, existing soil data, soil properties of interest, long-term intentions, scale of the study, and the statistical skills of people involved. Not all approaches to measuring and monitoring soil change adequately fulfil the three pillars (monitoring, mapping and modelling) of understanding spatio-temporal soil change, however, each approach has its own advantages and disadvantages.

One of the oldest techniques in assessing soil change is the use of space-for-time substitution (SFTS), which is also commonly known as paired-sites in soil science. This approach uses the assumption that soils in different spatial positions (usually across a fence) in the landscape correspond to different phases in the temporal response to a driving factor, such as landuse change (Pickett 1989; Hodson *et al.* 1998; Schjønning *et al.* 2009). This technique is typically used on small areas, but is relatively inexpensive and particularly useful where there is little or no data on soil properties for the area of interest (Pickett *et al.* 1989). In a strict sense, paired-site sampling is not a soil monitoring technique as soil samples are not collected and analysed at different time periods, but it can be a useful way of assessing how soil changes over time as a result of some driving factor.

Long-term soil experiments (LTSE) are studies conducted on an individual farm or research station where soil properties are periodically sampled over an extended period of decades to centuries. Although somewhat few in number due to the high costs and effort involved, LTSEs have greatly assisted in developing an understanding of change in soil properties and processes over relatively long time scales (Richter *et al.* 2007). Information obtained from LTSEs could be seen as quite limited due to the experiments being performed over such a small area, however, they have contributed greatly to the development of simulation models, such as the RothC carbon simulation model from the Rothamsted experiments. The biggest advantage of these experiments is that they provide reliable, densely sampled in time soil data regarding soil change over longer timescales than many other approaches (Richter *et al.* 2007).

The use of purpose built soil monitoring networks (SMN) is perhaps the most effective way of reliably assessing temporal change in soil condition over large areas. The quality, as well as the scale of true SMNs is highly varied, and ranges from sparse to dense data and from regional to national soil monitoring projects. While SMNs are invaluable in detecting soil change, initiating and maintaining SMNs is extremely time and resource consuming, and this is the largest reason for the relative lack of such networks around the globe. There are largely two approaches used when

sampling within SMNs; the use of design-based systems, which uses probabilistic sampling, and the use of model-based systems, which uses systematic or purposive sampling, with each of these approaches having their own advantages and appropriateness for purpose (de Gruijter *et al.* 2006).

While many countries do not have established SMNs, many often do have soil legacy data that can be utilised in detecting spatio-temporal variations in soil properties (Karunaratne *et al.* 2014). Soil legacy data in this review is defined as; any existing soil observations that can be used to detect, or assist in detecting spatio-temporal variation in soil. Legacy soil data is extremely broad in its nature and extent, and can range from local soil surveys to national soil databases. Utilising this legacy data often encompasses many challenges, including that the data was not collected for the purpose of soil monitoring, and thus often lacks proper statistical design (Marchant *et al.* 2012). Nevertheless, legacy data provides a great opportunity to understand spatial and temporal changes in soil at a low cost, and is often the only option in data-sparse areas of the world.

Just as soil legacy data is wide-ranging in its type, so too are the approaches to its use. Some studies use several time-series legacy datasets to infer soil change, eliminating the need to resample, while others use a combination of legacy data and re-sampling a specific area – either at the same sites as the legacy data, or new locations. Much of this legacy data is in the form of soil surveys, either originally collected to characterise soil units for traditional soil type mapping, or to conduct a general soil survey of a particular area (Karunaratne *et al.* 2014). Soil test databases can also be utilised to infer change in soil properties over time. These databases can contain thousands to millions of soil test results collected over a long timeframe and typically large areas. This approach is inexpensive as it requires no resampling, but there are many limitations and challenges typically not faced with other types of legacy data, such as numerous sources of bias and lack of location information (Saby *et al.* 2008a, Marchant *et al.* 2012). Another approach is analysing archived soil samples to understand and detect soil change. The source of these archived samples play a major part in the quality of these studies, with samples being used from LTSEs, soil surveys or even structured SMNs.

In soil monitoring studies, long-term changes are considered more important than short-term variations, and therefore long-term monitoring is more highly regarded as it provides better information regarding more permanent changes in soil condition. However, if the time steps are too far apart, important shorter-term variations may be missed and may disguise what has happened or has caused that change. It is thus preferential to monitor over the long-term with relatively short time steps (Hartemink 2006).

2.3.1. Space-for-time-substitution/chronosequences/paired-sites

There is much confusion in the literature about the exact definitions of space-for-time-substitution (SFTS), chronosequences, and paired-sites, particularly since many of these definitions have been developed in other disciplines of environmental science, and have since been re-appropriated for soil science. For example, Pickett (1989) refers to SFTS as “inferring a temporal trend from a study of different aged sites”, while Hartemink (2006) refers to SFTS studies as those which have an assumption that soils of a set of sites are the same. These definitions are clearly antagonistic. The terms chronosequence and paired-sites are also often not clearly defined and used synonymously or incorrectly by many studies (e.g. Tye *et al.* 2013; Kuraganova *et al.* 2014). Despite this, these terms all share a common theme; to infer a temporal trend from a study of a set of sites in different spatial positions, sampled once, and at the same time.

For the purpose of this review, an attempt to define each of these terms from a soil science perspective will be made based on the various definitions provided by the literature. In general, SFTS is used quite loosely, and is often treated as an umbrella term which encompasses chronosequences and paired-site sampling. In paired-site studies, sampling is undertaken at the same time from an undisturbed location (typically in a natural state), and an adjacent disturbed location (typically agricultural land). This is commonly referred to as across-the-fence, adjacent, or bio-sequential sampling in soil science. In paired-site sampling it is assumed that soils of cultivated and undisturbed land are equivalent, and that any observed changes in soil properties are because of dissimilarities in landuse and management (Hartemink 2006). In soil science, chronosequences are most often referred to as a set of sites that share many similar soil forming factors (climate, organisms, relief, and parent materials), but the soils are of different ages (Sauer 2015).

When analysing the impact of landuse or climatic fluctuations on soil change, it is vital that all other soil-forming factors remain constant. How can you disentangle the impact of landuse/climatic shifts on soil when the soils are deemed different? For this reason, chronosequences are not particularly useful to detect decadal changes, as dissimilarities in soil properties could be a result of the differing ages of parent materials at different sites, and not due to landuse, or climatic differences. Although uncommon, SFTS can be used to study the impact of climatic changes on soil properties. Areas with constant factors such as geology, plant species etc., but with different climatic gradients can be used to predict how future climatic conditions will influence soil change (Barraclough *et al.* 2015). This approach has several limitations such as many inherently dissimilar factors, and several assumptions (Fukami and Wardle 2005). Used much more commonly, is the use of SFTS to interpret the historical effects of landuse change on soil properties.

The use of these approaches to observe soil change can be quite useful, particularly when there is a lack of legacy data and when the time of landuse change is known. It is commonly used for smaller scale studies, and was historically the most commonplace approach of assessing temporal soil change (Pickett *et al.* 1989). Such studies commonly assess how soil has changed over relatively long time periods, ranging from decadal scales (Rampazzo *et al.* 1999a; Rampazzo *et al.* 1999b; Mudge *et al.* 2014), to centuries (Cattle *et al.* 1994; Tye *et al.* 2013).

A study in Nottinghamshire, UK, used SFTS/paired-sites to assess the impact of landuse change on soil fertility, pH and carbon over more than 200 years (Tye *et al.* 2013). Information regarding landuse changes dating back to 1781 was known, allowing a comparison of soil under woodland to soil that had been cleared and converted to cropping between the years of 1781 and 1881. The study showed that conversion to cropping had a significant effect on soil properties, with arable lands displaying lower SOC contents and higher pH levels in comparison to woodlands (Tye *et al.* 2013). Similar results were found in a study in south-eastern Australia, where an un-disturbed site of native vegetation was compared to a site that had been cleared in 1894 and had been regularly cropped since (Cattle *et al.* 1994). It was established that both negative (lower SOC, nitrogen, and macroporosity) and beneficial (more alkaline pH, lower salinity) changes in the soil had occurred as a result of long-term cropping. It was also highlighted that the heterogeneity of the soil had been significantly altered, with the cropping area showing much more spatially homogenous results in several soil properties, resulting from continuous ploughing (Cattle *et al.* 1994).

On a shorter time scale of decades, a study on the Austrian-Hungarian border used SFTS/paired-sites to compare undisturbed soils with adjacent agricultural soils of Austria and Hungary (Rampazzo *et al.* 1999a; Rampazzo *et al.* 1999b). The undisturbed plot had not been cultivated for over 50 years; the Austrian plot had been subjected to intensive smallholder farming, and the Hungarian plot subjected to extensive broad-scale farming, resulting in very different soil management practices over the previous 50 years. Results from both agricultural sites showed that there was a significant alteration of several important physical, chemical and biological soil properties compared to undisturbed sites, with large differences in soil condition between sites, driven by rigorous agricultural management. Agricultural areas showed decreases in SOC, loss of nutrients, reduced microbial activity, and higher bulk density, and underlined the sensitivity of important soil attributes to change through agricultural practices (Rampazzo *et al.* 1999a; Rampazzo *et al.* 1999b). A similar, but much more extensive study compiled a database consisting of soil results from SFTS/paired-site studies to assess how SOC levels were impacted by wide scale landuse change over the Russian Federation. The collapse of collective farming after 1990 caused the abandonment of 45 million hectares of arable land, which was found to have an overwhelmingly positive impact

on SOC levels, with an average of $0.96 \pm 0.08 \text{ Mg C ha}^{-1} \text{ yr}^{-1}$ being accumulated in the upper 20 cm of the soil profile from 1990 to 2009 (Kurganova *et al.* 2014).

The historic popularity of using SFTS to assess soil change can be attributed to its relatively low cost, simplicity, and immediate production of soil change information (Pickett *et al.* 1989). Changes in soil can happen very slowly in relation to human events and commencing long-term monitoring studies is often not feasible or practical (Richter and Markewitz 2001). Direct, continuous observations of soil are not always possible, and many of these studies fail to extend beyond decades, thus shifting the focus to indirect methods of inferring soil change (Mudge *et al.* 2014).

While there are multiple advantages of using SFTS, there are several shortcomings with this approach (Walker *et al.* 2010), and results are often interpreted in a too straightforward and simple manner (Sauer 2015). For paired-site studies, it is essential that compared sites are of the same soil type and that any detected differences in soil properties are due to management. This is often not true, and the uncultivated soil may have been of poorer quality, and consequently not cropped (Hartemink 2006). The inherent spatial variability of soil can also be easily confused with temporal soil dynamics (Mudge *et al.* 2014), and this issue is often amplified by the relatively few soil samples that are usually taken in these types of studies e.g. (Assad *et al.* 2013). Sites are also commonly selected without a statistical basis, and randomisation of samples is rare. While paired-sites are appropriate for evaluating the effects of landuse/management on soil change, they are not particularly useful in assessing climatic impacts on soil change, as sites have been subject to the same climatic conditions and no prior information is known. The heavy reliance on assumptions, and the fact that SFTS/paired-sites are frequently not used correctly, has led to misinterpretations about temporal soil dynamics, which has subsequently triggered recent criticism of the approach (Walker *et al.* 2010). While SFTS/paired-sites are suitable for studying temporal changes in some soil properties that are likely to be related in a predictable and temporally linear manner (e.g. carbon, pH), it is not ideal for others (e.g. soil microbial species composition and abundance)(Walker *et al.* 2010).

2.3.2 Long-term soil experiments

Long-term soil experiments (LTSE) can be described as; '*Field experiments with permanent plots that are periodically sampled to quantify soil change across time scales of decades to centuries*' (Richter *et al.* 2007). Long-term soil experiments that monitor soil properties for more than a century are rare, as these experiments are highly costly and are usually prone to abandonment over time.

The most cited LTSE, and the oldest of its kind, is the Rothamsted experiment in England. Soil has been regularly monitored at Rothamsted since 1843, particularly for the Broadbalk wheat

experiment, where there has been a strong focus on soil research (Rothamsted Research 2014). Perhaps the biggest factor in the success of the Rothamsted LTSEs, and the element that separates it from the rest, is the comprehensive archiving of soil samples for over 170 years, allowing contemporary soil analytical techniques to be applied. There are over 300,000 archived soil samples, providing an extremely valuable resource for assessing soil change over time in response to different management regimes, and climatic fluctuations (Rothamsted Research 2014).

There are several other notable LTSEs throughout the world, however few are as distinguished as the Rothamsted experiments. The Morrow Plots, Illinois, USA, were established in 1876, but the earliest archival soil and chemical data dates back to 1904, with regular soil assessments undertaken at least at 10-year intervals since. Soil archiving has been much less comprehensive at the Morrow plots, and consequently much of the assessment of soil change has relied on legacy soil analysis rather than contemporary laboratory analysis (Aref and Wander 1997). Another LTSE is the Calhoun long-term experiment, South Carolina, USA, which was initiated in 1957 on old cotton fields and soil samples have been taken and archived between 1962 and 2010 (Hayes *et al.* 2014). There are nearly 300 LTSE studies identified on a global inventory of International LTSEs compiled by Richter *et al.* (2014). Experimental duration and quality of these LTSEs varies significantly, with some more useful than others in advancing our understanding of soil change (Richter *et al.* 2007).

As with any approach in assessing temporal soil change, there are limitations and disadvantages with using LTSEs. The proper documentation of management treatments is essential to fully understand possible driving factors of soil change, and this information is often lost over time. Additionally, if there is no soil archiving and original data has to be relied upon, there are obvious issues between comparing results from modern laboratory techniques to those of the past. Additionally, it was historically not possible to measure certain soil properties. Soil archiving does present its own problems, as soil properties change when stored for long periods of time. It is known that biological properties are extremely volatile and almost impossible to measure after significant storage, however, more stable soil properties such as pH are also known to change in storage over time (Slattery and Burnett 1992). While LTSEs provide dense data, such experiments are performed on a relatively small land area, creating issues in regards to data extrapolation to broader scales and other locations (Clark 2002; Richter *et al.* 2007). Nevertheless, the most noteworthy limitation of LTSEs is the significant time period before useful data is produced, and the huge cost and effort involved with initiating and maintaining these experiments.

While many scientists are cautious to commence new LTSEs, there is no doubt that well-managed LTSEs will be fundamental in understanding long-term soil trends in the future and the

effect of soil management and climatic changes on soil change (Richter *et al.* 2007). They are particularly useful in assessing soil properties that change over longer time scales, for example, SOC content of temperate soils generally changes gradually, and can often take over 20 years to observe significant changes (Richter *et al.* 2007).

One of the most useful characteristics of LTSEs is that simulation models can be tested against the data (Richter *et al.* 2007). The RothC carbon cycle model was developed and calibrated from data collected from the Rothamsted experiment, and simulates SOC turnover, and allows for the impacts of temperature, moisture content, soil type, and plant cover on the turnover process (Coleman and Jenkinson 1999). The model has been used in a variety of climatic zones and allows SOC outputs to be calculated from known inputs, and vice versa (Coleman and Jenkinson 1999). The RothC model is highly regarded globally by scientists and has been used for major carbon assessment projects throughout the world and to project possible changes in SOC levels into the future (e.g. Smith *et al.* 2005).

2.3.3 Soil monitoring networks

Soil monitoring networks (SMNs) can be described as; '*A set of sites that was purposely constructed to document changes in soil characteristics through periodic assessment of an extended set of soil parameters*' (Morvan *et al.* 2008). Monitoring studies will be strictly referred to here as projects that were intentionally created for soil monitoring purposes, and do not refer to those studies that use legacy data from previous studies such as soil surveys. Many studies that use legacy datasets are often referred to as SMNs, however, by the definition used here are not. While paired sites, and LTSEs have several benefits, the limitations regarding spatial area covered must not be ignored. To achieve global soil security, we need to know about changes over larger areas.

Several countries are beginning to invest in developing structured SMNs, particularly as soil monitoring begins to be recognised as an integral component in managing global soil resources. While many countries monitor their soil resources to an extent, well-established and well-designed SMNs are somewhat uncommon. Those countries that do have well-established SMNs (e.g. France, Britain) often share several things in common; they all possess relatively small land areas, have developed economies, and much of the land is actively managed (Hartemink 2008). There is large variability in SMN quality throughout the world, with developed countries generally having dense data, and developing countries having sparser data (McBratney *et al.* 2003). Obviously, in countries that possess large areas of unmanaged land, such as Brazil, Russia and Australia, the implementation of a national SMN that covers the whole nation becomes a much more costly and less feasible option. Regional SMNs are quite rare, as the focus is generally on soil change at the national scale, and the cost and effort of implementing and maintaining a SMN for a single region is often

unjustified. While there are some differences between national and regional SMNs, the objectives to record the current soil condition status, and to monitor possible changes, largely remains the same (Mol *et al.* 1998).

It is crucial that SMNs are statistically well designed, as they must be sensitive to early changes in soil condition (Arrouays *et al.* 2000). Sampling and monitoring soil is expensive and laborious, and it is thus vital to select an appropriate sampling scheme before data collection commences (Karunaratne *et al.* 2014). There are two broad sampling strategies that can be used when developing a SMN; a design-based approach, which uses probability sampling, and a model-based approach, which uses purposive sampling (often on a grid) (Fig. 1), with each of these statistical approaches having their own strengths and limitations (de Gruijter *et al.* 2006; Black *et al.* 2008). There is ongoing debate concerning the effectiveness of probabilistic sampling compared to systematic sampling, however, the choice of sampling scheme largely depends on the soil properties of interest and the precision required (Morvan *et al.* 2008). Before choosing a sampling design, it is essential to develop a detailed specification of the aim of the monitoring study in statistical terms, as this will be crucial in decision making (de Gruijter *et al.* 2006; Brus *et al.* 2009). Generally speaking, model-based approaches are better suited to providing local estimates for mapping purposes, while design-based methods are better for estimating global means for a study area or particular landuse (de Gruijter *et al.* 2006; Morvan *et al.* 2008). Contrary to what reviews on SMNs around the world suggest (e.g. Morvan *et al.* 2008; van Wesemael *et al.* 2011), reviewed literature indicates that model-based SMNs are more prevalent than true design-based systems, particularly on a national scale. Many of these systems labelled as design-based, are simply 'stratified' approaches that are arbitrarily designed, not purposively developed for monitoring, or have not been resampled, and hence could not be considered true SMNs.

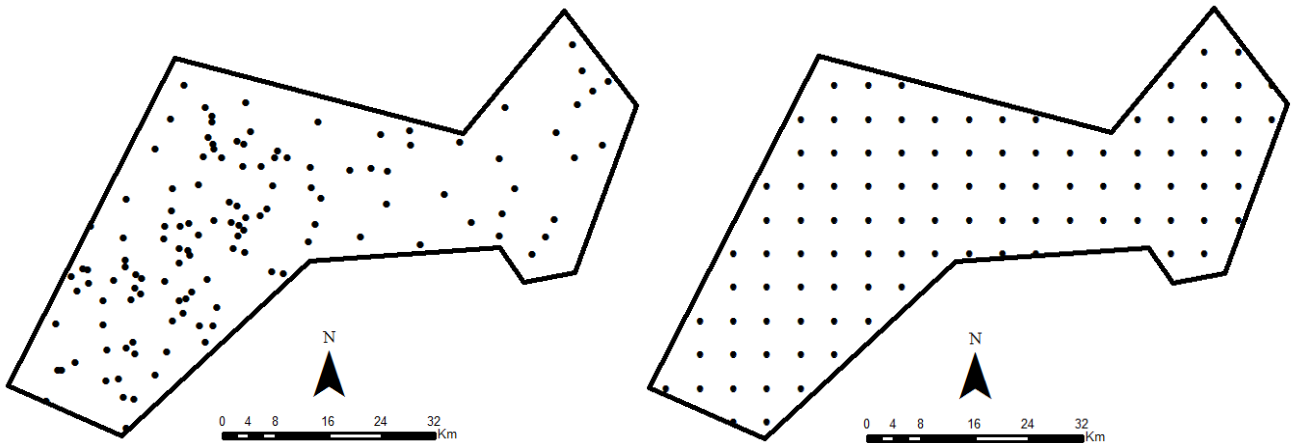


Fig. 2.1 – Examples of a random stratified design-based sampling scheme (left) and a systematic grid model-based sampling scheme (right) of a district in western NSW, Australia

2.3.3.1. Model-based sampling strategies

The sampling sites in model-based approaches are set out in a systematic way, and are particularly useful in mapping the spatial variation of soil properties (de Gruijter *et al.* 2006; Black *et al.* 2008). The use of a square or triangular grid for sample selection ensures an even coverage and is the most common approach for model-based methods. It is essential that some sites be assigned shorter-range points to ensure that the model that is computed from the data describes variation over short distances satisfactorily (de Gruijter *et al.* 2006; Black *et al.* 2008). In model-based approaches to sampling it is assumed that the variation of soil attributes derives from a random process that can be modelled statistically (Black *et al.* 2008). This assumption, rather than randomisation of the sampling scheme, provides the basis for estimation of variances around estimates or parameters (Black *et al.* 2008).

There are several countries that have used the model-based approach in the development of nation-wide SMNs, and the quality of these networks and density of data varies considerably. In the United Kingdom there are two major SMNs, one of which, the National Soil Inventory of England and Wales (NSIEW), uses a square grid sampling approach. The NSIEW was purposely constructed to assess and monitor the condition of topsoil (0-15 cm depth), with measurements taken from 1978-1983, and again from 1994-2003. The network covers England and Wales on a 5 km grid, providing a detailed representation of soils throughout the two countries. The study consists of 5,662 sites for soil sample collection, with 25 composited soil cores taken in a 20 × 20 m square site (Bellamy *et al.* 2005; Lark *et al.* 2006_b). Resampling occurred between at 12-25 year intervals at 40% of original sites, as this was the minimum proportion required in order to detect change in SOC with 95% confidence (Lark *et al.* 2006). Recorded soil data includes SOC, pH, texture information, P and K concentrations, and content of heavy metals (Bellamy *et al.* 2005). Landuse categories include

arable, rough grazing, coniferous and deciduous woodlands, lowland heath, and several others (Bellamy *et al.* 2005).

In the early 2000s, the NSIEW was the only soil inventory on such a scale anywhere in the world to have been resampled (Bellamy *et al.* 2005). The study found that carbon was lost from soils across England and Wales at a mean rate of $0.6\% \text{ yr}^{-1}$, relative to existing soil carbon content, over the survey period (Fig. 2). The rate of carbon loss was found to be faster in soils with higher carbon contents (Fig. 2), and landuse appeared to have no significant effect on the rate of carbon loss, suggesting that this loss could be attributed to the 0.5°C increase in air temperature over that period (Bellamy *et al.* 2005). A modelling study (Smith *et al.* 2007) later disproved this idea, and proposed that the temperature increase could only be accountable for a maximum loss of $0.08\% \text{ yr}^{-1}$ of SOC (Yan *et al.* 2011). This study suggested that this SOC loss was actually due to major trend shifts in land management and legacy landuse effects. Results regarding soil pH indicated that in general, English and Welsh soils were becoming less acidic, with modest, but widespread increases in pH being observed over a majority of the area (Fig. 3) (Kirk *et al.* 2010). An increasing pH could be linked to the increased application of lime on agricultural lands over the study period. Despite the NSIEW being statistically well designed with dense data, it is clear that the study does come with its limitations. Only topsoil (0-15 cm) was measured, disregarding any changes that have occurred over time in deeper subsoil layers of English and Welsh soils.

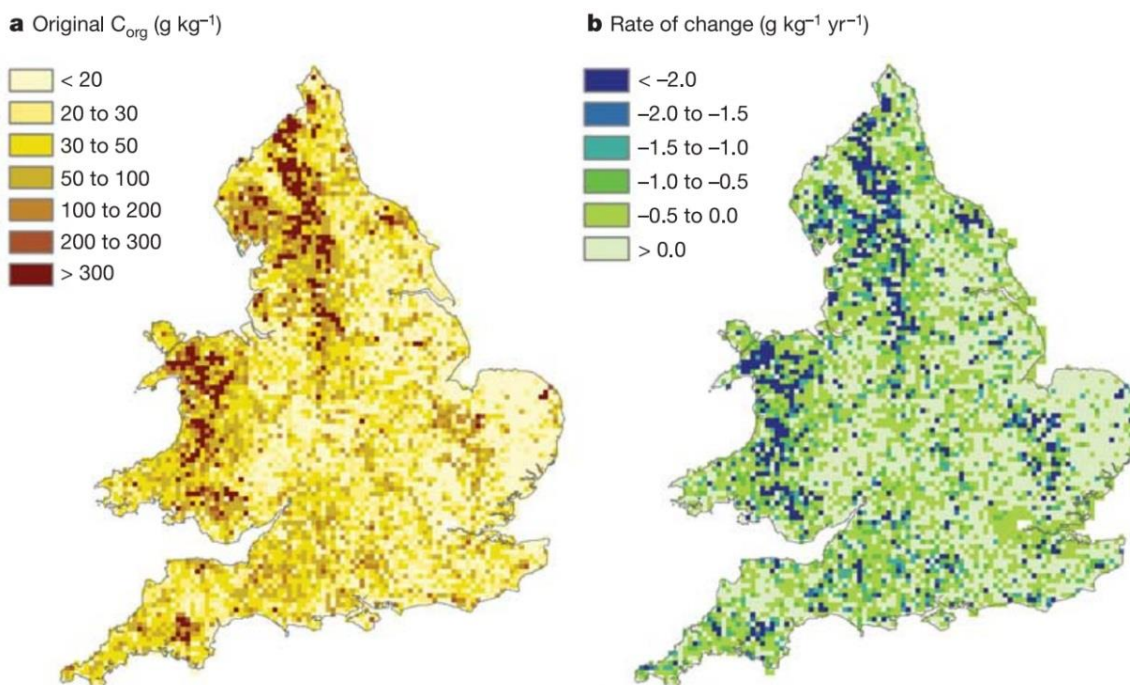


Fig. 2.2 – Changes in topsoil (0-15 cm) SOC contents across England and Wales between 1978 and 2003 – (a) SOC contents in original samples (1978-83), (b) predicted annual rates of change calculated from the changes over the different sampling intervals (Bellamy *et al.* 2005)

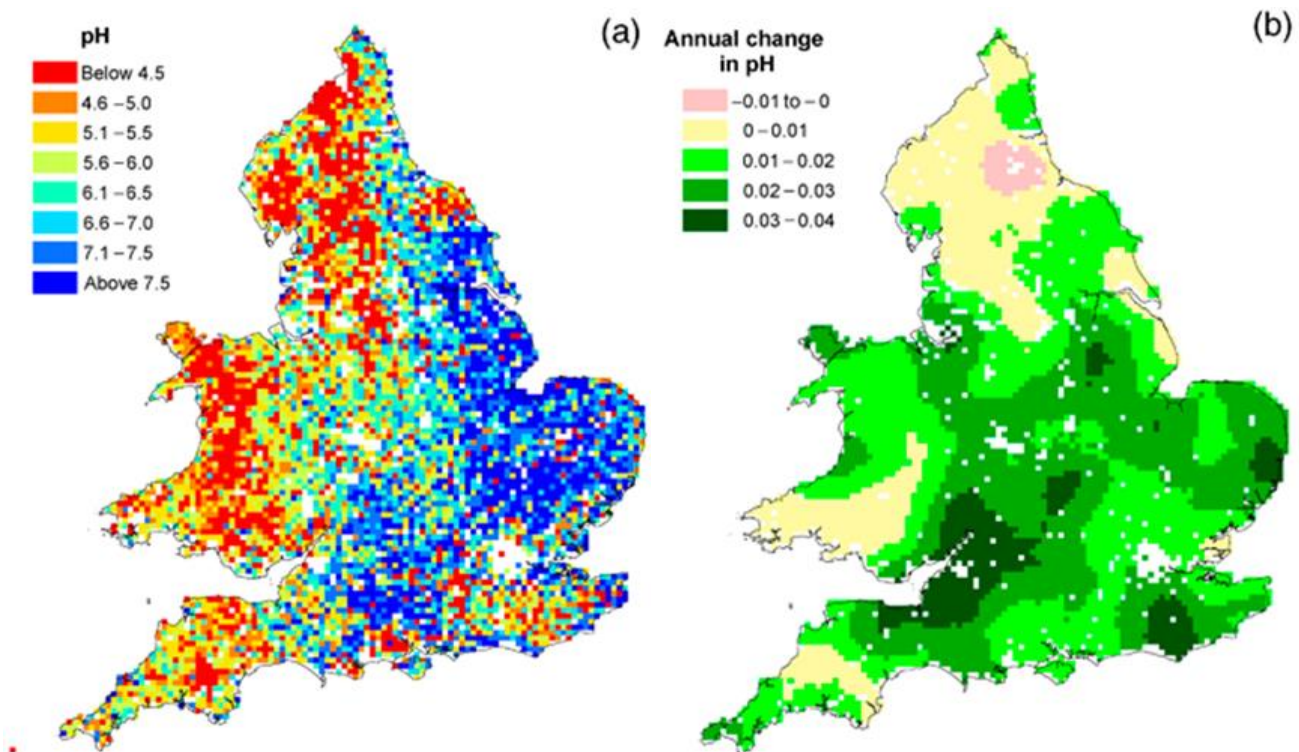


Fig. 2.3 – Changes in topsoil (0-15 cm) pH across England and Wales between 1978 and 2003 – (a) pH of original samples (1978-83), (b) predicted annual rates of change calculated from the changes over the different sampling intervals (Kirk *et al.* 2010)

The model-based approach has also been used in Denmark, where soil has been periodically monitored since 1986 (Adhikari *et al.* 2014). The nationwide 7 km soil-monitoring grid consists of 850 intersects, located on farmland, forests, and other land uses (Adhikari *et al.* 2014). In contrast to the NSIEW, soil was analysed to a depth of 1 m at 0.25 m intervals, providing a more insightful view into temporal soil change throughout the soil profile. Soil carbon was measured in 1986, 1997 and 2009, with results showing a significant decline of SOC in the top 0.5 m from 1997 to 2009. Soil type played a role in SOC dynamics, with an observed increase in SOC on sandy soils and decreases in loamy soils from 1986 to 2009. This effect is linked to landuse, as grasslands and dairy farms are more abundant on the sandy soils (Adhikari *et al.* 2014).

The large costs involved with developing national SMNs has often prevented their implementation, however, the increasing knowledge of the importance of soil monitoring has prompted many countries to set up nationwide SMNs in response to this. As the emphasis on this issue is quite new, many of these SMNs have not yet been resampled and therefore the temporal changes in soil properties cannot be reported. In France, the national soil monitoring network labelled 'Network Measures of Soil Quality' (RMQS), completed the first sampling campaign in 2008, and results regarding soil attribute change will be unknown until resampling occurs after 10 years (Jolivet *et al.* 2006). The French have developed a comprehensive model-based design SMN, with

2200 sites spread evenly across the nation on a 16 km² square grid. Composite samples of 25 subsamples were taken within a 20 x 20 m area at each sample site, for two horizons in the upper 50 cm of the soil profile (Meersmans *et al.* 2013). The RMQS measures a multitude of soil parameters, including SOC, pH and several trace elements, and all soils are archived, providing a valuable resource in the future (Antoni *et al.* 2007).

The model-based approach is also used to monitor soil in Germany and Mexico. In Germany, crop and grazing land is monitored for SOC on a grid (64 km² per site) to a depth of 1 m with composited samples. Initial sampling occurred in 2010, with the resampling due to occur at a 10 year interval in 2020. Mexico uses a grid-based approach, with 78 km² per site on both forest and non-forest land to a depth of 60 cm. Sampling occurred in 2003, with 20% of samples being resampled each year, resulting in a re-sampling frequency of 5 years (van Wesemael *et al.* 2011).

National SMNs are extremely useful, but harmonised soil change information across continents is even more favourable, particularly in Europe, where there is a high human population, and much of the land is actively managed. In 2009 a soil component was added to 'LUCAS', the Land Use/Cover Area frame Statistical Survey for Europe (Panagos *et al.* 2012). The original aim of the survey was to gather completely harmonised data on land use/cover and their temporal changes in the European Union. In 2009, 265,000 georeferenced sites on a 2 x 2 km grid were visited across Europe, with soil samples (0-30 cm) extracted from 22,000 of these points in order to improve the availability of harmonised data on soil properties in Europe (Fig. 4) (Panagos *et al.* 2012). The soil samples were analysed for a suite of soil properties (including; pH, SOC, CEC, texture, heavy metals etc.), and were processed by the same certified laboratory to ensure consistency (Panagos *et al.* 2012). Although not yet resampled, the initial 2009 sampling showed significant spatial trends in SOC content across Europe, with the colder, northern areas typically possessing higher SOC levels, compared to the warmer, drier southern areas (Fig. 4). While this study was not technically designed solely for soil monitoring, it still possesses a structured and somewhat statistically sound design (despite the absence of close pairs). It is less important that model-based monitoring networks are designed to specifically monitor soil, as this does not affect the structure and selection of sampling locations, whereas in design-based systems it is required that they are specifically designed to monitor soil as this affects the selection of sampling sites.

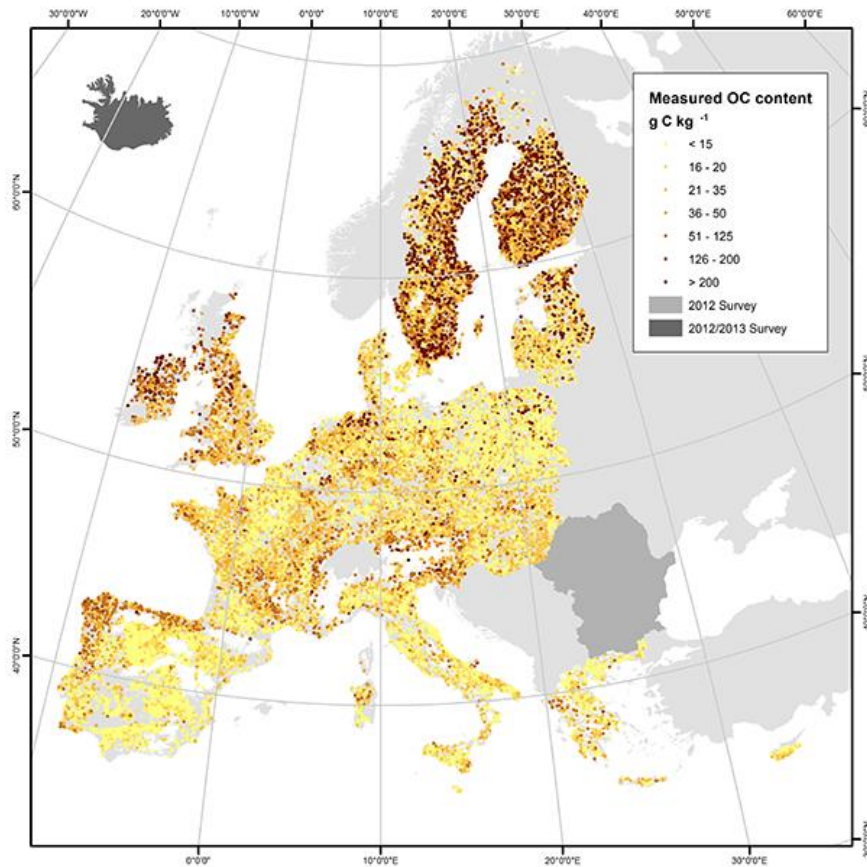


Fig. 2.4 – Measured SOC content (g C kg^{-1}) at survey points in the LUCAS topsoil survey (0-30 cm) (de Brogniez *et al.* 2015)

2.3.3.2. Design-based sampling strategies

In design-based approaches, the sample locations are selected using random sampling, rather than being systematically determined. These samples give an unbiased estimate of the mean and variance, as long as the probability of including each unit in the sample is known (Black *et al.* 2008; de Gruijter *et al.* 2006). In model-based approaches there is a risk of underrepresenting or completely missing particular landuses, soil types, and their combinations, whereas in design-based approaches it is ensured that each unit is appropriately represented (van Wesemael *et al.* 2011). It is generally for this reason that large countries with highly variable soil and landuse types choose design-based approaches (van Wesemael *et al.* 2011). In contrast to model-based approaches, design-based methods are good for producing regional estimates, but not necessarily for producing maps. Sample sites are generally revisited in model-based designs; however, this is less of a requirement in design-based systems where new locations within the same stratum can be revisited. This approach comes with its advantages and disadvantages, which will be discussed later on.

There are a few different methods within design-based sampling schemes, with stratified random sampling being the most popular (Black *et al.* 2008; Karunaratne *et al.* 2014). In stratified

random sampling the area is divided into sub-areas (strata), commonly defined by soil type or land use, or a combination of the two (McKenzie *et al.* 2002; Brus *et al.* 2009). Simple random sampling is applied in each stratum and sample sizes are generally selected based upon the areal extent of the stratum; however, they can be purposely selected depending on the aims and interests of the monitoring study (Karunaratne *et al.* 2014). As samples are randomly drawn from strata, rather than the whole population in stratified random sampling, the variance of the mean is considerably reduced (Black *et al.* 2008). Clustered sampling is another similar approach where sampling units are initially randomly selected within strata, but the actual sampling points are then purposively chosen within these sampling units, reducing travelling time and costs (Black *et al.* 2008).

The other major soil monitoring network in the United Kingdom, the Countryside Survey (CS), took the alternate approach to the NSIEW and used a design-based system with clustered sampling (Black *et al.* 2008; Emmett *et al.* 2007a; Emmett *et al.* 2007b). The CS was performed at a similar time to the NSIEW, with repeated soil surveys in 1978, 1998, and 2007, and also focused only on topsoil (0-15 cm depth) with physical, chemical and biological properties measured (Reynolds *et al.* 2013). The study showed important trends regarding landuse, and demonstrated that soil change is undoubtedly occurring in Great Britain on the human time scale (Reynolds *et al.* 2013). The study found that topsoil SOC levels in Great Britain slightly increased (8%) from 1978-1998, and slightly decreased (6.5%) from 1998-2007, with no significant change between 1978 and 2007 (Reynolds *et al.* 2013).

Although conducted at similar times, the results regarding SOC dynamics were markedly contrasting in the CS and the NSIEW, with no change in SOC found in the CS, and wide-scale decreases observed in the NSIEW. Whilst puzzling, these differences were attributed to the dissimilarities in design of the two studies, relating to statistical design, number of subsamples collected at each site, acceptable spatial precision for site relocation, and the methods of SOC determination (Kirk *et al.* 2013). Nevertheless, these significant differences in results of two large-scale SMNs of similar areas raises concerns and questions about the efficiency, effectiveness and reliability of SMNs in general. How could two supposedly statistically sound studies generate such contrasting results? A study by Black *et al.* (2008) identifies a superior sampling efficiency of the design-based approach used in the CS for detecting C stocks and change, suggesting its relative advantage to the model-based approach used in the NSIEW (Reynolds *et al.* 2013).

2.3.3.3. *Non-statistical monitoring programs*

Some SMNs use neither model, nor design-based sampling schemes, and haphazardly/arbitrarily choose sampling locations. The Netherlands Soil Monitoring Network (LMB) is an example of this and the design of this network could be considered a stratified convenience sample, or a stratified

arbitrary sample, as sampling locations were purposively selected, with no use of a systematic grid or randomisation (Brus *et al.* 2009). Sub-populations (strata) were created based on a combination of landuse and soil type, leaving 10 strata, and the number of sampling sites was not selected according to their weighted proportions and each stratum were allocated 20 sampling locations (Rutgers *et al.* 2008). The LMB commenced its first sampling campaign in 1993-1997, and completed its third in 2006-2010 (Wattel-Koekkoek *et al.* 2013). The original purpose was to assess pollutant concentrations in the soil, however, a larger suite of soil properties are now measured (Wattel-Koekkoek *et al.* 2013). This SMN has been less sensitive to observing changes in soil than initially anticipated, largely due to the poor statistical design (Wattel-Koekkoek *et al.* 2013).

In South Korea, a national SMN for agricultural lands was established in 1999. The sampling design is based on landuse, soil types and agro-climatic zones, comprising of 19 different strata (Minasny *et al.* 2012). Sampling sites were allocated to provinces/towns, and exact sampling points were purposively selected by pedologists, consequently making this study a stratified arbitrary sampling scheme (Minasny *et al.* 2012). A study by Minasny *et al.* (2012) used data from this SMN to assess the effects that both climate change and landuse change has had on SOC. The study found a linear increase in SOC from 1999 to 2007 in most areas (Fig. 5), with a correlation between increased SOC content and increased rainfall. There was no trend identified between temperature and SOC content, despite the increasing air temperatures of about 0.2°C per decade. The study demonstrates that agricultural practices can increase SOC storage in soils despite a warming climate (Minasny *et al.* 2012). It is critical that these conclusions be treated with caution, as the arbitrary sampling design, and lack of estimation variance and hypothesis test may lead to a misrepresentation of temporal shifts in SOC stocks. Similar results have been found in China (Yan *et al.* 2011) and Italy (Fantappiè *et al.* 2011), indicating that the effect of management supercedes that of temperature increase (Minasny *et al.* 2012).

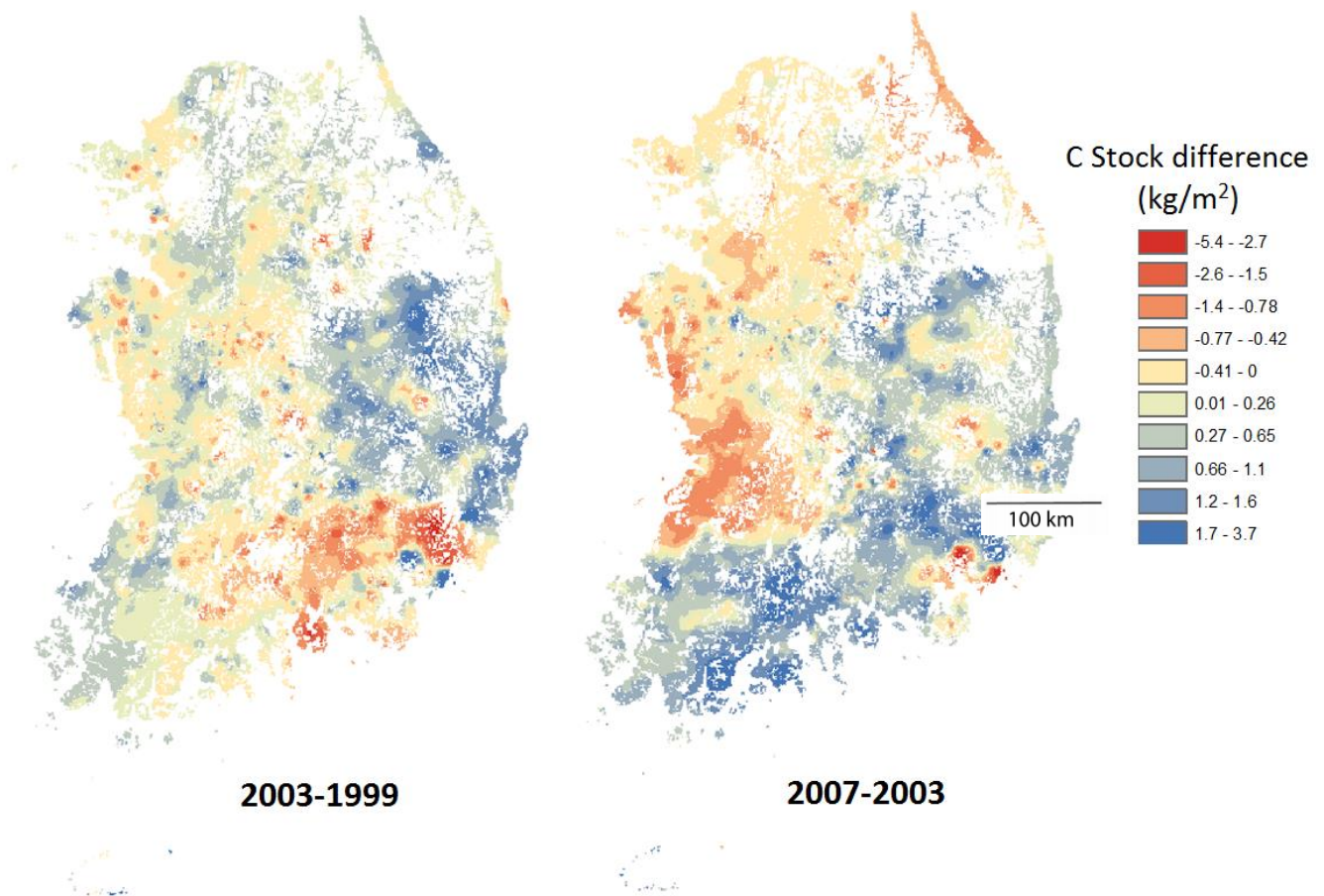


Fig. 2.5 – Maps of changes in topsoil (0-15 cm) C stocks in South Korean rice paddy fields from 1999-2003, and 2003-2007 (Minasny *et al.* 2012)

2.3.4. Legacy data

While the worth of SMNs is well known, many countries and regions do not have established SMNs due to the substantial costs and difficulty in implementing and sustaining them (Marchant *et al.* 2012; Karunaratne *et al.* 2014). Many countries and regions do, however, have legacy data that was previously collected for other purposes, such as soil test databases, characterising soil units for traditional soil type mapping or to conduct baseline soil surveys (Karunaratne *et al.* 2014). This legacy data often lacks proper statistical design as it was not collected for the purpose of soil monitoring, causing many challenges to arise in its use (Karunaratne *et al.* 2014). In Australia, practically all soil maps rely on purposive sampling, making it difficult to estimate their accuracy and precision (McKenzie *et al.* 2002). Legacy data quality is generally highly variable, and it is essential to ensure that the observations are truly representative of the study area (Marchant *et al.* 2012). Nevertheless, legacy data can prove invaluable in detecting and understanding spatial and temporal variations in soil properties at a low cost, with prompt results (Karunaratne *et al.* 2014).

Approaches to using soil legacy data to infer temporal changes in soil properties are diverse, with several methods demonstrated in the literature. Some studies use a baseline soil survey of a particular area and then resample the area, while others use multiple legacy datasets to infer change, eliminating the need to resample. Studies that resample an area can be differentiated by those that resample the same sample location, those that choose a new sampling location, and studies that use a combination of the two, which is largely dependent on statistical quality of legacy data and geo-referencing information.

2.3.4.1. Utilising legacy data without resampling

Some studies have assessed temporal change of soil properties through the use of multiple datasets from soil surveys, without resampling. This approach is more suitable for countries or regions that have poor soil data records and lack SMNs. In China, there is an absence of quantitative history of soil conditions beyond experimental plots, which prompted Lindert *et al.* (1996) to attempt to judge historic soil trends in China through the utilisation of several legacy datasets without taking actual measurements. Chinese soil profile data from three sampling periods in the 1930s, 1950s and 1980s were used to perform basic statistical tests to identify broad trends in soil degradation. The study found trends of declining SOC and N since the 1950s, but no declines in P and K in soil. The trends identified in this study should be treated with caution, as there is likely significant bias involved, particularly in regards to sample site selection. For example, soil attributes in the 1950s appeared to be better, although this could simply be due to better soils being sampled in that period, rather than actual soil trends (Lindert *et al.* 1996). A similar study was performed in China in 2010 showing the decline in pH of Chinese agricultural soils since the 1980s. The study used data from two national soil surveys, as well as data from long-term soil experiments and paired sites to evaluate changes in soil acidity (Guo *et al.* 2010).

A similar approach was used in Java, Indonesia, where changes in SOC were inferred through the use of multiple soil surveys from 1930 to 2010 and consisting of 2002 soil profile measurements (Minasny *et al.* 2011). Results indicated that SOC content dropped from 2.11% in 1930-1940 to 0.75% in 1970-1980, and then steadily increased to 1.18% in 2000-2010 (Fig. 6). The initial declines in SOC levels could be attributed to clearing and conversion of natural vegetation to arable land, while the steady increase from the 1970s and onwards could be attributed to improved yields and management practices. Trends in SOC content seem to be contrary to the increasing temperatures since the 1930s in Java, suggesting that anthropogenic activities are a stronger influence on SOC change than climatic factors. How well this data depicts actual decadal changes in SOC in Java is questionable, as landuse details were sporadically recorded and the majority of data was not geo-referenced and were allocated to the nearest town (Minasny *et al.* 2011). While this study could be

criticised on its limitations and sources of error, the reality is that there is no official monitoring network in Indonesia, and this legacy soil survey data is the best resource to observe the temporal changes in SOC stocks in Java (Minasny *et al.* 2011).

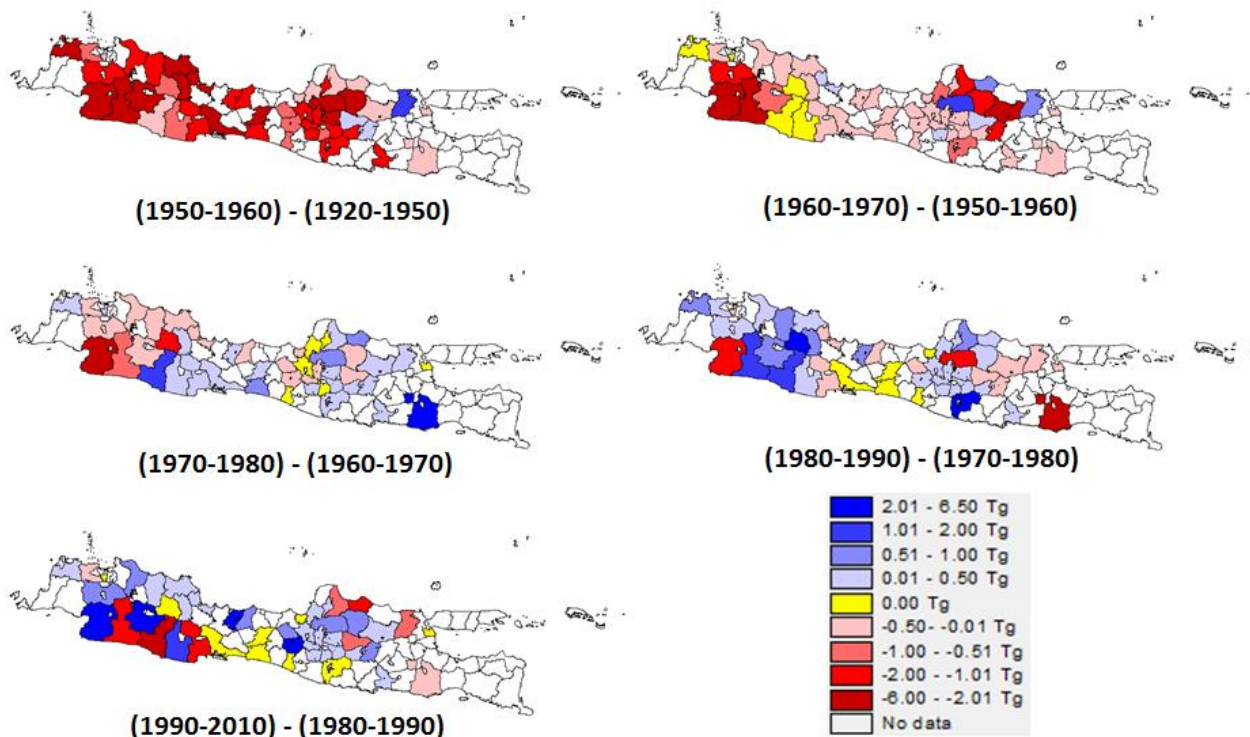


Fig. 2.6 – Changes in C stocks (Tg) at various intervals from 1920-2010 in Java, Indonesia, grouped into kabupaten (regions) (Minasny *et al.* 2011)

When using legacy data from several decades apart, there is almost certainly a difference in techniques of sampling and soil analyses between surveys, making it difficult to compare. While using multiple datasets to identify broad trends in soil attribute change over time can be extremely useful due to its low cost and immediate output, the limitations of this approach must not be ignored. This approach should not be used as the sole basis to make decisions on natural resource management and policy, and more tried and tested methodologies should be used in combination with this approach to ensure more realistic and accurate trends in soil change are identified.

2.3.4.2. Utilising legacy data with resampling

Resampling an area that has existing soil legacy data is another approach to monitoring soil condition. Although often not used on national studies, this approach is particularly useful on regional scales, as there are many regions that have had baseline soil surveys performed (Odeh *et al.* 2012). While this approach is similar to non-statistical SMNs, it differs in the intention of its development. Non-statistical SMNs were designed to return and resample, whereas there was no

intention to return and resample legacy data – it was created for another purpose other than monitoring temporal soil change.

3.4.2.1. New site

As legacy data is not designed for monitoring soil change, sample sites are often haphazardly or purposively selected, and thus the data lack proper statistical design (Marchant *et al.* 2012). For this reason, many studies opt to develop a new sampling scheme, and sample new sites to provide a more accurate and statistically sound representation of soil attributes of an area. New sites are also often selected due to the lack of site location information, as sampling location coordinates are not always recorded in legacy data, particularly those from several decades ago where geo-referencing technology was unavailable. While re-sampling at the same site is advantageous as comparisons can be made at a specific location and co-location of sites assists in detecting soil change, proper sampling design is of foremost importance. If sampling sites for the legacy data have been haphazardly chosen, we cannot be sure that this is a representative depiction of actual soil attributes (Karunaratne *et al.* 2014).

A study by Karunaratne *et al.* (2014) monitored the temporal change of soil carbon at a 10-year interval in northern NSW through the use of soil legacy data and newly collected data. Despite geo-referenced site locations being available, new sampling locations were chosen as the previous sites of the legacy data were not selected according to a specified design. The legacy data was originally taken in the year 2000 to develop soil-landscape maps and determine the spatial variation in soil properties, and data was collected by two different organisations, furthering its unreliability. Re-sampling occurred in a stratified random sampling approach based on landuse and soil type in 2010, and soil was collected at 0-10, 10-30 and 30-50 cm depths. In addition, a second sample was extracted approximately 30 m away and analysed separately at each site, as the short-range spatial variability of SOC is well-known, and it is essential to capture this (McBratney and Pringle 1999). The study found that after a decade between sampling campaigns there was no significant change in SOC content identified in the majority of the study area (Fig. 7). Much of the area had been developed for more than half a century, so it was hypothesised that SOC was reaching equilibrium and the change in SOC was decelerating (Karunaratne *et al.* 2014). The study demonstrated that temporal changes in SOC can be assessed using a combination of legacy data and resampling new locations, while still developing a better sampling scheme for future monitoring. While this approach is justified, the revisiting of some original sites that were deemed representative of the strata would have been invaluable, despite the arbitrary design of the original sites, and would provide an idea of the covariance between the two surveys and improved the precision of estimates of change.

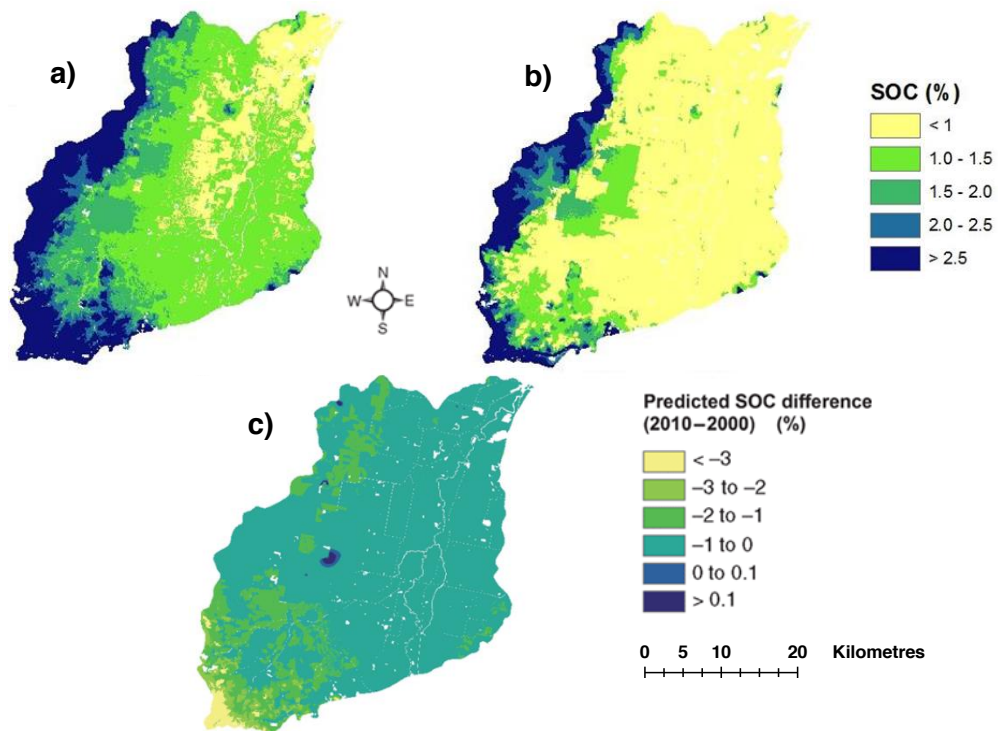


Fig. 2.7 – Map of predicted SOC for 2000 (a), 2010 (b), and predicted temporal SOC change from 2000-2010 (c) for the Cox's Creek Catchment, NSW, Australia (Karunaratne *et al.* 2014) Journal compilation © CSIRO 2014. <http://www.publish.csiro.au/nid/84/paper/SR13081.htm>

In China, a national soil survey that collected 1509 soil samples from cropland throughout the nation from 1979-1982 was used to assess temporal SOC change (Yan *et al.* 2011). Exact sampling locations were not recorded, rendering it impossible to resample the original sites. A follow-up survey in 2007-2008 was conducted, taking a total of 1394 cropland soil samples from all over the country (Yan *et al.* 2011). The study showed that the average SOC content in the top 20 cm increased by 0.22% per year between the two surveys, with this increase attributed to the large increases in crop yields since the 1980s. This study was not a paired resampling, and as with almost all studies that use legacy data, there were some issues that arose. The legacy data was obtained from pedogenetic horizons, rather than fixed depths, thus requiring SOC contents to be calculated at certain depths. The legacy data was also collected from single point samples, while the latter survey used composite samples, making it difficult to compare spatial soil variation over short distances (Yan *et al.* 2011).

3.4.2.2. Same site

Relatively rare, due to the typically poor statistical design of legacy datasets is the revisiting of the same locations to monitor soil change. A study that monitored salinity trends in the irrigated district of Flumen, Aragón, Spain over 24-years revisited sampled sites, but there were several issues in

regards to site identification, as some site information was lost between surveys. The legacy data had no specified statistical design, which raises questions about the accuracy of the representation of the soil attributes in the district. Soil samples were collected to a mean depth of ≈ 1 m, and sampling occurred in 1975, 1985/86 and 1999. Over the entire study period, it was found that soil salinity levels were decreasing in this irrigated district, contrary to expectations (Herrero and Pérez-Coveta 2005).

Interestingly, a study in Florida, USA, used a combination of selecting new sites according to a specified design, and revisiting historical legacy data locations to detect changes in SOC in the top 20 cm of the profile (Ross *et al.* 2013). The original legacy dataset consisted of 402 samples collected between 1965 and 1996, where sampling locations were selected purposively. In 2008-09 resampling occurred, with samples taken according to a stratified-random scheme based on soil type and landuse, with a 50% overlap from legacy data sites that were deemed to accurately represent the strata, giving a total of 302 sites. Modelling results from the study indicated that average SOC content had more than doubled over the experimental period, and that the area has acted as a net sink for atmospheric carbon in the previous 40 years. The revisiting and resampling of legacy data sites proved valuable, and allowed a co-located comparison, with 62% of co-located sites sequestering SOC over the 40 year period, showing an average increase of $0.8 \text{ g C m}^{-2} \text{ yr}^{-1}$ at each site. There was some indication that the inclusion of legacy data sites had included a degree of bias into the study, as there was a contrastingly large discrepancy in wetland soils, with a halving of SOC content in this soil type (Ross *et al.* 2013). Results from this study suggest that resampling legacy data sites that were selected according to 'expert knowledge' presents some unavoidable bias, but the advantages of co-located sites are invaluable in detecting soil change over several decades (Ross *et al.* 2013).

The time and resource consuming nature of our traditional soil survey methods makes it impossible to cover the whole earth to gain all the soil data that we need. It is well known that the world is abundant with soil data as a result of past soil surveys, but the majority of this data is not utilised and remains largely unused (Odeh *et al.* 2012). It is imperative that we utilise the data that we do have, and use new methods and technologies to extract as much information on the soil resource as possible (Hartemink *et al.* 2008).

2.3.4.3. Utilising archived soil samples

A study by Deng *et al.* (2013) took a different approach in monitoring soil change, and used archived soil samples. Archived samples from 1986 to 2009 were taken from the Danish national 7-km soil monitoring grid, and were analysed by both Vis-NIR spectroscopy and traditional laboratory (TL)

methods to monitor temporal changes in SOC. Both analytical techniques showed that topsoil (0-25 cm) and subsoil (25-50 cm) SOC levels had decreased slightly between 1986 and 2009. The study suggests that Vis-NIR is a promising new approach for monitoring spatial and temporal shifts in SOC at the national scale (Deng *et al.* 2013). The study also demonstrates that using archived soil samples to monitor soil change can be extremely valuable, particularly if archived samples are obtained from a structured SMN. There are some limitations with this technique, as only a certain suite of soil attributes can be measured. These largely include soil chemical properties, as many physical properties of soils are altered upon collection and biological activity is adversely affected over time. As aforementioned, the extended storage of archived soil samples has been found to have a significant effect on some soil chemical properties, for example, soil pH (Slattery and Burnett 1992; Miller *et al.* 2001).

2.3.4.4. Laboratory test database

Utilising existing soil data from routine soil testing by farmers, agronomists and fertiliser companies presents a promising opportunity for identifying temporal shifts in soil properties in agricultural areas (Chauveau *et al.* 2014). While many areas of the world have some form of legacy data derived from soil surveys, many districts, regions and nations also have a wealth of soil data analysed by laboratories for farmers and fertiliser companies that are stored in databases, typically consisting of thousands to millions of results. Using this data requires minimal operational costs and eliminates the need to resample, but the results have usually been open to several sources of bias, which are unescapable in any study that lacks a structured sampling strategy (Saby *et al.* 2008; Chauveau *et al.* 2014).

A French study utilised the wealth of existing soil data that was available from farmers and fertiliser companies and showed how these soil test results can be a valuable alternative approach in detecting changes in SOC, despite this data not being intended for soil monitoring purposes (Saby *et al.* 2008a). Results from tests of agricultural topsoils are stored in the national 'soil-test' database in France, where all data have been collected and analysed using standardised protocols by commercial laboratories (Saby *et al.* 2008a). This study used this data from the Franche-Comté region (16,202 km²), in eastern France, with results showing large losses in SOC from 1990-2004 (Fig. 8), with these losses being attributed to both changes in landuse and increasing temperatures during the survey period (Saby *et al.* 2008a).

Soil organic carbon content (g kg^{-1})

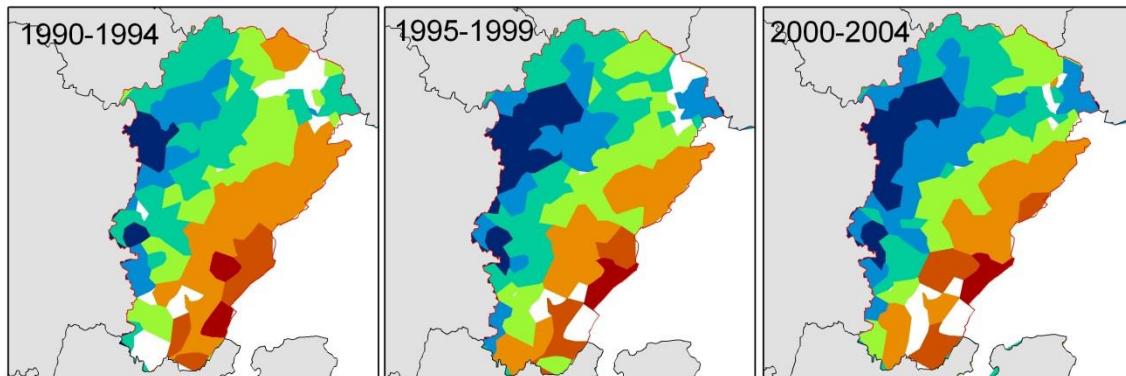


Fig. 2.8 – Spatial distribution of cantonal medians of SOC in topsoil (g/kg) for the three periods 1990–1994, 1995–1999 and 2000–2004 in the Franche-Comté region, France (white polygons lacked sufficient quantities of data) (Saby *et al.* 2008a)

A study in the Netherlands took a similar approach to analysing soil change, where changes in SOC contents of topsoils from grasslands and arable land from 1984–2004 were determined using a database with about 2 million SOC results from farmers (Reijneveld *et al.* 2009). The study showed that SOC contents of soils under both grassland and arable land are increasing (Fig. 9), with significant differences between regions. The results described here contrasts with those of other studies (e.g. Vleeshouwers and Verhagen 2002; Sleutel *et al.* 2003; Bellamy *et al.* 2005; Saby *et al.* 2008a), where SOC levels have been found to be declining in agricultural land in Europe (Reijneveld *et al.* 2009).

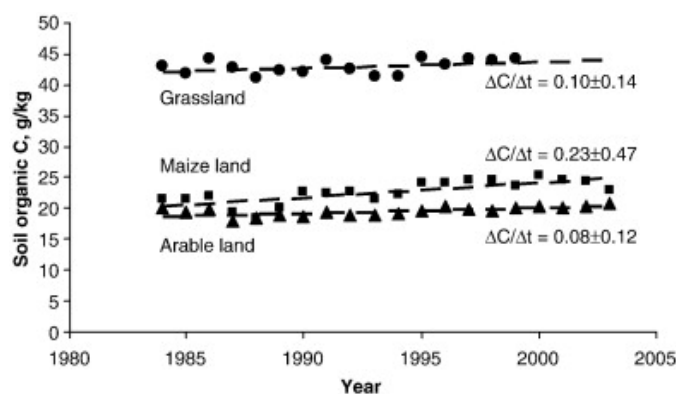


Fig. 2.9 – Changes in mean SOC contents of grassland (1984–2000), maize land (1984–2004) and arable land (1984–2004) in the Netherlands. The mean annual change in SOC is indicated as $\Delta C / \Delta t$, in g/kg/yr (Reijneveld *et al.* 2009)

Regional soil databases have also been used to detect temporal soil change. In Victoria, Australia, a database of 75,000 soil analyses requested by farmers between 1973 and 1993 were

used to observe temporal and spatial variations in topsoil pH, and topsoil available potassium (K) content (Marchant *et al.* 2012; 2015). Exact location of soil observations was not recorded and sample sites were identified by place names, or nearest towns. An overall acidification trend was observed throughout much of Victoria, with only small pockets of land predicted to have experienced an increase in soil pH (Fig. 10). The study also indicated widespread increases in available soil K across the state during the study period (Fig. 11), with a correlation between original concentration and the extent of K increase identified. Although the fitted models were considered noisy, spatial and temporal trends were still clear in the predicted maps, and the study demonstrated that it is possible to fit space-time variograms to soil properties gained from soil test databases (Marchant *et al.* 2012; 2015).

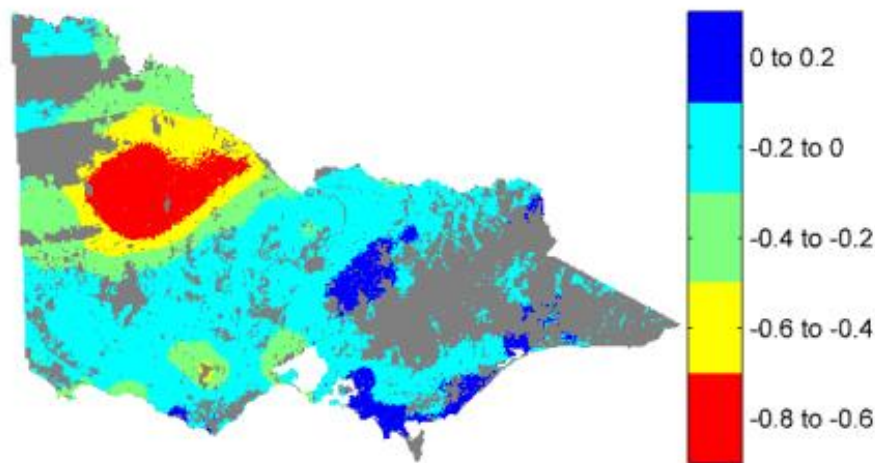


Fig. 2.10 – Change in topsoil pH of pasture and cropping soils across Victoria, Australia from 1973-1994 (Marchant *et al.* 2015)

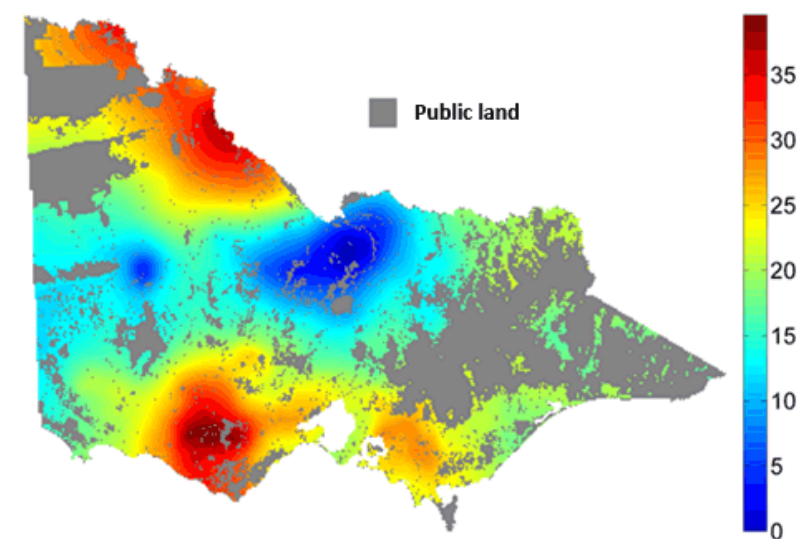


Fig. 2.11 – Change in topsoil available potassium in parts per million of pasture and cropping soils across Victoria, Australia from 1973-1994 (DPI 2012)

It is clear that the use of soil databases for monitoring purposes can be extremely valuable, as SMNs are costly to implement and can only predict rates of change after soil properties have been measured more than once. However, there are a number of challenges that often arise when using these soil databases, as they are not specifically designed for soil monitoring (Marchant *et al.* 2012). Farmers request soil samples to be analysed in laboratories to aid them in management decisions, with a specific incentive behind these analysis requests, whether that be concerns of crop nutrition, or potential soil degradation threats (e.g. acidification, salinity). There is little incentive for farmers to request the analysis of soil from native vegetation sites, and thus, data solely from agricultural land is available in these databases. Motivations to request soil tests vary according to changes in economics, climate, technology and regulatory factors, and cause obvious bias towards certain areas, and time periods, leading to deceptive results (Marchant *et al.* 2012). Laboratory soil databases also solely record the information that is relevant for the particular study, often leaving out crucial information required for soil monitoring. Exact sampling location is often not included in databases, and sampling locations are designated to the nearest town in many studies (e.g. Saby *et al.* 2008b; Reijneveld *et al.* 2009; Marchant *et al.* 2012). Information regarding land use is also sporadic, making it difficult to predict effects of land management on soil change (Marchant *et al.* 2012).

There are a number of regions and industries that have soil databases around the world. For example, the Australian cotton industry has funded a number of soil inventory projects, resulting in a large amount of soil data, unmatched by any other agricultural industry in Australia (Odeh *et al.* 2004). These databases need to be utilised, and despite the issues with using this data, it allows a comprehensive assessment of soil change with considerably low operational costs.

2.3.5. Summary of methods to monitor/measure soil change

Table 2.3 – Summary of methods to monitor/measure soil change - (Expense = low (*) – high (*****)

Method	Positives	Negatives	Typical scale	Expense
<i>Space-for-time-substitution/ paired-sites</i>	- Inexpensive - Immediate change information	- Many assumptions - Actual soil change not measured - inferred	Field/farm	**
<i>Long-term soil experiments</i>	- Spans over decades/centuries - Dense soil change data - Data can be used to develop models	- Expensive & vulnerable to loss - Soil analytical techniques change over time - Small area covered/data extrapolation issues	Field/farm	****
<i>Soil monitoring networks</i>	- Statistically sound - Built for purpose	- Expensive - Not feasible for large countries	National to continental	-
<i>Model-based SMNs</i>	- Better for local estimates - Good for DSM purposes/spatially spread out	- Possibility of missing whole strata	-	****
<i>Design-based SMNs</i>	- Better for global estimates - Strata ensure good representation of study area	- Not as good for DSM purposes	-	****
<i>Legacy data</i>	- Typically inexpensive - Time saving - Often only option to detect changes in soil	- Lacks proper statistical design - Essential information often missing - Unrepresented areas - Poor/unknown geographical locations	Field to regional to national to continental	-
<i>Without resampling</i>	- Inexpensive - No further sampling required	- Many assumptions and bias - Often no co-location of samples	-	*
<i>Resampling: new site</i>	- Statistically sound selection of new sites possible	- No co-location of samples	-	***
<i>Resampling: same site</i>	- Co-location of sites are useful in detecting change	- Original sites often arbitrarily selected	-	***
<i>Archived soil samples</i>	- New analytical/laboratory techniques can be used - Possibly statistically sound if taken from SMN	- Storage affects some soil attributes - May not be statistically sound	-	**
<i>Laboratory test database</i>	- Inexpensive - No further sampling required - Consists of thousands to millions of results - Typically covers large area	- Uncontrolled sampling strategy - Results affected current concerns/interest - Only topsoils from agricultural land - Often solely few soil fertility indicators	-	*

2.4. Challenges in monitoring soil change

The specific issues regarding different soil monitoring approaches were discussed in the previous section; however, there are several issues that consistently arise, regardless of the chosen approach.

2.4.1. Detecting a change in soil properties

The statistical approaches used to detect change in a soil property can be separated based on the statistics being estimated and the type of inference method being used (Karunaratne *et al.* 2014). These can largely be split into design- and model-based systems. Estimations of changes in a soil property can be categorised as global estimates (e.g. mean change of a soil property for a soil type/landuse), and local estimates (e.g. the way a soil property changes on a grid within soil type/landuse) (de Gruijter *et al.* 2006). When fitting separate models for two time periods in model-based studies, calculations of changes in a soil property are simply made by determining the differences at the pixel/block level at the different time periods. For design-based approaches, temporal changes in a soil property are calculated by determining the differences in means from the two time periods for each unit, for example a landuse or soil type. Design-based estimates are generally simple to compute, and are not influenced by subjective assumptions about a statistical model (Papritz & Webster 1995b). It is still possible for global estimates to be calculated in model-based systems. If we want to protect our soil resources and detect early signs of undesirable change, it is imperative that monitoring is sensitive in detecting change (Papritz & Webster 1995a).

Models can always be developed to predict a change in soil properties, whether this is from comparing data from two time periods, or comparing independently created digital soil maps. It is important that a confidence interval is placed around this predicted change, as it subsequently allows a hypothesis test to be performed. Despite the importance of this, there are very few studies where this is actually implemented. The general form for a hypothesis test is:

$$\begin{aligned} H_0: \mu_d &= 0; \\ H_1: \mu_d &\neq 0; \end{aligned} \tag{2.1}$$

where μ_d is some estimate of the change (equation 2), such as a mean or a prediction at a specific location.

$$\mu_d = \hat{y}_1 - \hat{y}_2, \tag{2.2}$$

where \hat{y}_1 is the prediction of the soil property for the observations at the first point in time, and \hat{y}_2 at the second point in time. To test this hypothesis, we need an estimate of the variance of the change, which can be expressed as:

$$\hat{V}(\Delta\hat{y}) = \hat{V}(\hat{y})_1 + \hat{V}(\hat{y})_2 - 2\hat{V}(\hat{y})_{1,2}, \quad (2.3)$$

where $\hat{V}(\Delta\hat{y})$ is the variance of the change in the soil property between two points in time, $\hat{V}(\hat{y})_1$ is the variance of the soil property for the observations at the first point in time, $\hat{V}(\hat{y})_2$ at the second point in time, and $\hat{V}(\hat{y})_{1,2}$ is the covariance between the measurements of soil properties at the two time periods (the covariance term). This equation can be used to calculate a test statistic, which is expressed as:

$$z = \frac{\hat{y}_1 - \hat{y}_2}{\sqrt{\hat{V}(\Delta\hat{y})}}, \quad (2.4)$$

where z is the z-statistic score, \hat{y}_1 is the value of the prediction of the soil property at time point one, \hat{y}_2 at the second time point, and $\hat{V}(\Delta\hat{y})$ is the contrast variance, which is obtained using equation 2.3. This equation presents both pitfalls and opportunities in monitoring. The pitfalls are the spatial variation of the soil property and the magnitude of the change, while the opportunity is the covariance between sampling campaigns.

2.4.1.1. Soil spatial variation

Equation (2.3) shows that a high estimation variance for the baseline or resampled survey can cause a large contrast variance, which can make it difficult to detect a statistically significant change. Soil can vary dramatically over very short distances; as summarised by Beckett and Webster (1971) in their soil variability review; “*up to half the variance within a field may already be present within any m² in it*” (McKenzie *et al.* 2002). This short-range spatial variability can be easily confused with temporal change, and there needs to be a sound statistical design and sufficient replication to distinguish relatively small temporal change from the larger spatial variation present (McKenzie *et al.* 2002; Grealish *et al.* 2011). This large natural variation also makes it difficult to distinguish whether observed soil change is real (Mol *et al.* 1998).

It does not matter whether we consider change for a soil type or region or across a region with a digital soil map. In digital soil mapping studies which actually present the prediction variance, they are often quite large. For example, in a study of SOC temporal dynamics by Karunaratne *et al.* (2014), when assessing the model quality, the predictions were found to be quite poor, but a good assessment of their

reliability in the form of prediction variance was shown. Knowing the prediction variance is essential in soil change studies as it is used in testing the statistical significance of the estimated change in a soil property. If baseline surveys have a high prediction variance, problems would arise in detecting significant change in all areas apart from those that have experienced the largest changes.

Soil spatial variability is influenced by many factors, such as climate and management. For example, practices such as cultivation are known to reduce the spatial heterogeneity of many soil properties, through mixing and homogenising the soil (Cattle *et al.* 1994). Furthermore, all soil properties are inherently spatially variable, but the extent of spatial variability differs greatly amongst individual soil attributes. This presents a verification challenge for soil change and monitoring studies, as sample size and network design may be appropriate to represent the spatial variation for some soil properties, but may be unsuitable for others, and this should be acknowledged in studies where there is more than one soil property of interest. As an example, soil texture is one attribute of soil that is likely to be less spatially variable than many others, such as SOC. It is well known that SOC is highly spatially variable and easily influenced by management, and thus requires high sampling density and longer time intervals to detect changes adequately (Saby *et al.* 2008b).

2.4.1.2. Magnitude of change

Both model- and design-based methods can estimate the change in a soil property, and its error, which is crucial in determining the statistical significance of an observed change. Gradual, small changes in soil properties are difficult to detect, particularly over small time periods. Given the link between sample size, the variance of prediction and the actual change, we can determine important information, such as smallest detectable difference (minimum detection limit), provided we have a sample size and variance of prediction. This is especially easy for design-based methods. A study by Bishop *et al.* (2013) calculated how many samples were required to estimate SOC concentration with a known precision, and to estimate a change in SOC. The results showed that inventory required fewer samples than was required for change detection and concluded that smaller changes in SOC are prohibitively expensive to detect. However, through the use of more efficient sample designs and cheaper measurement techniques, a reduction in sample size requirements and hence cost is possible.

Ensuring that there is adequate time intervals (e.g. 10 years) between sampling campaigns is essential, because change is challenging to detect over short time periods, even when dealing with severe cases of soil degradation (McKenzie and Dixon 2007; Grealish *et al.* 2011). Detecting differences when the rate of change is low is difficult, but this is further amplified when background levels of

variation in the soil property of interest is large. For example, Conant *et al.* (2001) showed that grasslands were sequestering SOC at a rate of 0.46 Mg C ha/year in the topsoil, but this was against large existing stocks of 30–80 Mg C ha⁻¹ (Conant *et al.* 2001; Conant and Paustian 2002).

2.4.1.3. Covariance between surveys and returning to sampling sites

One opportunity from equation (3) is to take advantage of the covariance term between surveys. When the covariance term is included there is a stronger relationship between the two surveys, which results in a smaller contrast variance, and subsequently greater sensitivity in predicting a statistically significant change. Whether a design-based or model-based approach is used, there is a requirement that some or all locations are resampled to estimate the covariance term. If this is not done, the covariance term is considered to be 0 and we have a conservative estimate of the contrast variance. A strong relationship between soil properties through time would be expected in most cases, and it would be advantageous to utilise this approach. A study that used model-based inference methods to map temporal changes in SOC by Liu *et al.* (2011) disregarded the covariance between the two surveys. If the temporal correlation between observations was not ignored, it could be assumed that the study would have produced better estimates of the change in SOC over time, a better estimation of the variance of change, and better sensitivity in detecting a statistically significant change (Karunaratne *et al.* 2014).

In studies that are interested in mapping a change in a soil property, a linear model of coregionalisation (LMCR) (Papritz and Webster 1994), or a bivariate linear mixed model (BLMM) (Marchant and Lark 2007) should be considered. In these approaches one model for both time periods/surveys is developed, and this accounts for the covariance between surveys. This is the ideal approach if we have co-located samples as we can estimate the cross-covariance, otherwise a pseudo-cross-semivariogram (Papritz and Flüher 1994) can be used in place of cross-semivariograms, but we assume the nugget cross-semivariance is 0, thus inflating the contrast variance. Overall, there are very few digital soil mapping studies that consider the contrast variance and hypothesis testing. Few still consider using BLMMs or LMCR to take advantage of the covariance term in equation (3).

It was previously believed that a SMN should consist of a fixed set of sampling locations, but there are several other reasons to steer away from this approach. Continuously sampling the same sites is an effective method to detect temporal variation and is the most statistically efficient design as it makes use of the covariance, reducing the estimate of the variance of change (Papritz and Webster 1995; Karunaratne *et al.* 2014). Despite this, no further spatial information about the study area is gained when returning to the same locations. Replacing a proportion of old sites with new locations at each sampling

time can be advantageous, as this increases information on soil spatial variation (de Gruijter *et al.* 2006; Mol *et al.* 1998). Land is also more likely to be managed differently if known to be included in a SMN, and disturbances at revisited sites may also cause some bias (Mol *et al.* 1998).

2.4.2. Operational issues and sample support

Soil change studies are wide-ranging in their choice of sample support, with many of these choices being strongly influenced by the aims of the study. Sampling by depth or horizon is one such option, with a majority of earlier studies choosing to sample by pedogenic horizons, as this prevented horizon mixing. Recently, there has been a move towards sampling via depth increments (e.g. Yan *et al.* 2011; Karunaratne *et al.* 2014) as sampling by horizons requires some amount of bias and lacks consistency. The development of equal-area quadratic splines means that predictions at certain depth increments or at horizon depths (if this data is available) can still be made (Bishop *et al.* 1999), allowing the comparison of different datasets. There has also recently been interest in the use of equivalent soil mass (ESM) sampling over sampling via depth or horizon. If bulk density has changed between monitoring time periods, or varies between treatments, there are issues regarding sampling by depth or horizon (Ellert and Bettany 1995). Soil mass increases as bulk density increases, and this can lead to an overestimation of some soil properties in higher bulk density soil, prohibiting direct and accurate comparisons being made (Wendt and Hauser 2013). While this is not an issue for all soil properties, it is particularly important for soil organic matter monitoring.

It is quite clear that the majority of SMNs and soil change studies solely focus on topsoil, often at depths of 20 cm or less, completely disregarding any change in subsoil properties. Part of the reason for this is due to the additional costs involved with sampling at these greater depths, as well as the perception that deeper soil horizons are much more stable and is not subject to change over short periods (Meersmans *et al.* 2009; Harrison *et al.* 2011). Some medium and long-term studies do not support this assumption, with significant changes over relatively short timescales observed at depths deeper than the topsoil (e.g. Adhikari *et al.* 2014).

Composite sampling deals with some of this large soil spatial variation and can be useful, but it is not always appropriate. Information is lost about a particular point when compositing samples, and there is a loss of sensitivity because of the dilution of samples. Composite samples provide an excellent estimate of the mean, but it does not give any information about the variation within the sampling area. One alternative is to include a short-range variation point, where a second sample is extracted at a short distance (e.g. 30-50 m) away from the original point and analysed separately, not composited.

Karunaratne *et al.* (2014) used this method on all of their sites; however, it can also be used on a proportion of sites. This approach gives a better idea of how soil spatially varies over short distances, and is useful in the fitting of spatially correlated models, without losing information about a particular point (Karunaratne *et al.* 2014).

2.4.3. Disentangling effects of climate & landuse

While we can predict/measure the change in a soil property over a period of time, it is difficult to determine whether this is an effect of management or climate. This issue is quite evident in the literature, where many studies fail to make a clear distinction between the driving cause of observed changes in soil properties (e.g. Bellamy *et al.* 2005, Minasny *et al.* 2012). Future work needs to focus on better understanding, and unravelling the effects of both, and on developing a model to attempt to explain the role and weighting that each driving factor plays in changing soil condition over time. The inclusion of undisturbed or naturally vegetated sites that can be deemed to be in a steady state of equilibrium can be treated somewhat as a control, and it can be assumed that any temporal changes observed in soil properties under such a landuse are a result of climate and not land management.

2.5. Modelling soil change into the future

An understanding of soil spatial variation (i.e. maps) and the capacity to identify and understand temporal soil change (i.e. monitoring studies) are two major components of soil knowledge necessary for sustainable land use and management. The final essential component of this is the ability to project the expected future condition of soils under changing landuses and future climates (i.e. process-based and simulation models) (Arrouays *et al.* 2014).

The inclusion of temporal soil modelling in soil monitoring studies is not inevitable (McKenzie *et al.* 2002), and this is evident in much of the reviewed literature. The use of models to simulate temporal soil properties is, however, a useful alternative or addition to the time and resource-consuming task of collecting soil data from soil monitoring exercises (Karunaratne *et al.* 2014). As we strive to move away from expensive and laborious soil surveys, it is clear that the future of soil resource assessment lies in modelling and predicting temporal soil change. The anticipation that there will be dramatic changes in both climate and landuse in the coming decades further emphasises the crucial need for temporal soil modelling exercises.

As previously mentioned, long-term soil experiments (LTSE) have greatly contributed to the development of temporal soil change models, such as the Rothamsted carbon model (RothC) developed from the Rothamsted experiments. While models have been developed for many different soil

properties and processes, the research focus has been primarily fixed on a select few. Unsurprisingly, developing and implementing SOC models has been of principal importance, evidenced by the hundreds of models developed within the last century (Manzoni and Porporato 2009). Erosion has also received significant attention, with several models developed and implemented over previous decades.

Broadly, there are two types of models available when predicting future temporal soil change; process-based models (often termed mechanistic or dynamic), and empirical models. Process-based models are typically more comprehensive and simulate detailed ecosystem processes that describe system behaviour, while empirical models are simpler, and use correlative relationships with a less detailed description of system processes (Adams *et al.* 2013). Inherently, there is some empirical information in all process-based models, as well as a link to process in all empirical models, with many models around the world using a hybrid approach (Adams *et al.* 2013). Process-based models are undoubtedly the most popular, and they work by simulating the processes in the soil and evaluating the probable effects of changes in management and climate on soil properties (Viaud *et al.* 2010). The preference for these types of models essentially derives from their ease of use, simplicity, adaptability, lengthy time steps, suitability in application to larger areas, and the ability to be used in combination with GIS software (Batlle-Aguilar *et al.* 2011; Stockmann *et al.* 2013). Most dynamic-mechanistic simulation models do not have a spatial component, but soil attribute maps can be combined with these models, where the model is run on each individual pixel of the map based on the likely scenario (e.g. increased temperature and/or landuse change) (Minasny *et al.* 2013). The scales at which these models are used are diverse, ranging from single fields, to entire continents.

A farm-scale study in north-western France combined two process-based models to evaluate soil redistribution (erosion processes) and soil carbon dynamics under landuse and climate change between 2010 and 2100 (Lacoste *et al.* 2014). The model, inclusive of soil redistribution, predicted a 25% reduction of SOC levels in the top 1.05 m of the study area between 2010 and 2100, while the model exclusive of soil redistribution predicted a 24% reduction. Similarly, Minasny *et al.* (2013) used a SOC map from 2010 of an area in the Hunter Valley, NSW, Australia, to estimate changes in SOC levels for the year 2020 as a result of all vineyards in the area being converted to unimproved pasture. The study predicted an increase in SOC levels for the area under this landuse change (Fig. 12). These models greatly assist in making better predictions of soil change in a landscape or region, but are typically under-utilised, with few studies simultaneously combining DSM and dynamic-mechanistic modelling (Walter *et al.* 2006; Minasny *et al.* 2013).

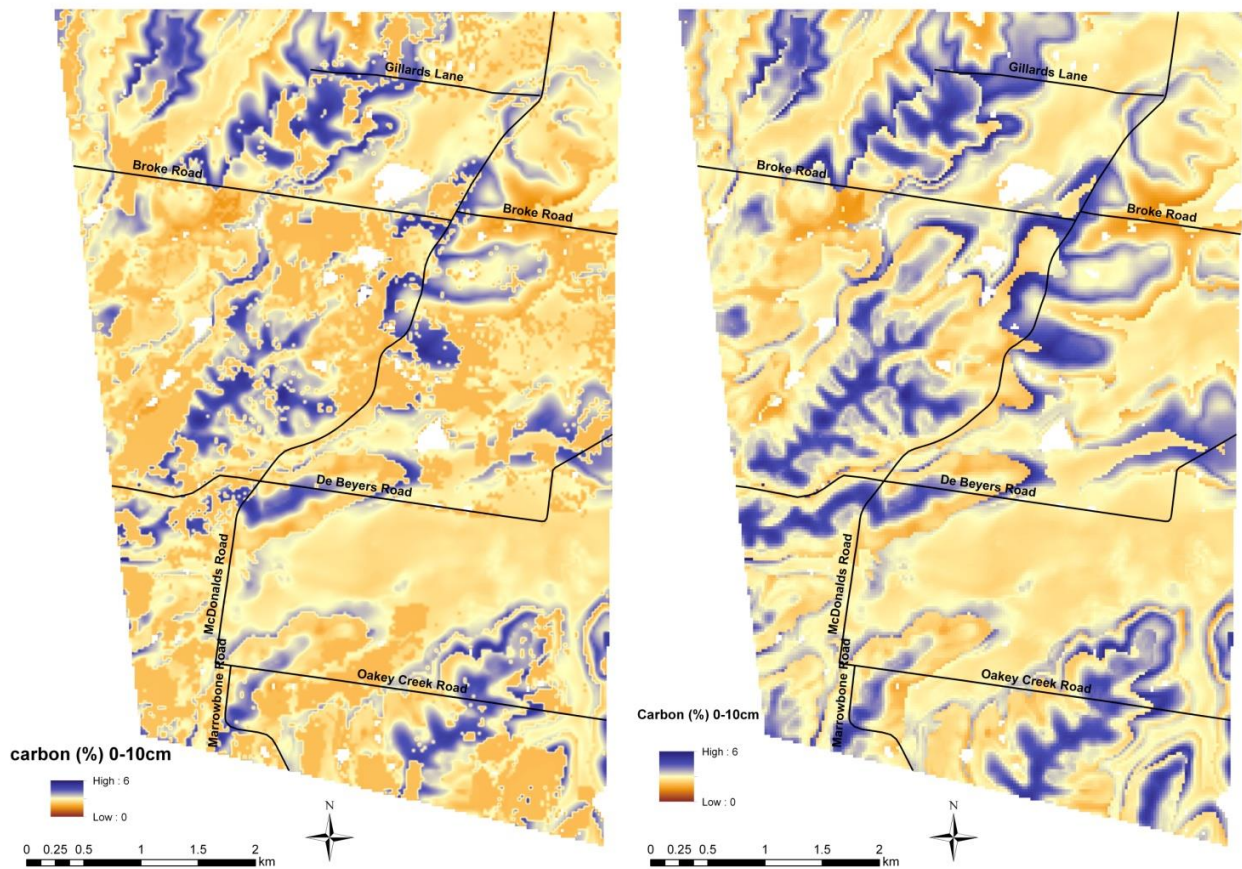


Fig. 2.12 – Predicted SOC stocks for an area in the Hunter Valley, NSW, Australia, in the year 2010 (a), and 2020 (b) (Minasny *et al.* 2013)

Projecting future change over small areas can be very useful, but this site-specific data provides little value in making inferences about larger areas (Milne *et al.* 2007). A comprehensive study that spanned across Europe assessed future changes in SOC stocks between 1990-2080 in cropland and grassland, using a process-based SOC model (RothC) with databases of soil, climate change, landuse change, and technology change (Smith *et al.* 2005). The combination of using several possible future climatic conditions and the inclusion of valuable landuse and technology information transforms the outputs of this study into extremely useful information. Maps produced indicated that under climate change-only scenarios there would be a substantial loss of SOC in most areas between 1990-2080 in both cropland and grassland (Fig. 13). On the contrary, the scenario that included climate change, technology change, and net primary production (NPP) changes showed considerable increases in SOC across most of Europe in both cropland and grassland (Fig. 13). This study indicates that climate change will be an important driver of SOC stock change, however, technology, landuse, and management

changes will also have a significant role to play, and provide an opportunity to counteract the negative impacts of increasing temperatures (Smith *et al.* 2005).

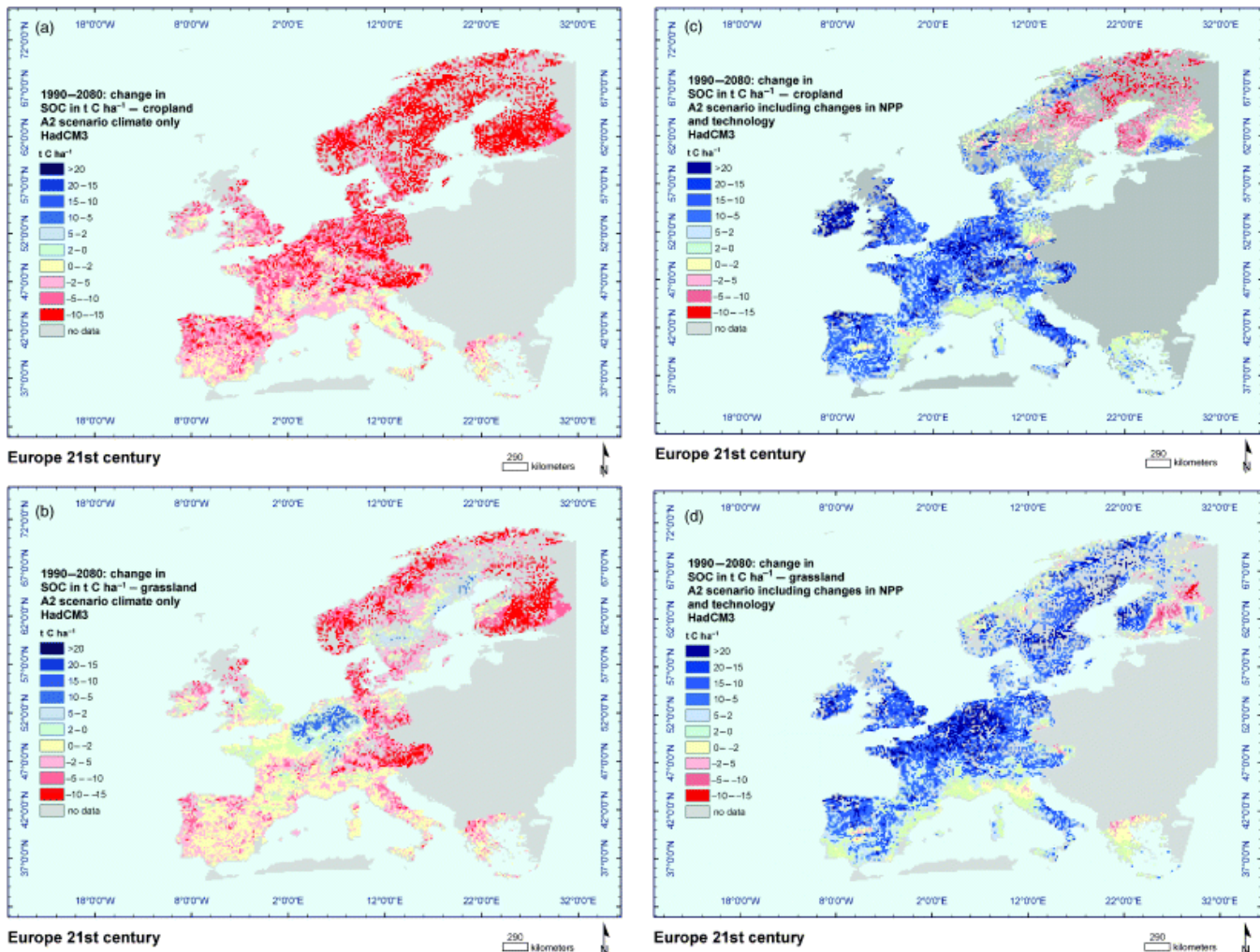


Fig. 2.13 – Predicted change in SOC in $t\ C\ ha^{-1}$ in cropland (a) and grassland (b) under climate-only changes across Europe from 1990-2080, and predicted change in SOC in $t\ C\ ha^{-1}$ in cropland (c) and grassland (d) under climate change, technology change, and net primary production (NPP) changes across Europe from 1990-2080 (Smith *et al.* 2005)

The second approach that can be used to predict likely changes in soil properties over time is the static-empirical model, using the Scorpan framework (McBratney *et al.* 2003). The Scorpan framework describes a relationship between soil and environmental factors, and includes; previous soil information (*S*), climate (*c*), organisms (*o*), topography (*r*), parent material (*p*), age (*a*), and spatial position (*n*). If we know any of the changes of the Scorpan factors over time (e.g. increased temperature or change in landuse), we can project the existing soil map forward. Minasny *et al.* (2013) demonstrated

this for a region in NSW, Australia, where future increases in temperature, and reductions in rainfall were used to estimate the future SOC levels for the year 2030, based on a SOC map from 1990. While this empirical approach is quick and practical, it does come with some limitations when compared with a dynamic simulation model, such as lack of feedback and potential extrapolation difficulties (Minasny *et al.* 2006). Nevertheless, the Scorpan approach has been locally calibrated, and any changes in the factors should reflect the actual changes in the soil (Minasny *et al.* 2013).

Historically, the empirical approach was more widely adopted in the development of predictive soil erosion models, with the Revised Universal Soil Loss Equation (RUSLE) and the original Universal Soil Loss Equation (USLE) undoubtedly being the most popular soil erosion models (Yu 2004). More recently, the popularity of process-based soil erosion models has increased (e.g. WEPP, LISEM, EUROSEM, GUEST). Common to most of these process-based models is the considerable increase in their functionality in terms of dealing with complex topography and landuse, sequences of management practices, realistic storm events and continuous weather sequences. More importantly, the physically-based models are capable of simulating spatial and temporal patterns of erosion as well as deposition. The ability to evaluate the effects of climate and landuse change, as well as the changes in management practices, is the main attraction of this new generation of erosion prediction technology.

2.6. Conclusions and future research needs

This review reinforces the cruciality of soil monitoring for understanding temporal soil change on the human timescale, and discusses the many approaches that are being used to do this. Approaches such as LTSEs and SFTS/paired-sites that are typically performed over small areas have their worth, but we ultimately need soil resource assessment over larger areas, i.e. regions, nations, continents. It is widely accepted that purpose-built SMNs are the most effective and accurate way of monitoring temporal soil change over large areas, with a select number of countries developing and implementing them. It is essential that SMNs have a specified design, and most countries have opted for the statistically-sound, model-based approach, but there are some that lack a structured, statistically sound design, making these networks less accurate and effective as hoped. This review argues that SMNs refers to statistically-sound, purpose-built networks, not those derived from legacy data or that have been adapted for soil monitoring that are often incorrectly labelled SMNs.

Despite the many benefits of SMNs, the massive effort and financial costs involved have precluded adoption in many parts of the world. There has been a recent trend in making use of soil data that already exists, and utilising this globally abundant soil legacy data can help overcome some of the

challenges faced in SMNs, and in turn detect temporal soil change. Using legacy data requires very little operational cost, is extremely time-efficient, and is often the only information to understand the historical state of soils in many areas. The potential to produce useful soil change information from legacy data is great; however, this is largely impacted by the varying usefulness, type and quality of legacy data. Studies have demonstrated that soil laboratory databases, baseline surveys and archived soil samples all provide promising opportunities to monitor soil change. It is essential we take advantage of the wealth of soil legacy data available to monitor soil, but it is imperative that our understanding of how to handle all types of legacy data is increased to heighten its value and so we can accurately detect temporal changes in soil condition.

Many soil monitoring studies use copious amounts of financial and labour resources, but the results and conclusions from these studies are often disappointing. The potential to transform land management data into extremely useful information is often lost due to lacking the inclusion of DSM and temporal modelling into the future. DSM is extremely effective at displaying and assisting in understanding large datasets of how soil varies spatially, and how soil properties have changed in space and time. There have been recent advances in DSM techniques, with increasingly accurate predictions of soil properties at unvisited locations able to be made, allowing baseline and change maps to be created with minimal operational cost. In addition, the use of available models to simulate future temporal shifts in soil properties is highly useful and cost-efficient, and should be included in soil monitoring studies. These models provide insight into how management and shifting climates will influence soil in years to come, but should be seen as an addition to soil monitoring, rather than a substitute. Furthermore, all estimation in both space and time need a prediction variance to perform hypothesis testing, and there are very few studies that include this in their statistical analysis of the data.

Soil monitoring studies vary considerably in their interests and concerns, but it is clear that there has recently been a strong focus on monitoring SOC stocks. While this is important, and often a good indicator of overall soil condition, there must not be a sole focus on this, and attention must not be taken away from other important soil degradation issues. Soil acidity, salinity, sodicity, and erosion are major global soil degradation issues and monitoring studies and SMNs should encompass these matters in their assessment. Additionally, it is evident from the reviewed literature that the majority of soil monitoring studies only regards the topsoil, often ignoring shifts in subsoil condition. High operational costs are largely to blame for this, but understanding subsoil change is crucial as this has significant implications for agricultural production. Issues in the subsoil are prominent in many parts of the world

(e.g. subsoil sodicity, acidity, salinity), and it has also been shown that an important amount of SOC is also stored at greater depths. Furthermore, most of the reviewed studies are focused in temperate and tropical areas, with very few considering change in arid and semi-arid areas.

There have been numerous monitoring studies that have made clear conclusions about how human activities have impacted soil over time, but few have done the same for climate-induced changes. The impacts of a shifting climate on the global soil resource are still poorly understood, and this is evidenced by the contrasting conclusions by monitoring studies about the effects of increasing atmospheric temperatures on SOC storage, where several studies found that SOC content of the studied area increased despite a warming climate (Reijneveld *et al.* 2009; Fantappiè *et al.* 2011; Yan *et al.* 2011; Minasny *et al.* 2012; Ross *et al.* 2013), while several others reported reduction in SOC content as the atmospheric temperature increased (Bellamy *et al.* 2005; Saby *et al.* 2008; Deng *et al.* 2013; Adhikari *et al.* 2014). This issue is furthered by the difficulty in disentangling the impacts of climate and landuse, and future studies need to work at overcoming this challenge.

It is unquestionable that soil monitoring is crucial in securing and protecting our global soil resources. The implementation of SMNs in countries where it is possible is preferable, but where it is not, legacy data to detect soil change is a promising option, particularly with growing improvements in operational aspects, statistical analysis, and the inclusion of geo-referenced co-located samples. The impact of expected shifts in climatic conditions on soil must not be disregarded, and must be analysed in conjunction with the influence of human activities in soil monitoring studies. With all this in place, the accurate representation of past and future changes in soil condition is possible, providing an encouraging opportunity to determine the direction, nature and rapidity of soil change, and the contributing driving factors.

References

- Adams, H.D., Williams, A.P., Xu, C., Rauscher, S.A., Jiang, X., McDowell, N.G. (2013) Empirical and process-based approaches to climate-induced forest mortality models. *Frontiers in Plant Science*, **4**, 438.
- Adhikari, K., Minasny, B., Greve, M.B., Greve, M.H. (2014) Constructing a soil class map of Denmark based on the FAO legend using digital techniques. *Geoderma*, **214-215**, 101-113.
- Allen, D.E., Singh, B.P., Dalal, R.C. (2011) Soil health indicators under climate change: a review of current knowledge. In: B.P. Singh, A.L. Cowie, K.Y., Chan (Eds.), *Soil Health and Climate Change*. (pp. 25-45) Dordrecht: Springer.
- Antoni, V., Perry-Giraud, C., Villanneau, E., Jolivet, C., Saby, N., Toutain, B., Arrouays, D. (2008) The detection of outliers in trace element content in soils based on the French Soil Quality Monitoring Network. In: Eurosoil 2008 Book of Abstracts (p. 85-86). Presented at Eurosoil 2008,

- Vienne, AUT (2008-08-25 - 2008-08-29). Vienne, AUT : University of Natural Resources and Applied Life Sciences. <http://prodinra.inra.fr/record/51573>
- Aref, S., Wander, M.M. (1997) Long-term trends of corn yield and soil organic matter in different crop sequences and soil fertility treatments on the morrow plots. *Advances in agronomy*, **62**, 153.
- Arrouays, D., Grundy, M.G., Hartemink, A.E., Hempel, J.W., Heuvelink, G.B.M., Hong, S.Y., Lagacherie, P., Lelyk, G., McBratney, A.B., McKenzie, N.J., Mendonca-Santos, M.D.L., Minasny, B., Montanarella, L., Odeh, I.O.A., Sanchez, P.A., Thompson, J.A., Zhang, G.L. (2014) Chapter Three - GlobalSoilMap: Toward a Fine-Resolution Global Grid of Soil Properties. *Advances in Agronomy*, **125**, 93-134.
- 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**, 456-469.
- Arrouays, D., Martin, S., Lepretre, A., Bourennane, H.(2000) Short-range spatial variability of metal contents in soil on a one hectare agricultural plot. *Communications in Soil Science and Plant Analysis*, **31**, 387-400.
- Assad, E.D., Pinto, H.S., Martins, S.C., Groppo, J.D., Salgado, P.R., Evangelista, B., Vasconcellos, E., Sano, E.E., Pavao, E., Luna, R., Camargo, P.B., Martinelli, L.A. (2013) Changes in soil carbon stocks in Brazil due to land use: paired site comparisons and a regional pasture soil survey. *Biogeosciences*, **10**, 6141-6160.
- Barracough, D., Smith, P., Worrall, F., Black, H.I.J., Bhogal, A. (2015) Is there an impact of climate change on soil carbon contents in England and Wales? *European Journal of Soil Science*, **66**, 451-462.
- Battle-Aguilar, J., Brovelli, A., Porporato, A., Barry, D.A. (2011) Modelling soil carbon and nitrogen cycles during land use change. A review. *Agronomy for Sustainable Development*, **31**, 251-274.
- Beckett, P.H.T., Webster, R. (1971) Soil variability: a review. *Soils and Fertilizers*, **34**, 1-15.
- Bellamy, P.H., Loveland, P.J., Bradley, I.B., Lark, R.M., Kirk, G.J.D. (2005) Carbon losses from all soils across England and Wales 1978-2003. *Nature*, **437**, 24-248.
- Bennett, H.H. (1929) Some Aspects of Soil Erosion as a National Problem. *Soil Science Society of America Journal*, **10**, 55-74.
- Bishop, T.F.A., McBratney, A.B., Laslett, G.M. (1999) Modelling soil attribute depth functions with equal-area quadratic smoothing splines. *Geoderma*, **91**, 27-45.
- Black, H., Bellamy, P., Creamer, R., Elston, D., Emmett, B., Frogbrook, Z., Hudson, G., Jordan, C., Lark, M., Lilly, A., Marchant, B., Plum, S., Potts, J., Reynolds, B., Thompson, R., Booth, P. (2008) Design and operation of a UK soil monitoring network. Environment Agency, Bristol, UK. <<http://www.envirobase.info/PDF/7577.pdf>>
- Blake, L., Goulding, K.W.T., Mott, C.J.B., Johnston, A.E. (1999) Changes in soil chemistry accompanying acidification over more than 100 years under woodland and grass at Rothamsted Experimental Station, UK. *European Journal of Soil Science*, **50**, 401-412.
- Blois, J.L., Williams, J.W., Fitzpatrick, M.C., Jackson, S.T., Ferrier, S. (2013) Space can substitute for time in predicting climate-change effects on biodiversity. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 9374-9379.
- Brevik, E.C. (2012) Soils and Climate Change: Gas Fluxes and Soil Processes. *Soil Horizons*, **53**, 12-23.
- Brus, D.J., Lamé, F.P.J., Nieuwenhuis, R.H. (2009) National baseline survey of soil quality in the Netherlands. *Environmental Pollution*, **157**, 2043-2052.
- Bui, E (2006) Chapter 2 - A Review of Digital Soil Mapping in Australia. In: Lagacherie, P., McBratney, A.B., Voltz, M. (Eds.), *Developments in Soil Science*, **31**, (pp. 25-37). Elsevier.
- Cattle, S.R., Koppi, A.J., McBratney, A.B. (1994) The effect of cultivation on the properties of a Rhodoxeralf from the wheat/sheep belt of New South Wales. *Geoderma*, **63**, 215-225.

- Chauveau, D., Saby, N.P.A., Orton, T.G., Lemerrier, B., Walter, C., Arrouays, D. (2014) Large-scale simultaneous hypothesis testing in monitoring carbon content from French soil database - A semi-parametric mixture approach. *Geoderma*, **219–220**, 117-124.
- Chhabra, R. (2002) Salt-affected soils and their management for sustainable rice production-key management issues: a review. *Agricultural Reviews*, **23**, 110-126.
- Clark, D.A. (2002) Are tropical forests an important carbon sink? Reanalysis of the long-term plot data. *Ecological Applications*, **12**, 3-7.
- Coleman, K., Jenkinson, D., (1999) RothC-26.3 – A model for the turnover of carbon in soil model description and windows users guide. Harpenden: University press.
- Conant, R.T., Paustian, K. (2002) Spatial variability of soil organic carbon in grasslands: implications for detecting change at different scales. *Environmental Pollution*, **116**, 127-135.
- Conant, R.T., Paustian, K., Elliott, E.T. (2001) Grassland management and conversion into grassland: effects on soil carbon. *Ecological Applications*, **11**, 343-355.
- Craine, J.M., Gelderman, T.M. (2011) Soil moisture controls on temperature sensitivity of soil organic carbon decomposition for a mesic grassland. *Soil Biology and Biochemistry*, **43**, 455-457.
- de Brogniez, D., Ballabio, C., Stevens, A., Jones, R.J.A., Montanarella, L., van Wesemael, B. (2015) A map of the topsoil organic carbon content of Europe generated by a generalized additive model. *European Journal of Soil Science*, **66**, 121-134.
- De Gruijter, J.J., Bierkens, M.F.P., Brus, D.J. & Kotters, M. (2006) Sampling for natural resource monitoring. Berlin: Springer.
- Deng, F., Minasny, B., Knadel, M., McBratney, A., Heckrath, G., Greve, M.H. (2013) Using Vis-NIR Spectroscopy for Monitoring Temporal Changes in Soil Organic Carbon. *Soil Science*, **178**, 389-399.
- Department of Primary Industries (DPI) Victoria (2012) Soil data matters. Available from: <http://vro.depi.vic.gov.au/dpi/vro/vrosite.nsf/pages/lwm_usfs-newsletter-edition5-november2012> [20 August 2015]
- Eldridge, D.J., Green, R.S.B., Dean, C. (2011) Climate change impacts on soil processes in rangelands. In: Singh, B.P., Cowie, A.L., Chan, K.Y. (Eds.), *Soil Health and Climate Change* (pp. 237-255). Dordrecht: Springer.
- Ellert, B.H., Bettany, J.R. (1995) Calculation of organic matter and nutrients stored in soils under contrasting management regimes. *Canadian Journal of Soil Science*, **75**, 529-538.
- Emmett, B., Black, H.I.J., Bradley, R.I., Elston, D., Reynolds, B., Creamer, R., Frogbrook, Z.L., Hudson, G., Jordan, C., Lilly, A., Moffat, A., Potts, J., Vanguelova, E., (2007a). National Soil Monitoring Network: Review and assessment study. Final report. SNIFFER, Project LQ09.
- Emmett, B.A., Frogbrook, Z.L., Chamberlain, P., Griffiths, R., Pickup, R., Reynolds, B., Rowe, E., Rowlands, P., Spurgeon, D., Wilson, J., Wood, C. (2007b). Countryside Survey Soils 2007: Design, power analyses and protocols. Centre for Ecology and Hydrology Project No. C03042
- Fantappiè, M., L'Abate, G., Costantini, E.A.C. (2011) The influence of climate change on the soil organic carbon content in Italy from 1961 to 2008. *Geomorphology*, **135**, 343-352.
- Frazer, L. (2009) Climate change: will warmer soil be as fertile? *Environmental Health Perspectives*, **117**, 59.
- Fukami, T., Wardle, D. A. (2005) Long-term ecological dynamics: reciprocal insights from natural and anthropogenic gradients. *Proceedings of the Royal Society of Biological Sciences*, **272**, 2105–2115.
- Ghassemi, F., Jakeman, A.J., Nix, H.A. (1995) Salinisation of land and water resources: human causes, extent, management and case studies. CAB international.
- Goidts, E., van Wesemael, B. (2007) Regional assessment of soil organic carbon changes under agriculture in Southern Belgium (1955–2005). *Geoderma*, **141**, 341-354.

- Grealish, G., Clifford, D., Wilson, P., Ringrose-Voase, A. (2011) National soil monitoring for soil pH and soil carbon – objectives, design, protocols, governance and reporting. CSIRO Land and Water, Science Report 05/11, CSIRO, Canberra.
- Groffman, P., Hardy, J., Fisk, M., Fahey, T., Driscoll, C. (2009) Climate variation and soil carbon and nitrogen cycling processes in a northern hardwood forest. *Ecosystems*, **12**, 927-943.
- Guo, J.H., Liu, X.J., Zhang, Y., Shen, J.L., Han, W.X., Zhang, W.F., Christie, P., Goulding, K.W.T., Vitousek, P.M., Zhang, F.S. (2010) Significant Acidification in Major Chinese Croplands. *Science*, **327**, 1008-1010.
- Guo, L.B., Gifford, R.M. (2002) Soil carbon stocks and land use change: a meta analysis. *Global Change Biology*, **8**, 345-360.
- Hamza, M.A., Anderson, W.K. (2005) Soil compaction in cropping systems - A review of the nature, causes and possible solutions. *Soil & Tillage Research*, **82**, 121-145.
- Harrison, R.B., Footen, P.W., Strahm, B.D. (2011) Deep soil horizons: contribution and importance to soil carbon pools and in assessing whole-ecosystem response to management and global change. *Forest Science*, **57**, 67-76.
- Hartemink, A. (2008) Soil Map Density and a Nation's Wealth and Income. In: A. Hartemink, A. B. McBratney, M.L. Mendonça-Santos (Eds.), *Digital soil mapping with limited data* (pp. 53-66). Netherlands: Springer.
- Hartemink, A.E. (2006) Assessing soil fertility decline in the tropics using soil chemical data. *Advances in Agronomy*, **89**, 179-225.
- Hayes, D.C., Stout, S.L., Crawford, R.F., Hoover, A.P. (2014) USDA forest service experimental forests and ranges: research for the long-term. Springer
- Haynes, R.J. (2008) Soil organic matter quality and the size and activity of the microbial biomass: their significance to the quality of agricultural soils. In: Q. Huang, P. Huang, A. Violante (Eds.), *Soil mineral microbe-organic interactions* (pp. 201-231). Berlin Heidelberg: Springer.
- Helyar, K.R., Cregan, P.D., Godyn, D.L. (1990) Soil acidity in New South Wales - current pH values and estimates of acidification rates. *Australian Journal of Soil Research*, **28**, 523-537.
- Herrero, J., Pérez-Coveta, O. (2005) Soil salinity changes over 24 years in a Mediterranean irrigated district. *Geoderma*, **125**, 287-308.
- Hodson, M.E., Langan, S.J., Kennedy, F.M., Bain, D.C. (1998) Variation in soil surface area in a chronosequence of soils from Glen Feshie, Scotland and its implications for mineral weathering rate calculations. *Geoderma*, **85**, 1-18.
- Hyvönen, R., Ågren, G.I., Linder, S., Persson, T., Cotrufo, M.F., Ekblad, A., Freeman, M., Grelle, A., Janssens, I.A., Jarvis, P.G., Kellomäki, S., Lindroth, A., Loustau, D., Lundmark, T., Norby, R.J., Oren, R., Pilegaard, K., Ryan, M.G., Sigurdsson, B.D., Strömberg, M., van Oijen, M., Wallin, G. (2007) The likely impact of elevated CO², nitrogen deposition, increased temperature and management on carbon sequestration in temperate and boreal forest ecosystems: a literature review. *New Phytologist*, **173**, 463-480.
- Islam, K.R., Weil, R.R. (2000) Land use effects on soil quality in a tropical forest ecosystem of Bangladesh. *Agriculture, Ecosystems & Environment*, **79**, 9-16.
- Jolivet, C., Boulonne, L., Ratié, C., Mouclier, M., Saby, N., Arrouays, D. (2008) The soil quality monitoring network for French soils (RMQS): state of progress and first results. In: Eurosoil 2008 Book of Abstracts (p. 307). Presented at Eurosoil 2008, Vienne, AUT (2008-08-25 - 2008-08-29). Vienne, AUT : University of Natural Resources and Applied Life Sciences.
- Jones, R.J.A., Verheijen, F.G.A., Reuter, H.I., Jones, A.R. (Eds) (2008) 'Environmental assessment of soil for monitoring. Vol. V: Procedures & Protocols.' EUR 23490 EN/5. (Office for the Official Publications of the European Communities: Luxembourg)

- Karunaratne, S.B., Bishop, T.F.A., Odeh, I.O.A., Baldock, J.A., Marchant, B.P. (2014) Estimating change in soil organic carbon using legacy data as the baseline: issues, approaches and lessons to learn. *Soil Research*, **52**, 349-365.
- Kay, B.D., Dexter, A.R., Rasiah, V., Grant, C.D. (1994) Weather, cropping practices and sampling depth effects on tensile strength and aggregate stability. *Soil and Tillage Research*, **32**, 135-148.
- Kirk, G.J.D., Bellamy, P.H., Lark, R.M. (2010) Changes in soil pH across England and Wales in response to decreased acid deposition. *Global Change Biology*, **16**, 3111–3119.
- Kirk, G.J.D., Bellamy, P.H., Scott, A., Emmett, B.A. (2013). Comparison of soil carbon changes across England and Wales estimated in the Countryside Survey and the National Soil Inventory. Rep. SP1101. Dep. for Environ., Food and Rural Aff., London.
- Kirschbaum, M.U.F. (1995) The temperature dependence of soil organic matter decomposition, and the effect of global warming on soil organic C storage. *Soil Biology and Biochemistry*, **27**, 753-760.
- Kobza, J., 1995. Soil monitoring system in Slovakia. *Environmental Monitoring Assessment*, **34**, 127-129.
- Kurganova, I., Lopes de Gerenyu, V., Six, J., Kuzyakov, Y. (2014) Carbon cost of collective farming collapse in Russia. *Global Change Biology*, **20**, 938-947.
- Lacoste, M., Minasny, B., McBratney, A., Michot, D., Viaud, V., Walter, C. (2014) High resolution 3D mapping of soil organic carbon in a heterogeneous agricultural landscape. *Geoderma*, **213**, 296-311.
- Lal, R. (1997) Degradation and resilience of soils. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **352**, 997–1010.
- Lark, R.M., Bellamy, P.H., Rawlins, B.G. (2006) Spatio-temporal variability of some metal concentrations in the soil of eastern England, and implications for soil monitoring. *Geoderma* **133**, 363-379.
- Lindert, P.H., Lu, J., Wanli, W. (1996) Trends in the soil chemistry of South China since the 1930s. *Soil Science*, **161**, 329-342.
- Magdoff, F., Weil, R.R. (2004) Soil organic matter in sustainable agriculture. Florida: CRC Press.
- Manzoni, S., Porporato, A. (2009) Soil carbon and nitrogen mineralization: Theory and models across scales. *Soil Biology and Biochemistry*, **41**, 1355-1379.
- Marchant, B.P., Lark, R.M. (2007) Estimation of linear models of coregionalization by Residual Maximum Likelihood. *European Journal of Soil Science*, **58**, 1506-1513.
- Marchant, B.P., Crawford, D.M., Robinson, N.J. (2012) Spatial and temporal prediction of soil properties from legacy data. In: B. Minasny, B.P. Malone, A.B. McBratney (Eds.), *Digital Soil Assessments and Beyond* (pp. 239-244). London: CRC Press.
- 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*. Vol. 25. No. 1. IOP Publishing, 2015.
- McBratney, A., Field, D.J., Koch, A. (2014) The dimensions of soil security. *Geoderma*, **213**, 203-213.
- McBratney, A., Pringle, M. (1999) Estimating average and proportional variograms of soil properties and their potential use in precision agriculture. *Precision Agriculture*, **1**, 125-152.
- McBratney, A.B., Mendonça Santos, M.L., Minasny, B. (2003) On digital soil mapping. *Geoderma*, **117**, 3-52.
- McKenzie, N.J., Henderson, B., McDonald, W. (2002) Monitoring Soil Change: Principles and practices for Australian conditions. CSIRO Land and Water, Technical Report 18/02, CSIRO, Canberra.
- Mckenzie, N.J., Dixon, J. (2007) Monitoring soil condition across Australia – recommendations from the expert panels. National Committee on Soil and Terrain for the National Land and Water Resources Audit
- Meersmans, J., Martin, M.P., Lacerce, E., Orton, T.G., De Baets, S., Gourrat, M., Saby, N.P.A., Wetterlind, J., Bispo, A., Quine, T.A., Arrouays, D. (2013) Estimation of Soil Carbon Input in France: An Inverse Modelling Approach. *Pedosphere*, **23**, 422-436.

- Meersmans, J., Van Wesemael, B., De Ridder, F., Fallas Dotti, M., De Baets, S., Van Molle, M. (2009) Changes in organic carbon distribution with depth in agricultural soils in northern Belgium, 1960–2006. *Global Change Biology*, **15**, 2739-2750.
- Miller, J.D., Duff, E.I., Hirst, D., Anderson, H.A., Bell, J.S., Henderson, D.J. (2001) Temporal changes in soil properties at an upland Scottish site between 1956 and 1997. *Science of The Total Environment*, **265**, 15-26.
- Milne, E., Al Adamat, R., Batjes, N., Bernoux, M., Bhattacharyya, T., Cerri, C., Coleman, K., Easter, M., Falloon, P. (2007) National and sub-national assessments of soil organic carbon stocks and changes: The GEFSOC modelling system. *Agriculture, Ecosystems & Environment*, **122**, 3-12.
- Minasny, B., McBratney, A.B., Hong, S.Y., Sulaeman, Y., Kim, M.S., Zhang, Y.S., Kim, Y.H., Han, K.H. (2012) Continuous rice cropping has been sequestering carbon in soils in Java and South Korea for the past 30 years. *Global Biogeochemical Cycles*, **26**, 3027.
- Minasny, B., McBratney, A.B., Malone, B.P., Wheeler, I. (2013) Digital Mapping of Soil Carbon. *Advances in Agronomy*, **118**, 1-47.
- Minasny, B., McBratney, A.B., Mendonça-Santos, M.L., Odeh, I.O.A., Guyon, B. (2006) Prediction and digital mapping of soil carbon storage in the Lower Namoi Valley. *Soil Research*, **44**, 233-244.
- Minasny, B., Sulaeman, Y., McBratney, A.B. (2011) Is soil carbon disappearing? The dynamics of soil organic carbon in Java. *Global Change Biology*, **17**, 1917-1924.
- Mol, G., Vriend, S.P., van Gaans, P.F.M. (1998) Future trends, detectable by soil monitoring networks? *Journal of Geochemical Exploration*, **62**, 61-66.
- Morvan, X., Saby, N.P.A., Arrouays, D., Le Bas, C., Jones, R.J.A., Verheijen, F.G.A., Bellamy, P.H., Stephens, M., Kibblewhite, M.G. (2008) Soil monitoring in Europe: A review of existing systems and requirements for harmonisation. *Science of The Total Environment*, **391**, 1-12.
- Mudge, P.L., Schipper, L.A., Baisden, W.T., Ghani, A., Lewis, R.W. (2014) Changes in soil C, N and $\delta^{15}\text{N}$ along three forest–pasture chronosequences in New Zealand. *Soil Research*, **52**, 27-37.
- Müller, C., Rütting, T., Abbasi, M.K., Laughlin, R.J., Kammann, C., Clough, T.J., Sherlock, R.R., Kattge, J., Jäger, H.J., Watson, C.J. (2009) Effect of elevated CO₂ on soil N dynamics in a temperate grassland soil. *Soil Biology and Biochemistry*, **41**, 1996-2001.
- Murphy, B.W., Packer, I.J., Cowie, A.L., Singh, B.P. (2011) Tillage and crop stubble management and soil health in a changing climate. In: Singh, BP, Cowie, AL, Chan, KY (eds) *Soil Health and Climate Change*. Springer, Dordrecht, pp 181-206
- Norby, R.J., Warren, J.M., Iversen, C.M., Medlyn, B.E., McMurtrie, R.E. (2010) CO₂ enhancement of forest productivity constrained by limited nitrogen availability. *Proceedings of the National Academy of Sciences*, **107**, 19368-19373.
- Odeh, I.O.A., Cattle, S.R., Triantafyllis, J., McBratney, A.B., Taylor, J.A. (2004) A simple soil database management assistant for Australian cotton industry. In 'SuperSoil 2004: Program and Abstracts for the 3rd Australian New Zealand Soils Conference'. The University of Sydney, NSW. (Ed. B Singh) p. 50. (ASSSI: Sydney)
- Odeh, I.O.A., Leenaars, J., Hartemink, A., Amapu, I. (2012) The challenges of collating legacy data for digital mapping of Nigerian soils. In: Minasny, B., Malone, B., McBratney, A.B., (Eds.), *Digital Soil Assessments and Beyond*. Taylor and Francis Group, Londres, UK 453-458.
- Oldeman, L.R. (Eds D.J. Greenland, I. Szabolcs (1994) 'The global extent of soil degradation.' (C a B International: Wallingford)
- Panagos, P., Meusburger, K., Alewell, C., Montanarella, L. (2012) Soil erodibility estimation using LUCAS point survey data of Europe. *Environmental Modelling & Software*, **30**, 143-145.
- Papritz, A., Flühler, H. (1994) Temporal change of spatially autocorrelated soil properties: optimal estimation by cokriging. *Geoderma*, **62**, 29-43.

- Papritz, A., Webster, R. (1995a) Estimating temporal change in soil monitoring: I. Statistical theory. *European Journal of Soil Science*, **46**, 1-12.
- Papritz, A., Webster, R. (1995b) Estimating temporal change in soil monitoring: II. Sampling from simulated fields. *European Journal of Soil Science*, **46**, 13-27.
- Paré, D., Boutin, R., Larocque, G.R., Raulier, F. (2006) Effect of temperature on soil organic matter decomposition in three forest biomes of eastern Canada. *Canadian Journal of Soil Science*, **86**, 247-256.
- Pickett, S.A. (1989) Space-for-Time Substitution as an Alternative to Long-Term Studies. In 'Long-Term Studies in Ecology.' (Ed. G Likens.) pp. 110-135. (Springer New York:
- Pimentel, D., Harvey, C., Resosudarmo, P., Sinclair, K., Kurz, D., McNair, M., Crist, S., Shpritz, L., Fitton, L., Saffouri, R., Blair, R. (1995) Environmental and Economic Costs of Soil Erosion and Conservation Benefits. *Science*, **267**, 1117-1123.
- Post, W.M., Emanuel, W.R., Zinke, P.J., Stangenberger, A.G. (1982) Soil carbon pools and world life zones. *Nature*, **298**, 156-159.
- Raison, J., Khanna, P.K. (2011) Possible impacts of climate change on forest soil health. In: Singh, B.P., Cowie, A.L., Chan, K.Y. (Eds.), *Soil Health and Climate Change*. Springer, Dordrecht, pp 257-285
- Rampazzo, N., Rajkai, K., Blum, W.E.H., Varallyay, G., Ubleis, T. (1999a) Effects of long-term agricultural land use on soil properties along the Austrian-Hungarian border. Part I. Soil mineralogical, physical and micromorphological parameters. *International Agrophysics*, **13**, 15-39.
- Rampazzo, N., Rajkai, K., Blum, W.E.H., Varallyay, G., Ubleis, T. (1999b) Effects of long-term agricultural land use on soil properties along the Australian-Hungarian border. Part II. Soil chemical, microbiological and zoological parameters. *International Agrophysics*, **13**, 171-183.
- Reijneveld, A., van Wensem, J., Oenema, O. (2009) Soil organic carbon contents of agricultural land in the Netherlands between 1984 and 2004. *Geoderma*, **152**, 231-238.
- Rengel, Z. (2011) Soil pH, soil health and climate change. In: Singh, B.P., Cowie, A.L., Chan, K.Y. (Eds.), *Soil Health and Climate Change*. Springer, Dordrecht, pp 69-85
- Reth, S., Reichstein, M., Falge, E. (2005) The effect of soil water content, soil temperature, soil pH-value and the root mass on soil CO₂ efflux – A modified model. *Plant and Soil*, **268**, 21-33.
- Reynolds, B., Chamberlain, P.M., Poskitt, J., Woods, C., Scott, W.A., Rowe, E.C., Robinson, D.A., Frogbrook, Z.L., Keith, A.M., Henrys, P.A., Black, H.I.J., Emmett, B.A. (2013) Countryside Survey: National “Soil Change” 1978–2007 for Topsoils in Great Britain—Acidity, Carbon, and Total Nitrogen Status. *Vadose Zone Journal*, **12**, 1-15.
- Richter, D.D., Hofmockel, M., Callaham, M., Powlson, D., Smith, P. (2014) Long-term soil experiments. Available from: < <http://iscn.fluxdata.org/Data/LTSEs/Pages/Map.aspx> > [20 August 2015]
- Richter, D.D., Bacon, A.R., Megan, L.M., Richardson, C.J., Andrews, S.S., West, L., Wills, S., Billings, S., Cambardella, C.A., Cavallaro, N., DeMeester, J.E., Franzluebbers, A.J., Grandy, A.S., Grunwald, S., Gruver, J., Hartshorn, A.S., Janzen, H., Kramer, M.G., Ladha, J.K., Lajtha, K., Liles, G.C., Markewitz, D., Megonigal, P.J., Mermut, A.R., Rasmussen, C., Robinson, D.A., Smith, P., Stiles, C.A., Tate, R.L., Thompson, A., Tugel, A.J., van Es, H., Yaalon, D., Zobeck, T.M. (2011) Human–Soil Relations are Changing Rapidly: Proposals from SSSA's Cross-Divisional Soil Change Working Group. *Soil Science Society of America Journal*, **75**, 2079-2084.
- Richter, D.D., Hofmockel, M., Callaham, 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.D., Markewitz, D. (2001) 'Understanding Soil Change: Soil Sustainability Over Millennia, Centuries, and Decades.' (Cambridge University Press:
- Ross, C.W., Grunwald, S., Myers, D.B. (2013) Spatiotemporal modeling of soil organic carbon stocks across a subtropical region. *Science of The Total Environment*, **461–462**, 149-157.

- Rothamsted Research (2014) Long-term experiments national capability. Available from: <<http://www.rothamsted.ac.uk/long-term-experiments-national-capability>> [11 August 2015]
- Rustad, L., Melillo, J., Mitchell, M., Fernandez, I., Steudler, P., McHale, P. (2000) Effects of Soil Warming on Carbon and Nitrogen Cycling. In: Mickler, R., Birdsey, R., Hom, J., (Eds.), *Responses of Northern U.S. Forests to Environmental Change*, (pp. 357-381). New York: Springer.
- Rutgers, M., Mulder, C., Schouten, A.J., Bloem, J., Bogte, J.J., Breure, A.M., Brussaard, L., de Goede R.G.M., Faber, J.H., Jagers op Akkerhuis, G.A.J.M., Keidel, H., Korthals, G.W., Smeding, F.W., ten Berg, C., van Eekeren, N., (2008) Soil ecosystem profiling in the Netherlands with ten references for biological soil quality. Rijksinstituut voor Volksgezondheid en Milieu RIVM,
- Saby, N.P.A., Arrouays, D., Antoni, V., Lemerrier, B., Follain, S., Walter, C., Schvartz, C. (2008a) Changes in soil organic carbon in a mountainous French region, 1990–2004. *Soil Use and Management*, **24**, 254-262.
- Saby, N.P.A., Bellamy, P.H., Morvan, X., Arrouays, D., Jones, R.J.A., Verheijen, F.G.A., Kibblewhite, M.G., Verdoodt, A.N.N., Üveges, J.B., Freudenschuß, A., Simota, C. (2008b) Will European soil-monitoring networks be able to detect changes in topsoil organic carbon content? *Global Change Biology*, **14**, 2432-2442.
- Sauer, D., (2015) Pedological concepts to be considered in soil chronosequence studies. *Soil Research*, **53**, 577–591.
- Schjørring, P., Heckrath, G., Christensen, B.T., Fakultet, A.U.D.J. (2009) Threats to Soil Quality in Denmark: A Review of Existing Knowledge in the Context of the EU Soil Thematic Strategy. (Aarhus University, Faculty of Agricultural Sciences, Department of Agroecology and Environment)
- Shirokova, Y., Forkutsa, I., Sharafutdinova, N. (2000) Use of Electrical Conductivity Instead of Soluble Salts for Soil Salinity Monitoring in Central Asia. *Irrigation and Drainage Systems*, **14**, 199-206.
- Singh, B., Odeh, I.O.A., McBratney, A.B. (2003) Acid buffering capacity and potential acidification of cotton soils in northern New South Wales. *Soil Research*, **41**, 875-888.
- Singh, G., Ram Babu, Narain, P., Bhushan, L.S. and Abrol, I.P., (1992). Soil erosion rates in India. *Journal of Soil & Water Conservation*, **47**, 97-99.
- 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.
- Sleutel, S., De Neve, S., Hofman, G., Boeckx, P., Beheydt, D., Van Cleemput, O., Mestdagh, I., Lootens, P., Carlier, L., Van Camp, N., Verbeeck, H., Vande Walle, I., Samson, R., Lust, N., Lemeur, R. (2003) Carbon stock changes and carbon sequestration potential of Flemish cropland soils. *Global Change Biology*, **9**, 1193-1203.
- Smith, J.O., Smith, P., Wattenbach, M., Zaehle, S., Hiederer, R., Jones, R.J.A., Montanarella, L., Rounsevell, M.D.A., Reginster, I., Ewert, F. (2005) Projected changes in mineral soil carbon of European croplands and grasslands, 1990–2080. *Global Change Biology*, **11**, 2141-2152.
- Smith, P., Chapman, S.J., Scott, W.A., Black, H.I.J., Wattenbach, M., Milne, R., Campbell, C.D., Lilly, A., Ostle, N., Levy, P.E., Lumsdon, D.G., Millard, P., Towers, W., Zaehle, S., Smith, J.U. (2007) Climate change cannot be entirely responsible for soil carbon loss observed in England and Wales, 1978–2003. *Global Change Biology*, **13**, 2605-2609.
- Stockmann, U., Adams, M.A., Crawford, J.W., Field, D.J., Henakaarchchi, N., Jenkins, M., Minasny, B., McBratney, A.B., Courcelles, V.D.R.D., Singh, K., Wheeler, I., Abbott, L., Angers, D.A., Baldock, J., Bird, M., Brookes, P.C., Chenu, C., Jastrow, J.D., Lal, R., Lehmann, J., O'Donnell, A.G., Parton, W.J., Whitehead, D., Zimmermann, M. (2013) The knowns, known unknowns and unknowns of sequestration of soil organic carbon. *Agriculture, Ecosystems & Environment*, **164**, 80-99.
- Stoorvogel, J.J., Smaling, E.M.A., Janssen, B.H. (1993) Calculating soil nutrient balances in Africa at different scales. *Fertilizer Research*, **35**, 227-235.

- Suarez, D.L. (2006) Inorganic Carbon: Land Use Impacts. In: Lal, R., (Eds.), *Encyclopedia of Soil Science-Second Edition - Vol. 1'* (pp. 895-897). Florida: CRC Press.
- Tang, C., Rengel, Z. (2003) Role of plant cation/anion uptake ratio in soil acidification. In: Rengel, Z. (Eds.), *Handbook of soil acidity* (pp. 57-81). New York: Marcel Dekker.
- Thonon, I., Klok, C. (2007) Impact of a changed inundation regime caused by climate change and floodplain rehabilitation on population viability of earthworms in a lower River Rhine floodplain. *Science of The Total Environment*, **372**, 585-594.
- Trenberth, K.E. (2011) Changes in precipitation with climate change. *Climate Research*, **47**, 123-138.
- Tugel, A.J., Herrick, J.E., Brown, J.R., Mausbach, M.J., Puckett, W., Hipple, K. (2006) Soil change, soil survey, and natural resources decision making: A blueprint for action. *Soil Science Society of America Journal*, **70**, 1416-1416.
- Tye, A.M., Robinson, D.A., Lark, R.M. (2013) Gradual and anthropogenic soil change for fertility and carbon on marginal sandy soils. *Geoderma*, **207–208**, 35-48.
- van Wesemael, B., Paustian, K., Andr n, O., Cerri, C.P., Dodd, M., Etchevers, J., Goidts, E., Grace, P., K tterer, T., McConkey, B., Ogle, S., Pan, G., Siebner, C. (2011) How can soil monitoring networks be used to improve predictions of organic carbon pool dynamics and CO₂ fluxes in agricultural soils? *Plant and Soil*, **338**, 247-259.
- Viaud, V., Angers, D.A., Walter, C. (2010) Toward Landscape-Scale Modeling of Soil Organic Matter Dynamics in Agroecosystems. *Soil Science Society of America Journal*, **74**, 1847-1860.
- Vleeshouwers, L.M., Verhagen, A. (2002) Carbon emission and sequestration by agricultural land use: a model study for Europe. *Global Change Biology*, **8**, 519-530.
- Walker, L.R., Wardle, D.A., Bardgett, R.D., Clarkson, B.D. (2010) The use of chronosequences in studies of ecological succession and soil development. *Journal of Ecology*, **98**, 725-736.
- Walter, C., Lagacherie, P., Follain, S., (2006) Integrating pedological knowledge into digital soil mapping. In: Lagacherie, P., McBratney, A.B., Voltz, M. (Eds.), *Developments in Soil Science* (pp. 281-615). Elsevier.
- Wattel-Koekkoek, E.J.W., van Vliet, M.E., Boumans, L.J.M., Ferreira, J., Spijker, J., van Leeuwen, T.C., (2013) Soil quality in the Netherlands for 2006-2010 and the change compared with 1993-1997: Results of the National Soil Quality Monitoring Network. RIVM Report, National Institute for Public Health and the Environment, Netherlands.
- Wendt, J.W., Hauser, S. (2013) An equivalent soil mass procedure for monitoring soil organic carbon in multiple soil layers. *European Journal of Soil Science*, **64**, 58-65.
- Wilson, B.R., Barnes, P., Koen, T.B., Ghosh, S., King, D. (2010) Measurement and estimation of land-use effects on soil carbon and related properties for soil monitoring: a study on a basalt landscape of northern New South Wales, Australia. *Soil Research*, **48**, 421-433.
- Yan, X., Cai, Z., Wang, S., Smith, P. (2011) Direct measurement of soil organic carbon content change in the croplands of China. *Global Change Biology*, **17**, 1487-1496.
- Yang, D., Kanae, S., Oki, T., Koike, T., Musiak, K. (2003) Global potential soil erosion with reference to land use and climate changes. *Hydrological Processes*, **17**, 2913-2928.

Chapter 3

3. Materials and methods – study area, soil datasets, soil property analysis and modelling approaches

3.1. Introduction

While the focus of each chapter in this thesis differs, all research chapters address soil data collected from the same two soil surveys within the same study area. The covariates used for digital soil modelling and mapping in these chapters, and the way that these covariates were processed, are also very similar. In addition, Linear Mixed Models (LMMs) were used to model different soil properties in space and time in the different research chapters, although these models varied in their complexity. Due to the large size of the soil datasets and the high number of samples in this study, visible near infrared (VisNIR) spectroscopic techniques were used to predict certain soil properties to fill in the gaps where traditional laboratory methods were time- and cost-limiting. To prevent repetition in the ensuing research chapters, this chapter describes the study area, the soil datasets, soil property analyses, the covariates used for mapping and modelling, and the approaches used to model these soil properties in space and time.

3.2. Study area

3.2.1. Physiography

The study area is a semi-arid landscape that surrounds the township of Hillston (-33.4853°, 145.5328°), which is situated along the Lachlan River in south-eastern Australia (Fig. 3.1). The study area is 2650 km² in size and is largely flat to slightly undulating, and primarily consists of partly active alluvial floodplains (Cay and Cattle 2005). There are some scarce sandstone outcrops, with the Lachlan Range being the area at the highest elevation (298 m) in the study region (Fig. 3.2). The main watercourse of the study area is the Lachlan River, which flows to the south-west. There are also two major creeks that flow off the Lachlan River in the study area; Merrowie Creek, which also flows to the south-west, and Willandra Creek, which flows to the north-west (Fig. 3.2). In addition, there are several ephemeral streams that flow during wet periods. The region is characterised by an abundance of groundwater located in aquifers.

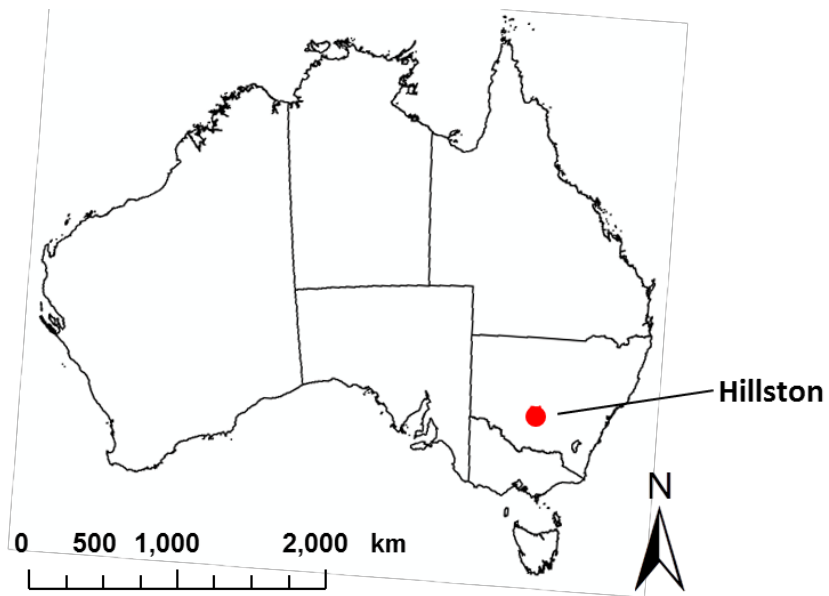


Fig. 3.1 – The location of the study area (Hillston) within Australia

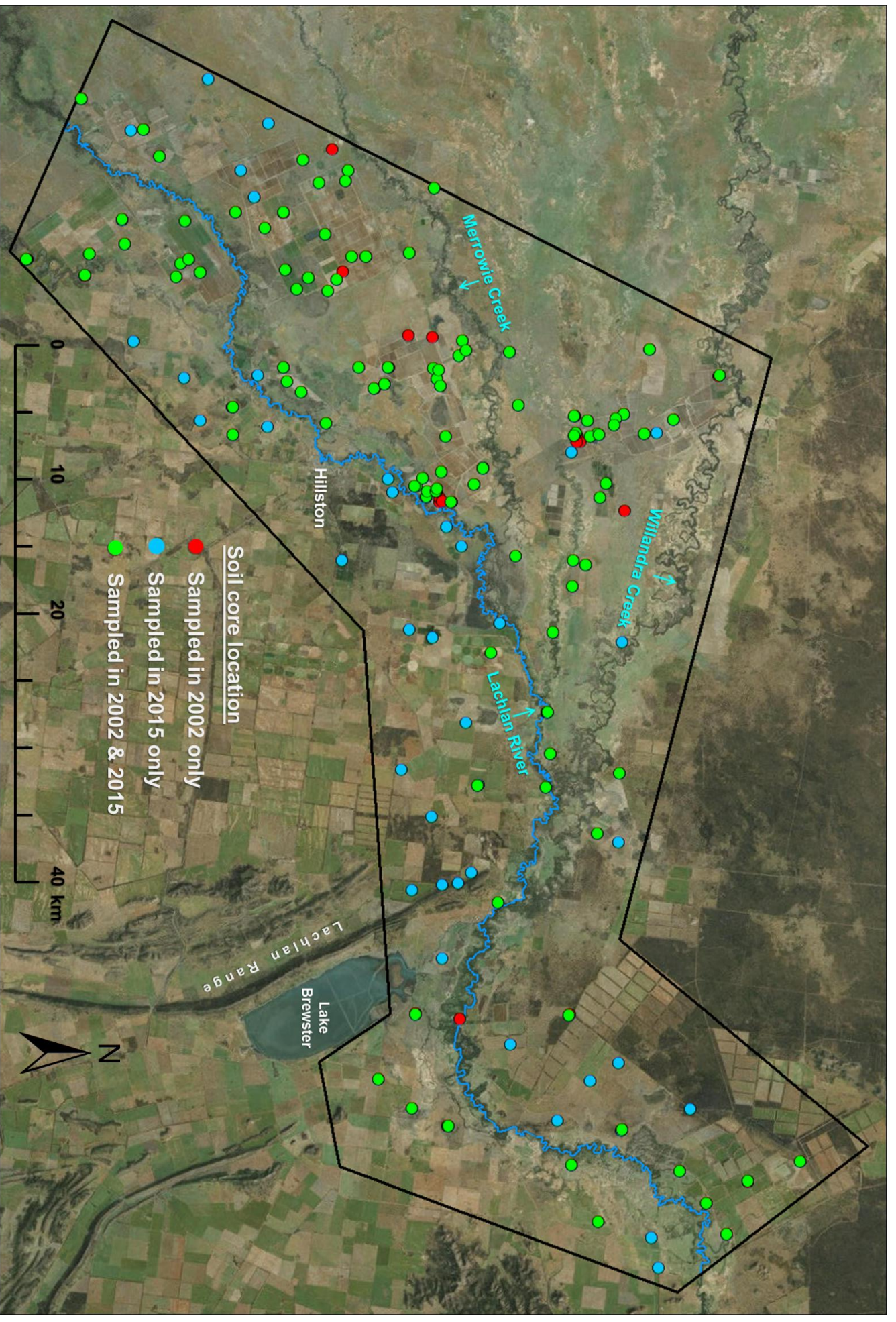


Fig. 3.2 – The locations of soil cores extracted in the 2002 and 2015 soil surveys overlaid on a satellite image with notable landmarks labelled

3.2.2. Soil types and geological units

Although no detailed soil map of the area currently exists, a detailed geological survey by Cameron *et al.* (2000) has mapped the area. The three most widespread geological units are Quaternary alluvial units of differing ages; Qa, Qaf and TQs, ranging from the most recent, to the oldest deposits, respectively (Fig. 3.3). These have a strong correlation with Grey Vertosols, Brown Vertosols, and Red Vertosols, respectively, as they would be classed according to the Australian Soil Classification (Isbell 2016). The Vertosols are located on the floodplains and along watercourses and are the most agriculturally productive soils of the study area, particularly for irrigated cotton production due to their high water holding capacity. There are also a number of less common and less agriculturally productive alluvial and aeolian soils with a typically sandier texture, and these are grouped together as ‘non-Vertosols’. These non-Vertosols are generally used for irrigated annual horticulture, dryland cropping and grazing.

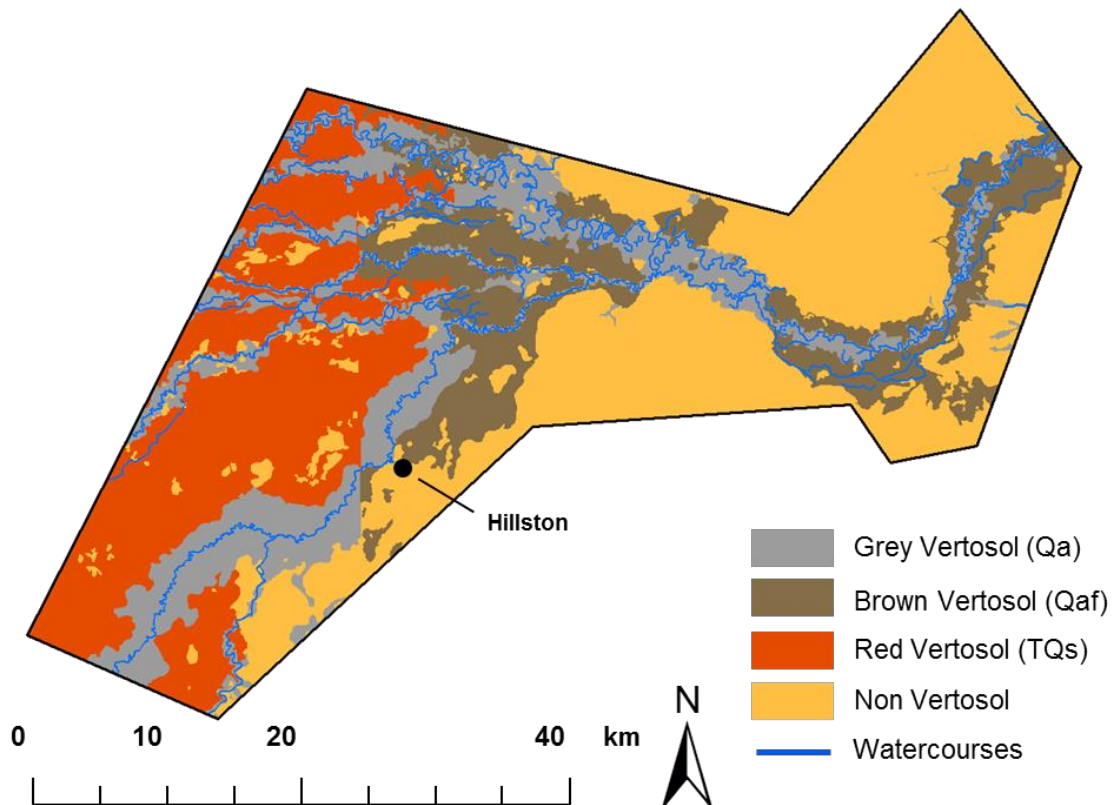


Fig. 3.3 – The broad soil types of the study area, derived from geological units of the Hillston 1:100,000 mapsheet (Cameron *et al.* 2000)

3.2.3. Climate and climatic fluctuations

Summers in the Hillston area are hot and winters cool, with a mean minimum temperature of 17.7°C in summer and 4.5°C in winter, and mean maximum temperatures of 31.2°C in summer and 15.9°C in winter (BOM 2017). Rainfall is evenly distributed throughout the year, with an average annual rainfall of 372 mm and an annual pan evaporation of 1919 mm (BOM 2017). In the recent decade or so, Hillston has undergone some severely fluctuating rainfall patterns (Fig. 3.4). From 2002 to 2009, Hillston was subjected to the worst drought ever recorded, with an average rainfall of 306 mm for these eight consecutive years (Fig. 3.4), and this meant that surface water for irrigation was limited. In 2010, the rainfall patterns changed abruptly, and Hillston received an average of 633 mm annually from 2010 through to the end of 2012 (Fig. 3.4). This was not simply a period of above-average rainfall, but was the wettest period that Hillston had ever experienced on record, and this meant that surface water for irrigation was abundant again for several years (BOM 2017).

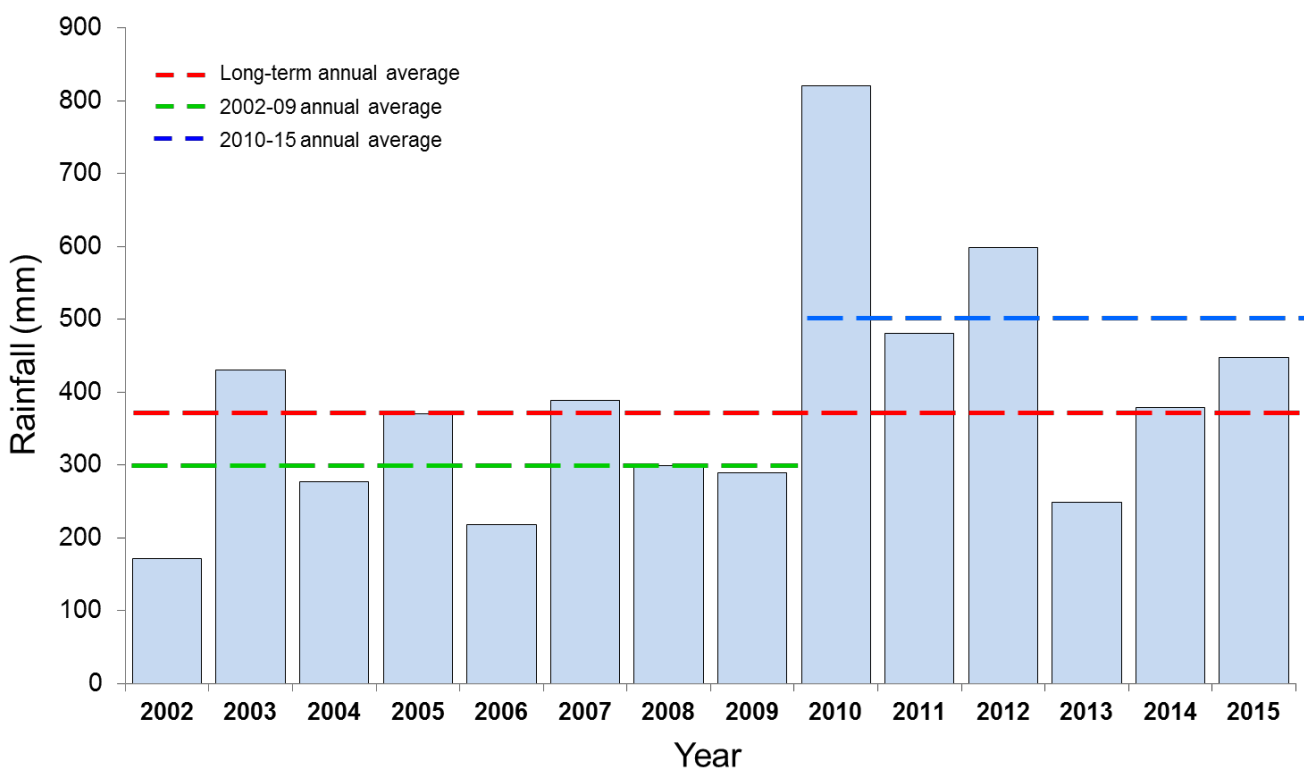


Fig. 3.4 – Annual rainfall received at Hillston from 2002 to 2015 (BOM 2017)

3.2.4. Land use and land use change

Patterns of land use and soil management in Hillston over the last decade or so have changed significantly (Fig. 3.5), with much of this due to the fluctuating rainfall and water availability. Despite the lengthy drought from 2002 to 2009 there has been an expansion of irrigated agriculture (Fig. 3.5), and this has been primarily due to increased utilisation of groundwater from aquifers. However, water availability varies for each individual water licence holder, as not all have access to groundwater and rely solely on seasonal availability of surface water. Historically, dryland cropping and extensive grazing were the main enterprises in Hillston, but since the mid 1990s, the most economically important agricultural crop for the Hillston district has been irrigated cotton (*Gossypium hirsutum*). The area under irrigated cotton has continued to expand in recent years, with several large areas recently developed for this. The other main land uses in the study area include horticulture, dryland cropping, rangeland grazing, and areas under native vegetation (Fig. 3.5).

Cotton is a summer crop, and is typically planted in September and harvested in May at Hillston. All cotton crops in Hillston are either furrow-irrigated, or overhead sprinkler-irrigated, as the climatic conditions render the area unsuitable for rain-fed production. The management practices for irrigated cotton are quite intensive, with large quantities of water applied (~1000-1400 mm or ~10-14 megalitres/ha of water in a season), constant cultivation of the soil, large fertiliser inputs, and continual trafficking of heavy machinery (Table 3.1). Cotton crops are commonly planted back-to-back with another cotton crop, or rotated with a dryland winter crop, such as wheat (*Triticum aestivum*), or a dryland summer crop, such as safflower (*Carthamus tinctorius*). During the 2002-09 Millennium Drought in Hillston, surface water was scarce, and only irrigated cotton farms that had access to groundwater (approximately 50% of the irrigated cotton farms in Hillston) were able to continue to irrigate, while the remaining irrigated cotton farms were primarily used for dryland cropping or left in fallow (Table 3.1). When rainfall patterns changed in 2010, there was an abundance of irrigation water available, and most growers in the Hillston district could carry on producing cotton for several years (Table 3.1).

A significant portion of the study area is also used purely for dryland cropping, with winter cereal crops typically grown (Fig. 3.5). Large parts of the area are used for very low intensity rangeland grazing, and this land use and areas of native vegetation, were grouped together as 'natural' due to their similarities (Fig. 3.5). This was done because there are minimal parts of the study area that are not used for agriculture, and the density of grazing from livestock and native/feral animals is largely analogous. The horticulture land use is diverse at Hillston, with a variety of irrigated annual and perennial plants

being grown, including garlic, betroot, watermelon, cherries, citrus and almonds. The area used for horticulture has been expanding in recent years (Fig. 3.5), largely driven by market prices and water shortages in different countries throughout the globe. In particular, the area covered by almond (*Prunus dulcis*) farms has increased, with several new locations developed for this, as well as a few large cotton farms that have re-allocated their water provisions and planted almond trees. Almond trees require large amounts of water, and the equivalent of ~1600-2000 mm (16-20 megalitres/ha) of irrigation is applied in a typical year in the Hillston district, which is even greater than cotton crops (Table 3.1).

Table 3.1 – The typical amount of water percolation that occurs on an annual basis, and throughout the study period (2002-15) for characteristic land uses in the Hillston district

Land use	Average rainfall (2002-2015)	Annual irrigation (2002-2009)	Annual irrigation (2010-2015)	Total water received (2002-2015)
Dryland cropping	~387 mm	0 mm	0 mm	~5,000 mm
Natural	~387 mm	0 mm	0 mm	~5,000 mm
Cotton (irrigated through drought)	~387 mm	~1000-1400 mm	~1000-1400 mm	~20,600 mm
Cotton (not irrigated through drought)	~387 mm	0 mm	~1000-1400 mm	~11,000 mm
Irrigated almonds	~387 mm	~1600-2000 mm	~1600-2000 mm	~28,400 mm

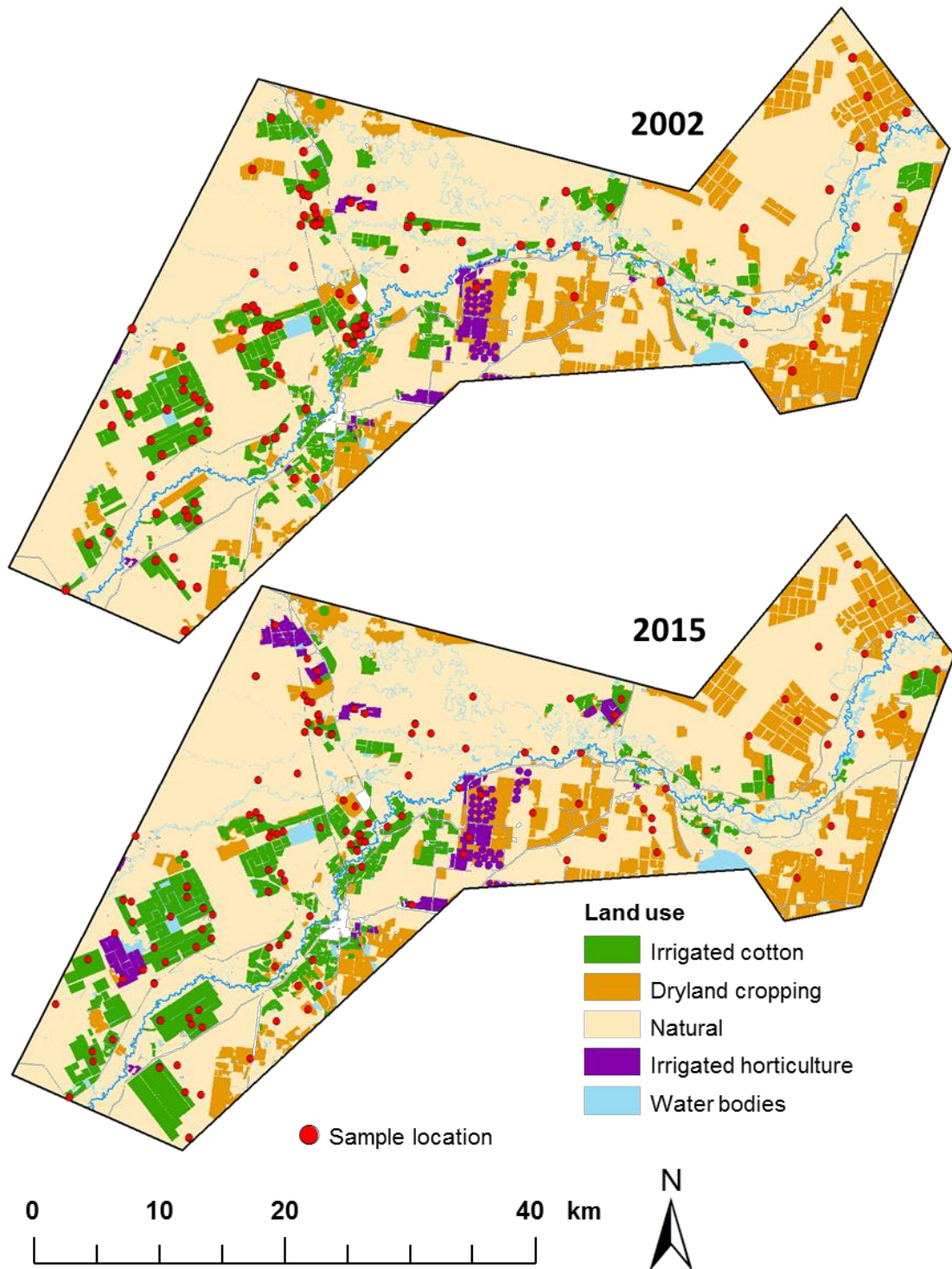


Fig. 3.5 – Soil core locations, and land use classes for the study area in the year 2002 (left), and 2015 (right) (Office of Environment and Heritage 2017)

3.3. Soil datasets

3.3.1. 2002 dataset

A soil survey was conducted in Hillston in 2002, with 115 soil cores extracted to 1.5 m (where possible) to assess the status of the soil resource, and to evaluate the effect of land use on soil condition in this semi-arid region (Fig. 3.2). Cores were then subsampled at six depth increments (0-0.2, 0.3-0.4, 0.55-0.65, 0.8-0.9, 1.1-1.2, 1.35-1.5 m). A purposive sampling design was adopted with strata developed based on soil type and land use. The location of sites within each soil type-land use complex was not selected probabilistically, and in many cases sites were chosen opportunistically based on access to fields, so there are no known-inclusion probabilities. The number of sampling sites was not allocated proportionally to the area of each soil type-land use complex, but has a bias towards the cotton land use, due to the aims of the original project.

3.3.2. 2015 dataset

A soil survey of the same study area was performed in 2015, with 160 soil cores extracted (Fig. 3.2). Many of the sample locations from 2002 were returned to ($n = 103$) using georeferenced coordinates, and new sites ($n = 37$) were also sampled to increase the number of sites for certain soil-landuse (SLU) complexes, and to fill in gaps for digital soil mapping purposes (Fig. 3.2). In addition, at 20 of the sites, a second soil core was extracted ≈ 30 -50 m away and analysed separately to investigate the extent of soil spatial variation over short distances. Due to the non-random nature of the sampling design in 2002, the 2015 design is also considered stratified non-random. While soil cores to 1.5 m were again collected, the subsample depth increments were changed (0-0.1, 0.1-0.3, 0.3-0.5, 0.5-0.8, 0.8-1.2, 1.2-1.5 m) to be completely inclusive of the whole soil profile.

3.4. Analysis of soil properties

Both sets of soil samples were analysed for a number of attributes using traditional laboratory methods, but due to the large number of samples and the time- and resource-consuming nature of these laboratory techniques, not all samples were analysed in this way. Consequently, archived soils from both surveys were scanned with visible near infrared (VisNIR) spectroscopy, and this information, combined with machine learning techniques were used to predict certain soil properties. This was not performed on all soil samples, but a subset, and was particularly useful for predicting soil properties that are laborious and costly to measure, such as exchangeable sodium percentage. Prior to any traditional laboratory or spectroscopic analysis, all of the soil samples were air-dried, ground and passed through a 2 mm sieve.

3.4.1. Traditional laboratory methods

3.4.1.1. Soil pH

Soil pH was analysed in 1:5 soil:water extracts using a PHM 83 Autocal™ pH meter with glass calomel electrode for 2002 samples, and a Mettler Toledo S220 SevenCompact™ pH/Ion meter for 2015 samples.

3.4.1.2. Soil electrical conductivity

To estimate salinity, soil electrical conductivity (EC) was determined for 1:5 soil:water extracts, using a CDM 83™ conductivity meter for the 2002 samples, and using a Mettler Toledo SevenCompact™ conductivity meter for the 2015 samples.

3.4.1.3. Soil exchangeable sodium percentage

In both the 2002 and 2015 datasets, exchangeable base cation (Ca^{2+} , Mg^{2+} , K^+ and Na^+) contents were determined for samples that were extracted with alcoholic 1 M ammonium chloride (pH 8.5), and then analysed by atomic absorption spectrometry (Rayment and Lyons 2011). Prior to this, samples were pre-treated to remove soluble salts by using a combination of 60% aqueous ethanol and 20% aqueous glycerol. The effective cation exchange capacities (ECEC) were estimated by summing the exchangeable basic cations, and exchangeable sodium percentages (ESP) were calculated by dividing the exchangeable sodium contents by the ECEC, and multiplying by 100.

3.4.1.4. Soil carbon – total, organic and inorganic

Prior to laboratory analysis, all samples were tested for the presence of soil inorganic carbon (SIC). A ~1 g subsample of ground soil was placed on a ceramic plate and a few drops of 1 M hydrochloric acid (HCl) were placed directly onto the sample. Any sample that showed effervescence was considered to contain calcium carbonate, the most prominent form of SIC in these soils. An additional subsample (~10 g) was then taken and finely ground (<53 μm) using a Fritsch Mortar Grinder Pulverisette 2 (Fritsch, Germany) for 4 minutes at 50-60 Hz frequency. Soil total carbon (STC) was determined by the combustion method with the Leco1 CHN analyser for 2002 samples, and the Elementar vario MAX CNS for 2015 samples. Soil organic carbon (SOC) for 2002 samples was determined by treating samples with 2 M HCl to remove inorganic carbon, and then analysing by the Leco1 CHN analyser (Tiessen and Moir 1993). For 2015 samples, SOC was determined by the Walkely-Black method, which is a wet oxidation technique that uses chromic acid (Walkley and Black 1934). To estimate SIC, the difference between STC and SOC was used. The Elementar vario MAX CNS and the Leco1 CHN analyser are very similar in their analytical

approach, and both use the combustion technique. For 2002 samples, SOC was determined immediately, however, STC was determined from archived samples 13 years later.

3.4.2. Visible near infrared (VisNIR) predictions

While pH and EC are rapid and inexpensive to analyse in the laboratory, the estimation of CEC, ESP and the different components of soil carbon (organic and inorganic) are considered to be quite labour intensive. Many of the 2002 soil samples were archived and stored ($n = 385$), and these were scanned by VisNIR, along with all of the 2015 soil samples ($n = 906$). Spectroscopic measurement was made with an Agrispec portable spectrophotometer with a contact probe attachment, on dried and ground samples (Analytical Spectral Devices, Boulder, Colorado). To reduce signal-to-noise ratios of the spectra, three scans of each sample were performed, from which an averaged reflectance spectrum was derived. Calibration of the instrument was made with a Spectralon white tile and was re-calibrated after every 15 scans, or five samples.

3.4.2.1. Soil exchangeable sodium percentage

All samples from the 2002 survey were analysed for ESP and ECEC by traditional laboratory methods, but only a subset ($n = 138$) were analysed in the 2015 dataset. This number was chosen due to budgetary constraints. Instead, both ESP and ECEC were directly predicted using visible near infrared (VisNIR) spectroscopic techniques in those samples from the 2015 survey that were not measured by traditional laboratory techniques. To ensure that this subset of 138 samples represented the remaining 768 samples from the 2015 survey appropriately, conditioned Latin hypercube (CLHC) sampling was used (Minasny and McBratney 2006). Initially, a principal component analysis (PCA) was performed on the spectra from the VisNIR wavelengths (350-2500 nm) of the 906 samples from the 2015 survey. The first two principal components (PCs) from this PCA explained >95% of the cumulative variation. We aimed to select sites based on the whole profile, so the first two PCs of all sampling depths were used as input criteria for the CLHC, which ensured that all of the soil profile was considered during site selection. Land use and soil type was also an input in the CLHC to ensure that each soil-land use complex of the study area was represented.

The training dataset consisted of the 385 samples that were scanned by VisNIR from the 2002 survey, as well the 138 subset of samples from the 2015 survey, and this was used to predict onto the remaining 768 un-analysed samples from 2015 using Cubist models, which is a machine learning technique. The predictor variables included wavelengths from 500-2450 nm, averaged into segments of

ten nm, and the mid-depth of the sample, to ensure that the depth of the sample was taken into account when predicting.

To test the prediction quality of the Cubist models, 75% of the dataset was used as calibration, and the remaining 25% was used as validation, as is common practice when undertaking this procedure. These two datasets were selected by using the VisNIR spectra and ESP values as inputs in a Latin hypercube, and this ensured that both the validation and calibration datasets were appropriately represented. During the model testing, it was ensured that other depth samples from the same soil core for that year were not included in the calibration dataset when validating. It is assumed there is a strong correlation between these different depths of the same core and this may bias results.

3.4.2.2. Soil organic and inorganic carbon

Only the 0-0.1, 0.1-0.3 and 0.3-0.5 m sampling depths were considered in the chapters that focused on soil carbon. While all samples were analysed qualitatively for the presence of SIC, a portion of these samples ($n = 118$) that had SIC present were only analysed for STC, and as a result, had an unknown quantity of SIC and SOC. The combustion method of measuring SOC is useful and rapid in samples that lack SIC, but when SIC is present, a lengthy process of carbonate dissolution is required, as described in the methodology in the previous section. As a result, soil spectroscopic techniques were used to make predictions of SOC and SIC in these 118 soil samples. Archived soil samples from both the 2002 and 2015 soil surveys ($n = 517$) were scanned by VisNIR. The calibration dataset consisted of 399 samples from both 2002 ($n = 148$) and 2015 ($n = 251$) surveys to predict on the remaining 118 samples. Cubist models were used to predict SOC and SIC in the 118 un-analysed samples. The predictor variables included VisNIR wavelengths from 500-2450 nm, averaged into segments of ten, and the mid-depth of the sample, to ensure that the depth was taken into account when predicting. In addition to this, soil pH (1:5 soil: water) and STC (measured by traditional laboratory methods) were included as predictor variables, as this data was available for all samples.

To test the prediction quality of the Cubist models, 75% of the dataset was used as calibration, and the remaining 25% was used as validation. These datasets were selected by performing a Latin hypercube of the VisNIR spectra, pH, STC, mid-depth, and the response variables to ensure that both the validation and calibration datasets were appropriately represented. The calibration and validation dataset splits of 75% and 25%, respectively, is also extremely similar to the dataset split when making actual predictions. When model testing, 299 (75%) samples are used to predict 100 (25%), and in actual

predictions, 399 samples were used to predict 118. For each soil property, model quality was tested for four variations of inputs. These included VisNIR and mid-depth; VisNIR, mid-depth and pH; VisNIR, mid-depth and STC; and finally VisNIR, mid-depth, pH and STC.

3.5. Harmonising sampling depths

The 2002 and 2015 soil surveys had dissimilar sampling depths, and to standardise this, equal-area quadratic smoothing splines (Bishop *et al.* 1999) were used on both the laboratory-measured, and VisNIR-predicted soil attribute values. The 2015 soil sampling intervals (0-0.1, 0.1-0.3, 0.3-0.5, 0.5-0.8, 0.8-1.2, 1.2-1.5 m) were chosen as the standard depths to be fully inclusive of the whole soil profile. Where necessary, the same procedure was also used on the 2015 samples, such as when a full 1.5 m core could not be extracted.

3.6. Collation and processing of covariates for modelling and mapping

3.6.1. Covariates for modelling and mapping

A suite of predictor variables for the study area were gathered based on the *SCORPAN* approach for the creation of models and digital soil maps (McBratney *et al.* 2003). These covariates can be broadly separated into four types; terrain attributes, gamma radiometric products, soil type, and land use information (Table 3.2).

Table 3.2 – List and specifications of environmental covariates used in model development

Covariate	Data source	Original resolution	Mean	Min	Max
Land use ¹	OEH, NSW	1:100,000	4 classes	-	-
Geological units ¹	Geoscience Australia	1:100,000-1:250,000	3 classes	-	-
Elevation	DEM	30 m	129.33	101.1	297.7
Slope	DEM	30 m	0.01	0	0.80
Length-slope factor ²	DEMh	30 m	0.11	0	50.35
Length-slope factor ³	DEMh	30 m	0.05	0	29.57
MRRTF	DEMh	30 m	4.76	0	6.83
MRVBF	DEMh	30 m	7.14	0	8.82
Topographic wetness index	DEMh	30 m	14.91	8.66	31.17
Wetness index ²	DEMh	30 m	25.42	8.59	28.31
Wetness index ³	DEMh	30 m	11.08	4.23	28.85
Weathering index	Geoscience Australia	90 m	3.70	0.88	7.20
Total dose	Geoscience Australia	90 m	84.40	35.40	162.3
Dose	Geoscience Australia	90 m	46.34	-2.58	124.0
K	Geoscience Australia	90 m	1.15	-0.21	3.75
Th	Geoscience Australia	90 m	9.72	-0.07	24.80
U	Geoscience Australia	90 m	1.24	-0.33	3.75
Th/K	Geoscience Australia	90 m	8.56	0.88	47.48
U ² /Th	Geoscience Australia	90 m	0.17	-0.01	0.89
U/K	Geoscience Australia	90 m	1.09	0.05	10.67
U/Th	Geoscience Australia	90 m	0.13	0.01	1.49

¹Categorical covariates; ²Derived from individual calculation in SAGA GIS; ³Derived from basic terrain analysis calculation in SAGA GIS

Gamma radiometrics data for the study area at a spatial resolution of 90 m was obtained from the Geophysical Archive Data Delivery System - Geoscience Australia. Low-pass filtering was used on all radiometric products (Minty *et al.* 2009). The digital elevation model (DEM) at 30 m resolution was obtained from the National Elevation Data Framework - Geoscience Australia. Both the hydrologically corrected DEM (DEMh) with the removal of artificial sinks, and original DEM, were used to generate a suite of terrain attributes using the software program SAGA GIS. Elevation and slope information was generated from the original DEM, and length-slope (LS) factor, multi-resolution valley bottom flatness (MRVBF), multi-resolution ridge-top flatness (MRRTF), and topographic wetness index (TWI) were generated using DEMh.

As detailed earlier, there are four broad classes of soil type (Fig. 3.3), however, for the purpose of this study the Grey and Brown Vertosols have been grouped together. This was done due to the similar traits and close proximity of the Grey and Brown Vertosols, as well as the different levels of detail of the three original geological mapping surveys undertaken by Cameron *et al.* (2000). As a result, this left three classes of soil type; Grey/Brown-, Red-, and non-Vertosols. Land use maps for the study area in 2002 were obtained from the Office of Environment and Heritage (OEH), NSW Government, and the classification was simplified into four broad land use classes; cotton, horticulture, dryland cropping, and

natural (OEH 2017). The class 'natural' consists of both grazed and native vegetation areas, as described previously. As there had been no survey of land use in Hillston since the early 2000s, it was required that land use information for 2015 be independently collected during the sampling campaigns. This involved querying individual landholders about past and current land use and resulted in a strong understanding of the management practices that growers implement and the patterns of land use change in the Hillston region. A large majority of growers in the study area were surveyed, and land use information was consequently thorough and up-to-date. The areas were classified into the same land use classes as 2002 (Fig. 3.5). This categorical information of soil type and land use were subsequently used as a covariate in the modelling and mapping of the various soil properties for the subsequent research chapters of this thesis.

3.6.2. Principal component analysis

Overall, there were nineteen numerical covariates derived from terrain attributes and gamma radiometrics that could potentially be used as predictor variables for modelling. However, because many of these would be insignificant in the development of models, and the culling of redundant covariates would be time-consuming, a principal component analysis (PCA) was performed on all numerical covariates. Covariates were not transformed before the analysis, but were scaled and centred prior to the PCA. The principal components (PCs) that cumulatively explained 95% of the variation were kept as potential predictor variables, and the remaining PCs were discarded. To investigate the impact that each original input variable (covariate) had on the individual principal components, the component loadings were analysed.

3.7. Linear mixed models (LMMs)

When soil monitoring studies possess a time-series of co-located data, as we have here in this study, there are opportunities to maximally benefit from the data and improve the understanding of spatio-temporal variation of soil properties. Co-located data for the purpose of this thesis refers to soil information from the same spatial location that varies in some other aspect. For example, the same soil property of the same soil core at a different depth increment, or the same soil property at the same depth increment taken at the same location at a different time point. When the latter is available, this can be particularly beneficial in soil monitoring studies, as there is often a strong correlation through time between these samples.

A modelling/mapping approach that can take advantage of this correlation is multivariate linear mixed models (MLMMs). In MLMMs, the correlation between co-located sites can be utilised to strengthen predictions at non-co-located sites (Webster and Oliver 2001; Lark and Papritz 2003), and there is particular benefit from this approach when one soil survey possesses more sample points than the other. In this approach, one model can be created for the multiple time points, rather than creating separate models for each time point. Creating models in this way typically allows for more precise estimates, and results in soil maps that are coherent and have a logical connection through time as the spatio-temporal patterns of the response variables are utilised. Despite the obvious advantages, many studies that model/map multiple responses disregard the covariance between co-located samples either because they used a modelling approach that cannot incorporate this into their predictions, e.g. a machine learning approach (Akpa *et al.* 2014), or they used univariate geostatistical methods (Ross *et al.* 2013).

In Chapter 4, 5 and 7 of this thesis, bivariate linear mixed models (BLMMs) are used to model two response variables in conjunction, with these two variables being a soil property value at a standard depth in 2002, and in 2015, and this allows the correlation through time to be utilised. In Chapter 6, a different approach is taken and vertical spatial correlation is accounted for in addition to this temporal correlation. Here we use multivariate linear mixed models (MLMMs) as we are modelling four response variables simultaneously, with these four variables being the soil property value at sampling depth *A* in 2002, sampling depth *A* in 2015, sampling depth *B* in 2002, and sampling depth *B* in 2015. This approach is particularly advantageous when one sampling depth has more samples than the other, and results in pedologically sensible soil maps.

In linear mixed models, the response variable is modelled as a combination of fixed and random effects. In this study the fixed effects are the linear relationship between the response variable and the covariates, and the random effects describe the spatial correlation of the residuals. In cases where there is no trend (no fixed effects/covariates), this is analogous to co-kriging or using a linear model of co-regionalisation (LMCR) in classical geostatistics (Papritz and Flüher 1994; Bishop and Lark 2007). The structure of BLMMs and MLMMs allows different covariates to be used for each response variable. It also permits the spatial auto-correlation of the residuals for each response variable to be modelled, as well as the cross-correlation between these. Modelling of the cross-correlation allows for more precise estimates of the response variable at each time point or time point/depth, as well as the subsequent changes over time (Papritz and Flüher 1994; Lark and Papritz 2003). There are several advantages of

using bivariate and multivariate LMMs, as opposed to univariate approaches when data jointly co-varies, as univariate approaches ignore any correlation between variables. In this thesis an exponential semivariogram model was used for the random effect terms in the models.

3.7.1. Model selection and map production

The combination of covariates to be included in the final model was determined by backwards elimination. The initial model had separate fixed effects for each time point (i.e. all PCs, land use and soil type) and Wald tests were performed to identify redundant covariates, where those with the highest p-value (least significant) were removed. This was continued until all predictors had a p-value less than 0.1. Categorical data was kept in the model if at least one of the levels was deemed significant. Once the model trends were determined, those numerical predictor variables (PCs) that were common to all response variables were analysed to assess if the upper and lower confidence intervals around the parameter estimates overlapped. If this was the case, they were merged to have a common covariate for all response variables, and the model was then re-run to test the significance. More simply, this could lead to a common partial regression coefficient for both time periods. Merging these overlapping covariates creates a more simplified model. After this, the model was used to predict onto the covariate grid of the study area (100 m resolution) using the EBLUP (Empirical Best Linear Unbiased Predictor) approach and digital soil maps were then produced. All analyses were performed in R (R Development Core Team 2017).

3.7.2. Assessment of model quality

The quality of predictions from the BLMMs were tested by using leave-one-site-out cross-validation (LOSOCV), where both time points for one site were removed from the dataset and predicted using the remaining dataset. This was repeated for each site. The same was done for the MLMs, however, both time points and all depths for one site was removed in this approach. This entails fitting one model, and then validating the model by removing all data for a particular location and then predicting at that location. This is repeated at all locations. The performance of the prediction variances and uncertainty was assessed through the mean and median of the standardized squared prediction errors (SSPE). When prediction variance matches the actual errors the mean would be equal to 1.0 and the median equal to 0.455 (Lark 2000; Orton *et al.* 2014). Lin's concordance correlation coefficient (LCCC), was used as an assessment of model quality as it is the fit of the observed and predicted values to the 1:1 line, and is unit-less, making it useful for comparing between models (Lin 1989). In addition, the RMSE (root mean

square error) was used to analyse the quality of predictions, and the mean of the residuals was used to analyse the bias (over- or under-fitting) of model predictions.

3.7.3. Change calculation and statistical significance of change

It is important that a prediction interval is placed around predicted change, as it subsequently allows a hypothesis test to be performed. Despite the importance of this, there has been very few soil monitoring studies where this has actually been implemented. The hypotheses we are testing are:

$$H_0: \mu_d = 0; \quad 3.1)$$

$$H_1: \mu_d \neq 0;$$

where u_d is some estimate of the change (Equation 3.2), such as a mean for a region or a prediction at a specific location;

$$\mu_d = \hat{y}_2 - \hat{y}_1, \quad 3.2)$$

where \hat{y}_1 is the prediction of the soil property for the first point in time and \hat{y}_2 at the second time point. To test this hypothesis, we need an estimate of the variance of the change, which can be expressed as:

$$\hat{V}(\Delta\hat{y}) = \hat{V}(\hat{y})_1 + \hat{V}(\hat{y})_2 - 2(\hat{V}(\hat{y})_{1,2}), \quad 3.3)$$

where $\hat{V}(\Delta\hat{y})$ is the variance of the change (contrast variance) in the soil property between two points in time, $\hat{V}(\hat{y})_1$ is the variance of the soil property for the observations at the first point in time, $\hat{V}(\hat{y})_2$ at the second point in time, and $\hat{V}(\hat{y})_{1,2}$ is the covariance between the measurements of soil properties at the two time periods (the covariance term). Whether it is the BLMM approach we use in Chapter 4, 5 or 7, or the MLMM approach we use in Chapter 6, the equations presented here remain the same, as the difference we are interested in is the change between 2002 and 2015 at a single standard depth interval. Equation 3.3 presents both pitfalls and opportunities in monitoring. The pitfalls are the spatial variation of the soil property and the magnitude of the change, while the opportunity is the covariance between sampling campaigns (Filippi *et al.* 2016). When the covariance term is included and there is a strong relationship between the two surveys, a smaller contrast variance results, subsequently allowing greater sensitivity in predicting a statistically significant change. If we did not use BLMMs we would not be able to estimate the covariance term ($\hat{V}(\hat{y})_{1,2}$), and would have a conservative test. Many studies ignore the correlation at co-located sites and disregard the covariance term and contrast variance, and

very few consider the statistical significance of the temporal change in soil properties (Ross *et al.* 2013). The statistical significance of change can be analysed using the z-statistic:

$$z = \frac{\hat{y}_2 - \hat{y}_1}{\sqrt{\hat{V}(\Delta\hat{y})}}, \quad 3.4)$$

where z is the z-statistic score, \hat{y}_1 is the value of the prediction of the soil property at time point one, \hat{y}_2 at the second time point, and $\hat{V}(\Delta\hat{y})$ is the contrast variance, which is obtained using Equation 3.3. For a two-tailed test, a z-statistic score greater than (in a positive case) or less than (for a negative case) ± 1.28 , ± 1.65 , ± 1.96 , and ± 2.58 represents a prediction interval of 80%, 90%, 95%, and 99% respectively.

3.8. Concluding remarks

The subsequent research chapters focus on individual soil properties, where the datasets and modelling approaches used is based on the materials and methods that have been described in this chapter.

References

- Akpa, SIC, Odeh, IOA, Bishop, TFA, Hartemink, AE (2014) Digital Mapping of Soil Particle-Size Fractions for Nigeria. *Soil Science Society of America Journal* **78**, 1953-1966.
- Bishop, TFA, Lark, RM (2007) A landscape-scale experiment on the changes in available potassium over a winter wheat cropping season. *Geoderma* **141**, 384-396.
- Bishop, TFA, McBratney, AB, Laslett, GM (1999) Modelling soil attribute depth functions with equal-area quadratic smoothing splines. *Geoderma* **91**, 27-45.
- Bureau of Meteorology - BOM (2017) Monthly rainfall – Hillston airport. Available from: <http://www.bom.gov.au/climate/averages/tables/cw_075032_All.shtml> [25 April 2017]
- Cameron RG, Colquhoun GP, Meakin NS (2000) Provisional Hillston 1:100 000 Geological Series Sheet. Geological Survey of New South Wales, Sydney, Australia
- Cay, ER, Cattle, SR (2005) The effects of landforming on soil profile characteristics of an irrigated cotton-producing area of southeastern Australia. *Soil and Tillage Research* **84**, 76-86.
- Filippi, P, Minasny, B, Cattle, SR, Bishop, TFA (2016) Monitoring and Modeling Soil Change: The Influence of Human Activity and Climatic Shifts on Aspects of Soil Spatiotemporally. *Advances in Agronomy* **139**, 153-214.
- Isbell R (2016) The Australian Soil Classification, CSIRO Land and Water, Canberra.
- Lark, RM (2000) A comparison of some robust estimators of the variogram for use in soil survey. *European Journal of Soil Science* **51**, 137-157.
- Lark, RM, Papritz, A (2003) Fitting a linear model of coregionalization for soil properties using simulated annealing. *Geoderma* **115**, 245-260.
- Lin, LIK (1989) A concordance correlation coefficient to evaluate reproducibility. *Biometrics* **45**, 255-268.
- McBratney, AB, Mendonça Santos, ML, Minasny, B (2003) On digital soil mapping. *Geoderma* **117**, 3-52.
- Minasny, B, McBratney, AB (2006) A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers & Geosciences* **32**, 1378-1388.
- Minty B, Franklin R, Milligan P, Richardson M, Wilford J (2009) The Radiometric Map of Australia. *Exploration Geophysics* **40**, 325-333.
- Office of Environment and Heritage - OEH (2017) NSW Landuse 2007. NSW Government, Australia

- Orton, TG, Pringle, MJ, Page, KL, Dalal, RC, Bishop, TFA (2014) Spatial prediction of soil organic carbon stock using a linear model of coregionalisation. *Geoderma* **230–231**, 119-130.
- Papritz, A, Flühler, H (1994) Temporal change of spatially autocorrelated soil properties: optimal estimation by cokriging. *Geoderma* **62**, 29-43.
- Rayment, G. E., Lyons, D. J. (2011). Soil chemical methods: Australasia, 3. CSIRO publishing.
- R Development Core Team (2017) R: A language and environment for statistical computing, R version 3.2.1. R Foundation for Statistical Computing, Vienna, Austria.
- Ross, CW, Grunwald, S, Myers, DB (2013) Spatiotemporal modeling of soil organic carbon stocks across a subtropical region. *Science of The Total Environment* **461–462**, 149-157.
- Tiessen, H., Moir, J.O., 1993. Soil chemical analyses: total organic carbon. In: Carter, M.R. (Ed.), Soil Sampling and Methods of Analysis. Lewis Publishers, Florida, USA, pp. 187–200.
- Walkley, A, Black, IA (1934) An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil science* **37**, 29-38.
- Webster, R, Oliver, MA, (2001) Geostatistics for Environmental Scientists, John Wiley & Sons, Chichester, UK

Chapter 4

4. Monitoring and modelling spatio-temporal changes in topsoil and subsoil pH at decadal scales in an irrigated semi-arid landscape, using bivariate linear mixed models

Abstract

Intensive agricultural management practices and increasingly fluctuating rainfall have the potential to significantly impact the status and change of important soil properties. This study looks at the change in soil pH during a 13-year period in a semi-arid irrigated cotton-growing region in the lower Lachlan River valley in southern NSW, Australia under various land uses. Two soil surveys – one from 2002 and the other from 2015 – that had many of the same sites revisited and resampled, were used in conjunction with bivariate linear mixed models (BLMMs) to model soil pH status and change, at six depth increments to 1.5 m depth. The BLMM approach resulted in models with high predictive power and low prediction variance, likely due to the utilisation of the correlation of pH values at co-located sites through time. Results revealed an overall acidification trend throughout the soil profile, particularly in the subsoil. Despite this, only isolated areas of the study area were predicted to have experienced a statistically significant change in soil pH over time. This acidification is presumably due to management practices associated with irrigated cotton production and increased water flow throughout the profile due to continuous irrigation and periods of very high rainfall. Future work should look at acidification due to the dissolution of carbonates under high water percolation through the soil profile. Further investigation into utilising the correlation from other co-located soil information should also be considered, such as different soil depths and soil properties in combination with the different time points used in this study. While acidification in the highly alkaline soils that cover much of the study area is of no great concern at this time, further acidification needs to be monitored, particularly in those soil types that have a lower pH buffering capacity.

4.1. Introduction

Soil pH is a crucial component in determining crop growth and development. The optimum pH range varies for each crop species, but is broadly considered to be from 5.5 to 9.0, with indications of toxicities or deficiencies outside these bounds (Hazelton and Murphy 2007). Highly alkaline soils are particularly common in arid and semi-arid landscapes, as there is minimal leaching and high evaporation levels, resulting in a concentration of basic cations in the soil (Dregne 1976). While alkalinity may be a common problem for arid and semi-arid landscapes, the process of soil acidification is one of the largest soil degradation issues that the world faces (Oldeman 1994). Unlike some soil issues that are typically constrained to the topsoil, such as soil organic carbon loss, soil alkalinity and acidity are considered important issues throughout the whole soil profile, but particularly in the subsoil. Despite this, there are very few soil monitoring studies that focus on monitoring subsoil change of any soil property.

There are several driving factors for the acidification of soil, and these fluctuate depending on climate, land use, relief, and inherent soil properties. Soil acidification can be driven by the removal or displacement of basic cations, which is typically caused by excessive leaching in areas continuously irrigated, or removed when plant biomass is continuously harvested and not returned to the soil (Helyar *et al.* 1990). A more common cause of acidification in intensive cropping situations is through continuous application of nitrogen fertilisers, which causes an excess of H⁺ ions in soil (Helyar *et al.* 1990). Large amounts of water percolating through soil also encourages the dissolution of soil carbonates, which can hasten soil acidification (Suarez 2006). Arid and semi-arid areas often have large quantities of natural carbonates found in the soil, and several of these regions throughout the world are being, or have been used for irrigated agriculture, such as Israel, California, and the Aral Sea in Central Asia.

There has also been recent expansion of irrigated cotton production in the semi-arid regions of eastern Australia. While these areas possess typically low rainfall, they often have an abundance of groundwater and surface water available for irrigation. Irrigated cotton is considered one of the most intense agricultural crops grown in Australia, with large amounts of water applied, considerable fertiliser inputs, constant cultivation of the soil, and continual trafficking of heavy machinery. The semi-arid regions of eastern Australia also characteristically possess alkaline soils with substantial quantities of pedogenic lime (CaCO₃) and other carbonates. This combination of large nitrogen fertiliser inputs, high irrigation rates, and the presence of pedogenic carbonates suggests that there is potential for an

acidifying trend in the alkaline soils of these semi-arid regions under irrigated cotton production, but there has been minimal monitoring of soil condition in these regions (Holland 2014).

The realisation that soil is fundamental in overcoming the many global environmental challenges that the world is currently facing has led to increased efforts into monitoring soil resources (McBratney *et al.* 2014; Filippi *et al.* 2016). Soil monitoring studies often possess ‘co-located’ soil information, and this presents opportunities to maximally benefit from the available soil data. Co-located data for the purpose of this study, refers to soil information from the same spatial location, that varies in some other aspect. For example, a different depth increment of the same soil core and the same soil property, or the same soil depth increment of the same soil property taken at the same location at a different time point. Utilising the latter is particularly beneficial in soil monitoring studies, as there is often a strong correlation through time between these samples. Using multivariate linear mixed models (MLMMs) to model soil properties can take advantage of these useful spatio-temporal soil patterns, and the correlation can be utilised to strengthen predictions at non-co-located sites (Webster and Oliver 2001; Lark and Papritz 2003). In this approach, one model can be created for the multiple time points, rather than creating separate models for each time point. Creating models in this way typically allows for more precise estimates, and there is particular benefit from this approach when one soil survey possesses more sample points than the other. In addition, soil maps created using this approach are coherent and have a logical connection through time, as the spatial patterns of the maps are influenced by samples from the corresponding time point. Despite the obvious advantages, many studies that model/map multiple responses disregard the covariance between co-located samples either because they used a modelling approach that cannot incorporate this into their predictions, e.g. a machine-learning approach (Akpa *et al.* 2014), or they used univariate geostatistical methods (Ross *et al.* 2013).

Most studies that monitor soil acidification over time focus on the topsoil only, and the attention is generally on the transition from a neutral to acidic soil pH (Guo *et al.* 2010; Marchant *et al.* 2015). In addition, few studies consider formal hypothesis testing to analyse the statistical significance of any observed changes in a soil property, which can lead to misinterpretations of the magnitude of the change (Papritz and Flühler 1994). In this study, bivariate linear mixed models (BLMMs) are used to model soil pH at six depth increments to 1.5 m in 2002 and 2015, and to monitor change between these two time points in a semi-arid irrigated cotton-growing district in southern NSW, Australia. The aim of the study is to analyse the impact that different land uses and recent shifts in rainfall patterns and water

availability have on the trajectory of change in soil pH at decadal scales, and to examine the statistical significance of these changes.

4.2. Materials and methods

4.2.1. Study area and changes in climate and land use

This study uses soil data collected from a semi-arid region surrounding the township of Hillston, in south-west NSW, as described in greater detail in Chapter 3. The most widespread soils in the study area are Grey, Brown and Red Vertosols, which are found on the alluvial floodplains, and these are particularly well-suited for irrigated cotton production. The study area also possesses other soils of alluvial and aeolian origins that typically have a more sandy texture, and these are generally used for irrigated horticulture, dryland cropping and grazing. In the recent decade or so, Hillston has undergone some severely fluctuating rainfall patterns. Hillston was subjected to an extended period of drought from 2002-09 with 20% below average rainfall, forcing some cotton producers to adapt to a long period of reduced water availability. In 2010, the rainfall patterns changed abruptly, and Hillston received approximately 70% above-average rainfall through to the end of 2012, making irrigation water more available again (BOM 2017). Patterns of land use and soil management in Hillston over the last decade have changed significantly, with much of this due to the fluctuating rainfall and water availability. Historically, dryland cropping and extensive grazing were the main enterprises in Hillston, but there has been a significant expansion of irrigated cotton and horticulture since the original soil survey in 2002. This expansion of irrigated agriculture has occurred despite the lengthy period of drought, and is primarily due to increased utilisation of groundwater from aquifers.

4.2.2. Soil datasets

A soil survey was conducted in Hillston in 2002, with 115 soil cores extracted to 1.5 m (where possible) and then subsampled at six depth increments (0-0.2, 0.3-0.4, 0.55-0.65, 0.8-0.9, 1.1-1.2, 1.35-1.5 m). A soil survey of the same study area was performed in 2015, with 160 soil cores extracted. Many of the sample locations from 2002 were returned to ($n = 103$) using georeferenced coordinates, and new sites ($n = 37$) were also sampled. In addition, at 20 of the sites in the 2015 survey, a second soil core was extracted ≈ 30 -50 m away and analysed separately to investigate the extent of soil spatial variation over short distances. While soil cores to 1.5 m were again collected in the 2015 survey, the subsample depth increments were changed (0-0.1, 0.1-0.3, 0.3-0.5, 0.5-0.8, 0.8-1.2, 1.2-1.5 m) to be completely inclusive of the whole soil profile. Soil pH was analysed in 1:5 soil:water extracts using a PHM 83 Autocal™ pH meter with glass calomel electrode for 2002 samples, and using a Mettler Toledo S220 SevenCompact™

pH/Ion meter for 2015 samples. Because of the differences in sampling depths in the 2002 and 2015 surveys, equal-area quadratic smoothing splines (Bishop *et al.* 1999) were used to standardise the 2002 soil pH data to the same depths as the 2015 samples.

4.2.3. Covariates and principal component analysis

A suite of predictor variables for the study area were gathered based on the *SCORPAN* approach (McBratney *et al.* 2003), and this is described in greater detail in Chapter 3. These covariates can be broadly separated into four types; terrain attributes, gamma radiometric products, soil type, and land use information. The soil type map consisted of three classes; Grey/Brown Vertosols, Red Vertosols, and non-Vertosols. There are four classes of land use used in this study, including; cotton, dryland, natural, and horticulture. Due to the large number of numerical covariates derived from terrain attributes and gamma radiometrics ($n = 19$), a principal component analysis (PCA) was performed on these to reduce this. The principal components (PCs) that cumulatively explained 95% of the variation were kept as predictor variables, and the remaining PCs were discarded. This assists in streamlining the selection of the combination of covariates to use in the final models. To investigate the impact that each original input variable (covariate) had on the individual principal components, the component loadings were analysed.

4.2.4. Modelling and mapping

4.2.4.1. Bivariate linear mixed models (BLMMs)

In linear mixed models, the response variable is modelled as a combination of fixed and random effects. In this study the fixed effects are the linear relationship between the response variable (pH) and the covariates, and the random effects describe the spatial correlation of the residuals. In this study, we model two response variables, soil pH in 2002 and soil pH in 2015, so we are using multivariate linear mixed models (MLMMs) or bivariate linear mixed models (BLMMs). In cases where there is no trend (no fixed effects/covariates) this is analogous to co-kriging or using a linear model of co-regionalisation (LMCR) in classical geostatistics (Papritz and Flüher 1994; Bishop and Lark 2007). Most commonly, these types of models are used to model different soil properties that co-vary together in space (Orton *et al.* 2014), but in this case we wish to model the same soil property collected at two time points.

The structure of a MLMM allows different covariates to be used at each time point (each survey). It also permits the spatial auto-correlation of the residuals for each time point to be modelled, as well as the cross-correlation between the time points. Modelling of the cross-correlation through time allows for more precise estimates of pH for each time point, as well as the subsequent change in

pH (Papritz and Flühler 1994; Lark and Papritz 2003). There are several advantages of using bivariate and multivariate LMMs, as opposed to univariate approaches when data jointly co-varies, as univariate approaches ignore any correlation between variables. We used an exponential semivariogram model for the random effect terms in our models.

4.2.4.2. Model selection and map production

One model for the pH of each soil layer at both time points was created, giving a total of six models. The combination of covariates to be included in the final model was determined by backwards elimination. The initial model had separate fixed effects for each time point (i.e. all PCs, land use and soil type) and Wald tests were performed to identify redundant covariates, where those with the highest p -value (least significant) were removed. This was continued until all predictors had a p -value less than 0.05. Categorical variables were kept in the model if at least one of the levels was deemed significant. Once the model trends were determined, those numerical predictor variables (PCs) that were common to both the 2002 trend and the 2015 trend were analysed to see if the upper and lower confidence intervals around the parameter estimates overlapped. If this was the case, they were merged to have a common covariate for both surveys, and the model was then re-run to test the significance. Merging these overlapping covariates creates a more simplified model. After this, the model was used to predict onto the covariate grid of the study area (100 m resolution) using the EBLUP (Empirical Best Linear Unbiased Predictor) approach and digital soil maps were then produced.

4.2.4.3. Assessment of model quality

The quality of the BLMMs produced was tested by using leave-one-site-out cross-validation (LOSOCV). This entails fitting one model, and then validating the model by removing all data for a particular location and then predicting at that location. This is repeated at all locations. The performance of the prediction variances and uncertainty was assessed through the mean and median of the standardized squared prediction errors (SSPE). When prediction variance matches the actual errors the mean would be equal to 1.0 and the median equal to 0.455 (Lark 2000; Orton *et al.* 2014). Lin's concordance correlation coefficient (LCCC), was used as an assessment of model quality as it is the fit of the observed and predicted values to the 1:1 line, and is unit-less, making it useful for comparing between models (Lin 1989). In addition, the RMSE (root mean square error) was used to analyse the quality of predictions, and the mean of the residuals was used to analyse the bias (over- or under-fitting) of model predictions.

4.2.4.4. Change calculation and statistical significance of change

It is important that a prediction interval is placed around predicted change, as it subsequently allows a hypothesis test to be performed. Despite the importance of this, there are very few studies where this is actually implemented. The hypotheses we are testing are:

$$H_0: \mu_d = 0; \quad 4.1)$$

$$H_1: \mu_d \neq 0;$$

where $\Delta\hat{y}$ is some estimate of the change (Equation 3.2), such as a mean for a region or a prediction at a specific location:

$$\Delta\hat{y} = \hat{y}_2 - \hat{y}_1, \quad 4.2)$$

where \hat{y}_1 is the prediction of the soil property for the first point in time and \hat{y}_2 at the second time point. To test this hypothesis, we need an estimate of the variance of the change, which can be expressed as:

$$\hat{V}(\Delta\hat{y}) = \hat{V}(\hat{y})_1 + \hat{V}(\hat{y})_2 - 2(\hat{V}(\hat{y})_{1,2}), \quad 4.3)$$

where $\hat{V}(\Delta\hat{y})$ is the variance of the change (contrast variance) in the soil property between two points in time, $\hat{V}(\hat{y})_1$ is the variance of the soil property for the observations at the first point in time, $\hat{V}(\hat{y})_2$ at the second point in time, and $\hat{V}(\hat{y})_{1,2}$ is the covariance of soil properties at the two time periods (the covariance term). Equation 4.3 presents both pitfalls and opportunities in monitoring. The pitfalls are the spatial variation of the soil property and the magnitude of the change, while the opportunity is the covariance between sampling campaigns (Filippi *et al.* 2016). When the covariance term is included and there is a strong relationship between the two surveys, a smaller contrast variance results, subsequently allowing greater sensitivity in predicting a statistically significant change. If we did not use BLMMs we would not be able to estimate the covariance term ($\hat{V}(\hat{y})_{1,2}$), and would have a conservative test. Many studies ignore the correlation at co-located sites and disregard the covariance term and contrast variance, and very few consider the statistical significance of the temporal change in soil properties (Ross *et al.* 2013). The statistical significance of change can be analysed using the z-statistic:

$$z = \frac{\hat{y}_2 - \hat{y}_1}{\sqrt{\hat{V}(\Delta\hat{y})}}, \quad 4.4)$$

where z is the z-statistic score, \hat{y}_1 is the value of the prediction of the soil property at time point one, \hat{y}_2 at the second time point, and $\hat{V}(\Delta\hat{y})$ is the contrast variance, which is obtained using Equation 4.3. For a two-tailed test, a z-statistic score greater than (in a positive case) or less than (for a negative case)

± 1.28 , ± 1.65 , ± 1.96 , and ± 2.58 represents a prediction interval of 80%, 90%, 95%, and 99% respectively. The BLMM approach, where the same soil property is measured at the same location at two time points, has the limitation that it is often difficult to return to the exact site, even when spatial coordinates are available. Nevertheless, the accuracy of GPS devices has substantially improved, allowing this limitation to be largely overcome.

4.3. Results

4.3.1. Soil pH summary statistics

The soils of the Hillston cotton growing district are neutral to strongly alkaline, and pH generally rises with increasing depth, with some pH values exceeding 10 in the subsoil (Table 4.1). There has been an overall acidification trend in both the topsoil and subsoil in the 13-year period between the two surveys. All land uses experienced an acidifying trend over time at all sampling depths, apart from the dryland land use where there was an increase in soil pH at the 0.3-0.5, 0.5-0.8, and 0.8-1.2 m depths (Table 4.1).

Table 4.1 – Mean soil pH for all sites at different sampling depths, time points and land uses

Statistic	Depth	Cotton		Dryland		Natural		Horticulture	
		2002	2015	2002	2015	2002	2015	2002	2015
Mean	0-0.1 m	7.91	7.70	7.21	6.93	7.26	6.88	7.93	7.48
	0.1-0.3 m	8.25	7.99	7.62	7.85	7.81	7.65	8.01	7.68
	0.3-0.5 m	8.78	8.58	8.20	8.65	8.56	8.24	8.35	8.08
	0.5-0.8 m	9.06	8.79	8.66	8.73	8.86	8.29	8.88	8.10
	0.8-1.2 m	8.99	8.77	8.89	8.93	8.80	8.42	9.01	8.49
	1.2-1.5 m	8.73	8.56	9.00	8.84	9.40	8.27	8.75	8.38
Minimum	0-0.1 m	6.67	6.68	6.33	5.84	6.08	5.36	7.71	5.50
	0.1-0.3 m	7.07	6.68	6.35	6.35	6.48	5.06	7.57	5.44
	0.3-0.5 m	7.38	7.40	6.02	7.32	6.67	5.02	7.83	6.23
	0.5-0.8 m	7.38	7.22	7.06	7.78	7.28	5.31	8.60	7.07
	0.8-1.2 m	7.78	7.53	7.70	7.75	7.16	5.77	8.56	7.75
	1.2-1.5 m	7.48	7.66	8.28	7.42	9.12	6.63	8.75	7.36
Maximum	0-0.1 m	9.08	8.78	8.97	8.69	8.85	8.28	8.32	9.03
	0.1-0.3 m	9.51	9.05	9.03	9.31	9.16	9.19	8.32	9.28
	0.3-0.5 m	9.74	9.43	9.56	9.60	9.69	9.40	8.76	9.50
	0.5-0.8 m	9.71	9.72	9.80	9.42	10.00	9.58	9.29	9.19
	0.8-1.2 m	9.83	9.70	9.81	9.57	10.05	9.53	9.49	9.12
	1.2-1.5 m	9.86	9.34	9.86	9.63	9.64	9.62	8.75	9.08
Standard deviation	0-0.1 m	0.587	0.486	0.817	0.868	0.744	0.672	0.336	0.928
	0.1-0.3 m	0.550	0.555	0.838	0.896	0.647	0.942	0.386	1.058
	0.3-0.5 m	0.604	0.448	1.088	0.582	0.699	0.970	0.473	0.896
	0.5-0.8 m	0.486	0.494	0.919	0.394	0.547	0.829	0.364	0.701
	0.8-1.2 m	0.437	0.477	0.773	0.418	0.641	0.647	0.463	0.417
	1.2-1.5 m	0.546	0.511	0.677	0.566	0.261	0.579	NA	0.574
<i>n</i>	0-0.1 m	60	65	14	21	38	61	3	13
	0.1-0.3 m	60	62	14	20	38	61	3	12
	0.3-0.5 m	60	62	14	20	37	61	3	12
	0.5-0.8 m	60	62	12	19	33	60	3	12
	0.8-1.2 m	60	56	11	19	23	60	3	12
	1.2-1.5 m	53	56	4	17	3	52	1	12

While Table 4.1 shows the pH statistics for all sites measured in both surveys, the profile plots in Fig. 4.1 are produced from co-located sites only – i.e. those sites that have been sampled in both 2002 and 2015. This provides us with more precise information regarding the change in soil pH at particular locations and within different land uses and geological units, and eliminates soil information for those sites only sampled at one time point, which would bias results. There were very few soil samples for natural sites in the 1.2-1.5 m depth interval in 2002; hence the plots only extend to 1.2 m. It can be seen that different geological units and land uses have experienced different patterns of soil pH change (Fig.

4.1). Non-Vertisols experienced the largest decrease in pH in the topsoil, followed by Red Vertisols and then Grey/Brown Vertisols. This also reflects what would be expected to be the soils with the lowest, to highest pH buffering capacity, respectively. In contrast, the subsoils of both Red Vertisols and Grey/Brown Vertisols had experienced the largest decrease, with non-Vertisols displaying the smallest decrease in pH. The profile plot of paired-sites by land use shows that while sites under cotton had experienced a consistent decrease at paired-sites throughout the profile, there was a less uniform pattern at sites under the natural land use. At natural sites no change was observed in the 0.1-0.3 m depth, but a considerable mean decrease of 0.5 pH units was observed in the 0.5-0.8 m depth. All other layers in the plot had a comparable decrease in pH to cotton sites (Fig. 4.1).

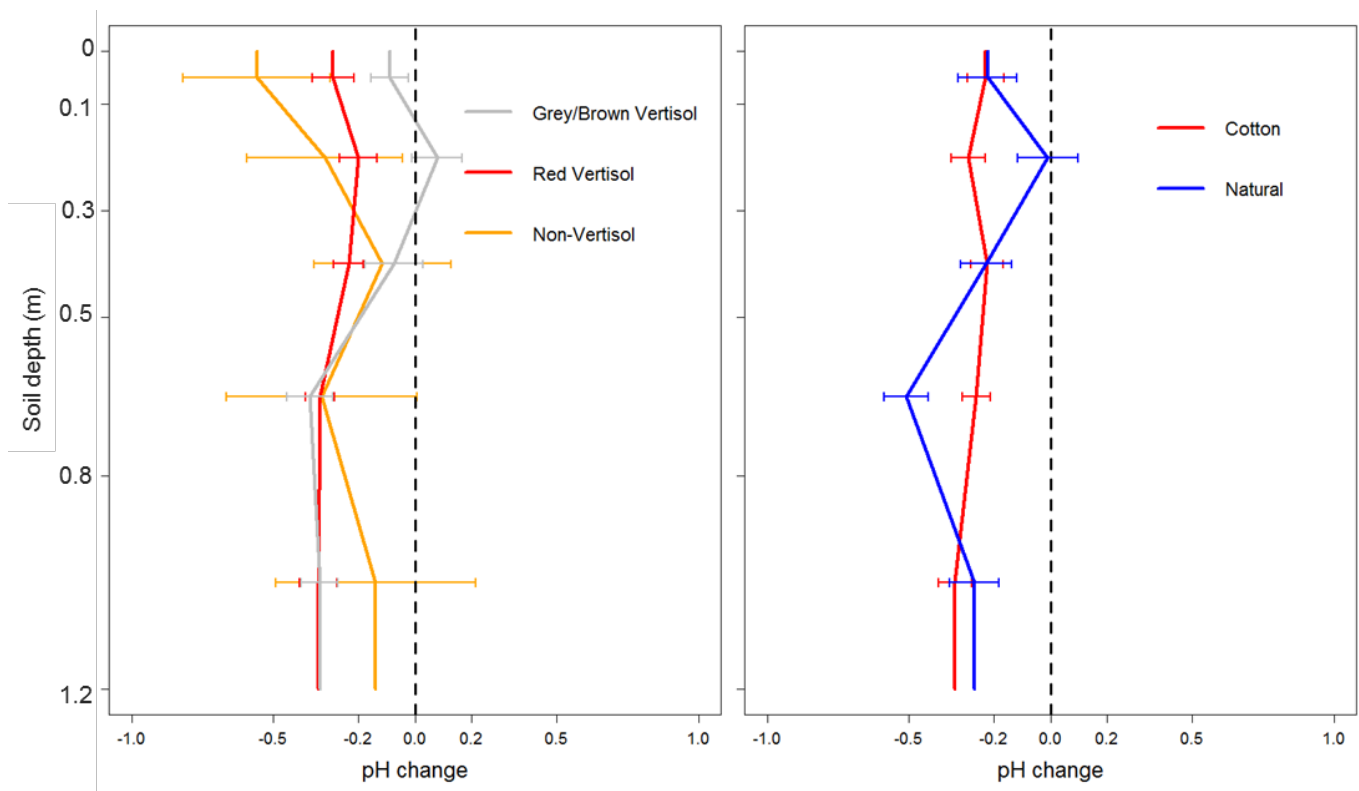


Fig. 4.1 – Mean change in soil pH (1:5 soil:water) from 2002 to 2015 of co-located sites based on different soil types (left) and land use (right) to the depth of 1.2 metres (with standard error bars)

At co-located sites, soil pH measured in 2002 was highly correlated with soil pH measured in 2015 for all six soil layers (Fig. 4.2). The clear positive correlation suggests that a statistical modelling approach that makes use of the joint variability of these two variables, such as the BLMM, could be advantageous (Orton *et al.* 2014). This is particularly beneficial when one time point has more soil

samples than the other as, for example, in the 0.8-1.2 m depth layer, where the 2002 survey has 97 samples and the 2015 survey has 147 samples (Table 4.1).

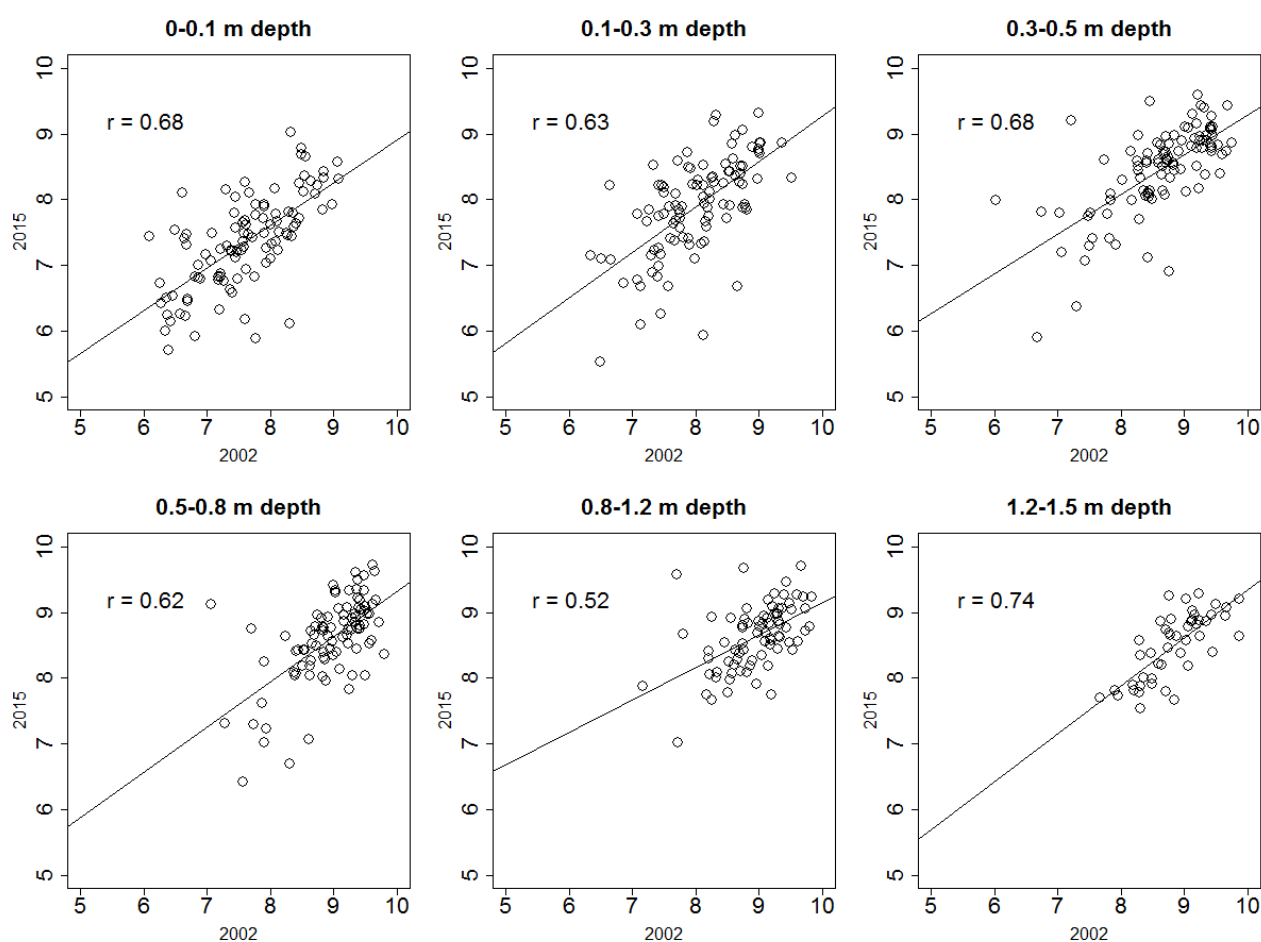


Fig. 4.2 – Correlation of 2002 with 2015 soil pH at co-located samples in the six sampling depths, with the line of best fit and r values (Pearson correlation) included

4.3.2. Principal component analysis (PCA) of numerical predictor variables

The first seven principal components (PCs) explained up to 95% of the variation, and the remaining twelve PCs were discarded (Table 4.2). The impact that each original input variable had on the individual principal components (component loadings) were investigated, with those covariates that contributed the most possessing values furthest from 0 (Table 4.3). As an example, PC1 is largely driven by radiometric data.

Table 4.2 – The proportion of variance that each principal component explains, the cumulative variance, and the standard deviation from the principal component analysis (PCA) of the 19 numerical covariates

Principal component	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Standard deviation	2.53	1.99	1.50	1.37	1.25	1.09	0.80
Proportion of variance	0.34	0.21	0.12	0.10	0.08	0.06	0.04
Cumulative proportion	0.34	0.55	0.66	0.76	0.85	0.91	0.95

Table 4.3 – Results from the principal component analysis (PCA) showing the proportion (loading) that each input variable contributes to the first seven principal components

Covariate	PC1	PC2	PC3	PC4	PC5	PC6	PC7
1. Elevation	-0.25	0.21	0.08	0.08	-0.26	0.24	0.29
2. Slope	-0.19	0.29	-0.29	0.13	0.10	0.04	-0.09
3. Length-slope factor ¹	-0.17	0.32	-0.28	0.24	0.16	0.02	-0.28
4. Length-slope factor ²	-0.16	0.31	-0.28	0.25	0.17	0.02	-0.30
5. MRRTF	0.05	-0.21	-0.14	-0.26	0.42	-0.43	0.02
6. MRVBF	0.25	-0.21	0.00	-0.08	0.11	-0.08	-0.65
7. Topographic wetness index	0.08	-0.13	0.31	0.54	0.27	-0.03	0.08
8. Wetness index ¹	0.15	-0.08	0.26	0.06	-0.40	0.35	-0.50
9. Wetness index ²	0.08	-0.12	0.31	0.54	0.27	-0.02	0.08
10. Weathering index	-0.31	0.00	0.26	-0.04	-0.12	-0.27	-0.10
11. Total dose	0.36	0.12	-0.14	0.10	-0.14	-0.10	0.10
12. Dose	0.36	0.11	-0.15	0.10	-0.12	-0.08	0.09
13. K	0.34	-0.05	-0.27	0.09	-0.01	0.14	0.13
14. Th	0.33	0.11	-0.14	0.16	-0.25	-0.24	0.05
15. U	0.31	0.29	0.09	-0.08	0.07	0.02	0.05
16. Th/K	-0.04	0.26	0.19	0.10	-0.33	-0.60	-0.08
17. U ² /Th	0.24	0.32	0.18	-0.16	0.18	0.14	0.03
18. U/K	0.05	0.39	0.33	-0.15	0.08	-0.17	-0.07
19. U/Th	0.11	0.31	0.28	-0.26	0.33	0.22	0.00

¹Derived from individual calculation in SAGA GIS; ²Derived from basic terrain analysis calculation in SAGA GIS

4.3.3. Bivariate linear mixed models

The soil pH data and the residuals were normally distributed in all instances, so no transformation was required prior to model development. The covariates included in the different models, and the parameters for the fixed-effect models are shown (Table 4.4). While the covariates included in each model varied, there were some predictor variables that appeared multiple times. Land use and soil type, as well as PC2, PC6, and PC7 were common to many of the final models (Table 4.4). It was common for the same principal components to be included in both the 2002 and 2015 trend for each layer/model. The upper and lower prediction intervals for these common PCs overlapped in all situations, and thus they were merged. The significance in the model was then checked and in all cases they were statistically significant ($p < 0.05$), so they were retained in the model.

Table 4.4 – Covariates (predictor variables) included in the models and the coefficient matrix (estimated fixed effects) for the six soil layers

Covariate		0-0.1 m layer	0.1-0.3 m layer	0.3-0.5 m layer	0.5-0.8 m layer	0.8-1.2 m layer	1.2-1.5 m layer	
2002	Intercept ¹	7.87	8.30	8.83	8.99	8.88	8.67	
	Land use	Dryland	-0.28	-0.17	0.06			0.69
		Natural	-0.74	-0.47	-0.14			0.23
		Horticulture	-0.33	-0.60	-0.58			0.12
	Soil type	Non-Vertisol	-0.10	-0.53	-0.71	-0.64		0.40
		Red Vertisol	0.21	0.10	0.06	0.11		0.14
	Principal components	PC1					0.06	
		PC2						
		PC3			0.08			
		PC4						
		PC5			-0.12	-0.12	-0.10	
		PC6						
		PC7			-0.16	-0.14		0.17
2015	Intercept ¹	7.83	8.26	8.75	8.74	8.72	8.63	
	Land use	Dryland	-0.36	0.53	0.65	0.36	0.22	0.15
		Natural	-0.64	-0.20	-0.18	-0.23	-0.21	-0.27
		Horticulture	-0.21	-0.27	-0.24	-0.34	-0.40	-0.37
	Soil type	Non-Vertisol	-0.71	-1.17	-0.86	-0.60		
		Red Vertisol	-0.13	-0.16	-0.09	0.06		
	Principal components	PC1						-0.06
		PC2	0.08					
		PC3						
		PC4						
		PC5						
		PC6					0.10	
		PC7						
Merged covariates ²	Principal components	PC1			0.05			
		PC2		0.08	0.14	0.11	0.11	0.07
		PC3						
		PC4						
		PC5		-0.10				
		PC6	0.07	0.10		0.09		0.14
		PC7	-0.17	-0.16			-0.09	

¹The intercept represents the cotton land use and Grey/Brown Vertisol, as this is displayed alphabetically in the model output

²Principal components common to both the 2002 and 2015 were merged if the upper and lower 95% confidence interval around each parameter estimate overlapped

The residuals are the difference between the predicted and observed values based on the fixed effect terms of the BLMMs. The random effects relate to the semivariogram models fitted to the residuals. The nugget in 2002 ($c_{0,1,1}$) is generally larger than the nugget in 2015 ($c_{0,2,2}$), and the cross-correlation nugget is smaller than both of these in all models (Table 4.5). The smallest estimated distance parameter was in the 1.2-1.5 m sampling depth BLMM (216 m), while the largest was in the

0.3-0.5 m sampling depth BLMM (26666 m). This shows that the 1.2-1.5 m sampling depth BLMM is the noisiest or hardest to model based on the available covariates, and is spatially correlated over short distances (Table 4.5).

Table 4.5 – Summary of the random effects from the pH BLMMs at different sampling depths – the auto-correlation nugget ($c_{0,1,1}$ and $c_{0,2,2}$), the cross-correlation nugget ($c_{0,1,2}$), the auto-correlation structural semivariance ($c_{1,1,1}$ and $c_{1,2,2}$), the cross-correlation structural semivariance ($c_{1,1,2}$), and the range (ϕ , or ϕ). The range is presented in metres, and is multiplied by 3. The subscript 1 refers to the 2002 survey, and the subscript of 2 refers to the 2015 survey.

Random effects	0-0.1 m	0.1-0.3 m	0.3-0.5 m	0.5-0.8 m	0.8-1.2 m	1.2-1.5 m
$c_{0,1,1}$	0.25	0.21	0.31	0.00	0.05	0.00
$c_{0,2,2}$	0.13	0.17	0.25	0.05	0.16	0.04
$c_{0,1,2}$	0.03	0.05	0.15	0.00	0.09	0.00
$c_{1,1,1}$	0.17	0.13	0.77	0.29	0.22	0.24
$c_{1,2,2}$	0.21	0.30	1.08	0.30	0.07	0.22
$c_{1,1,2}$	0.19	0.20	0.91	0.20	0.02	0.19
ϕ (m)*	973	1336	26666	656	962	216

*Effective range for exponential model is three times the distance parameter

The final models were quite consistent in their quality for each soil layer and all models had low levels of bias (Table 4.6). Most fitted models displayed standardised squared prediction errors (SSPE) that were close to the expected values of a mean of 1.0 and a median of 0.455, however, the 0.3-0.5 sampling depth had a median 0.29, which suggests the presence of outliers (Table 4.6). Values for Lin's concordance correlation coefficient (LCCC) showed that there was generally a good fit for all models, with the topsoil being the best predicted. The root mean squared error (RMSE) also revealed low values which ranged from 0.50-0.67 pH units (Table 4.6).

Table 4.6 – Quality statistics of the models for the six soil layers based on leave one out cross-validation

Soil layer	LCCC	RMSE	Bias	Mean (SSPE)	Median (SSPE)
0-0.1 m	0.57	0.61	0.00	1.03	0.44
0.1-0.3 m	0.47	0.65	0.00	1.05	0.41
0.3-0.5 m	0.44	0.67	-0.01	1.02	0.29
0.5-0.8 m	0.55	0.53	-0.01	1.01	0.32
0.8-1.2 m	0.44	0.50	0.02	1.02	0.45
1.2-1.5 m	0.48	0.50	0.00	1.00	0.53

4.3.4. Soil pH maps and modelled data

The 0-0.1 m depth maps reveal the lowest soil pH values for the whole profile on average. The impact of land use is clear, with cotton farms typically possessing higher pH values. The change map showed a predicted decrease of greater than 0.5 pH units for modest parts of the study area from 2002 to 2015, particularly the eastern side where the sandier soils are situated (Fig. 4.3). In the 0.1-0.3 m depth maps, soil pH values are generally higher than those of the surface soil, and the extent of modelled change in pH is not as great (Fig. 4.3). The change map showed that while small areas underwent decreases in pH, there were also small pockets where there was a predicted increase (Fig. 4.3).

It is clear that the subsoil pH is substantially higher than the upper layers (Fig. 4.4). In the 0.3-0.5 m depth maps there is an obvious correlation with soil type and pH, and the areas at higher elevation possess the lowest pH values, with the alluvial floodplains possessing the highest (Fig. 4.4). The change map shows that several different pockets of the study area were predicted to have experienced acidification of 0.5-1 pH units, but there were also some areas where there was a modelled increase in pH. The 0.5-0.8 m depth maps revealed similar spatial patterns, however, the change map showed more widespread acidification trends of 0.5-1 pH units.

The pH maps of the deepest layers (0.8-1.5 m) reveal some very high pH levels (Fig. 4.5). The 0.8-1.2 m maps show quite uniform pH values throughout the study area, with the rocky outcrops again showing the lowest pH levels. The change map in this layer displayed quite widespread acidification of 0.5-1 pH units over the 13-year period, particularly on the alluvial floodplains. Again, there was a predicted increase in pH in those areas of higher elevation from 2002 to 2015 (Fig. 4.5). In the deepest sampling depth (1.2-1.5 m) there are very few samples in the eastern half of the study area in 2002, and the maps must be viewed with this in mind. Widespread acidification is evident in the change maps across much of the study area, with some areas predicted to have experienced a decrease of 1-2 pH units. Interestingly, the 1.2-1.5 m depth maps show that soil pH is higher in the sandier soils than those situated on the alluvial floodplains, but in contrast, the maps from the initial four layers to 0.8 m showed that these soils generally had a lower soil pH than the soils on the floodplains (Fig. 4.5).

The rate of pH change from model predictions across the study area for different land uses ranged from -0.07 pH unit yr^{-1} to 0.03 pH unit yr^{-1} (Table 4.7). The dryland land use had the most variable rates of change throughout the profile, while the cotton land use had the most consistent acidification trends.

Table 4.7 – Rate of change on a per year basis by land use from modelled maps of whole study area during the 2002-2015 period (pH unit yr⁻¹)

Soil depth	Cotton	Dryland	Natural	Horticulture
0-0.1 m	-0.014	-0.046	-0.019	-0.019
0.1-0.3 m	-0.019	-0.010	-0.003	-0.010
0.3-0.5 m	-0.015	+0.033	-0.018	-0.000
0.5-0.8 m	-0.018	+0.013	-0.037	-0.053
0.8-1.2 m	-0.019	+0.013	-0.029	-0.044
1.2-1.5 m	-0.019	-0.065	-0.056	-0.051

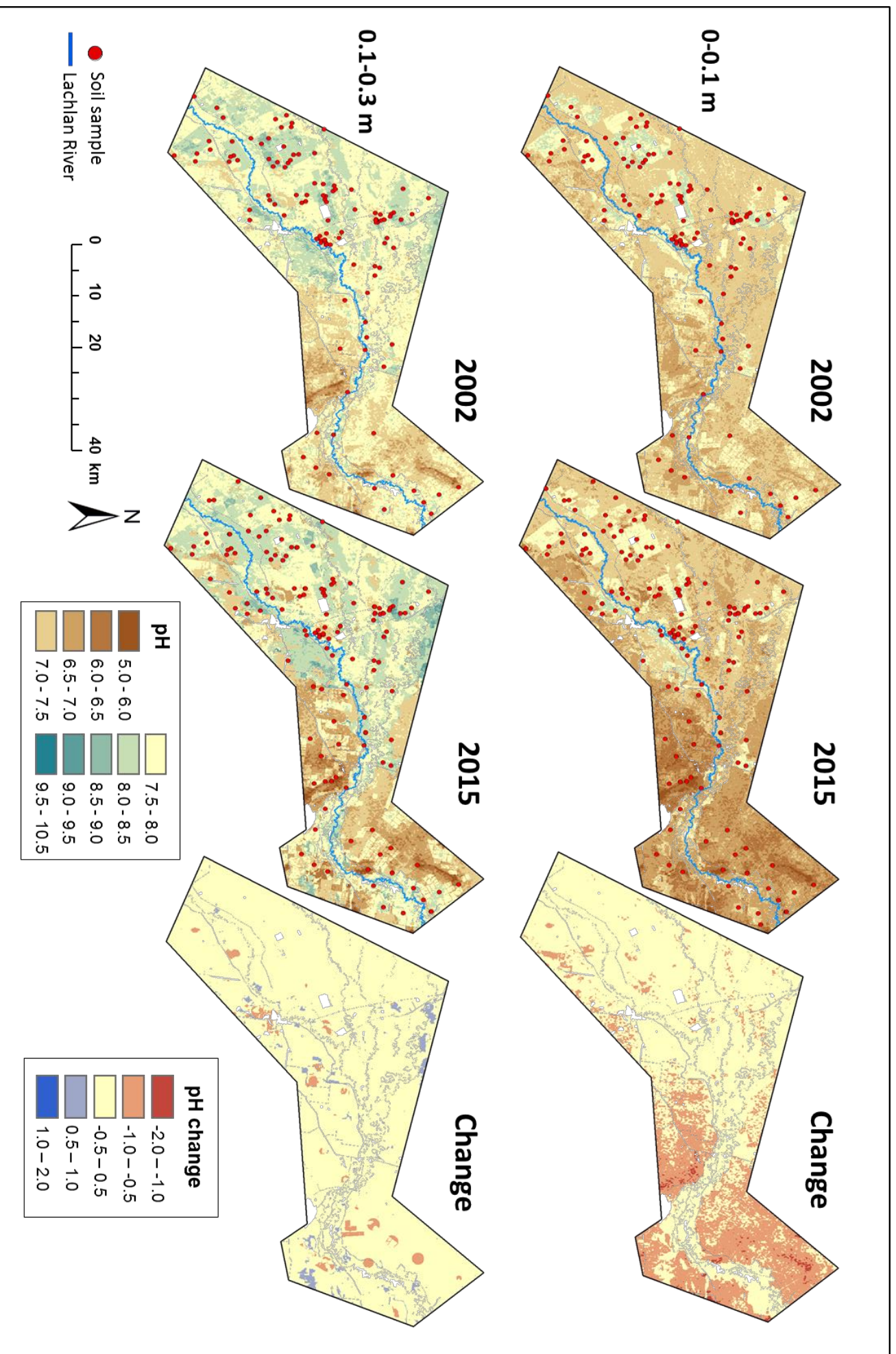


Fig. 4.3 – Maps of soil pH values (1:5 soil: water) for the 0-0.1 and 0.1-0.3 m depths in 2002 and 2015 overlain with sample locations, and the change in soil pH in this time period (2015-2002)

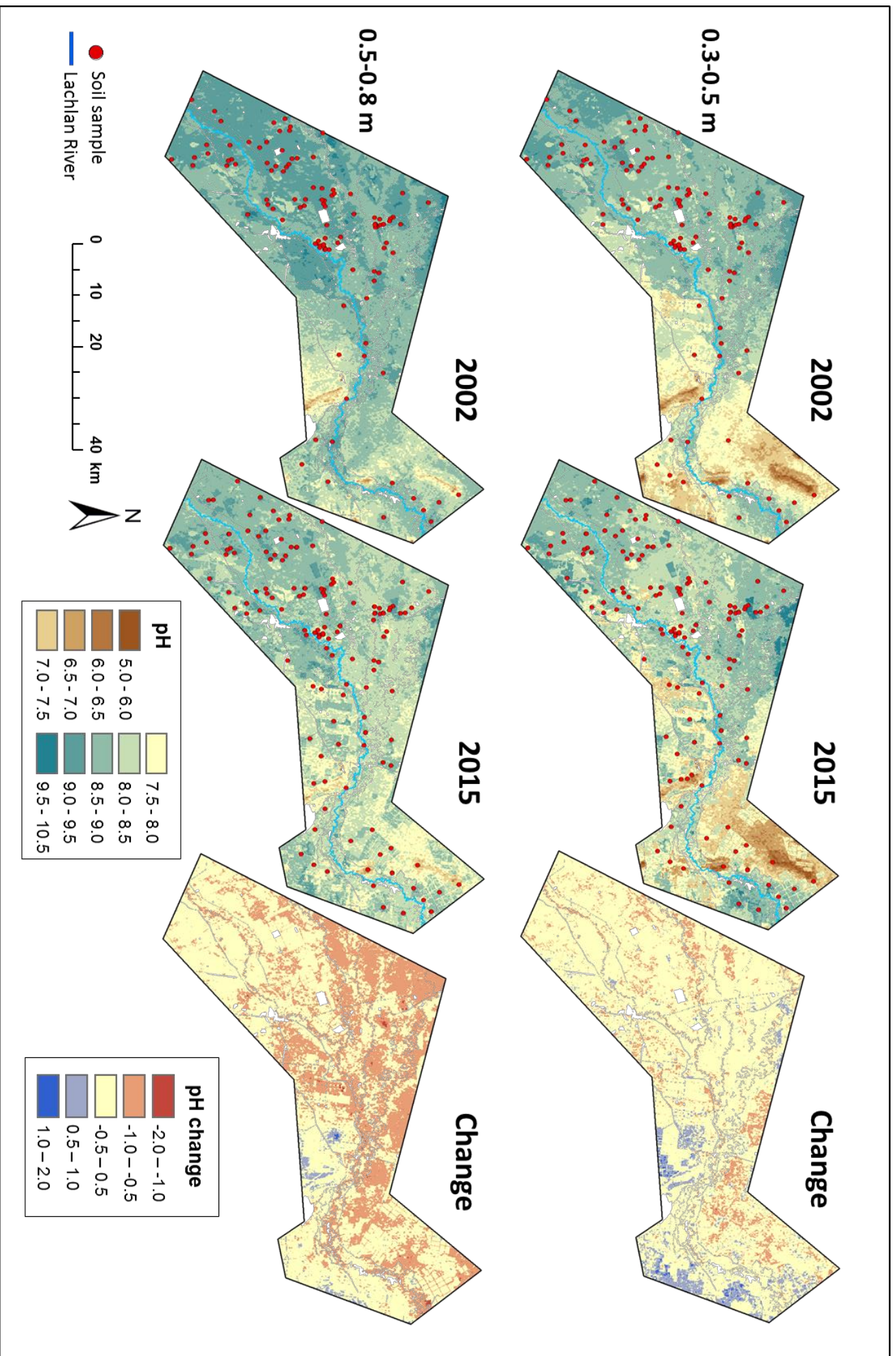


Fig. 4.4 – Maps of soil pH values (1:5 soil: water) for the 0.3-0.5 and 0.5-0.8 m depths in 2002 and 2015 overlain with sample locations, and the change in soil pH in this time period (2015-2002)

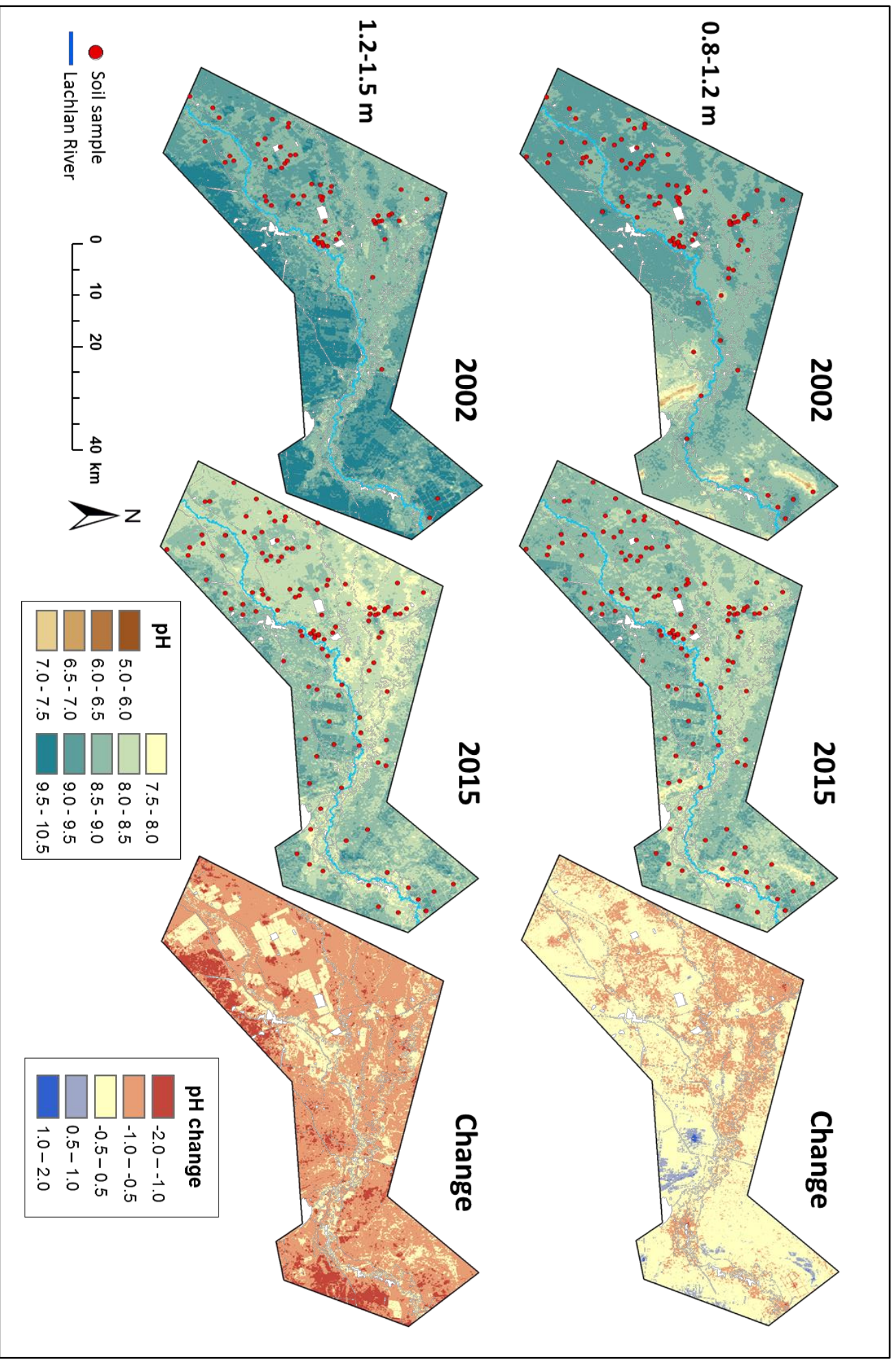


Fig. 4.5 – Maps of soil pH values (1:5 soil: water) for the 0.8-1.2 and 1.2-1.5 m depths in 2002 and 2015 overlain with sample locations, and the change in soil pH in this time period (2015-2002)

4.3.5. Statistical significance of change in soil pH

The maps of the z-scores for the 0-0.1 m depth increment show that there are some very small areas that experienced a significant decrease in soil pH (p -value <0.20) (Fig. 4.6). For the 0.1-0.3 and 0.3-0.5 m depths, it is clear that almost no statistically significant change in pH was detected. The 0.5-0.8 and 0.8-1.2 m z-statistic maps are quite similar and reveal some areas where a statistically significant decrease (p -value <0.20) in soil pH was predicted, as well as an isolated area with an increase in pH. The final layer of 1.2-1.5 m shows the most widespread and statistically significant decreases in pH spread throughout the study area with reasonably sized areas predicted to have experienced acidification (p -value <0.10). Different land uses underwent varying severities and extents of statistically significant soil acidification (Table 4.8). The most extensive acidification occurred in the lower horizons (0.5-1.5 m) and was particularly widespread in soils under the natural land use, as well as horticulture (Table 4.8). All of these results show that overall, only isolated areas of the study area had experienced a statistically significant decrease in soil pH, apart from the deepest layer where it was quite extensive. It is clear that despite some of the predicted change maps (Fig. 4.3, 4.4, 4.5) depicting widespread decreases in soil pH, this does not always link with the statistical significance of the change.

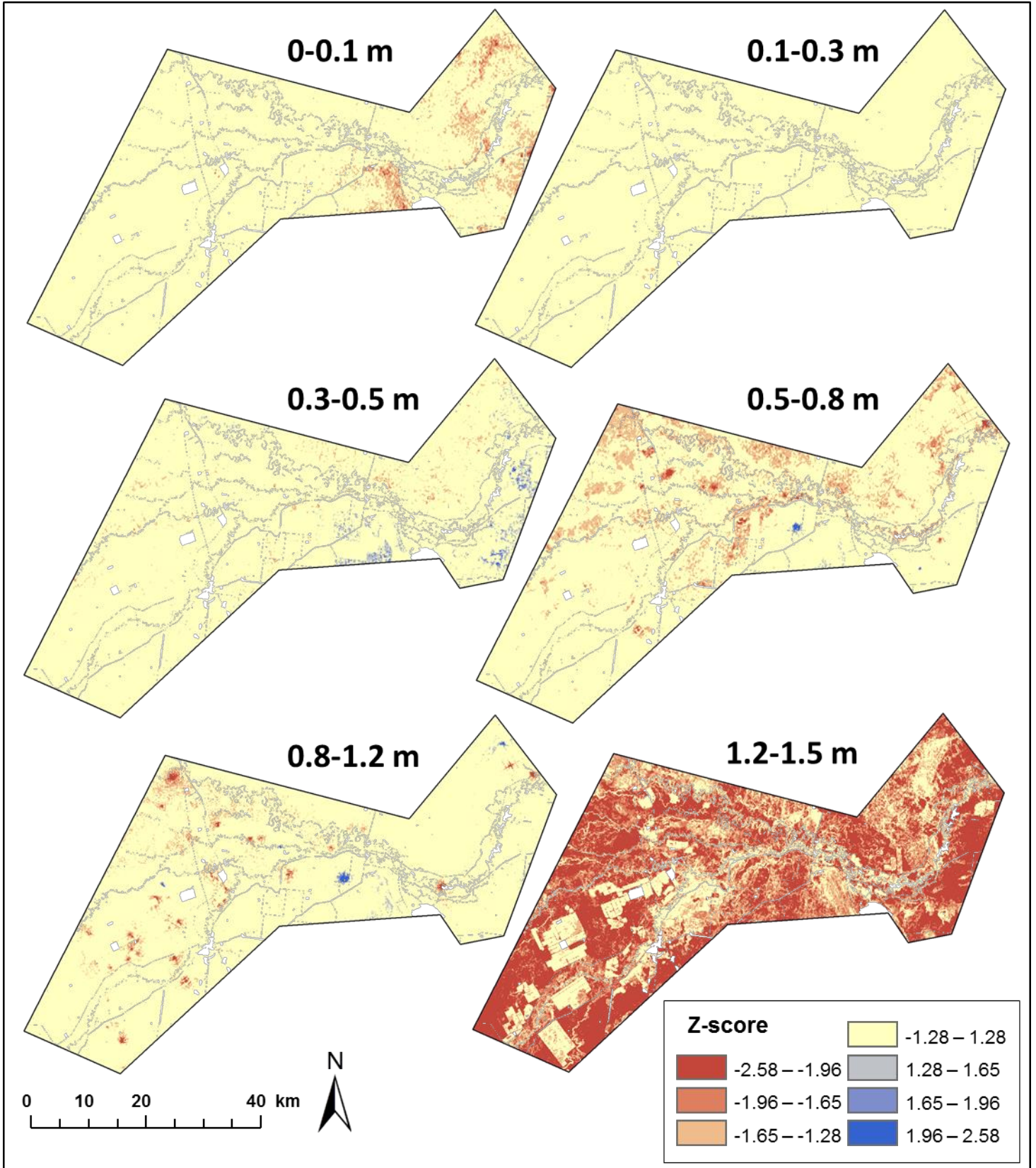


Fig. 4.6 – Maps of z-scores showing the statistical significance of the change in soil pH for each sampling depth

Table 4.8 – Area extent of predicted statistically significant decreases in soil pH by landuse across the study area at different levels of significance

Significance level		P < 0.20			P < 0.10			P < 0.05			P < 0.01		
z-score value		-1.28			-1.65			-1.96			-2.58		
Soil layer	Land use	¹ ha	² Total %	³ LU %	¹ ha	² Total %	³ LU %	¹ ha	² Total %	³ LU %	¹ ha	² Total %	³ LU %
0-0.1 m	Cotton	-	-	-	-	-	-	-	-	-	-	-	-
	Dryland	784	<1	2.0	7	<0.01	<0.1	-	-	-	-	-	-
	Natural	1289	<1	<1	52	<0.1	<0.1	1	<0.01	<0.01	-	-	-
0.1-0.3 m	Horticulture	-	-	-	-	-	-	-	-	-	-	-	-
	Cotton	-	-	-	-	-	-	-	-	-	-	-	-
	Dryland	-	-	-	-	-	-	-	-	-	-	-	-
0.3-0.5 m	Natural	-	-	-	-	-	-	-	-	-	-	-	-
	Horticulture	-	-	-	-	-	-	-	-	-	-	-	-
	Cotton	-	-	-	-	-	-	-	-	-	-	-	-
0.5-0.8 m	Dryland	32	<0.1	<1	4	<0.01	<0.1	1	<0.01	<0.01	-	-	-
	Natural	17	<0.01	<0.1	8	<0.01	<0.1	3	<0.01	<0.01	-	-	-
	Horticulture	2502	<1	1.4	267	<1	<1	80	<0.1	<0.1	16	<0.01	<0.01
0.8-1.2 m	Cotton	1022	<1	14.3	151	<0.1	2.1	36	<0.1	<1	9	<0.01	<1
	Dryland	133	<0.1	<1	7	<0.01	<0.1	1	<0.01	<0.01	-	-	-
	Natural	78	<0.1	<1	14	<0.01	<0.1	4	<0.01	<0.1	-	-	-
1.2-1.5 m	Horticulture	1414	<1	<1	220	<0.1	<1	57	<0.1	<0.1	5	<0.01	<0.01
	Cotton	686	<1	9.6	129	<0.1	1.8	40	<0.1	<1	2	<0.01	<0.1
	Dryland	27	<0.1	<0.1	1	<0.01	<0.01	1	<0.01	<0.01	-	-	-
1.2-1.5 m	Natural	22428	8.8	56.8	8418	3.3	21.3	748	<1	1.9	1	<0.01	<0.01
	Horticulture	80276	31.7	45.1	17516	6.9	9.8	2537	1.0	1.4	72	<0.1	<0.1
	Cotton	2090	<1	29.2	125	<0.1	1.7	1	<0.01	<0.1	-	-	-

¹Number of hectares, ²percentage of total study area and, ³percentage of total land use area predicted to have experienced a statistically significant decrease in soil pH

4.4. Discussion

4.4.1. Soil pH, acidification, and driving factors

The optimum soil pH range for cotton is 5.5 to 7 (Hazelton and Murphy 2007), and this study revealed some very high pH levels in the subsoil, suggesting that there are potential limitations to crop productivity in the Hillston district. Roots of cotton crops typically go beyond 1 metre, but an unsuitable pH can inhibit vertical root growth, which results in a smaller volume of soil that is accessible by the crop, and consequently a smaller amount of water and nutrients that can be utilised. The monitoring and modelling of soil pH in this study indicates that there has been an overall acidification trend at all sampling depths to various degrees in the Hillston district since the original 2002 survey. An acidification trend in soil is commonly regarded as undesirable (Helyar *et al.* 1988), but in this case the naturally very high soil pH, particularly in the subsoil, makes acidification less of an issue. However, if acidification was

to continue there is a potential in the distant future for undesirable effects, such as the inhibition of nutrient uptake by crops. This is even more crucial for cotton-growing regions, as cotton is particularly sensitive to soil acidity.

The observed soil acidification trend in this study is presumed to be at least partly due to the intensive management practices of irrigated cotton production. Cotton production in Australia is particularly renowned for high, and often excessive applications of N fertiliser to the soil (Rochester 2011), and this could have contributed to the soil acidification observed in our study. A study that observed widespread decreases in topsoil pH (-0.015 to -0.04 pH unit yr^{-1}) across China from 1980 to 2010, primarily attributed this acidification to the increased use of nitrogen fertilisers in agricultural soils (Guo *et al.* 2010). Additionally, the large amounts of irrigation water percolating through the profile in soils under cotton production in this study is also likely to have leached basic cations out of the upper soil profile and hastened the dissolution of soil inorganic carbonates, causing the soil pH to decrease (Jenny 1980). The irrigation requirement for cotton in the southern growing-regions of Australia ranges from 10-14 ML/ha a season (equivalent to 1000-1400 mm of rainfall), and is about triple to quadruple the average annual rainfall of Hillston. Soils of the Hillston district are known to possess an abundance of naturally-occurring lime (CaCO_3), particularly in the subsoil (Cay and Cattle 2005). The dissolution of carbonates is therefore a probable cause of the large decreases in soil pH in the 1.2-1.5 m depth. This layer experienced the most widespread and statistically significant pH decrease of all sampling depths.

It should be noted that very few studies map soil at depths below 1 m, and very few have considered monitoring subsoil pH. Despite this, topsoil pH has been monitored at decadal scales by various approaches: at the regional level using a database of soil test results requested by farmers (Marchant *et al.* 2015), and at a national scale using data from two soil surveys (Guo *et al.* 2010). A more comparable study to ours that was situated in the Namoi, Macintyre, and Gwydir cotton-growing valleys of northern NSW by Singh *et al.* (2003) observed similar trends of soil acidification. These areas are physiographically similar to the lower Lachlan valley, with extensive floodplains supporting mainly Vertisol soil types which are commonly alkaline and contain pedogenic lime in the subsoil. The mean annual rainfall of the northern cotton-growing valleys (~ 600 - 650 mm) is almost double that of Hillston (367 mm) (BOM 2017), and hence the soil is more accustomed to higher water percolation. Rather than surveying soil pH at two time points, the authors measured the pH and pH buffering capacity of the topsoil (0-0.1 m) to calculate acidification rates into the future, with estimations that it would take between 10 and 417 years (-0.002 to -0.1 pH unit yr^{-1}) for soil pH to decrease by 1 unit (Singh *et al.*

2003). These acidification rates are similar to those found in the topsoil (0-0.1 m) of our study (-0.01 to -0.05 pH unit yr⁻¹). The authors attributed the acidification to being mainly a function of land use, with larger acid inputs assumed for cultivated fields compared with un-cultivated fields. However, the authors also acknowledged the potential contribution of soil carbonate dissolution to this acidification, with this assumed to be due to the increased water percolation from irrigation. While most monitoring studies on soil pH focus on the transition of acidic-neutral soils to acidic soils, our study and that of Singh *et al.* (2003) focused on the acidification of alkaline soils.

The naturally occurring pedogenic calcium carbonate (CaCO₃) is assumed to be responsible for much of the high soil pH levels found in the study area. However, the presence of some very alkaline samples (pH >9.5) in the subsoil indicates that this could also be due to the presence of other carbonates, such as sodium carbonate (Na₂CO₃) and sodium bicarbonate (NaHCO₃). Sodium carbonate in particular has a much higher pH level (11.26 per 100 mM at 25°C) than calcium carbonate (pH 9.91 per 100 mM at 25°C) and is also much more soluble in water (Brečević and Nielsen 1989). The original 2002 study also revealed very high exchangeable sodium percentages (ESP) in the subsoil, with the majority of values above 15% (Onus *et al.* 2003), and this is another indication that these sodium-derived carbonates are likely to be widespread in the study area. The combination of high pH and high solubility suggests that the dissolution of sodium carbonate is an important factor in the acidifying trends of soil observed in this study.

Nevertheless, our study found that not only had soil acidification occurred under the more intensive irrigated land uses (irrigated cotton and horticulture), but also, in many cases, under the non-irrigated land uses (dryland cropping and 'natural'). It is believed that the three very wet years experienced in Hillston from 2010 to 2012 had contributed to this decrease in pH by encouraging the dissolution of carbonates under these non-irrigated land uses. Carbonates in soils under irrigated land uses had likely undergone some amount of dissolution when irrigation practices were initially introduced decades ago, whereas areas under non-irrigated land uses have only recently been exposed to large amounts of water percolating through the soil. Overall, this suggests that appreciable soil change can occur over short time periods as a result of significantly high amounts of rainfall. In addition, the less intense dryland land use also drives acidification to some extent through the application of acidifying fertilisers, and removal of basic cations through continuous harvesting of plant biomass.

While irrigated cotton production is almost entirely carried out on Vertisols, the non-irrigated land uses are typically situated on the non-Vertisols (sandier soils). Vertisols are known to possess a

higher pH buffering capacity and a greater ability to counteract acidification than the sandier, non-Vertisol soils found in the study area (Helyar *et al.* 1990; Singh *et al.* 2003; Cay and Cattle 2005). Fig. 4.1 also supports this, where the greatest decreases in pH in the upper section of the profile was experienced in the non-Vertisols. It also shows that the Grey/Brown Vertisols experienced the smallest change in pH at these depths, which is what would be expected of these younger soil types, with the Red Vertisols experiencing an intermediate level of soil pH change. This combination of reasons could explain the decreases in pH under the ‘natural’ land use observed in this study. Nevertheless, it is difficult to conclusively disentangle the impacts of land use and climate when undertaking a soil monitoring/change study.

4.4.2. Spatial patterns of soil pH maps

The spatial patterns of the soil pH maps produced in this study reflect both land use and soil type. Management is expected to have a greater influence on the topsoil, and this is reflected by the stronger correlation of pH with land use in the shallower layers of the soil profile (0-0.5 m). It was also clear that land use generally had more of an impact on pH in 2015 than in 2002 (Table 4.4), as land use was not a significant predictor in the sampling depths from 0.5-1.2 m in 2002. This is logical, as irrigated cotton production became prevalent in Hillston in the 1990s and was considered a relatively new industry in 2002. Similarly, soil type played a significant role in model development in the upper three layers of the soil profile, but had less of an influence in the subsoil (0.5-1.5 m). In the topsoil, there is considerable contrast in properties between the soil types, with the non-Vertisols typically being more acidic with a sandier texture, but there is often less contrast in the subsoil of these differing soil types. This could explain the relatively high importance of soil type in driving spatial patterns of soil pH in the upper 0.5 m (Table 4.4).

4.4.3. Statistical significance of pH change

The z-statistic score is calculated using the contrast variance, as described in Equation 4, and in this study we have also included the covariance term from Equation 3. When the covariance term is included, the contrast variance is reduced, which subsequently gives greater sensitivity in predicting a statistically significant change. In this study, the 1.2-1.5 m sampling depth revealed the most widespread statistically significant decreases in soil pH (Fig. 4.6). This predicted change may appear to be due to the smaller number of samples at this depth, but calculation of the z-statistic takes this into account, and the confidence intervals surrounding this would consequently be much larger. The 0.5-0.8 and 0.8-1.2 m z-statistics maps also revealed pockets of land where statistically significant decreases in pH were

predicted, indicating that the subsoil experienced greater levels of acidification than the topsoil (Fig. 4.6).

The maps of soil pH and the standalone change maps (2015-2002) showed that large parts of the study area had experienced a decrease in pH throughout the soil profile, with a few isolated areas of a modelled increase in soil pH. Despite this, these maps do not indicate if these changes have been statistically significant. The z-statistic maps, however, show this and reinforce the importance of placing a prediction interval around the predicted change. Take the 0.5-0.8 and 0.8-1.2 m depth increments for example, where the standalone change maps (Fig. 4.4 and 4.5) show large and widespread shifts in soil pH, and compare these with the z-statistic maps (Fig. 4.6), which showed that only a small area actually experienced a statistically significant change in pH. This is a good example of why standalone predicted change maps must be viewed with caution, and why placing a prediction interval around predicted change is crucial to ensure that the importance of the change is properly represented. Ignoring the prediction intervals may lead to the implementation of poorly informed policy decisions and management techniques. There are still very few digital soil mapping and monitoring studies that consider hypothesis testing and the statistical significance of the change in a soil property.

4.4.4. Bivariate linear mixed models – benefits and opportunities

Overall, the six BLMMs had high predictive power and low prediction variance. The fitted models displayed mean and median standardised squared prediction errors (SSPE) that were close to the expected values of 1.0 and 0.455, respectively, which suggests that the prediction variances represented the actual errors. The maps also displayed a logical connection between different time points for the same depth, which is likely due to the use of our BLMM, and in particular modelling the cross-correlation between surveys in our random effects term. There are no studies, to our knowledge, that utilise this co-variance through time to create digital maps of any soil property, with all studies opting for univariate approaches. Karunaratne *et al.* (2014) attempted to use a BLMM through time on soil survey carbon data for a region sampled twice over a decade, however, the optimum bivariate model found no significant cross correlation in the random effects between the two surveys. Because of this, the covariance term was considered to be zero, and two separate univariate LMMs were fitted for each time point (Karunaratne *et al.* 2014).

The varying quality of models at different depths could be attributed to a few characteristic limitations. The number of samples available is crucial in determining model quality, with the deepest

layer possessing the fewest samples; the shallowest having the most, and a gradient in between. It is generally accepted that there is a positive correlation between the number of samples available and model quality. Another important factor is the original standard sampling depths of the 2002 survey. For example, considerable assumptions are made when interpolating the 0.3-0.4 m depth from the 2002 data to the standard 0.3-0.5 m depth used. In contrast, the 2002 depths of 0.8-0.9 and 1.1-1.2 m would have resulted in a more accurate interpolation to the standard 0.8-1.2 m depth. This theory is supported by the lower model quality of the 0.3-0.5 m depth layer, and the higher model quality of the 0.8-1.2 m depth layer. This could also explain the lack of a clear pattern in model quality throughout the soil profile.

In this study we have utilised co-located soil information from different time points to create separate soil maps, and thus account for spatial correlation through time. However, there are further opportunities to utilise spatial correlation from other soil depths or other soil properties to improve our models and predictions. A study in Nigeria created digital soil maps of texture at six different soil layers to 2 m depth, using machine-learning approaches (Akpa *et al.* 2014). Instead of creating a unique model for each soil layer as we have done in our study, one model was developed for all soil layers for a particular soil property, and soil sampling depth was included as a predictor variable. This approach dramatically improved prediction accuracy as the abundance of information from other layers was utilised. However, the covariance between these co-located samples cannot be incorporated into the machine-learning approach. Having one model for multiple depths creates soil maps that pedologically make sense, as the same covariates are used, and information from depths above and below the depth of interest are utilised. In a more complex way, Orton *et al.* (2014) used a linear model of co-regionalization (LMCR) to model two soil properties at three depths, resulting in six variables in the LMCR. This meant that the correlation between the two soil properties, and the correlation between the different depths, was utilised. Results showed improved predictions, particularly where missing samples occurred (Orton *et al.* 2014).

There is potential in the future to use multivariate LMMs where one model is created for multiple time points, at multiple depths, and for multiple soil properties. The main limitation to this approach is the amount of computing power required to develop such a model. Modelling the spatial co-variation between multiple variables is a much greater assignment than modelling only one variable, leading to significant practical impediments (Orton *et al.* 2014).

Finally, our maps were created on a 100 m grid and predicted on point support, but there is the opportunity to predict on the block support. Blocks can be square or irregular shapes, such as agricultural fields and consist of averaged values of all response variables in the block area. In contrast, when predicting on point support, the value of the response at a single point is used, and then this is visualised into a square grid when maps are created (Bishop and Lark 2007). There is much literature on predicting on a block support (Goovaerts 1997) and how it results in smaller prediction variances, however very few studies adopt this approach (Bishop *et al.* 2015). Block kriging would be beneficial in monitoring studies as it would result in more statistical power for the hypothesis testing. An example by Bishop and Lark (2007) analysed a potassium fertilizer experiment using geostatistics at different spatial supports (point, field, and farm), and recent work by Bishop *et al.* (2015) showed experimentally that predictions on the block support were more accurate than those on the point support (Bishop *et al.* 2015). Given this improvement, it would seem wise to consider the correct prediction support of maps of change, especially as predictions on block supports of a field or landscape unit are probably more useful for guiding management decisions (Vaysse *et al.* 2017). For example, a grower may be more interested in the average change in their fields, rather than a point on a 100 m grid.

4.5. Conclusions & future directions

A mild soil acidification trend was observed across the cotton-growing region of Hillston in the lower Lachlan catchment between the years 2002 and 2015. However, only isolated areas experienced a statistically significant decrease in pH, and these significant decreases in pH were more pronounced at deeper layers of the soil profile. There are very few studies that consider monitoring the change in subsoil pH, or the change of any property in the subsoil. The likely driving factors of this soil acidification are management practices associated with irrigated cotton production (e.g. N fertilizer application) and increased water percolation through the soil profile (e.g. base cation leaching, and carbonate dissolution) due to continuous irrigation and periods of very high rainfall. While acidification of these highly alkaline soils is of no great concern at this time, possible further acidification still needs to be monitored, particularly in the more acidic topsoil and in those soil types with a lower pH buffering capacity. Overall, the six bivariate linear mixed models (BLMMs) developed had high predictive power and low prediction variance, and these are likely aided by the utilisation in the modelling of the correlation of pH values at co-located sites. Further work should focus on not solely utilising the correlation between the same soil property at different time points at co-located sites, but also using

the correlation with other sources of information, such as different soil depths or different soil properties.

References

- Akpa, SIC, Odeh, IOA, Bishop, TFA, Hartemink, AE (2014) Digital Mapping of Soil Particle-Size Fractions for Nigeria. *Soil Science Society of America Journal* **78**, 1953-1966.
- Bishop, TFA, Horta, A, Karunaratne, SB (2015) Validation of digital soil maps at different spatial supports. *Geoderma* **241**, 238-249.
- Bishop, TFA, Lark, RM (2007) A landscape-scale experiment on the changes in available potassium over a winter wheat cropping season. *Geoderma* **141**, 384-396.
- Bishop, TFA, McBratney, AB, Laslett, GM (1999) Modelling soil attribute depth functions with equal-area quadratic smoothing splines. *Geoderma* **91**, 27-45.
- Brečević, L, Nielsen, AE, (1989) Solubility of amorphous calcium carbonate. *Journal of Crystal Growth* **98**, 504-510.
- Bureau of Meteorology - BOM (2017) Monthly rainfall – Hillston airport. Available from: <http://www.bom.gov.au/climate/averages/tables/cw_075032_All.shtml> [25 April 2017]
- Cameron RG, Colquhoun GP, Meakin NS (2000) Provisional Hillston 1:100 000 Geological Series Sheet. Geological Survey of New South Wales, Sydney, Australia
- Cay, ER, Cattle, SR (2005) The effects of landforming on soil profile characteristics of an irrigated cotton-producing area of southeastern Australia. *Soil and Tillage Research* **84**, 76-86.
- Dregne H (1976) Soils of arid regions. Elsevier, Amsterdam
- Filippi, P, Minasny, B, Cattle, SR, Bishop, TFA (2016) Monitoring and Modeling Soil Change: The Influence of Human Activity and Climatic Shifts on Aspects of Soil Spatiotemporally. *Advances in Agronomy* **139**, 153-214.
- Goovaerts P (1997) Geostatistics for Natural Resources Evaluation. Oxford University Press, New York.
- Guo, JH, Liu, XJ, Zhang, Y, Shen, JL, Han, WX, Zhang, WF, Christie, P, Goulding, KWT, Vitousek, PM, Zhang, FS (2010) Significant Acidification in Major Chinese Croplands. *Science* **327**, 1008-1010.
- Hazelton P, Murphy B (2007) Interpreting Soil Test Results: What do all the Numbers mean? CSIRO Publishing, Collingwood, Victoria.
- Helyar KR, Hochman Z, Brennan JP (1988) The problem of acidity in temperate area soils and its management. In 'National Soils Conference Review Papers, 1988'. (Ed. J. Loveday.) pp. 22-54. (Aust. Soc. Soil Sci. Inc.: Univ. of Western Australia, Perth.)
- Helyar, KR, Cregan, PD, Godyn, DL (1990) Soil acidity in New South Wales - current pH values and estimates of acidification rates. *Australian Journal of Soil Research* **28**, 523-537.
- Jenny H (1980) The soil resource: origin and behavior. Springer-Verlag, New York.
- Karunaratne, SB, Bishop, TFA, Odeh, IOA, Baldock, JA, Marchant, BP (2014) Estimating change in soil organic carbon using legacy data as the baseline: issues, approaches and lessons to learn. *Soil Research* **52**, 349-365.
- Lark, RM (2000) A comparison of some robust estimators of the variogram for use in soil survey. *European Journal of Soil Science* **51**, 137-157.
- Lark, RM, Papritz, A (2003) Fitting a linear model of coregionalization for soil properties using simulated annealing. *Geoderma* **115**, 245-260.
- Lin, LIK (1989) A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 255-268.
- Marchant BP, Crawford DM, Robinson NJ (2015) What can legacy datasets tell us about soil quality trends? Soil acidity in Victoria. IOP Conference Series: Earth and Environmental Science. 25 No. 1. IOP Publishing.

- McBratney, A, Field, DJ, Koch, A (2014) The dimensions of soil security. *Geoderma* **213**, 203-213.
- McBratney, AB, Mendonça Santos, ML, Minasny, B (2003) On digital soil mapping. *Geoderma* **117**, 3-52.
- Minty B, Franklin R, Milligan P, Richardson M, Wilford J (2009) The Radiometric Map of Australia. *Exploration Geophysics* **40**, 325-333.
- Office of Environment and Heritage - OEH (2017) NSW Landuse 2007. NSW Government, Australia
- Oldeman, LR (Eds DJ Greenland, I Szabolcs (1994) 'The global extent of soil degradation.' (C a B International: Wallingford)
- Onus A, Cattle S, Odeh I (2003) How do Lachlan Valley cotton soils compare to northern NSW? *Australian Cottongrower* **24**, 28–30.
- Orton, TG, Pringle, MJ, Page, KL, Dalal, RC, Bishop, TFA (2014) Spatial prediction of soil organic carbon stock using a linear model of coregionalisation. *Geoderma* **230–231**, 119-130.
- Papritz, A, Flühler, H (1994) Temporal change of spatially autocorrelated soil properties: optimal estimation by cokriging. *Geoderma* **62**, 29-43.
- Rochester, IJ (2011) Assessing internal crop nitrogen use efficiency in high-yielding irrigated cotton. *Nutrient Cycling in Agroecosystems* **90**, 147-156.
- Ross, CW, Grunwald, S, Myers, DB (2013) Spatiotemporal modeling of soil organic carbon stocks across a subtropical region. *Science of The Total Environment* **461–462**, 149-157.
- Singh, B, Odeh, IOA, McBratney, AB (2003) Acid buffering capacity and potential acidification of cotton soils in northern New South Wales. *Soil Research* **41**, 875-888.
- Suarez, D.L. (2006) Inorganic Carbon: Land Use Impacts. In: Lal, R., (Eds.), *Encyclopedia of Soil Science- Second Edition - Vol. 1'* (pp. 895-897). Florida: CRC Press.
- Vaysse, K, Heuvelink, GBM, Lagacherie, P (2017) Spatial aggregation of soil property predictions in support of local land management. *Soil Use and Management* **33**, 299-310.
- Webster, R, Oliver, MA, (2001) *Geostatistics for Environmental Scientists*, John Wiley & Sons, Chichester, UK

Chapter 5

5. Monitoring changes in topsoil and subsoil salinity and sodicity in an irrigated semi-arid landscape over a 13-year period

Abstract

Soil salinity and sodicity are two of the most widespread and limiting soil constraints for agriculture globally, and are particularly significant in arid and semi-arid landscapes. This study monitors the change in soil electrical conductivity (EC) and exchangeable sodium percentage (ESP) between 2002 and 2015 using data from soil surveys collected in a semi-arid irrigated cotton-growing region in the lower Lachlan River valley in southern NSW, Australia. Bivariate linear mixed models (BLMMs) were used to create digital maps of soil properties using co-located soil information and correlation through time at six standard depth increments for EC (0-0.1, 0.1-0.3, 0.3-0.5, 0.5-0.8, 0.8-1.2, 1.2-1.5 m), and two depth increments for ESP (0-0.1, 0.8-1.2 m). The models in this study showed high predictive power, and the advantages of using bivariate linear mixed models to model soil properties through time compared to univariate approaches are demonstrated. Traditional laboratory methods were used to measure EC, but visible near infrared (VisNIR) spectroscopy techniques were used in combination with traditional methods to predict soil ESP values to high accuracy (0.79 LCCC independently-validated). The study area possessed low levels of soil salinity, and temporal shifts in EC were detected. Some degree of statistically significant changes in EC were detected at all sampling depths, however, the most widespread changes occurred at the 1.2-1.5 m depth where statistically significant increases and decreases in salinity were detected. Overall, no appreciable change was observed in soils under the natural land use at most depths, but some areas under irrigated cotton production experienced a desalination trend. In contrast, soil EC under irrigated perennial horticultural production was predicted to have increased over time, with this attributed to continual irrigation, the use of poorer quality irrigation water, and the use of fertilisers that contain salts. It was apparent that the varying trajectories of soil EC under the different irrigated land uses were due to the varying quantity and quality of irrigation water used over the 13-year study period. Soil sodicity at Hillston was found to be generally low to moderate in the upper 0.5 m of the soil profile, but very high in the lower part of the profile; 0.5-1.5 m depths, leading to potentially undesirable impacts on soil structure. A trend of increasing soil sodicity through time was observed in the study area. The impact of land use on changes in soil ESP was less clear than that of EC, but most of the statistically significant increases occurred in areas under irrigated cotton and horticultural production. The cause of this increase in sodicity is not clear, although it is likely due to the continual addition of sodium to the soil system through irrigation and fertilisers. These statistically significant changes must be viewed with caution, as a portion of the 2015 samples were predicted by VisNIR spectroscopic methods, and hence contain a greater amount of error than samples measured by traditional laboratory methods. The observed trend of increasing ESP and decreasing EC is concerning for areas under irrigated agriculture, as their relative proportions in a soil determine whether a soil disperses or stays aggregated upon irrigation. Additionally, there is the dilemma that the management solution to reduce the salinity levels of soil (irrigate with high quality water) is antagonistic to the management solution to reduce the impact of sodic soil on crop production (irrigate with saline-sodic water). It is essential that soil salinity and sodicity both continue to be monitored in conjunction to examine the direction of any future changes, and that growers are aware of the salinity and sodicity status of their soil when making decisions on irrigation.

5.1. Introduction

Soil salinity and sodicity are two of the most important and widespread soil constraint and degradation issues in Australia and throughout the globe (Shahid 2013; Mora *et al.* 2017). Soil salinity has typically received most of the spotlight in Australian agriculture, but soil sodicity far overshadows salinity in regards to impact on agriculture and land area covered (Northcote and Skene 1972). Arid and semi-arid ecosystems cover ~40% of the Earth's surface, and are particularly known for possessing saline and sodic soils, largely due to the minimal leaching of salts from the soil profile (Ford *et al.* 1993; Corwin *et al.* 2003; Rengasamy 2006; Jalali *et al.* 2008; Hulugalle *et al.* 2013). Despite receiving low rainfall, these areas often have an abundance of groundwater and surface water available for irrigated agriculture, and this has recently started to be utilised for irrigated cotton (*Gossypium hirsutum*) production in eastern Australia.

Excessively saline soils can detrimentally impact crop production, but soil salinity can also positively impact the physical properties of a soil, as it encourages particles to bind together into aggregates (McKenzie and Orange 2003; Warrence *et al.* 2003). In contrast, soil sodicity encourages dispersion of soil, as excessive sodium ions on exchange sites weaken bonds between clay particles (Rengasamy and Olsson 1991). The relative proportion of salinity and sodicity in a soil is integral, as this determines whether a soil will disperse or stay aggregated upon wetting (McNeal 1968). In addition, the salt concentration of applied irrigation water has an important role on the physical structure of soils, with saline irrigation water encouraging flocculation, and less saline water promoting dispersion (Warrence *et al.* 2003).

While soil salinity and sodicity are often inherent issues in arid and semi-arid landscapes, these issues can be further accentuated by agricultural management practices. Salt content of a soil can increase through movement from other parts of the soil, such as being washed down from above, or rising from soil beneath due to a rising water table. This connectivity is why it is crucial to monitor multiple depths of the soil concurrently, although this is rarely considered. The addition of salts to soil from external sources is another method of increasing salinity and sodicity, and this often occurs through application of certain fertilisers (e.g. potash) and by means of irrigation. The Australian cotton industry is dominated by irrigated production, accounting for as much as 79% of the production area, and 90% of total lint produced between 2009 and 2014 (Roth 2014). During a typical growing season of an irrigated cotton crop the equivalent of ~1000-1400 mm of water is applied, and with this roughly 1.5 tonnes per hectare of salt is deposited into the soil within the root zone (McKenzie and Orange 2003). In

general, sodium is the dominant cation in this deposited salt, causing potentially detrimental problems of both soil salinity and sodicity (McKenzie and Orange 2003). The quality of irrigation in Australia can be highly variable, with groundwater typically possessing higher sodium levels than that of water extracted from rivers (Speirs *et al.* 2011). During the 2000s, much of eastern Australia experienced the 'Millennium Drought', which meant that surface water for irrigation was scarce and only groundwater was available. However, the emergence of La Niña in 2010 resulted in several consecutive years of high rainfall and floods in some locations, making surface water for irrigation more available again.

While there are a few examples of studies that have monitored spatio-temporal changes in soil salinity (Herrero and Pérez-Coveta 2005), very few consider doing so for sodicity. This is not due to a lack of interest in sodicity, but rather because the electrical conductivity (EC) of a soil is easily measured by traditional laboratory techniques, while analysing exchangeable sodium percentage (ESP) by traditional laboratory methods is both extremely time and resource intensive. There has been much recent focus and development on using cheap and rapid soil spectroscopic techniques to analyse soil properties as a way of overcoming this common issue. This has largely been successful, however, the accuracy of predictions using soil spectroscopic techniques varies considerably among different soil properties (Viscarra Rossel *et al.* 2006). Nevertheless, there is promise for using these approaches to predict soil ESP and other difficult-to-measure soil properties, and this is likely to encourage monitoring studies to focus on environmentally- and agriculturally-important soil properties that are often disregarded due to their cost prohibitive nature.

It is anticipated that the recent shifts in rainfall and the variability of irrigation water quantity and quality, combined with the intensive management practices of irrigated cotton production may have caused significant shifts in soil condition in arid and semi-arid landscapes in eastern Australia. There is particular interest in the state of the soil in the relatively new southern cotton growing regions of NSW, as there has been very little monitoring on the impact that cotton production has had in these areas (Holland 2014). Long-term studies of the spatio-temporal changes in soil salinity and sodicity are scarce, and while these issues are important throughout the soil profile, most studies solely focus on the topsoil. This study monitors the change in soil electrical conductivity (EC) at six depth increments to 1.5 m and exchangeable sodium percentage (ESP) at two depth increments (0-0.1 and 0.8-1.2 m) from 2002 to 2015 in a semi-arid irrigated cotton-growing region in southern NSW, Australia. Due to cost limitations, visible near infrared (VisNIR) spectra are combined with traditional laboratory methods to predict soil ESP on a portion of samples, and the benefits and limitations of this are discussed. The

primary aim of the study is to analyse the impact that different land uses, management practices, and recent shifts in rainfall patterns have on the trajectory of change in soil salinity and sodicity.

5.2. Materials and methods

5.2.1 Soil datasets

As described in greater detail in Chapter 3, soil samples from the same 115 soil cores extracted in the soil survey conducted at Hillston in 2002 were used, along with samples from the 160 soil cores extracted from the soil survey conducted in 2015. Many of these cores ($n = 103$) were sampled at the same spatial location in the different surveys using georeferenced coordinates. The subsampling depths of soil cores in 2002 were; 0-0.2, 0.3-0.4, 0.55-0.65, 0.8-0.9, 1.1-1.2, and 1.35-1.5 m, while cores from 2015 were subsampled into; 0-0.1, 0.1-0.3, 0.3-0.5, 0.5-0.8, 0.8-1.2, and 1.2-1.5 m depth increments. Because of the differences in these sampling depths, equal-area quadratic smoothing splines (Bishop *et al.* 1999) were used on soil EC and ESP data for the 2002 samples to standardise this to the 2015 sampling depths.

5.2.2. Laboratory analyses

All of the samples were air-dried, ground and passed through a 2 mm sieve. To estimate salinity, soil electrical conductivity (EC) was determined for 1:5, soil: water extracts, using a CDM 83™ conductivity meter for the 2002 samples, and using a Mettler Toledo SevenCompact™ conductivity meter for the 2015 samples. In both the 2002 and 2015 datasets, exchangeable base cation (Ca^{2+} , Mg^{2+} , K^+ and Na^+) contents were determined for samples that were extracted with alcoholic 1 M ammonium chloride (pH 8.5), and then analysed by atomic absorption spectrometry (Rayment and Lyons 2011). Prior to this, samples were pre-treated to remove soluble salts by using a combination of 60% aqueous ethanol and 20% aqueous glycerol. The effective cation exchange capacities (ECEC) were estimated by summing the exchangeable basic cations, and exchangeable sodium percentages (ESP) were calculated by dividing the exchangeable sodium contents by the ECEC and multiplying by 100.

5.2.3. Visible near infrared (VisNIR) spectroscopy analyses

While EC is rapid and inexpensive to analyse by traditional laboratory methods, the estimation of ESP and ECEC is considered to be quite labour intensive. Consequently, not all soil samples were analysed for ESP and ECEC by traditional laboratory methods in the 2015 dataset, despite all 2002 soil samples being fully analysed by this approach (Fig. 5.1). Instead, both ESP and ECEC were directly predicted using visible near infrared (VisNIR) spectroscopic techniques in those samples from the 2015 survey that were

not measured by traditional laboratory techniques (Fig. 5.1). The training dataset consisted of 385 samples from the 2002 survey, as well as 138 samples from the 2015 survey, and this was used to predict onto the remaining 768 un-analysed samples from 2015. Spectroscopic measurement was made with an Agrispec portable spectrophotometer with a contact probe attachment on dried and ground samples (Analytical Spectral Devices, Boulder, Colorado). To reduce signal-to-noise ratios of the spectra, three scans of each sample were performed, from which an averaged reflectance spectrum was derived. Calibration of the instrument was made with a Spectralon white tile and was re-calibrated after every 15 scans, or five samples.

To ensure that this subset of 138 samples represented the remaining 768 samples from the 2015 survey appropriately, conditioned Latin hypercube (CLHC) sampling was used (Minasny and McBratney 2006). Initially, a principal component analysis (PCA) was performed on the spectra from the VisNIR wavelengths (350-2500 nm) of the 906 samples from the 2015 survey. The first two principal components (PCs) from this PCA explained >95% of the cumulative variation. We aimed to select 20 out of the 160 sites from 2015 based on the whole profile, so the first two PCs of the first four sampling depths (as not all cores had six sampling depths) were used as input criteria for the CLHC, which ensured that most of the soil profile was considered during site selection. Land use and soil type was also used as an input in the CLHC to ensure that each soil-land use complex of the study area was represented. An additional 10 samples were selected at the 0-0.1 and 0.8-1.2 m sampling depth by a similar approach. The training dataset ($n = 523$) was then used to directly predict soil ESP on the un-analysed samples ($n = 768$) using Cubist models, which is a machine learning technique. The predictor variables included wavelengths from 500-2450 nm, averaged into segments of ten nm, and the mid-depth of the sample, to ensure that the depth of the sample was taken into account when predicting.

To test the prediction quality of the Cubist models, 75% of the dataset was used as calibration, and the remaining 25% was used as validation, as is common practice when undertaking this procedure. These two datasets were selected by using the VisNIR spectra and measured ESP values as inputs in a Latin hypercube, and this ensured that both the validation and calibration datasets were appropriately represented. During model testing, it was ensured that other depth samples from the same soil core for that year were not included in the calibration dataset when validating. It is assumed there is a strong correlation between these different depths of the same core and this may bias results.

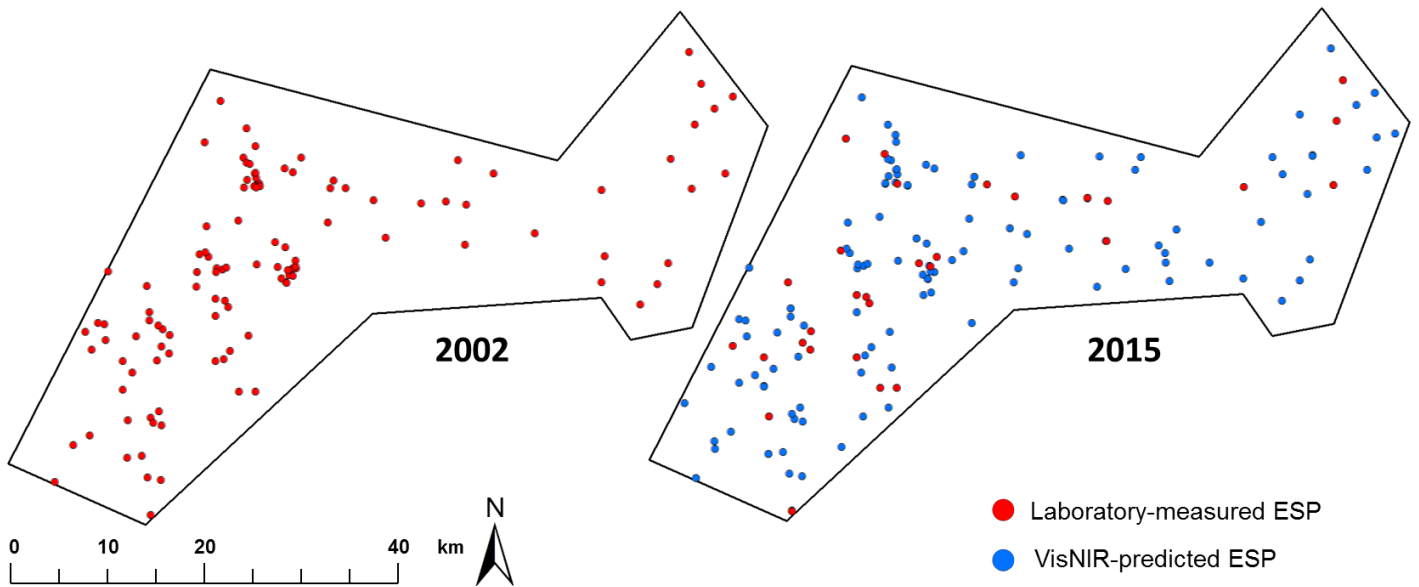


Fig. 5.1 – Locations of soil cores extracted in the 2002 and 2015 soil surveys, and sites analysed by traditional laboratory methods, and predicted from VisNIR spectra

5.2.4. Modelling and mapping

5.2.4.1. Predictor variables

Various spatial covariates were collated to use as predictor variables in the development of the soil EC and ESP models. These predictor variables can be roughly categorised into four types; terrain attributes, gamma radiometric data, soil type and land use information. Because there were 19 individual numerical covariates (terrain attributes and gamma radiometrics), a principal component analysis (PCA) was performed and the main PCs were used as potential predictors. The covariates used in this study and the preparation and processing of these covariates is described in greater detail in Chapter 3.

5.2.4.2. Bivariate linear mixed models (BLMMs) and statistical significance of change

Bivariate linear mixed models (BLMMs) were used to model the change in soil EC and ESP. The theory is discussed in detail in Lark and Papritz (2003), and further details of the modelling approach used here are presented in Chapter 3 and 4. One model for each soil property at each sampling depth for both time points was created, allowing the covariance and correlation at co-located sites through time to be utilised, which results in more precise estimates (Lark and Papritz 2003). There are many advantages to using this approach as opposed to univariate approaches for soil monitoring studies, however, very few studies consider using this approach (Papritz and Flühler 1994).

Most soil monitoring studies do not place a prediction interval around predicted change in soil attributes, but this is crucial in determining whether the change observed is statistically significant (Filippi *et al.* 2016). Failing to do this can lead to incorrect interpretations of the magnitude and accuracy of actual change in soil properties. The variance of the change is integral in calculating the statistical significance of change, and is expressed as:

$$\hat{V}(\Delta\hat{y}) = \hat{V}(\hat{y})_{t_1} + \hat{V}(\hat{y})_{t_2} - 2(\hat{V}(\hat{y})_{t_1,t_2}), \quad 5.1$$

where $\hat{V}(\Delta\hat{y})$ is the variance of the change in the soil property between two points in time, $\hat{V}(\hat{y})_{t_1}$ is the variance of the soil property for the observations at the first point in time, $\hat{V}(\hat{y})_{t_2}$ at the second point in time, and $\hat{V}(\hat{y})_{t_1,t_2}$ is the covariance between the measurements of soil properties at the two time periods. The covariance term is considered to be zero in univariate approaches, however, it is included in BLMMs, which results in a stronger relationship between the two surveys, a smaller contrast variance, and subsequently greater sensitivity in predicting a statistically significant change. For a two-tailed test, a z-statistic score of greater than (in a positive case), or less than (for a negative case) ± 1.28 , ± 1.65 , ± 1.96 , and ± 2.58 represents a confidence interval of 80%, 90%, 95%, and 99% respectively. These values are used in figures 5.10 and 5.11 to represent the level of significance of the change predicted.

5.2.4.3. Model selection and map production

In this chapter, maps for all six depths for soil EC are displayed, however, only the 0-0.1 m and 0.8-1.2 m depths are shown for soil ESP as these depths had fewer ESP values predicted by VisNIR and had more samples measured for ESP by traditional laboratory methods. One model for each of the layers for EC ($n = 6$) and each layer for ESP ($n = 2$) at both time points was created, giving a total of 8 models, and 16 individual soil property maps. Soil property data was statistically transformed if the residuals lacked a normal distribution. The combination of predictor variables to be included in the final model was determined by backwards elimination. Predictor variables with the highest p -value were removed from the model, and this continued until all predictor variables in the model possessed a p -value less than 0.05. Categorical data was kept in the model if at least one of the levels was deemed significant. Principal components (PCs) that were common to both the 2002 trend and the 2015 trend were then analysed to determine if the upper and lower confidence intervals overlapped. If this was the case, they were merged and the model was then re-run to test the significance. After this, the model was used to predict onto the covariate grid of the study area (100 m resolution) using the EBLUP (Empirical Best Linear Unbiased Predictor) approach and digital soil maps were then produced. For the soil data that

was transformed by applying a square root, back-transformation was achieved by simply squaring the transformed predictions. For the data that had a natural logarithmic applied, the transformed predictions were back-transformed from the following:

$$y_i = \exp(x_i + (\frac{\sigma^2_i}{2})), \quad 5.2$$

where y_i is the back transformed prediction at point i , x_i is the transformed prediction at point i , and σ^2_i is the prediction variance at point i . The quality of the models was tested by using leave-one-site-out cross-validation (LOSOVCV). The standardised squared prediction errors (SSPE) were analysed to see if the mean and median values were close to the expected values of 1.0 and 0.455, respectively (Lark 2000; Orton *et al.* 2014). Lin's concordance correlation coefficient (LCCC), was also used as a tool for assessing model quality as it is unit-less, making it useful for comparing between different the models at different depths where the values of the response variable have a different magnitude.

5.2.5. Determination of salinity and sodicity ratings

To simplify the interpretation of the EC maps (Fig. 5.6; 5.7; 5.8), soil salinity ratings and the corresponding EC values based on suggestions provided by Hazelton and Murphy (2007) were used, and are shown in Table 5.1. The ratings and EC values for soils with different proportions of clay is shown, but 40-60% clay content would be a representative range for most of the soils in the study area. This allowed the magnitude of soil salinity at the various depths to be easily interpreted, and also simplifies the identification of areas that contain crop-limiting salinity levels. Different crop species have a different reponse to levels of soil salinity, and the legend values selected for this study essentially show that any areas of the EC maps that are blue (>1.00 dS/m) could potentially limit cotton production (Hazelton and Murphy 2007). Information from Hazelton and Murphy (2007) was similarly used to guide the selection of legend values for the ESP maps (Fig. 5.9), as can be seen in Table 5.2. Any soil with an ESP value >6% could be considered sodic, and thus detrimental impacts on soil structure are likely to occur upon wetting (Table 5.2).

Table 5.1 – Soil salinity ratings for soils of varied proportions of clay, and the legend values used in the salinity maps in this study – adapted from Hazelton and Murphy (2007)

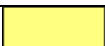











Soil salinity rating	10-20% Clay	20-40% Clay	40-60% Clay	60-80% Clay	Legend range (EC dS/m)	Legend colour
Very low	<0.07	<0.09	<0.12	<0.15	0-0.10	
Low	0.07-0.15	0.09-0.19	0.12-0.24	0.15-0.30	0.10-0.25	
Medium	0.15-0.34	0.19-0.45	0.24-0.56	0.30-0.70	0.25-0.50	
High	0.34-0.63	0.45-0.76	0.56-0.96	0.70-1.18	0.50-1.00	
Very high	0.63-0.93	0.76-1.21	0.96-1.53	1.18-1.87	1.00-1.50	
Extreme	>0.93	>1.21	>1.53	>1.87	>1.50	

Table 5.2 – Soil sodicity ratings and the legend values used in the sodicity maps in this study – adapted from Hazelton and Murphy (2007)

Soil sodicity rating	Legend range (%)	Legend colour
Non-sodic	0-2	
Non-sodic	2-4	
Marginally-sodic	4-6	
Sodic	6-10	
Highly sodic	10-15	
Extremely sodic	>15	

5.3. Results

5.3.1. Soil EC summary statistics

Soil in the majority of the Hillston cotton growing district contains low levels of salinity, and soil EC generally increases with increasing soil depth (Table 5.3). It is also clear that land use has an impact on both the EC levels of soil and the magnitude of change over time. The baseline (2002) EC levels of soils under irrigated cotton production were more than double those under the natural land use in the topsoil (0-0.1 m). An overall desalination trend was observed throughout the profile in soils under irrigated cotton production in the 13-year study period, but no appreciable change in EC was observed in soils of natural sites at most depths (Fig. 5.2). Soils under horticultural production experienced a mean increase in salinity at all sampling depths, as well as those under dryland from 0.3 m and deeper (Table 5.3). The deepest sampling depth of 1.2-1.5 m possessed particularly high EC values, with no discernible differences between land uses at this depth (Table 5.3). The influence that soil type had on the changes in EC over the study period was less clear than that of the influence of land use (Fig. 5.2). Consistent decreases in soil EC were observed for Red Vertosol soils, while slight, but consistent increases in the subsoil were observed for the Grey/Brown Vertosols (Fig. 5.2). There appeared to be minimal change in

EC in the upper half of the soil profile for non-Vertosols, however, larger increases in EC were observed in the lower half (Fig. 5.2).

Table 5.3 – Electrical conductivity (dS/m 1:5 soil: water) statistics for all sites at different sampling depths, time points and land uses

		EC (dS/m 1:5 soil: water)								
		Depth (m)	Cotton		Dryland		Natural		Horticulture	
			2002	2015	2002	2015	2002	2015	2002	2015
<i>n</i> =		0-0.1	60	65	14	21	38	61	3	13
		0.1-0.3	60	62	14	20	37	61	3	12
		0.3-0.5	60	62	14	20	37	61	3	12
		0.5-0.8	60	62	12	19	33	60	3	12
		0.8-1.2	60	56	11	19	23	60	3	12
		1.2-1.5	54	56	4	17	3	52	1	12
Mean		0-0.1	0.22	0.16	0.13	0.09	0.10	0.10	0.09	0.32
		0.1-0.3	0.22	0.15	0.14	0.15	0.15	0.13	0.10	0.22
		0.3-0.5	0.24	0.20	0.18	0.28	0.25	0.21	0.13	0.23
		0.5-0.8	0.36	0.31	0.36	0.51	0.43	0.46	0.19	0.39
		0.8-1.2	0.61	0.48	0.45	0.72	0.58	0.75	0.27	0.63
		1.2-1.5	1.10	0.77	0.74	0.90	0.83	1.07	0.31	1.08
Standard deviation		0-0.1	0.16	0.05	0.08	0.05	0.04	0.05	0.03	0.41
		0.1-0.3	0.10	0.06	0.07	0.13	0.11	0.10	0.03	0.16
		0.3-0.5	0.12	0.08	0.14	0.22	0.28	0.19	0.06	0.17
		0.5-0.8	0.26	0.13	0.30	0.40	0.48	0.42	0.11	0.25
		0.8-1.2	0.54	0.30	0.43	0.56	0.61	0.81	0.12	0.69
		1.2-1.5	1.11	0.77	0.53	0.64	0.28	1.03	NA	1.27
Minimum		0-0.1	0.06	0.07	0.04	0.02	0.02	0.02	0.06	0.10
		0.1-0.3	0.06	0.07	0.04	0.03	0.03	0.01	0.07	0.06
		0.3-0.5	0.06	0.07	0.03	0.02	0.02	0.01	0.08	0.05
		0.5-0.8	0.06	0.10	0.04	0.03	0.02	0.02	0.08	0.09
		0.8-1.2	0.06	0.06	0.04	0.12	0.02	0.02	0.14	0.05
		1.2-1.5	0.06	0.04	0.28	0.21	0.50	0.05	0.31	0.15
Maximum		0-0.1	0.93	0.32	0.33	0.2	0.19	0.23	0.12	1.64
		0.1-0.3	0.66	0.34	0.26	0.46	0.19	0.53	0.14	0.65
		0.3-0.5	0.54	0.44	0.54	0.74	1.14	0.86	0.19	0.59
		0.5-0.8	1.54	0.69	0.98	1.25	1.84	2.31	0.30	0.81
		0.8-1.2	2.74	1.34	1.30	1.94	2.32	3.90	0.38	2.64
		1.2-1.5	4.21	3.89	1.44	2.32	1.01	4.33	0.31	3.48

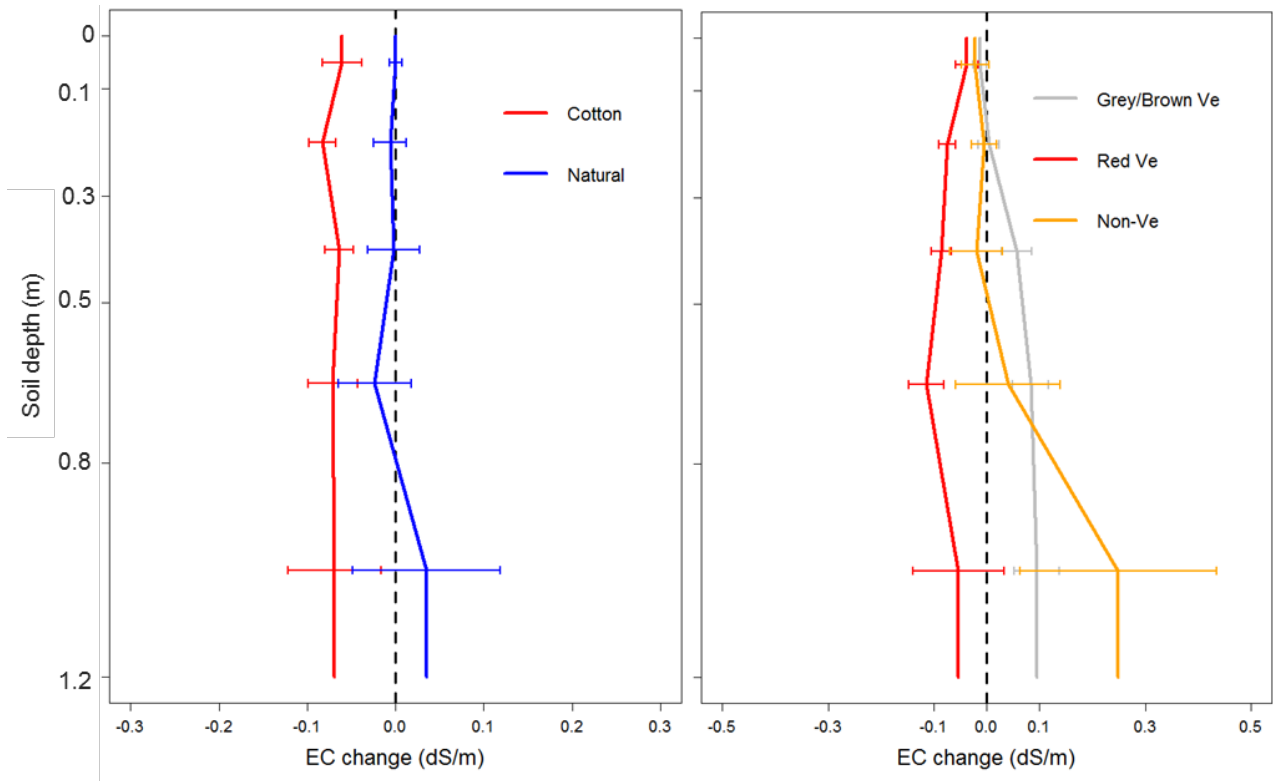


Fig. 5.2 – Mean change in soil EC (dS/m) from 2002 to 2015 of co-located sites by land use (left) and soil type (right) to 1.2 metres with standard error bars (Ve = Vertosol)

5.3.2. Soil ESP prediction by VisNIR

The quality statistics of the Cubist model with VisNIR spectra as predictor variables shows that ESP could be predicted to high accuracy. The validation dataset (25%) was predicted with the calibration dataset (75%) to an accuracy of 0.79 LCCC (Lin's concordance correlation coefficient), 0.65 R^2 , and an RMSE (root mean square error) of 4.6% (Fig. 5.3).

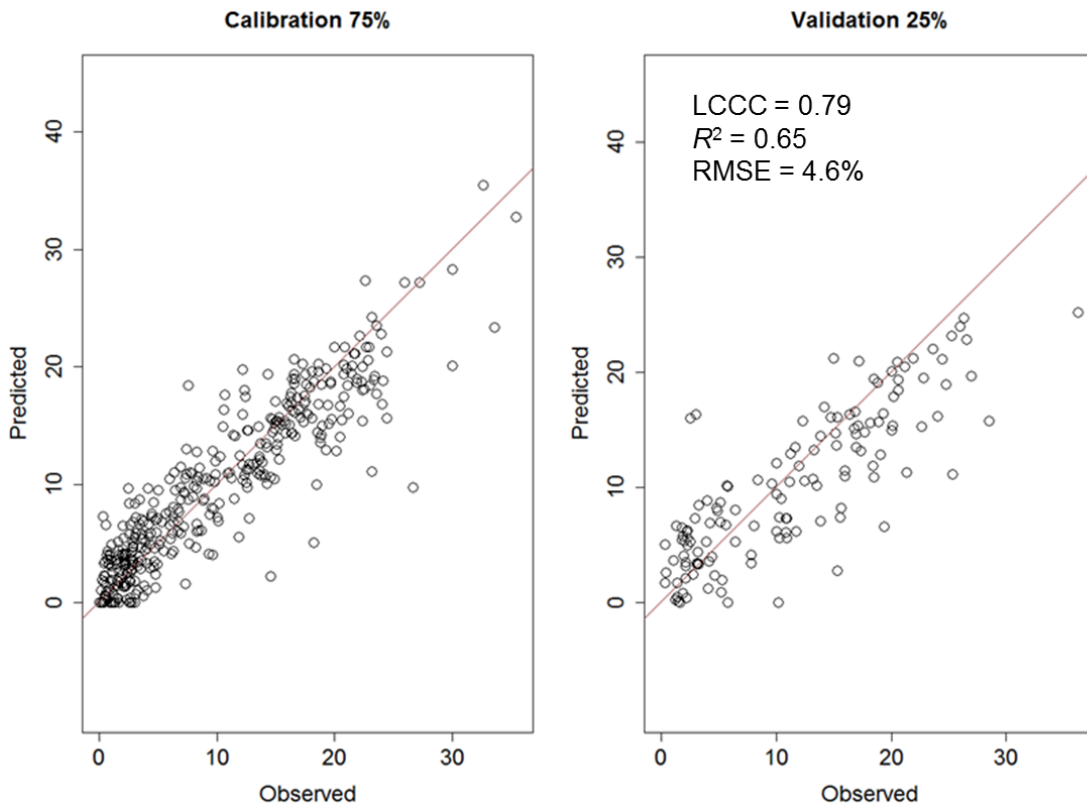


Fig. 5.3 – Observed and predicted ESP (%) of calibration (left) and validation (right) datasets predicted with VisNIR spectra and Cubist models (1:1 line overlaid)

5.3.3. Soil ESP summary statistics

In general, the soils of the Hillston district are non-sodic to marginally-sodic in the upper 0.5 m of the soil profile, with mean levels ranging from 2.7% to 8.9%, and highly sodic in the subsoil (0.5-1.5 m) with mean ESP levels ranging from 12.3% to 15.2% (Table 5.4). The VisNIR-predicted ESP values generally reflected the laboratory measured values well for 2015 samples (Table 5.4). Widespread trends of increasing soil ESP at most depths in the study area between 2002 and 2015 were observed (Fig. 5.4). The impact of land use on soil sodicity levels was not as obvious as it was for soil salinity, and ESP increased under all land uses at most depths (Table 5.4). The main soil types in Hillston exhibit different average levels of soil ESP throughout the profile (Fig. 5.4). The Red Vertosols consistently contained the highest ESP values at all depths in the soil profile, followed by the Grey/Brown Vertosols, with the non-Vertosols having the lowest ESP (Fig. 5.4).

Table 5.4 – Soil exchangeable sodium percentage (%) statistics for 2002 and 2015 laboratory-measured (Lab) samples, and 2015 VisNIR-predicted (Pred) samples at different soil depths, time points and land uses

		Exchangeable Sodium Percentage (%)											
		Cotton			Dryland			Natural			Horticulture		
	Depth (m)	2002 Lab	2015 Lab	2015 Pred	2002 Lab	2015 Lab	2015 Pred	2002 Lab	2015 Lab	2015 Pred	2002 Lab	2015 Lab	2015 Pred
n =	0-0.1	60	12	53	14	4	17	38	13	48	3	1	12
	0.1-0.3	60	7	55	14	3	17	38	9	52	3	1	11
	0.3-0.5	60	7	55	14	3	17	37	9	52	3	1	11
	0.5-0.8	60	7	55	12	3	16	33	9	51	3	1	11
	0.8-1.2	60	11	45	11	4	15	23	13	47	3	1	11
	1.2-1.5	53	6	50	4	3	14	3	9	43	1	1	11
Mean	0-0.1	2.7	2.2	3.7	3.9	3.2	2.3	1.8	3.4	2.5	1.3	0.3	2.3
	0.1-0.3	4.8	5.3	5.5	5.2	7.5	4.2	3.1	7.3	3.8	1.1	2.6	2.9
	0.3-0.5	8.7	9.6	12.7	8.5	13.5	8.5	5.7	11.3	8.5	2.3	7.3	8.5
	0.5-0.8	13.3	15.3	15.3	13.3	15.5	12.2	9.0	15.5	10.7	5.7	6.5	11.4
	0.8-1.2	14.9	19.6	17.2	13.9	17.5	14.5	10.9	15.8	13.0	6.5	7.2	12.5
	1.2-1.5	13.9	15.7	16.7	18.4	14.3	19.9	16.1	14.6	13.7	8.9	7.7	15.5
Standard deviation	0-0.1	1.77	0.71	1.75	1.81	1.48	1.81	3.28	3.74	1.72	0.98	NA	2.29
	0.1-0.3	3.07	3.19	2.15	2.01	2.41	3.33	4.13	6.67	2.12	0.27	NA	2.10
	0.3-0.5	5.94	7.20	4.42	3.69	1.18	5.20	6.16	7.46	5.62	0.79	NA	6.43
	0.5-0.8	6.80	10.82	4.89	6.42	3.33	6.15	7.15	9.13	5.04	1.83	NA	6.96
	0.8-1.2	6.22	9.61	5.72	7.63	4.18	5.57	7.18	8.17	5.66	1.75	NA	7.94
	1.2-1.5	5.22	9.78	5.52	4.06	8.64	7.21	2.88	7.31	5.06	NA	NA	6.50
Minimum	0-0.1	0.4	1.1	1.2	0.0	1.2	0.0	0.1	0.3	0.0	0.1	0.3	0.0
	0.1-0.3	0.3	1.8	1.1	0.3	4.8	0.0	0.3	1.0	0.0	0.3	2.6	0.4
	0.3-0.5	0.9	1.3	5.1	0.2	12.2	0.0	0.2	2.7	0.0	1.1	7.3	0.0
	0.5-0.8	2.0	3.2	5.3	0.5	13.0	2.1	0.4	3.1	0.0	3.9	6.5	0.0
	0.8-1.2	2.3	3.6	0.0	0.4	11.9	4.5	0.4	0.8	0.0	3.2	7.2	2.9
	1.2-1.5	2.1	2.5	3.6	14.1	4.4	7.5	9.5	4.3	0.0	0.7	7.7	4.6
Maximum	0-0.1	10.6	3.3	9.2	5.8	4.7	5.1	17.4	12.1	7.3	1.9	0.3	7.7
	0.1-0.3	13.3	9.4	10.9	6.0	9.5	10.4	21.4	18.9	8.7	0.8	2.6	5.4
	0.3-0.5	28.1	20.0	24.4	12.2	14.5	21.4	27.6	20.5	24.4	2.5	7.3	19.0
	0.5-0.8	30.3	33.6	25.8	20.4	19.3	25.8	29.7	28.5	21.0	7.2	6.5	24.7
	0.8-1.2	25.0	35.5	28.5	20.1	20.8	24.7	33.3	24.8	26.5	6.6	7.2	31.6
	1.2-1.5	25.5	23.3	31.7	23.1	20.0	33.1	23.2	24.4	22.5	0.7	7.7	24.4

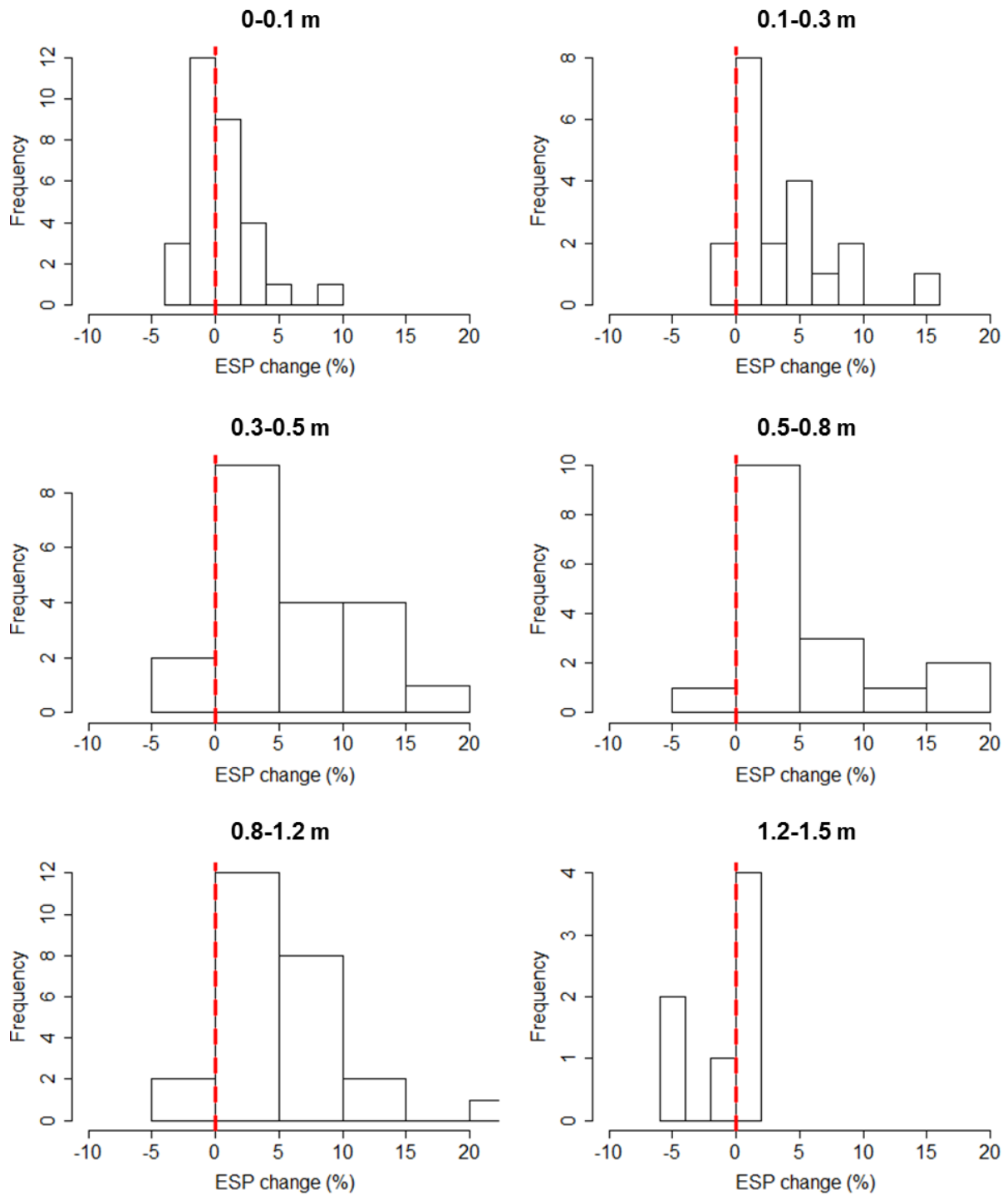


Fig. 5.4 – Change in soil exchangeable sodium percentage (%) of laboratory measured samples only at co-located sites between 2002 and 2015 for the six sampling depths

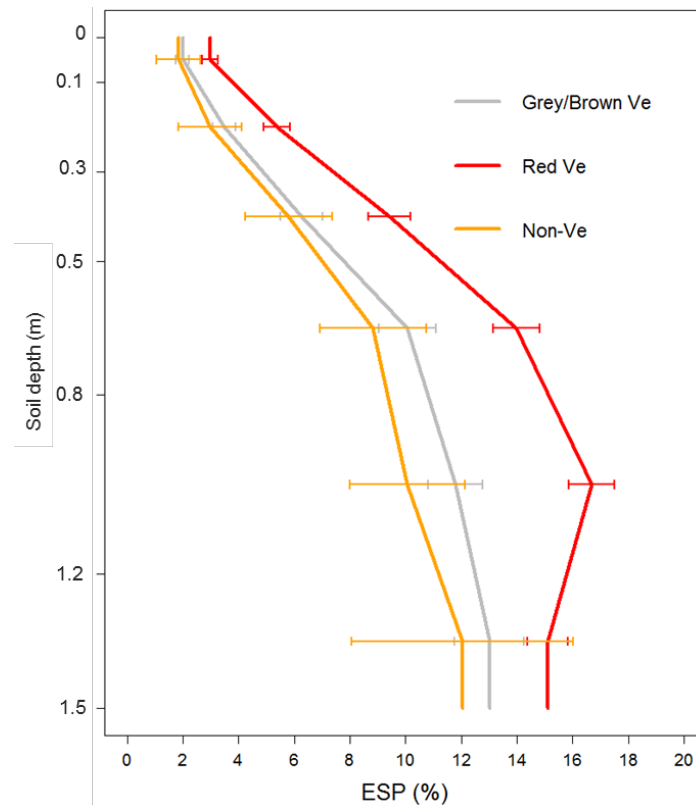


Fig. 5.5 – Mean soil exchangeable sodium percentage (%) of laboratory measured samples from 2002 by soil type to 1.5 metres with standard error bars (Ve = Vertosol)

5.3.4. Bivariate linear mixed models

The predictor variables that were used in the final BLMMs varied for each sampling depth and soil property, but there were particular variables that were consistently important. For the EC models, PC2 was an important predictor in all depths, and land use and soil type were included for the first three layers, but were less important in the lower depths. In the ESP models, land use, soil type, and PC2 were important predictors for both depths (0-0.1 and 0.8-1.2 m). It was common for similar predictor variables to be included in both the 2002 and 2015 trend for each model. Further detail about results from the PCA of numerical predictor variables, and the proportion that each original predictor variable contributes to different principal components is described in Chapter 4.

The soil EC data was transformed with natural logarithmic at all sampling depths, because the data and the residuals lacked a normal distribution in all instances. For the ESP data, it was only required to transform the topsoil (0-0.1 m), and this was done using a square root transformation. The final models generally possessed mean and median standardised squared prediction errors (SSPE) that were

close to the expected values of 1.0 and 0.455, respectively (Table 5.5). The ESP models had high LCCC values, and EC also showed a good fit throughout the soil profile, with the topsoil EC showing the best fit of 0.59 LCCC (Table 5.5).

Table 5.5 – Model quality statistics for EC and ESP BLMMs based on leave-one-out cross-validation

Attribute	Soil depth	LCCC	RMSE	Bias	Mean (SSPE)	Median (SSPE)
EC	0-0.1 m	0.59	0.44*	0.00	1.01	0.52
	0.1-0.3 m	0.48	0.50*	0.00	1.03	0.46
	0.3-0.5 m	0.44	0.60*	-0.01	1.03	0.44
	0.5-0.8 m	0.47	0.65*	-0.01	1.01	0.46
	0.8-1.2 m	0.41	0.80*	-0.04	1.01	0.46
	1.2-1.5 m	0.45	0.85*	-0.01	1.01	0.51
ESP	0-0.1 m	0.50	0.54*	0.01	1.06	0.44
	0.8-1.2 m	0.54	5.48	-0.24	1.01	0.42

* RMSE values are redundant as values were transformed

5.3.5. Modelled maps

5.3.5.1. Soil EC modelled maps

In the topsoil (0-0.1 m), the alluvial soils along the Lachlan River and areas on the floodplains possessed the highest EC values (Fig. 5.6). The change map for the 0-0.1 m depth reveals that changes in soil EC over time is primarily located on irrigated farms, with no notable changes predicted under the natural or dryland land uses. Some cotton farms experienced a decrease in EC, and several areas under irrigated horticultural production underwent an increase in EC at the 0-0.1 m depth (Fig. 5.6). The 0.1-0.3 m depth revealed very similar EC levels to the 0-0.1 m maps, as well as similar levels and patterns of change over the 13-year period, but a greater number of irrigated cotton farms were predicted to have experienced a desalination trend (Fig. 5.6).

The 0.3-0.5 m depth displayed comparable spatial patterns of soil EC to those depths above (Fig. 5.7). Similarly, no change in EC under the natural and dryland land uses was observed at this depth, and the same desalination trend was observed at some cotton farms, while the majority of horticultural farms continued to experience an increase in soil EC at the 0.3-0.5 m depth (Fig. 5.7). The 0.5-0.8 m depth maps are quite contrasting to those of the upper three depths, as the spatial patterns are clearly not as strongly driven by land use. The EC values are also substantially higher than the preceding layers. The change map showed that there were predicted decreases in EC in the south-west of the study area, with a predicted increase in EC in the north and north-east (Fig. 5.7).

The 0.8-1.2 m depth maps of EC also show that there is little influence of land use on the spatial patterns of soil EC. The pattern of temporal EC change is similar to that of the 0.5-0.8 m depth; however it is more widespread (Fig. 5.8). The EC maps of the final depth (1.2-1.5 m) show very high EC values in some locations, with considerable portions of the study area possessing EC values greater than 1.5 dS/m (Fig. 5.8). Both large decreases and increases in soil EC were predicted over time at this depth. The magnitude of EC levels are much higher than those of the upper soil profile at this depth, so the change maps must be viewed with this in mind (Fig. 5.8).

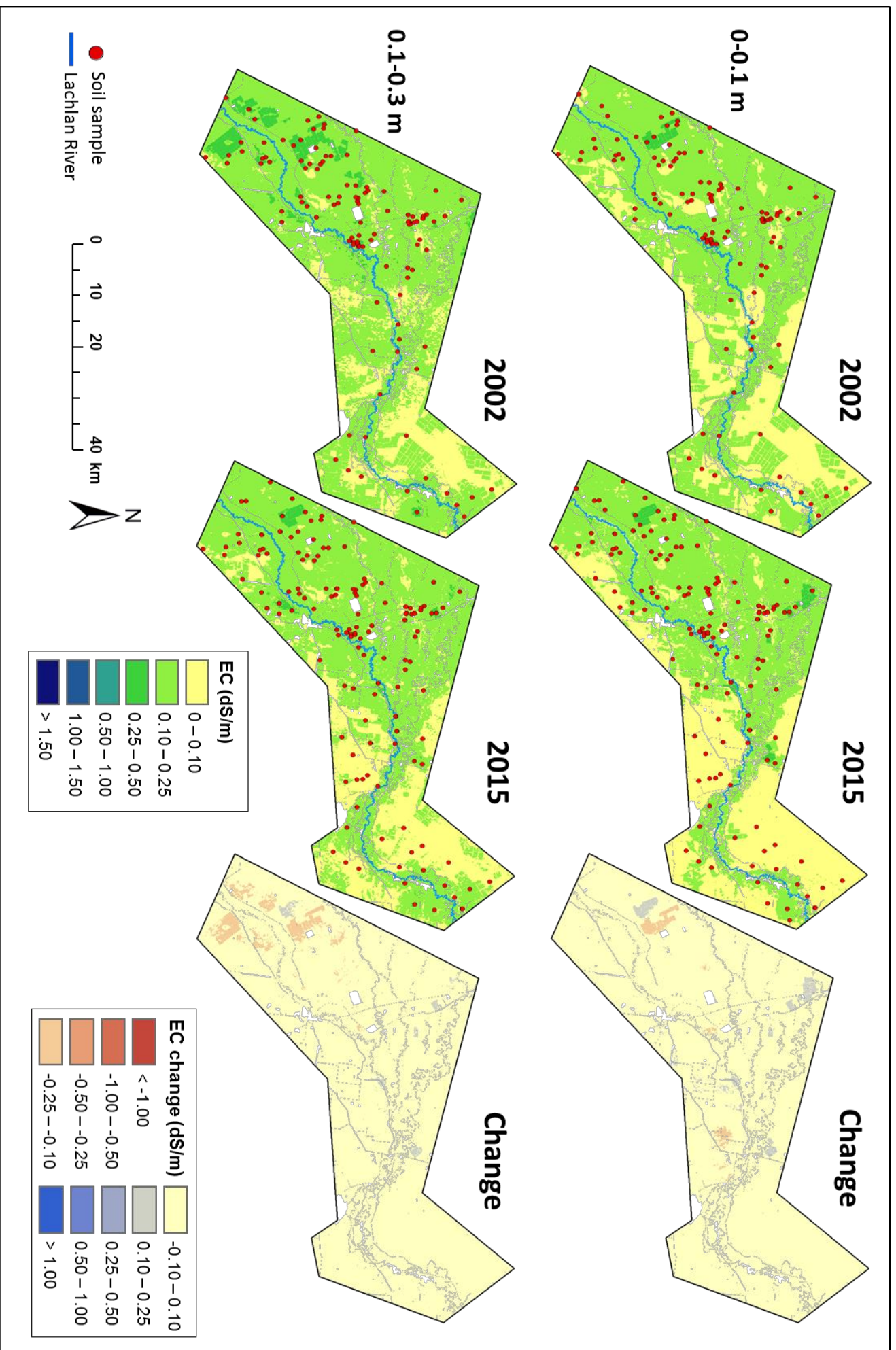


Fig. 5.6 – Electrical conductivity (EC) for 2002 and 2015 for the 0-0.1 m and 0.1-0.3 m depths and the EC change (ds/m 1:5 soil: water)

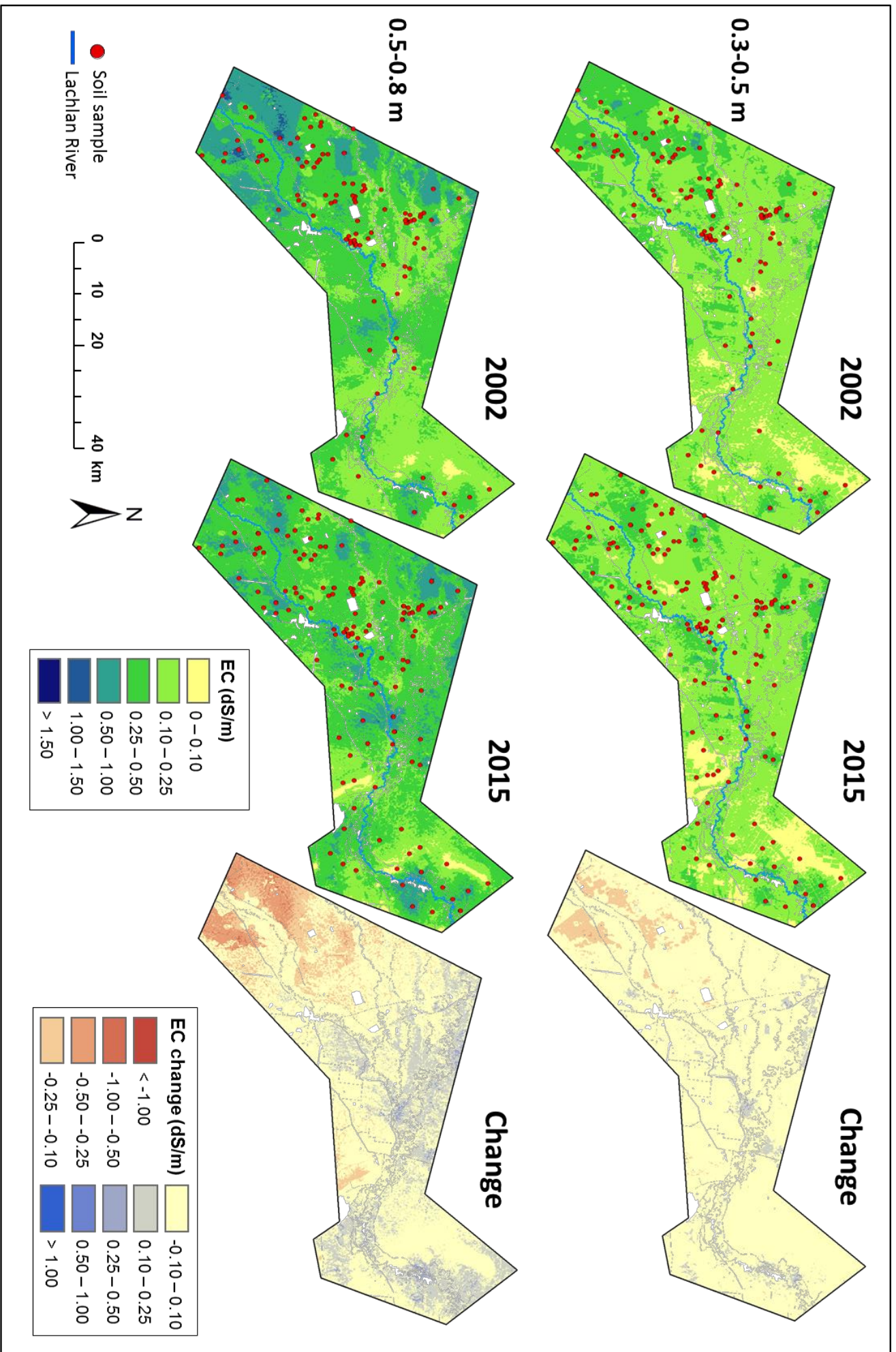


Fig. 5.7 – Electrical conductivity (EC) for 2002 and 2015 for the 0.3-0.5 m and 0.5-0.8 m depths and the EC change (ds/m 1:5 soil: water)

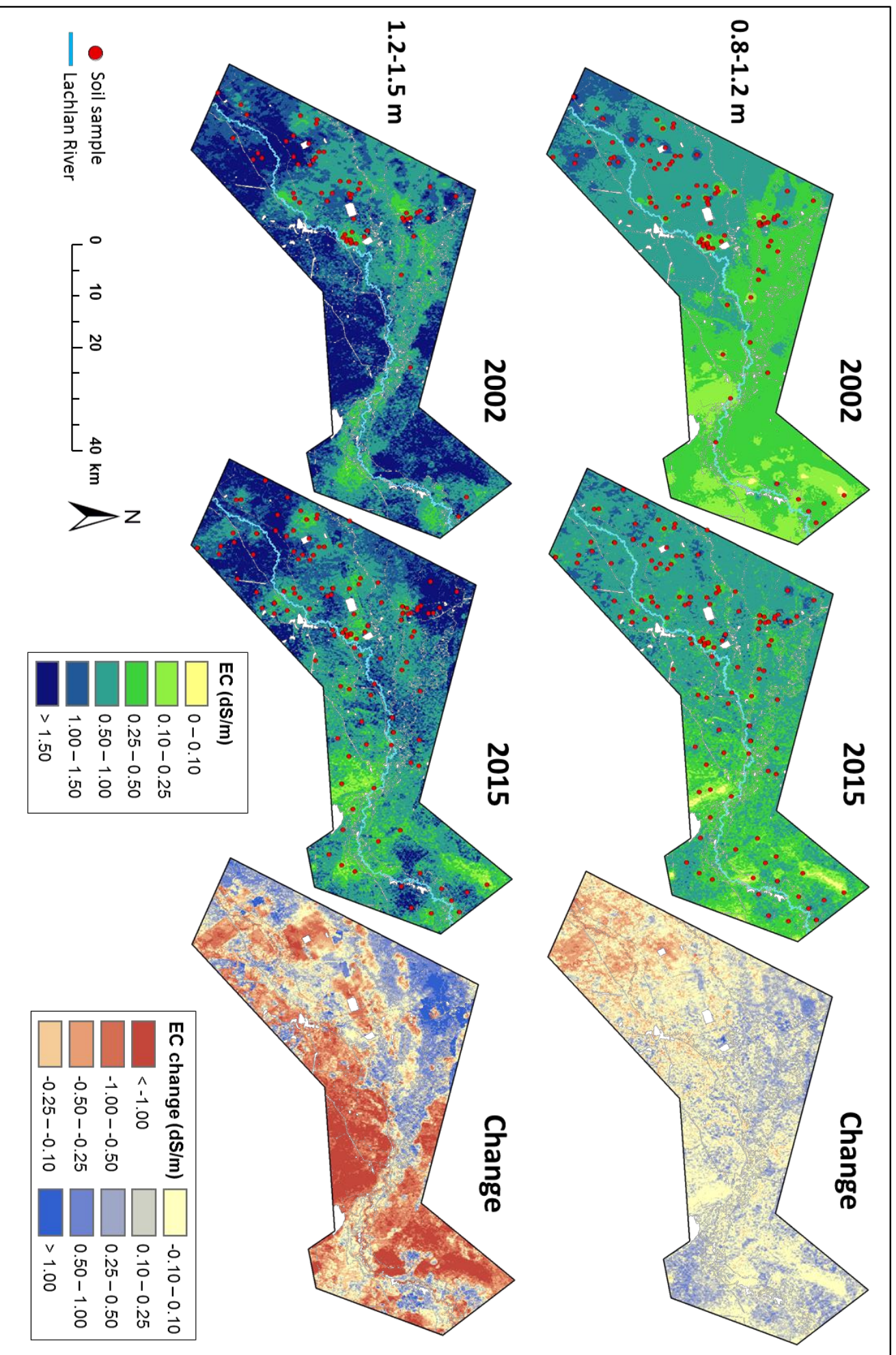


Fig. 5.8 – Electrical conductivity (EC) for 2002 and 2015 for the 0.8-1.2 m and 1.2-1.5 m depths and the EC change (ds/m 1:5 soil: water)

5.3.5.2. Soil ESP modelled maps

The different magnitudes of ESP values between the two different sampling depths are stark (Fig. 5.9). In the 0-0.1 m depth, higher ESP values were found on the alluvial floodplains west of the Lachlan River, although only small parts of the study area had ESP values greater than 6% at the 0-0.1 m depth at both time points. The change map reveals that parts of the study area was predicted to have experienced an increase in ESP between the two surveys, with most of this occurring in the cotton-growing and natural areas west of the Lachlan River. There were also some very isolated areas that were predicted to have undergone a decrease in ESP in the 0-0.1 m depth. The 0.8-1.2 m depth maps display a similar spatial pattern of ESP to those of the topsoil, with the higher ESP values located on the alluvial floodplains, however, almost the whole study area possesses ESP values above 6%, apart from some isolated areas at higher elevations (Fig. 5.9). The change map shows that large parts of the study area was predicted to have experienced an increase in ESP over time, with this not confined to any particular location (Fig. 5.9). While the impact of land use is still somewhat discernible in these ESP maps, it is apparent that land use does not drive the spatial patterns of ESP as intensely as it does in the EC maps (Fig. 5.9).

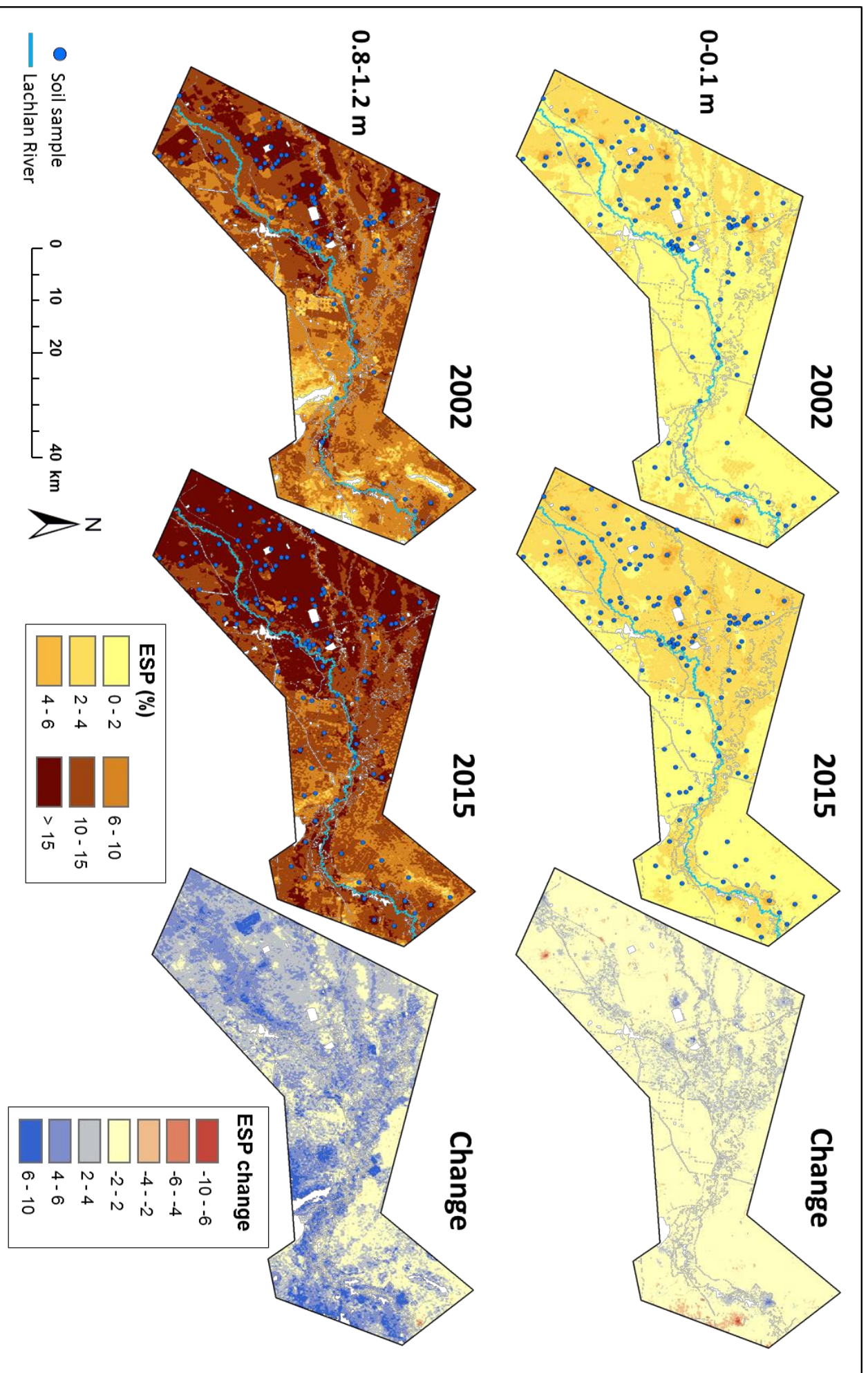


Fig. 5.9 – Exchangeable sodium percentages (ESP) for 2002 and 2015 for the 0-0.1 m and 1.2-1.5 m depths and change in ESP (%)

5.3.6. Statistical significance of predicted change

5.3.6.1. Soil EC

The six sampling depths to 1.5 m analysed in this study underwent varying directions, amounts, and degrees of statistically significant changes in EC (Fig. 5.10). Most of the statistically significant increases observed at the different sampling depths were confined to areas under irrigated horticultural production. Much of the statistically significant decreases in EC were located on irrigated cotton farms, as well as areas at higher elevation. The 1.2-1.5 m depth showed the most widespread statistically significant (p -value <0.05) change in EC, with both increases and decreases in soil salinity predicted across considerable parts of the study area (Fig. 5.10). While these z-score maps of temporal EC change indicate that both statistically significant increases and decreases have occurred, this does not necessarily mean that this change has been large enough to be of agricultural importance. This is particularly true for the topsoil (0-0.1 m), where no parts of the study area possessed EC values that would impact plant production (Fig. 5.10).

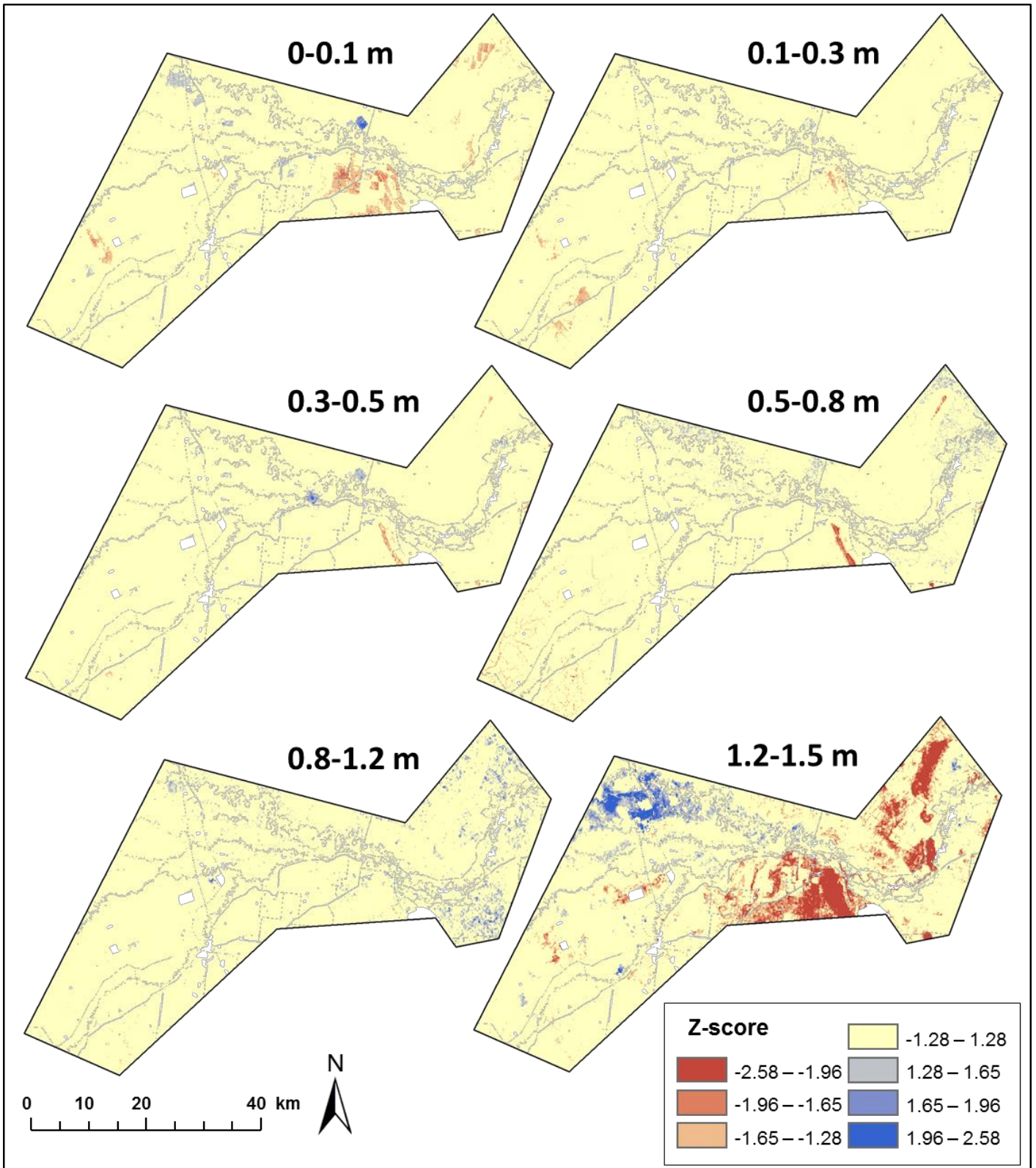


Fig. 5.10 – Z-scores showing the statistical significance of the change in EC (dS/m 1:5 soil: water) for the six different sampling depths

5.3.6.2. Soil ESP

While the standalone change maps showed widespread increases in soil ESP in both the 0-0.1 m and 0.8-1.2 m depths (Fig. 5.9), only isolated parts of the study area underwent a statistically significant change in soil sodicity over time at these sampling depths (Fig. 5.11). The 0-0.1 m depth experienced relatively more change than the 0.8-1.2 m depth, with several locations predicted to have undergone a statistically significant increase in ESP (p -value 0.05-0.10). There seemed to be no clear pattern of the locations of this statistically significant change at either depth, however, most of this change occurred on either irrigated cotton farms, or soils under irrigated horticultural production.

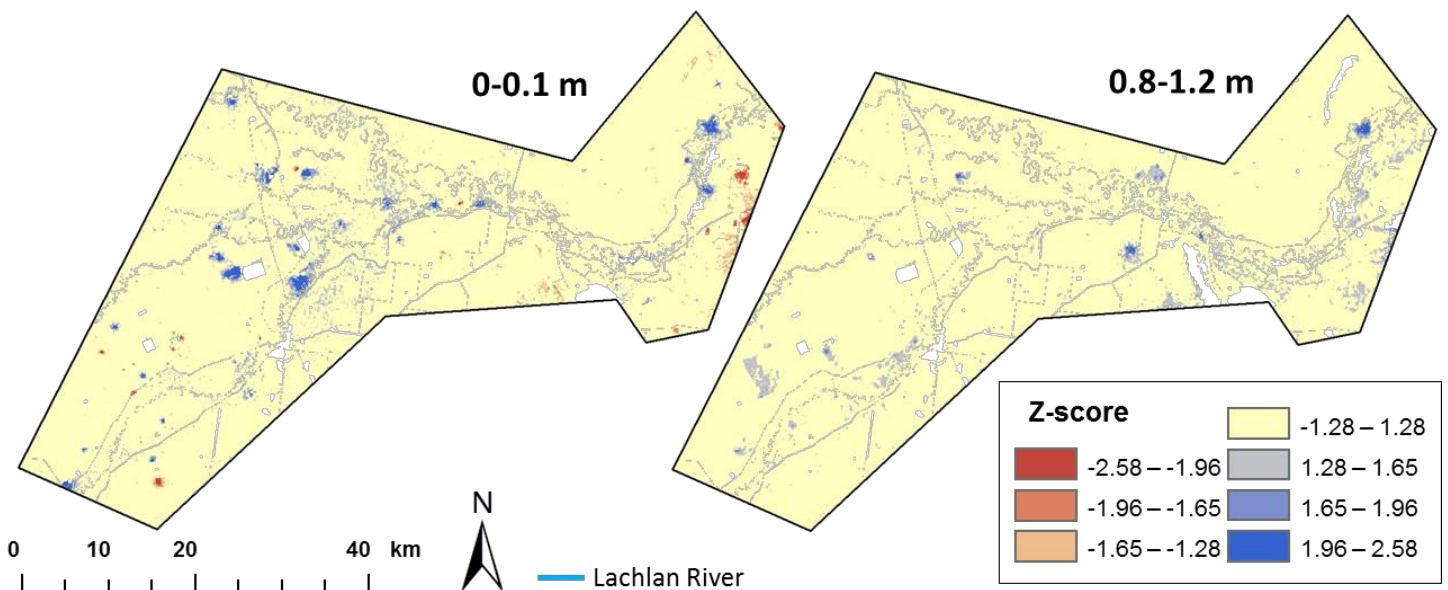


Fig. 5.11 – Z-scores showing the statistical significance of the change in ESP for the 0-0.1 m and 0.8-1.2 m sampling depths

5.4. Discussion

5.4.1. Changes in soil properties and possible driving factors

5.4.1.1. Salinity

While many semi-arid regions throughout the world possess soils with typically high levels of salinity, the soils of the Hillston district were found to be generally non-saline. Soil salinity increased as depth increased, which is typical, and only isolated parts of the study area at the deeper subsoil layers had EC values that would inhibit crop growth. Cotton is considered relatively tolerant to soil salinity, and can endure EC values up to 1.03 dS/m in a medium clay soil before impacts on plant growth are experienced (Hazelton and Murphy 2007). The legends used in figures 5.6, 5.7 and 5.8 are based on different soil

salinity ratings defined in Hazelton and Murphy (2007), and essentially show that any areas of these maps that are blue (>1.00 dS/m) could potentially limit cotton production. Important crops that are rotated with cotton in the Hillston district, such as wheat, are less tolerant of soil salinity, and EC values greater than 0.80 dS/m in medium clay soils can impact yield (Hazelton and Murphy 2007).

The 2002 results showed that initial topsoil (0-0.1 m) EC levels were more than double in irrigated cotton sites compared to natural sites, and the outline of some cotton farms can be seen in the 2002 EC maps. This suggests that irrigated cotton production had caused an increase in topsoil salinity from natural levels initially. This trend, however, did not continue into 2015, with a mean decrease in EC for most sites under cotton production in the topsoil. This desalination trend was also observed throughout much of the soil profile for cotton sites, however, no appreciable changes in soil salinity were observed for natural sites. In contrast, soils under irrigated horticulture generally experienced an increase in soil EC throughout the soil profile over time. While the standalone change maps for EC often showed widespread shifts for the various sampling depths, it was clear that not all of this was statistically significant when analysed (Fig. 5.10). In the three sampling depths in the upper 0.5 m of the soil profile, most of the statistically significant decreases occurred in areas under irrigated cotton production, and all of the statistically significant increases occurred under irrigated horticulture. The most widespread statistically significant changes occurred in the 1.2-1.5 m depth, where a large cotton farm in the north-west of the study area that had recently shifted to irrigated perennial horticulture underwent a significant increase in soil EC. Statistically significant decreases were also predicted under some irrigated cotton areas in the southwest of the study region, as well as in soils of the sandstone outcrops at high elevations in the deepest layer. These changes in soil salinity are likely linked to contrasting agricultural management practices, as it was clear that land use had a substantial impact on the degree and direction of change in soil salinity.

The lack of change in soils under the natural land use is somewhat expected, as it is generally accepted that these parts of the landscape are in equilibrium, and that insignificant shifts in EC would occur over a 13-year time period. In contrast, soil salinity is expected to be temporally less stable under irrigated land uses, as irrigation has the capacity to either increase or decrease salinity levels, even over very short time periods. Studies have found that the introduction of irrigation can markedly decrease soil salinity levels (Cetin and Kirda 2003), as well as increase salinisation (Shirokova *et al.* 2000) over time periods as little as two years. The quantity and quality of water applied plays an important role in determining the direction and severity of these changes. Increased irrigation can cause salts to be

flushed out of the soil profile, however, when irrigation water is of low quality and contains high concentrations of salt, there is a potential to increase soil salinity from baseline levels.

Australian groundwaters are dominated by sodium chloride and typically contain higher concentrations of salt than surface/river water, but, this can vary in both space and time (Speirs *et al.* 2011). Contrastingly, groundwater in the Hillston region is considered to be of high quality, and often contains lower levels of salt than surface water extracted from the Lachlan River. A study by Speirs *et al.* (2011) found that the groundwater at Hillston contained an EC level of 0.45 dS/m, and an SAR of 3.38 mmol/L^{1/2}, although, it is known that groundwater quality varies throughout the study area. Data from the NSW Government suggests that surface water quality is slightly lower, with an average EC of 0.54 dS/m during the study period (2002 to 2015). Surface water quality was also found to be highly variable, with a minimum EC of 0.10 dS/m observed in 2010, and a maximum of 1.04 dS/m in 2004. It is possible that the use of saltier, poorer quality water in irrigated cotton farms leading up to the 2002 survey could be the reason for the higher baseline soil salinity levels observed in the upper 0.5 m of the soil profile compared to natural sites.

The contrasting shifts in soil EC of irrigated cotton and irrigated horticultural production is likely linked to the different management and irrigation practices implemented by these different land uses. The horticulture land use is primarily made up of irrigated perennial orchards that require continual irrigation for survival, and thus these farms possess high security licences that are minimally affected by rainfall patterns and irrigation water availability. High EC values and statistically significant temporal increases in EC were observed in soils under the horticulture land use in this study, and this provides an insight into the impact of continuously irrigating soil in this semi-arid region. While this salinisation trend may be a result of irrigation water adding salt to the soil, the use of certain fertilisers that contain salt is also another plausible process (Darwish *et al.* 2005). Fertilisers that contain salts are typically not used in cotton production, but are very common in horticultural production, particularly on the almond and citrus orchards that are abundant in the Hillston district. This could be a reason for the observed high levels of salinity and modelled increases over time at most sampling depths under this land use. Irrigated perennial horticulture is expanding in the Hillston district, and the results found in this study suggest that soil condition under this land use should continue to be monitored.

Over the 13-year study period, the source and the availability of irrigation water were inconsistent due to the lengthy drought and periods of high rainfall, and this varied for each water licence holder that produced cotton. During the 2002-09 Millennium Drought in Hillston, surface water

was scarce, and only irrigated cotton farms that had access to groundwater (approximately 50% of the total irrigated area of the Hillston district) were able to continue to irrigate, while the remaining irrigated cotton farms were primarily used for dryland cropping or left in fallow. When rainfall patterns changed abruptly in 2010, there was suddenly an abundance of irrigation water available to growers in the Hillston district for several years. Despite this varied irrigation management of cotton farms, most of the change detected under irrigated cotton farms in our study indicated that a decrease in salinity had occurred at the various depths, with only minor areas predicted to have increased over time. We do not have detailed data about the quantity and quality of irrigation water applied on specific cotton fields/farms in the study period, which makes it difficult to fully understand the effects of this on changes in soil salinity. In addition, without soil salinity measurements at the junction of the shift in water availability in 2009-2010, it is challenging to understand the impact of management practices and water dynamics. Nevertheless, the changes in soil salinity under irrigated cotton and horticultural farms indicate that the quality and quantity of irrigation water, the application of salt-containing fertilisers, and the variable rainfall patterns have had a discernible effect on soil salinity levels in the Hillston district from 2002 to 2015.

Comparable temporal trends in soil salinity in Vertosols have also been detected in an irrigated cotton-growing system in southern Queensland from 1996-2014 (Melland *et al.* 2016). The authors determined that the use of marginal quality groundwater for irrigation had increased soil salinity levels initially, but over the subsequent two decades of higher quality surface water irrigation, much of the accumulated salt had leached out of the upper root zone and beyond one metre. While much of this leaching of salt was attributed to the use of higher quality surface water instead of groundwater, the significant floods experienced in 2010 to 2011 had also likely played an important role in the desalination (Melland *et al.* 2016). Similar soil desalination trends have also been observed in a semi-arid region of Spain as a result of irrigation, where sites were sampled and revisited at three time points over a 24-year period, showing that soil salinity levels decreased to the depth of 1 metre (Herrero and Pérez-Coveta 2005).

To manage soil salinity and prevent further salinisation at Hillston, the best quality irrigation water available should be used. When water quality is significantly low, it may require mixing with other available water sources to reduce salt input into the soil and to reduce the negative impact on crops. Reducing the salt input into the system, rather than relying on leaching of salts is a more favourable long-term approach, as this may eventually cause salinity in the broader landscape (Melland *et al.* 2016).

Despite this, using higher quality water could have potentially detrimental impacts on soil structure, as will be discussed shortly. In addition, the use of fertilisers that contain significant amounts of salt in horticultural production should be monitored, as the continued use of this may continually increase salinity levels in the soil.

5.4.1.2. Sodicty

According to the Australian Soil Classification (ASC), a sodic soil is characterised by possessing an ESP of 6% or higher, however, the World Reference Base (the international soil classification system) and Soil Taxonomy (that of the United States of America), consider a soil to be sodic if it possesses an ESP of 15% or greater. Our study found that sodicity levels in the topsoil were only limiting by Australian standards in small sections of the study area, however, large parts possessed very high levels of subsoil sodicity by both Australian and International standards. While the soils in the Hillston district are inherently sodic, it is likely that agricultural management practices have impacted the degree of sodicity. When water is applied to sodic soils they swell excessively, the clay particles disperse, causing structural collapse, and ultimately waterlogging and permeability issues in soils. This structural degradation can be further accentuated by mechanical disturbance that is common practice in the Australian cotton industry (Rengasamy and Olsson 1991).

Widespread trends of increasing soil ESP in both the topsoil and subsoil were found in the study area between 2002 and 2015, but only isolated areas underwent a statistically significant increase in ESP over time in the 0-0.1 and 0.8-1.2 m depths (Fig. 5.11). There were no clear differences in ESP change under the different land uses, however, most of the statistically significant increases in ESP occurred in areas under irrigated cotton production or irrigated horticulture. When viewing these z-score maps of temporal ESP change, it must be taken into account that a portion of the 2015 soil sample data was predicted with VisNIR, and that there is a greater amount of error with this. It is possible to include the error of the VisNIR predictions of ESP for each corresponding sampling depth in the calculation of the contrast variance (Equation 5.1). However, doing so would provide a conservative estimate of the change, as not all of the samples were predicted by VisNIR, and this may consequently mask some of the statistically significant change. Despite this, the sites that were sampled in both 2002 and 2015, and measured by traditional laboratory methods showed mean increases in soil ESP in the five depths to 1.2 m (Fig. 5.4). This suggests that the increases in ESP displayed in the maps (Fig. 5.9; 5.11) are not just an extrapolation of the models, but reflect actual changes that have occurred in the levels of soil sodicity at Hillston. While there are a few long-term studies that monitor spatio-temporal changes in soil salinity

over a region, there are no studies, to our knowledge, that focus on soil sodicity, despite it being a widespread and severely limiting global soil degradation issue (Rengasamy and Olsson 1991). This is likely due to the time and resource consuming nature of measuring soil ESP. There are, however, some studies that have mapped soil sodicity at one time point at the field (Corwin *et al.* 2003) and regional scale (Ford *et al.* 1993).

Soil salinity has been a relatively minor issue for the Australian cotton growing industry, but soil sodicity has caused dramatic losses in production, often undiagnosed (McKenzie and Orange 2003). The Vertosol soil types used for cotton production in Australia are very deep, meaning that the majority of salt added from irrigation water is washed past the plant-root-zone. Despite this, sodium ions often attach to clay particles and displace other ions when water percolates through the soil, constantly increasing the sodicity of the soil (Rengasamy and Olsson 1991), and this is further accentuated by irrigating with poor quality water (Rengasamy and Walters 1994). A small-scale study in Iran demonstrated this, where increases in soil ESP occurred as a result of application with poor quality wastewater (Jalali *et al.* 2008). While the increase in ESP in soils under irrigated cotton production observed in our study can be logically explained by the process described above, the reasons for the observed increase in ESP in paired-sites under the natural land use in this study is less clear.

The observed trend of decreasing soil salinity and increasing soil sodicity in the cotton growing district of Hillston creates concern about the potential future impacts on the soil resources and crop productivity. As mentioned previously, the ratio of soil salinity (EC) and sodicity (ESP) determines whether a soil will stay aggregated or disperse when irrigated. When the soil ESP increases, or soil EC decreases, there is a greater potential for soil to swell and expand upon wetting (McNeal 1968). In addition, The EC and SAR values of the irrigation water used also impacts this. Irrigating with higher salinity water can add more salt to the soil system and have a detrimental impact on crop growth, but it tends to cause clay particles to stay together and maintain soil structure (Sumner 1993). In contrast, if salinity of the applied water is low relative to soil EC and ESP, swelling and dispersion of clay particles occurs, resulting in a degraded soil structure (Shainberg *et al.* 1981; Rengasamy and Olsson 1991; Warrence *et al.* 2003). In semi-arid irrigated landscapes, the salt concentration of irrigation water is usually quite high, however, rainfall contains minimal salts, and when this percolates through soil it can have adverse effects on soil physical properties, even when soils possess a low ESP (Rengasamy and Olsson 1991). The recent patterns of lengthy drought, followed by prolonged periods of high rainfall and

flooding in the semi-arid regions of eastern Australia creates concern about potential impacts on the physical deterioration of these sodic soils (Shainberg *et al.* 1981).

Soil salinity is typically more easily altered and ameliorated than soil sodicity, and in deep alluvial soils, continuous flushing of soil with high quality water can reduce soil salinity, but often cannot reduce the levels of sodicity (Warrence *et al.* 2003). Attempts to ameliorate subsoil sodicity are challenging, and typically involve short term methods to reduce the negative effects on crop production (Rengasamy and Olsson 1991). The lack of an effective amelioration approach and typically poor diagnosis of sodicity, suggests that the issue of soil sodicity will continue to grow, and with this, the solutions for curtailing sodicity will become increasingly difficult (Rengasamy and Olsson 1991). Nonetheless, the management solution to reduce the salinity levels of soil and prevent further sodification (irrigate with high quality water) is antagonistic to the management solution to reduce the impact of sodic soil on crop production (irrigate with saline-sodic water). It is therefore important for growers that use irrigation to know the values of EC and ESP of their soils, and to use this to make a decision on the quality of irrigation water to apply.

5.4.2. Prediction of ESP by VisNIR

Exchangeable sodium percentage is generally not well predicted by VisNIR spectra, as it does not have a primary response in the VisNIR region (Zornoza *et al.* 2008). Studies have reported poor prediction accuracies ranging from 0.09 to 0.44 R^2 for exchangeable sodium (Stenberg *et al.* 2010). In spite of this, we were able to predict ESP with high accuracy (0.65 R^2 or 0.79 LCCC), with several plausible reasons for these high quality predictions. Rather than identifying specific areas in the spectra that correspond to exchangeable sodium, our models may be recognising areas that correspond to other soil properties that are correlated with ESP. Clay content and CEC of soil are known to be well predicted by VisNIR spectroscopic techniques and generally assumed to have a good correlation with ESP (Viscarra Rossel *et al.* 2006), however, we found very poor correlations (Pearson's) of ESP with CEC ($r = 0.05$) and clay content ($r = 0.08$) in our study, which is confounding. VisNIR identifies colour very well, and it is possible that the visible colour part of the spectra is contributing to these accurate predictions of ESP. At Hillston, there is a strong correlation between soil colour and ESP, with the Red Vertosols typically possessing yellow subsoils that are highly sodic, while the less-sodic soils are typically darker in colour (Grey/Brown) (Fig. 5.5). In addition, high reflectance in the visible band also occurs as a result of the presence of white carbonates. The presence of carbonates would likely be positively correlated with high exchangeable calcium and therefore likely to be negatively correlated with high ESP.

In addition to this, the large and comprehensive calibration dataset available for the study area undoubtedly contributed to the quality of the predictions. This is likely accentuated by the resampling of the same sites in the second survey, which consequently increases the probability that a similar soil would be present in the calibration dataset. Despite the accurate predictions of ESP in this study, the limitations of using spectroscopic techniques in soil monitoring must be acknowledged. While it is generally assumed that ESP is poorly predicted due to the lack of a spectral signature, and that any good predictions are a result of correlation with other well-predicted soil properties, we observed that both the predicted data and laboratory measured data showed increases in ESP over time. Time was not included as a predictor variable, which suggests that there must be some aspect of the soil that has changed over time that is correlated with ESP. Overall, the results presented here suggest that using VisNIR spectra and Cubist models to predict soil ESP shows promise for using this as a viable approach of rapidly detecting soil ESP in the future.

In our soil mapping approach, we have treated the laboratory-measured and the spectroscopic-predicted values as equal. It would be logical to give greater weighting to the laboratory measured data as it is a more accurate representation of actual ESP values in the soil. The different uncertainty in the measurement sources could be treated more formally in the spatial modelling, as suggested by Orton *et al.* (2014), and should be considered in future work.

5.4.3. Bivariate linear mixed models

5.4.3.1. Model quality and future opportunities

The BLMMs created in this study had high predictive power with low prediction variance. The eight fitted BLMMs displayed mean and median standardised squared prediction errors (SSPE) that were close to the expected values of 1.0 and 0.455, respectively, suggesting that the prediction variances represented actual errors. We have taken advantage of the co-location and correlation of soil information from different time points in our approach to create digital soil maps that have a logical connection through time. There are very few studies that utilise this co-variance through time to create digital maps of any soil property, with most studies opting for univariate approaches. There are, however, further opportunities to utilise correlation from other co-located soil data, such as other soil depths within a soil core/sampling location or other soil properties to improve our models and predictions (Orton *et al.* 2014). There is potential in the future to use multivariate LMMs where one model is created for multiple time points, at multiple depths, and for multiple soil properties. The main

limitation to this approach is the amount of computing power required to develop such a model (Orton *et al.* 2014).

5.4.3.2. Comparison of univariate and bivariate linear mixed models

One of the biggest advantages of using BLMMs through time for soil monitoring, as we have done in this chapter, compared to univariate approaches is that more coherent and logical soil maps are often produced as there is a greater cohesion between the two time points. This is because when a site has only been sampled at one time point, the soil information from the sampled time point at that site is utilised in the unsampled time point. Fig. 5.12 shows two topsoil (0-0.1 m) EC maps for 2002 and 2015 produced by separate univariate linear mixed models (ULMMs). The clear differences at the two time points can be seen in the north-east and centre of the maps (Fig. 5.12). In 2002, these areas were predicted to contain high EC values (blue), and in 2015 these areas were predicted to possess low EC values (yellow). In 2002, there was a lack of soil cores extracted at these locations and this caused the model to make assumptions of EC values and extrapolate predictions. These areas possess distinct differences in covariate values to the rest of the study area; they are at higher elevation and contain typically sandier soils. When we compare the maps produced from ULMMs (Fig. 5.12) to those produced from BLMMs (Fig. 5.6), we can see that this issue is non-existent for the BLMM approach. This is because the soil information from 2015 at these locations is utilised in the 2002 maps. The white areas within Fig. 5.12 are NA values, as no predictions were made at these locations. This was done as there was no soil samples used for modelling that represented the covariates at these sites.

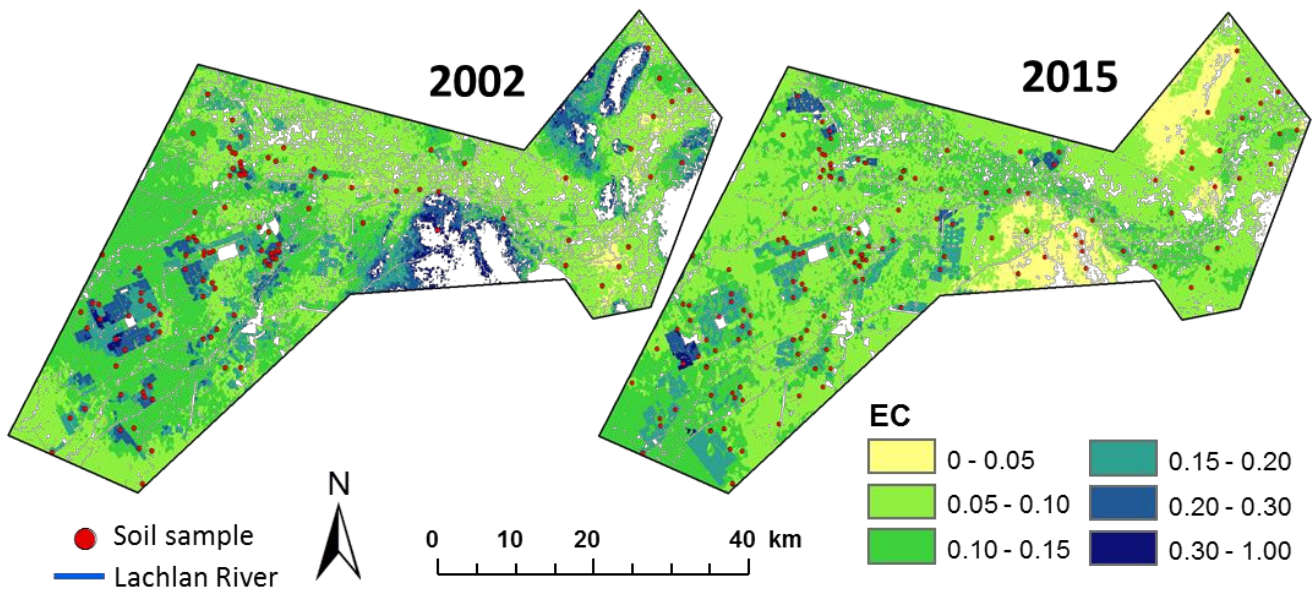


Fig. 5.12 – Soil electrical conductivity (EC) values for (1:5 soil: water dS/m) for 0-0.1 m sampling depths in 2002 (left) and 2015 (right) produced from separate univariate linear mixed models (ULMMs)

5.5. Conclusions & future directions

This study showed that the semi-arid cotton-growing region of Hillston in the lower Lachlan catchment generally had low levels of soil salinity, despite some high and potentially limiting EC levels in the subsoil. Results showed that the study area experienced both increases and decreases in soil salinity over the survey period. While all depths experienced statistically significant change to a degree, the most widespread changes occurred in the 1.2-1.5 m depth. Land use played an important role on the trajectory and degree of change in salinity at the various depths, and no considerable changes were observed in soils under the natural land use. Overall, a trend of a decrease in salinity was observed in some areas under irrigated cotton production, with this attributed to the leaching of salts out of the profile through the application of high quality irrigation water. In contrast, several irrigated horticultural farms experienced an increase in soil EC over time at various sampling depths, with this linked to the continual use of fertilisers that contain salts, as well as the use of poorer quality irrigation water. Irrigated horticulture is expanding in the study area, and soil salinity under this land use should continue to be monitored. Overall, it was clear that the varying trajectories of soil EC under irrigated land uses were due to the varying quantity and quality of irrigation water used over the 13-year study period. Soil sodicity at Hillston was found to be generally low to moderate in the upper 0.5 m of the soil profile, but very high in the lower 0.5-1.5 m depths, leading to potentially undesirable impacts on plant production.

A trend of increasing soil sodicity through time was observed in the study area. While the impact of land use on changes in soil ESP was less clear than that of EC, most of the statistically significant change occurred in areas under irrigated cotton and horticultural production. These statistically significant changes must be viewed with caution, as a portion of the 2015 samples were predicted by VisNIR spectroscopic methods. The observed trend of increasing ESP and decreasing EC is concerning for areas under irrigated agriculture. The relative values of sodicity and salinity are crucial in determining soil structural integrity upon irrigation, and the changes observed in this study could lead to heightened waterlogging issues and losses in plant productivity. It is therefore essential that soil salinity and sodicity both continue to be monitored in conjunction to analyse the direction of any future changes. More detailed information on the quantity and quality of irrigation water applied on irrigated farms should be considered in any similar studies in the future, as this would be extremely useful in understanding the influence of different irrigation management on temporal shifts of both soil salinity and sodicity. Very few studies monitor soil salinity and sodicity, particularly in the subsoil, and further research in this needs to be performed, as the subsoil is often where these soil constraints are most limiting to agricultural production. Our study showed that soil ESP could be predicted to high accuracy using VisNIR spectra and Cubist models, which shows promise for using this as a viable approach of rapidly detecting soil ESP in the future. The spatio-temporal bivariate linear mixed models (BLMMs) used to model soil EC and ESP in this study showed high predictive power, and the advantages of using BLMMs compared to univariate approaches were demonstrated. Opportunities to utilise other co-located data, such as additional soil depths or other soil properties was identified and should be taken into consideration for future research.

References

- Bishop, TFA, McBratney, AB, Laslett, GM (1999) Modelling soil attribute depth functions with equal-area quadratic smoothing splines. *Geoderma* **91**, 27-45.
- Bureau of Meteorology - BOM (2017) Monthly rainfall – Hillston airport. Available from: <http://www.bom.gov.au/jsp/ncc/cdio/weatherData/av?p_nccObsCode=139&p_display_type=dataFile&p_startYear=&p_c=&p_stn_num=075032> [25 April 2017]
- Cameron, RG, Colquhoun, GP, Meakin, NS, 2000. Provisional Hillston 1:100 000 Geological Series Sheet. Geological Survey of New South Wales, Sydney, Australia.
- Cetin, M, Kirda, C (2003) Spatial and temporal changes of soil salinity in a cotton field irrigated with low-quality water. *Journal of Hydrology* **272**, 238-249.
- Corwin, DL, Kaffka, SR, Hopmans, JW, Mori, Y, van Groenigen, JW, van Kessel, C, Lesch, SM, Oster, JD (2003) Assessment and field-scale mapping of soil quality properties of a saline-sodic soil. *Geoderma* **114**, 231-259.

- Darwish, T, Atallah, T, El Moujabber, M, Khatib, N (2005) Salinity evolution and crop response to secondary soil salinity in two agro-climatic zones in Lebanon. *Agricultural Water Management* **78**, 152-164.
- Filippi, P, Minasny, B, Cattle, SR, Bishop, TFA (2016) Monitoring and Modeling Soil Change: The Influence of Human Activity and Climatic Shifts on Aspects of Soil Spatiotemporally. *Advances in Agronomy* **139**, 153-214.
- Ford, G, Martin, J, Rengasamy, P, Boucher, S, Ellington, A (1993) Soil sodicity in Victoria. *Soil Research* **31**, 869-909.
- Hazelton, P, Murphy, B (2007) Interpreting soil test results - What do all the numbers mean? CSIRO Publishing, Collingwood, Australia.
- Herrero, J, Pérez-Coveta, O (2005) Soil salinity changes over 24 years in a Mediterranean irrigated district. *Geoderma* **125**, 287-308.
- Hulugalle, NR, Weaver, TB, Finlay, LA (2013) Soil water storage, drainage, and leaching in four irrigated cotton-based cropping systems sown in a Vertosol with subsoil sodicity. *Soil Research* **50**, 652-663.
- Jalali, M, Merikhpour, H, Kaledhonkar, MJ, Van Der Zee, SEATM (2008) Effects of wastewater irrigation on soil sodicity and nutrient leaching in calcareous soils. *Agricultural Water Management* **95**, 143-153.
- Lark, RM (2000) Estimating variograms of soil properties by the method-of-moments and maximum likelihood. *European Journal of Soil Science* **51**, 717-728.
- Lark, RM, Papritz, A (2003) Fitting a linear model of coregionalization for soil properties using simulated annealing. *Geoderma* **115**, 245-260.
- McKenzie, D., Orange, P.L.M., (2003) Salinity and sodicity – what’s the difference. *The Australian Cottongrower*, **24**, 28.
- McNeal, BL (1968) Prediction of the Effect of Mixed-Salt Solutions on Soil Hydraulic Conductivity. *Soil Science Society of America Journal* **32**, 190-193.
- Melland, A., Biggs, A., Silburn, M., Shaw, R., (2016) Changes in soil salinity, sodicity and nitrate under irrigated cotton. *The Australian Cottongrower* **37**, 26-31.
- Minasny, B, McBratney, AB (2006) A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers & Geosciences* **32**, 1378-1388.
- Mora, JL, Herrero, J, Weindorf, DC (2017) Multivariate analysis of soil salination-desalination in a semi-arid irrigated district of Spain. *Geoderma* **291**, 1-10.
- Northcote, K.H. and Skene J.K.M., (1972) Australian soils with saline and sodic properties. CSIRO Soil Publication 27.
- Orton, TG, Pringle, MJ, Page, KL, Dalal, RC, Bishop, TFA (2014) Spatial prediction of soil organic carbon stock using a linear model of coregionalisation. *Geoderma* **230–231**, 119-130.
- Papritz, A, Flühler, H (1994) Temporal change of spatially autocorrelated soil properties: optimal estimation by cokriging. *Geoderma* **62**, 29-43.
- Rayment, G. E., Lyons, D. J. (2011). Soil chemical methods: Australasia, 3. CSIRO publishing.
- Rengasamy, P, Olsson, K (1991) Sodicity and soil structure. *Soil Research* **29**, 935-952.
- Rengasamy, P., Walters, L., (1994) Introduction to soil sodicity. Cooperative Research Centre for Soil and Land Management.
- Rengasamy, P (2006) World salinization with emphasis on Australia. *Journal of Experimental Botany* **57**, 1017-1023.
- Roth, G., (2014) Australian grown cotton sustainability report. Cotton Research and Development Corporation and Cotton Australia.
- Shahid, SA (2013) Developments in Soil Salinity Assessment, Modeling, Mapping, and Monitoring from Regional to Submicroscopic Scales. In 'Developments in Soil Salinity Assessment and

- Reclamation: Innovative Thinking and Use of Marginal Soil and Water Resources in Irrigated Agriculture.' (Eds SA Shahid, MA Abdelfattah, FK Taha.) pp. 3-43. (Springer Netherlands: Dordrecht)
- Shainberg, I, Rhoades, JD, Prather, RJ (1981) Effect of Low Electrolyte Concentration on Clay Dispersion and Hydraulic Conductivity of a Sodic Soil. *Soil Science Society of America Journal* **45**, 273-277.
- Shirokova, Y, Forkutsa, I, Sharafutdinova, N (2000) Use of Electrical Conductivity Instead of Soluble Salts for Soil Salinity Monitoring in Central Asia. *Irrigation and Drainage Systems* **14**, 199-206.
- Speirs, SD, Cattle, SR, Melville, GJ (2011) Impact of sodium adsorption ratio of irrigation water on the structural form of two Vertosols used for cotton production. *Soil Research* **49**, 481-493.
- Stenberg, B, Viscarra Rossel, RA, Mouazen, AM, Wetterlind, J (2010) Chapter Five - Visible and Near Infrared Spectroscopy in Soil Science. *Advances in Agronomy* **107**, 163-215.
- Sumner, ME (1993) Sodic soils-New perspectives. *Soil Research* **31**, 683-750.
- Viscarra Rossel, RA, Walvoort, DJJ, McBratney, AB, Janik, LJ, Skjemstad, JO (2006) Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* **131**, 59-75.
- Zornoza, R, Guerrero, C, Mataix-Solera, J, Scow, KM, Arcenegui, V, Mataix-Beneyto, J (2008) Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. *Soil biology & biochemistry* **40**, 1923-1930.

Chapter 6

6. Simultaneous 4D space-time predictions of soil organic carbon by multivariate linear mixed models

Abstract

Land management and fluctuations in rainfall patterns have the potential to significantly alter the levels of organic carbon in a soil. This study looks at the change in soil organic carbon (SOC) content between 2002 and 2015 in a semi-arid irrigated cotton-growing region in the lower Lachlan River valley in southern NSW, Australia under various land uses. While traditional laboratory methods of determining SOC content are generally simple, this becomes more challenging when carbonates are present in the soil; such is commonly found in semi-arid areas. Consequently, this study uses visible near infrared (VisNIR) spectra to predict SOC and soil inorganic carbon (SIC) contents of samples, and the impact of including soil pH and soil total carbon data as predictor variables was evaluated. Both SOC and SIC content could be predicted with high accuracy using this approach, and the inclusions of soil pH and SOC content as predictor variables dramatically improved prediction accuracy when combined with VisNIR spectra. One multivariate linear mixed model (MLMM) was used to spatio-temporally model four response variables of SOC content in the study area – two sampling depths (0-0.1, and 0.3-0.5 m), and two time points (2002 and 2015). This resulted in a four-dimensional model (horizontal and vertical space, and time) to monitor the spatio-temporal shifts of SOC content. Using this approach, SOC content was predicted with high accuracy, with a Lin's concordance correlation coefficient (LCCC) of 0.68 when tested with leave-one-site-out cross-validation (LOSOCV). These high quality results are likely due to the utilisation of the correlation and covariance between samples extracted at different time points, as well as the correlation and covariance between different depth increments of the same soil core. No considerable change in subsoil (0.3-0.5 m) SOC content was observed, but topsoil (0-0.1 m) maps of SOC content revealed a trend of a statistically significant increase in the northeast of the study area. This increase in SOC content appeared to be due to the extended wet period experienced in the study area from 2010 to 2012, rather than a function of land management. Irrigated cotton production did not appear to increase SOC content from natural levels, and small, isolated areas under this land use underwent a statistically significant decrease in SOC content during the study period. This suggests that while irrigated cotton production increases the amounts of plant biomass that enter the soil, the negative effects of tillage counteracts this. Overall, the results from this study indicate that there is promise for using other readily available soil data in combination with VisNIR spectra to improve the predictions of different soil properties. The advantages of using MLMMs to monitor soil properties in space and time is also shown in terms of the coherence of the predictions, and the increased sensitivity of detecting a statistically significant change in soil properties over time is a particular benefit.

6.1. Introduction

Arid and semi-arid regions possess intrinsically low levels of soil organic carbon (SOC). This is largely due to insufficient precipitation in these areas for ample plant growth, resulting in small amounts of biomass entering the soil (Jobbágy and Jackson 2000). It is generally perceived that conventional cropping practices deplete natural SOC levels (Studdert and Echeverria 2000), however, when baseline values are so inherently low, cropping could be considered a tool for SOC accumulation. Irrigated cotton production is common in the semi-arid regions of NSW and QLD, Australia, and this land use is characterised by enormous inputs of water, which consequently results in some contrasting land uses in these areas. While irrigated cotton production increases the amount of plant biomass that enters the soil, the constant cultivation that is common practice in this industry would likely have an antagonistic effect. Over 99% of the cotton produced in Australia is genetically modified (GM), and growers are contractually obligated to regularly cultivate their soil to prevent weeds developing resistance to herbicides, and to inhibit pests developing resistance to insecticides produced by the inserted genes. While no-till and minimum-till management is generally better for maintaining SOC than conventional methods of cultivation, this is simply not an option for conventional cotton growers.

In soils that do not contain inorganic carbon (SIC), traditional laboratory methods of SOC content determination are generally similar to those for soil total carbon (STC) content determination. However, when carbonates are present in soil samples, such as in arid and semi-arid regions, the determination of SOC content becomes more challenging. The use of spectroscopic tools, such as visible near infrared (VisNIR), to predict the different components of soil carbon (SOC and SIC) has been extensively studied in recent years due to the cost and labour savings that come with these approaches (Miklos *et al.* 2010). Both SIC and SOC content are generally well predicted by these approaches (Viscarra Rossel *et al.* 2006), as there are specific VisNIR wavelengths that are heavily associated with these different components of soil carbon (Cécillon *et al.* 2009). Most studies solely use VisNIR spectra as input variables for predicting SOC and SIC content (Viscarra Rossel *et al.* 2006), however, there are opportunities to combine VisNIR spectra with other useful and readily available soil information to further improve predictions. Soil properties such as STC content and pH are relatively easily determined by traditional laboratory methods, and including this information as predictor variables could be advantageous, as these soil properties are often strongly correlated with both SOC and SIC content.

Values of SOC content that are sampled at the same location at two different time points are often strongly correlated with each other. Additionally, there is often a strong correlation with SOC

values sampled at different depth increments within the same soil core. When spatio-temporally modelling SOC content over a study region, it is possible to utilise these correlations to improve the accuracy of SOC predictions by using a multivariate linear mixed model (MLMM). Rather than creating a separate model for each sampling depth and each time point, it is possible to simultaneously model the same soil property at multiple sampling depths, at multiple time points. While some studies have created a single model to simultaneously model multiple soil depths at one time point, such as Akpa *et al.* (2014) who used a machine learning approach, very few studies have adopted a geostatistical modelling approach where the correlation between these variables is fully utilised through a linear model of coregionalisation or linear mixed model framework (e.g. Orton *et al.* 2014). Adopting an approach that takes advantage of both this vertical and temporal correlation is particularly useful where one sampling depth possesses more samples than another (e.g. if only a topsoil sample was collected), or when one survey possesses more soil cores than another, respectively. In addition, by modelling the entire profile we are likely to achieve coherent predictions which make pedological sense.

In this study, SOC and SIC content of samples are predicted using Cubist models, and the impact of combining VisNIR spectra data with soil pH and STC data as predictor variables is analysed. This study aims to investigate the impact of contrasting land uses and shifts in rainfall patterns on SOC levels from 2002 to 2015 in a semi-arid cotton-growing region in south-western NSW, Australia. In the previous chapters of this thesis we have used bivariate linear mixed models (BLMMs) to utilise the temporal correlation of co-located information, however, in this chapter we use multivariate linear mixed models (MLMMs) to account for vertical spatial correlation in addition to this. Using this MLMM, SOC is simultaneously modelled as four response variables (two depths, two time points), resulting in a four-dimensional model (eastings, northings, depth, and time) to monitor the spatio-temporal shifts.

6.2. Materials and methods

6.2.1. Soil datasets

This study uses soil data collected from a semi-arid area surrounding the township of Hillston, NSW as described in further detail in Chapter 3. Soil samples from 113 locations from a soil survey conducted in 2002 are used, as well as 160 soil cores from a soil survey from 2015. Many of the same sites were sampled in both surveys ($n = 103$), as the locations were georeferenced. The subsampling intervals in the surveys differed, with the 2002 survey sampled at 0-0.2 and 0.3-0.4 m, and the 2015 sampled at 0-0.1, 0.1-0.3 and 0.3-0.5 m. In this chapter, only the 0-0.1 and 0.3-0.5 m sampling depths are considered, as the 0.1-0.3 m samples from 2015 were not analysed by traditional laboratory methods due to

budgetary constraints. Because of the differing sampling depths between the two surveys, equal-area quadratic smoothing splines (Bishop *et al.* 1999) were used to standardise the soil attribute data from the 2002 survey to the depths of the 2015 survey (0-0.1 and 0.3-0.5 m).

6.2.3. Soil analyses

Although SIC values are not used for modelling and mapping in this chapter, the methods of the traditional laboratory analysis of SIC, and the method and results from the spectroscopic analysis of SIC are presented in this chapter to prevent repetition in the ensuing chapter.

6.2.3.1. Traditional laboratory methods

All of the soil samples were air-dried and then ground through a 2 mm sieve. Prior to laboratory analysis, all samples were tested for the presence of soil inorganic carbon (SIC). A ~1 g subsample of ground soil was placed on a ceramic plate and a few drops of 1 M hydrochloric acid (HCl) were placed directly onto the sample. Any sample that showed an effervescence reaction was considered to contain calcium carbonate, the most prominent form of SIC in these soils. An additional subsample (~10 g) was then taken and finely ground (<53 μm) using a Fritsch Mortar Grinder Pulverisette 2 (Fritsch, Germany) for 4 minutes at 50-60 Hz frequency. Soil total carbon (STC) content was determined by the combustion method with the Leco1 CHN analyser for 2002 samples, and the Elementar vario MAX CNS for 2015 samples. Soil organic carbon content for 2002 samples was determined by treating samples with 2 M HCl to remove inorganic carbon, and then analysing by the Leco1 CHN analyser (Tiessen and Moir 1993). For 2015 samples, SOC content was determined by the Walkely-Black method, which is a wet oxidation technique that uses chromic acid (Walkley and Black 1934). To estimate SIC content, the difference between STC and SOC contents was used. The Elementar vario MAX CNS and the Leco1 CHN analyser are very similar in their analytical approach, and both use the combustion technique. For 2002 samples, the SOC content was determined immediately, however, STC was determined from archived samples 13 years later.

6.2.3.2. Visible near infrared (VisNIR) spectroscopy predictions

While all samples were assessed for the presence of SIC, a portion of these samples ($n = 118$) that had SIC present were only analysed for STC content, and as a result, had an unknown quantity of SIC and SOC. The combustion method of measuring SOC content is useful and rapid in samples that lack SIC, however, when SIC is present, a lengthy process of carbonate dissolving is required, as described in the

methodology above. As a result, soil spectroscopic techniques have been used in this study to make predictions of SOC and SIC contents in these 118 un-analysed soil samples.

Archived soil samples from both the 2002 and 2015 soil surveys ($n = 517$) were scanned by visible near infrared (VisNIR) with an Agrispec portable spectrophotometer with a contact probe attachment (Analytical Spectral Devices, Boulder, Colorado) on the dried and ground soil samples. To reduce signal-to-noise ratios of the spectra, three scans of each sample were performed, from which an averaged reflectance spectrum was derived. Calibration of the instrument was made with a Spectralon white tile and was re-calibrated after every 15 scans, or five samples. The dataset in which we had both a SOC and SIC content value from traditional laboratory methods consisted of 399 samples from both the 2002 ($n = 148$) and 2015 ($n = 251$) surveys. Cubist models were used to predict SOC and SIC content, with predictor variables including VisNIR wavelengths from 500-2450 nm, averaged into segments of ten, and the mid-depth of the sample, to ensure that the depth was taken into account when predicting. In addition to this, soil pH (1:5 soil: water) and STC content (measured by traditional laboratory approaches described in the methodology above) were included, as this data was available for all samples.

Data splitting was used to analyse the prediction quality of the Cubist models created. To test this, 75% of the dataset was used as calibration, and the remaining 25% was used as validation. These datasets were selected by performing a Latin hypercube of the VisNIR spectra, pH, STC, mid-depth, and the response variables to ensure that both the validation and calibration datasets were appropriately represented. The calibration and validation dataset splits of 75% and 25%, respectively, is also similar to the dataset split when making actual predictions. When testing for prediction quality, 299 (75%) samples were used to predict 100 (25%), and in actual predictions, 399 (77%) samples were used to predict 118 (23%). For each soil property, model quality was tested for four variations of inputs. These included VisNIR and mid-depth; VisNIR, mid-depth and pH; VisNIR, mid-depth and STC content; and finally VisNIR, mid-depth, pH and STC content. The statistics used to test the model quality included Lin's concordance correlation coefficient (LCCC), mean square error (MSE), root mean square error (RMSE), bias (mean of the residuals), and R^2 , with this being tested on both the calibration and validation datasets.

6.2.4. Models and mapping

6.2.4.1. Multivariate linear mixed model

One multivariate linear mixed model (MLMM) was used to model SOC at both time points (2002 and 2015), and both depths (0-0.1 and 0.3-0.5 m) simultaneously. As discussed in the previous chapters of this thesis, the covariance between co-located samples extracted at different time points can be calculated, but in this approach there is also a covariance term between co-located samples at different depth increments of the same soil core. Despite this, we are only interested in the change in SOC content at each sampling depth between 2002 and 2015, so the simple pair-wise contrast variance (Equation 6.1) can be used to calculate the z-statistic to test the statistical significance of any observed change. The temporal contrast variance is expressed as:

$$\hat{V}(\Delta\hat{y}) = \hat{V}(\hat{y})_{t1} + \hat{V}(\hat{y})_{t2} - 2(\hat{V}(\hat{y})_{t1,t2}), \quad 6.1)$$

where $\hat{V}(\Delta\hat{y})$ is the variance of the change (contrast variance) in the soil property between two points in time, $\hat{V}(\hat{y})_{t1}$ is the variance of the soil property for the observations at the first point in time, $\hat{V}(\hat{y})_{t2}$ at the second point in time, and $\hat{V}(\hat{y})_{t1,t2}$ is the covariance of soil properties at the two time periods (the covariance term). This is calculated separately for each depth of interest. While the vertical covariance is not used in these calculations, it is still taken advantage of in the modelling process to strengthen predictions at non co-located sites, and to ensure they are coherent with each other.

6.2.4.2. Covariates, model selection, map production, and assessment of model quality

The covariates used for the modelling and mapping in this study are discussed in detail in Chapter 3. In summary, the covariates consisted of four broad types; gamma radiometrics data, digital elevation model terrain attributes, and maps of soil type and land use. Due to the large number of numerical covariates ($n = 19$), a principal component analysis was performed. As described in the previous chapters of this thesis, the combination of covariates to be included in the final model was determined by backwards elimination. The initial model had separate fixed effects for each of the four response variables (i.e. all PCs, land use and soil type). Wald tests were performed to identify redundant covariates, where those with the highest p -value (least significant) were removed, and this was continued until all predictors had a p -value less than 0.05. After the final model was determined, it was used to predict onto the covariate grid of the study area (100 m resolution) using the EBLUP (Empirical Best Linear Unbiased Predictor) approach and digital soil maps were then produced. No back-transformation of the response variables was necessary, as the SOC content dataset and the residuals

displayed a normal distribution and did not require statistical transformation. Leave-one-site-out cross-validation (LOSOCV) was used to test the quality of predictions, where both depths at both time points for one site was removed from the dataset and predicted using the remaining dataset, and this was repeated for each site. Model quality was then evaluated with Lin's concordance correlation coefficient (LCCC), root mean square error (RMSE), bias (mean of the residuals), and the standardised squared prediction errors (SSPE).

6.3. Results

6.3.1. Visible near infrared (VisNIR) spectroscopy predictions

Both SOC and SIC content of samples displayed a moderate to strong relationship (Pearson's correlation) with the predictor variables of STC content and soil pH (Table 6.1). The independently-validated statistics also showed that both SOC and SIC content of samples could be predicted with high accuracy using spectroscopic techniques (Fig. 6.1; Fig. 6.2; Table 6.2). Overall, SOC content was predicted with greater accuracy than SIC content, and the different combinations of model inputs had a clear impact on the prediction quality for both SOC and SIC content. Lin's concordance correlation coefficient (LCCC) was primarily used for the assessment of model quality, as it is the fit of the 1:1 line of the observed and predicted values. It is also unit less, which makes it useful for comparing different models of the same soil property, as well as comparing models for different soil properties.

For SOC content, it was clear that the inclusion of additional soil property data improved prediction results considerably, improving the predictions on the validation dataset from an LCCC of 0.81 for the simplest model (model A), to an LCCC of 0.94 for the model that included all predictor variables (model D) (Table 6.2; Fig. 6.1). It can be seen in Fig. 6.1 that model D predicted SOC to ~100% accuracy for multiple soil samples (where the points lie exactly on the 1:1 line) in both the calibration and validation plot, whereas this did not occur for model A. For SOC content, the inclusion of pH alone with VisNIR and mid-depth (model B) did not improve predictions, but the inclusion of STC content with VisNIR and mid-depth (model C) improved predictions significantly. Organic carbon content had a particularly high positive correlation (r) with total carbon content (0.86) and a weaker negative correlation with pH (-0.36), which explains their relative importance in the prediction of SOC content (Table 6.1). Overall, the inclusion of both pH and STC content together with VisNIR and mid-depth (model D) resulted in the best model.

The inclusion of STC content and pH data as predictor variables with VisNIR and mid-depth was even more effective in improving SIC content predictions with Cubist models (Table 6.2; Fig. 6.2). VisNIR and mid-depth only models (model A) predicted the validation dataset very poorly, with an LCCC of 0.35, whereas the model with the full suite of input variables (model D) predicted the validation dataset to an accuracy of 0.83 LCCC (Table 6.2). While the inclusion of pH alone (model B) did not improve SOC content predictions, it made a noteworthy improvement from the simplest model (model A) in SIC content predictions to predict the validation dataset to an accuracy of 0.52 LCCC (Table 6.2). There was a slightly higher correlation (r) with pH and SIC content (0.39) than there was with pH and SOC content (-0.36), although the correlation with STC and SIC contents was much weaker (0.35) than with STC and SOC contents (0.86) (Table 6.1). Overall, the Cubist models that included VisNIR spectra, mid-depth, pH and STC content as predictor variables proved to be the most accurate at predicting both SOC and SIC contents. Consequently, the full suite of predictor variables (model D) were used for final SOC and SIC content predictions for further data analysis, modelling and mapping.

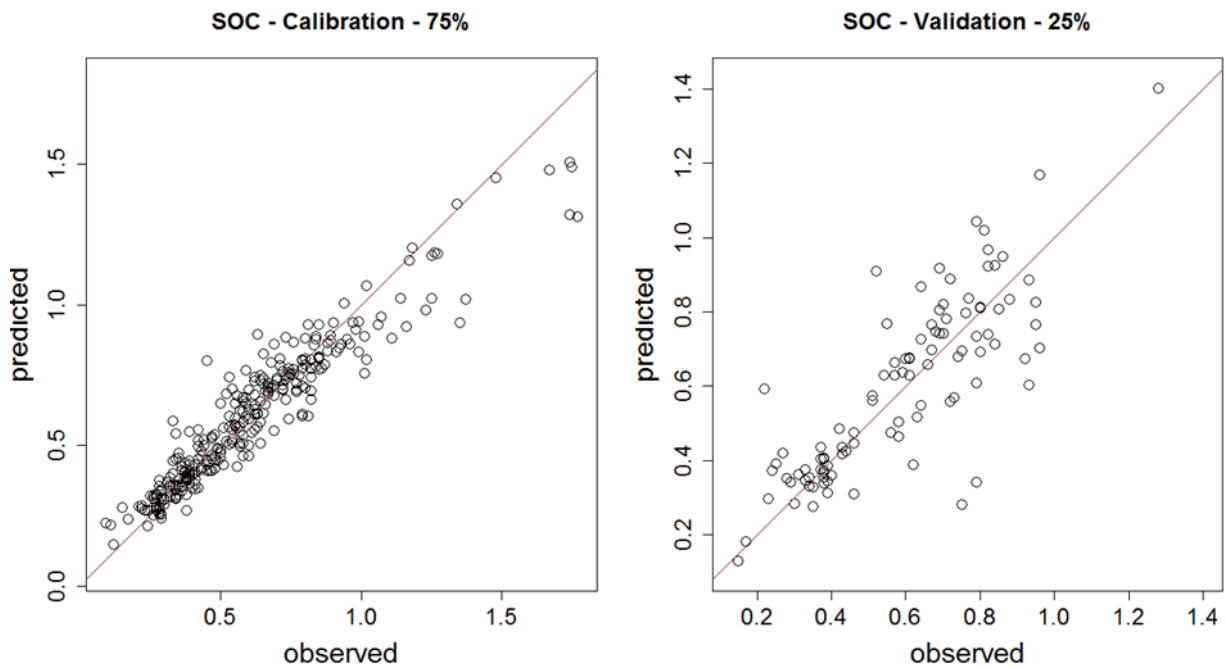
Table 6.1 – Pearson’s correlation (r) between response variables (SOC and SIC content) and predictor variables (STC content and pH)

	SOC	SIC	STC	pH
SOC	1	-	-	-
SIC	-0.15	1	-	-
STC	0.86	0.35	1	-
pH	-0.36	0.39	-0.18	1

Table 6.2 – Model quality statistics for SOC and SIC content predictions from Cubist models with varying combinations of input variables

Attribute	Model	Input variables	Dataset	LCCC	R ²	RMSE	MSE	bias
Soil Organic Carbon (SOC)	A	VisNIR + mid-depth	Calibration	0.93	0.90	0.10	0.01	0.00
			Validation	0.81	0.67	0.14	0.02	0.01
	B	VisNIR + mid-depth + pH	Calibration	0.93	0.89	0.10	0.01	0.00
			Validation	0.81	0.67	0.14	0.02	0.01
	C	VisNIR + mid-depth + TC	Calibration	0.97	0.96	0.06	0.00	0.00
			Validation	0.90	0.84	0.09	0.01	0.02
	D	VisNIR + mid-depth + pH + TC	Calibration	0.98	0.98	0.05	0.00	0.00
			Validation	0.94	0.89	0.07	0.01	0.00
Soil Inorganic Carbon (SIC)	A	VisNIR + mid-depth	Calibration	0.83	0.85	0.07	0.00	0.00
			Validation	0.35	0.14	0.12	0.01	0.02
	B	VisNIR + mid-depth + pH	Calibration	0.86	0.87	0.06	0.00	0.00
			Validation	0.52	0.29	0.11	0.01	0.01
	C	VisNIR + mid-depth + TC	Calibration	0.94	0.97	0.04	0.00	0.00
			Validation	0.82	0.73	0.06	0.00	0.01
	D	VisNIR + mid-depth + pH + TC	Calibration	0.97	0.98	0.03	0.00	0.00
			Validation	0.83	0.70	0.07	0.01	0.01

VisNIR + mid-depth



VisNIR + mid-depth + pH + total carbon

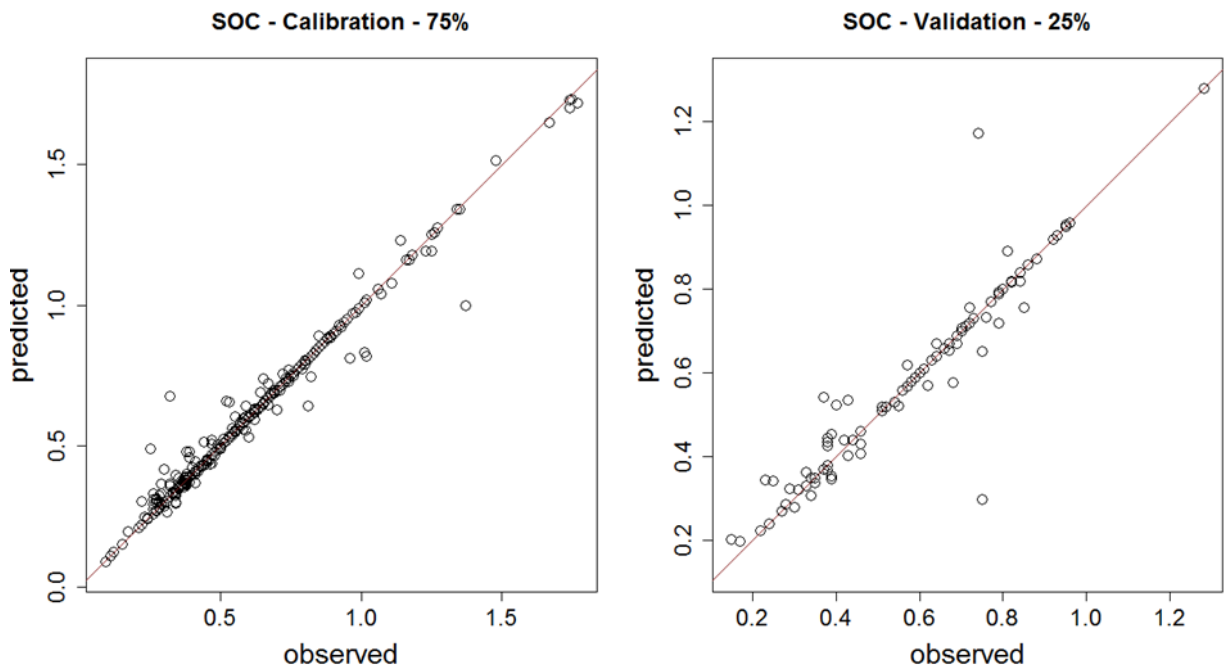
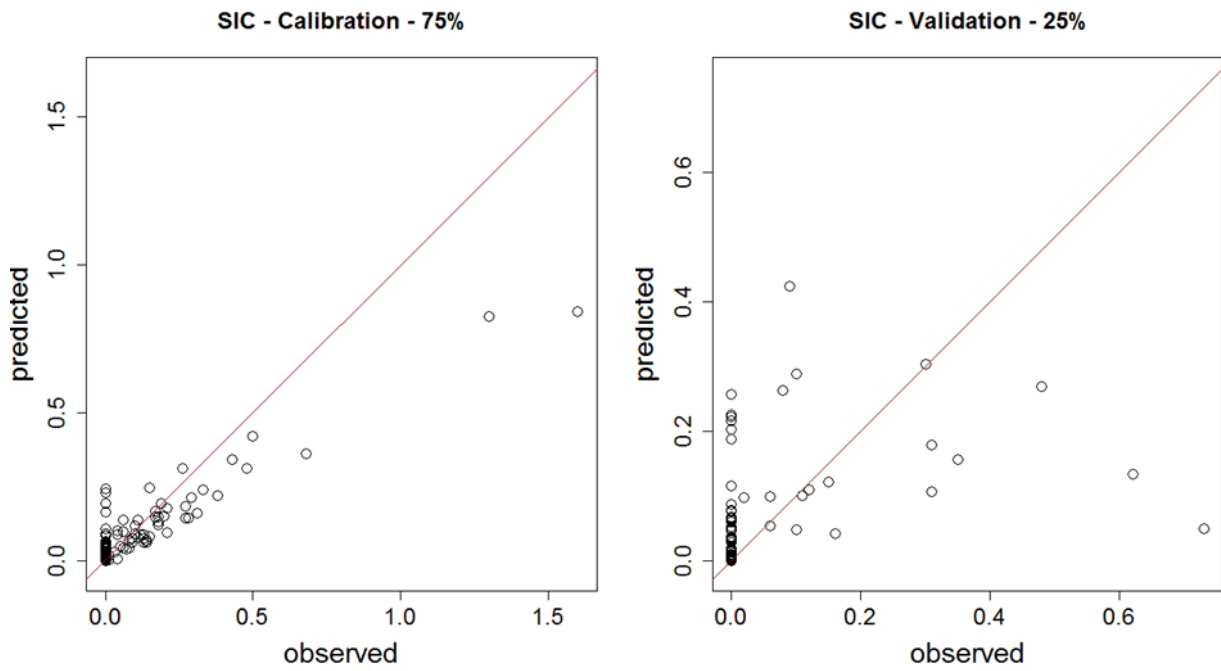


Fig. 6.1 – Observed vs. predicted soil organic carbon (SOC) content (%) values of calibration (left) and validation (right) datasets predicted with VisNIR spectra and mid-depth (top); and VisNIR spectra, mid-depth, soil pH and soil total carbon contents (bottom) as predictor variables in Cubist models

VisNIR + mid-depth



VisNIR + mid-depth + pH + total carbon

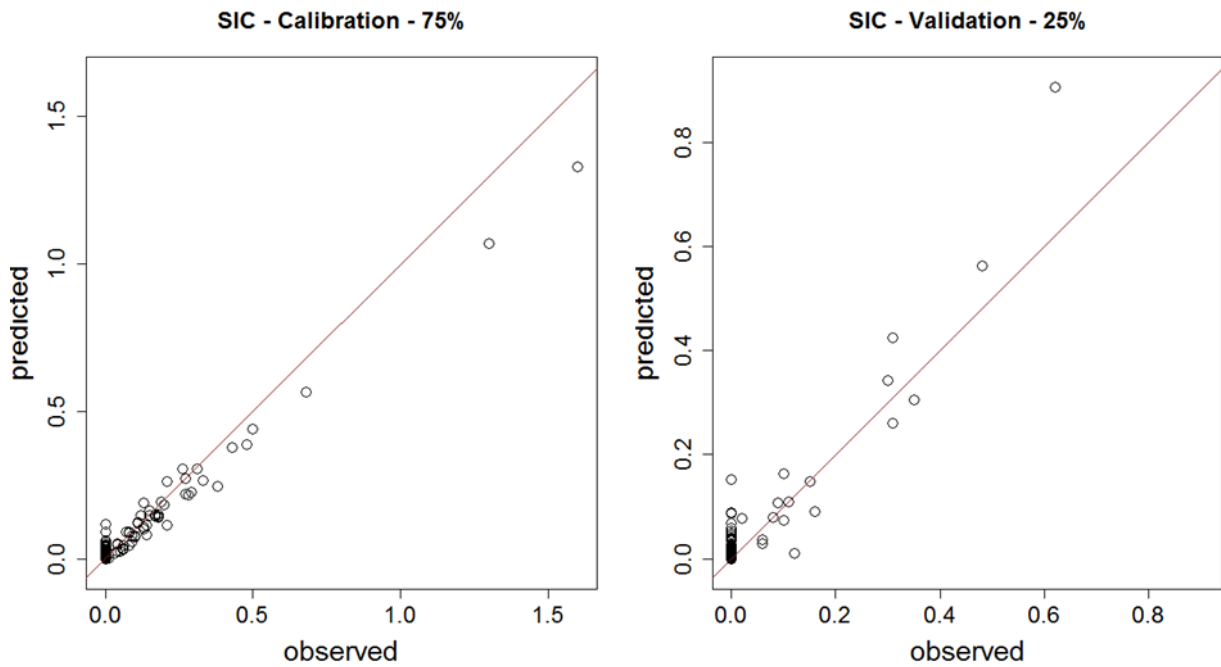


Fig. 6.2 – Observed vs. predicted soil inorganic carbon (SIC) content (%) values of calibration (left) and validation (right) datasets predicted with VisNIR spectra and mid-depth (top); and VisNIR spectra, mid-depth, soil pH and soil total carbon contents (bottom) as predictor variables in Cubist models

6.3.2. Soil organic carbon summary statistics

Both the laboratory measured and VisNIR-predicted samples are considered together in this section, due to the high accuracy of the predictions. Overall, SOC contents at Hillston are low, with values ranging from 0.22% to 1.87% in the 0-0.1 m sampling depth, and from 0.03% to 1.09% in the 0.3-0.5 m sampling depth (Table 6.3). In the topsoil (0-0.1 m), the cotton sites had slightly lower mean SOC content (0.72%) than natural sites (0.79%) in 2002. In 2015, mean SOC content increased by 0.05% to 0.77% in sites under the cotton land use, and equally increased by 0.05% to 0.84% in sites under the natural land use (Table 6.3). A mean decrease in SOC content for sites under the dryland land use was observed in the 0-0.1 m depth, shifting from 0.78% in 2002, to 0.69% in 2015. Contrastingly, the horticulture land use had a mean increase in SOC content in the 0-0.1 m depth, with an increase from 0.61% to 0.85% during the 13-year period, although there were very few samples under this land use extracted in 2002 (Table 6.3). In the 0.3-0.5 m sampling depth, SOC content was generally low, with mean values for the different land uses ranging from 0.31-0.40%. A mean increase in SOC content was observed in the subsoil under the dryland land use, with no considerable changes in sites under the cotton or natural land use (Table 6.3).

Table 6.3 – Combined laboratory-measured and VisNIR-predicted soil organic carbon content (%) statistics of all samples for different depth increments, time points and land uses

Statistic	Depth	Cotton		Dryland		Natural		Horticulture	
		2002	2015	2002	2015	2002	2015	2002	2015
Mean	0-0.1 m	0.72	0.77	0.78	0.69	0.79	0.84	0.61	0.85
	0.3-0.5 m	0.35	0.35	0.31	0.40	0.37	0.40	0.43	0.36
Minimum	0-0.1 m	0.22	0.34	0.36	0.34	0.29	0.33	0.42	0.52
	0.3-0.5 m	0.08	0.03	0.12	0.16	0.09	0.11	0.41	0.21
Maximum	0-0.1 m	1.46	1.59	1.87	0.96	1.48	1.77	0.95	1.23
	0.3-0.5 m	0.85	0.83	0.64	0.85	0.79	1.09	0.47	0.56
Standard deviation	0-0.1 m	0.259	0.197	0.440	0.166	0.263	0.329	0.293	0.285
	0.3-0.5 m	0.134	0.118	0.139	0.211	0.128	0.145	0.032	0.108
n =	0-0.1 m	60	65	13	21	37	61	3	13
	0.3-0.5 m	58	62	12	19	35	61	3	12

While many sites were sampled in both 2002 and 2015, the number of sites sampled in each survey differed, and this confounds the trends in SOC content observed in Table 6.3. The mean statistics of all sampled sites (Table 6.3) showed an increase in SOC content of 0.05% for both cotton and natural sites in the topsoil, but mean data from those sites only sampled in both 2002 and 2015 does not necessarily affirm this trend. This paired data suggests that almost no change in SOC content had

occurred over the study period, with sites under cotton decreasing by 0.02%, and sites under the natural land use undergoing a decrease of 0.01% (Table 6.4). Individual sampling sites exhibited both increases and decreases in SOC content over time, but the majority of sites experienced only slight shifts in SOC content (Fig. 6.4).

Table 6.4 – Mean change in SOC content (%) from 2002 to 2015 for co-located sites

	Cotton	Natural
0-0.1 m	-0.02%	-0.01%
0.3-0.5 m	-0.01%	+0.03%

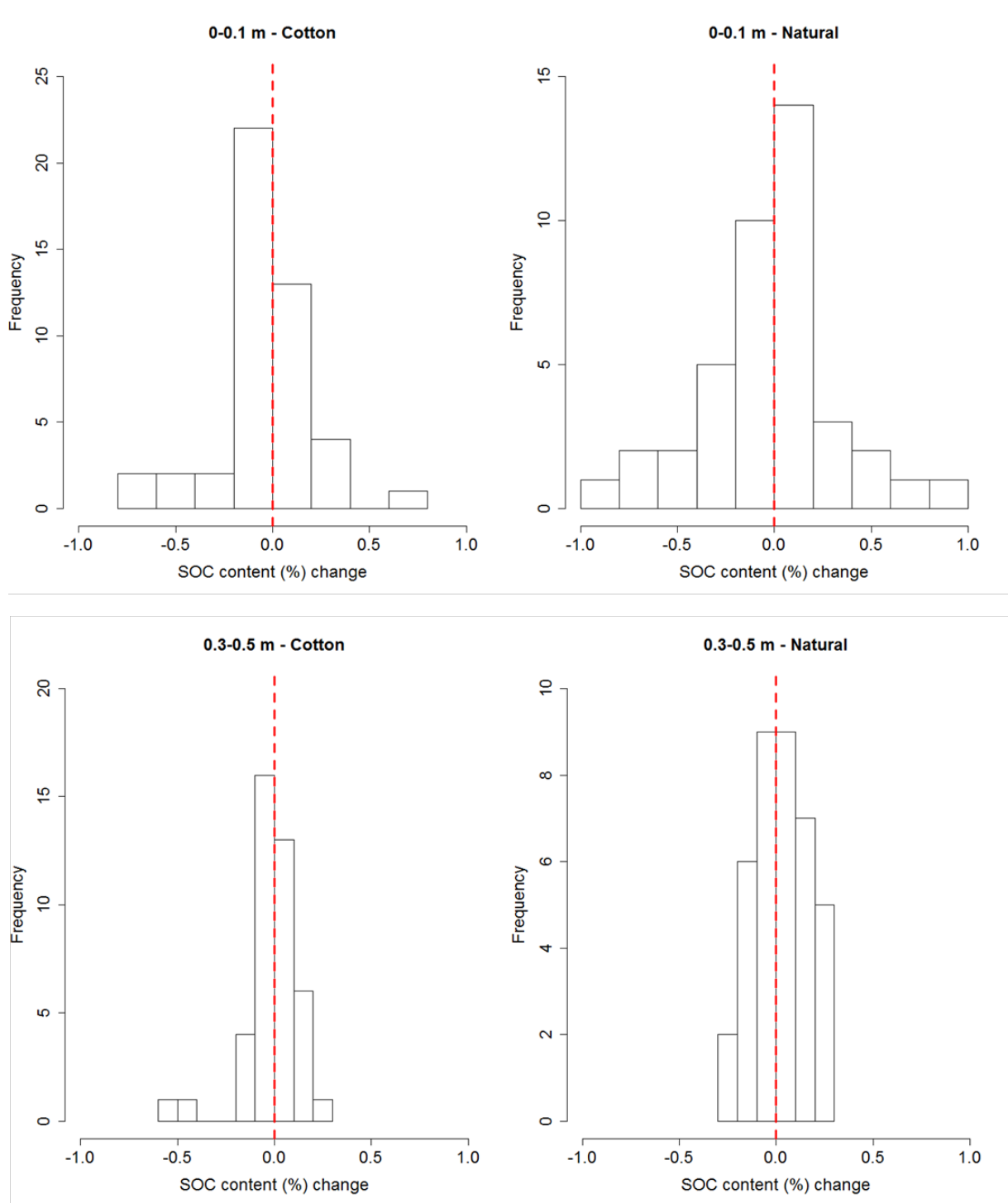


Fig. 6.3 – Temporal change in soil organic carbon (SOC) contents at sites sampled in both 2002 and 2015 surveys under the cotton (left) and natural (right) land use at the 0-0.1 m (top) and 0.3-0.5 m sampling depth (bottom)

6.3.3. Multivariate linear mixed model

6.3.3.1. Final model and model quality

The combination of covariates used as predictors for each individual response variable in the final fitted MLMM varied, although land use was an important predictor for three of the four response variables (Table 6.5). The final fitted MLMM could predict the four response variables of SOC content with high accuracy when tested with LOSOCV (Table 6.6; Fig. 6.4). The Lin's concordance correlation coefficient (LCCC) value for the fitted model was high, with a value of 0.68, and the root mean squared error (RMSE) was low, with a value of 0.21%, which indicates that the quality of model predictions were high. The fitted model had a mean standardised squared prediction error (SSPE) close the expected value of 1.0, and a median that was slightly lower than the expected value of 0.455 (Table 6.6), which suggests the presence of some outliers.

Table 6.5 – Included covariates in the final MLMM of SOC content

Depth	Year	Covariates
0-0.1 m	2002	Land use + PC3 + PC5
	2015	Land use + soil type + PC1 + PC2 + PC6
0.3-0.5 m	2002	Land use + PC3 + PC6
	2015	Soil type + PC2

Table 6.6 – Quality statistics of the soil organic carbon content MLMM using LOSOCV

Sampling depth	LCCC	RMSE	Bias	Mean (SSPE)	Median (SSPE)
0-0.1 & 0.3-0.5 m	0.68	0.21	-0.01	0.99	0.34

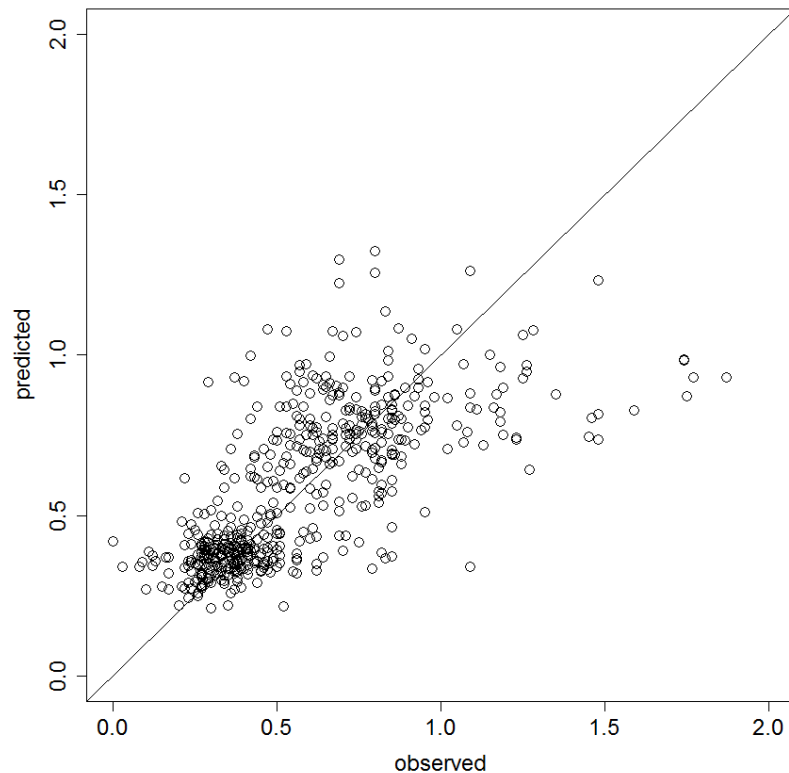


Fig. 6.4 – Observed vs. predicted SOC content (%) using LOSOCV and predicted with the MLMM

6.3.3.2. Soil organic carbon content maps

The maps of SOC content in the cotton-growing region of Hillston show that the topsoil (0-0.1 m) contains considerably larger quantities of SOC than the subsoil (0.3-0.5 m), as expected (Fig. 6.5). The variability of SOC content is also much higher in the topsoil, with values ranging from approximately 0.2 to 2.0%, whereas the majority of values in the subsoil ranged from 0.2 to 0.6%. The topsoil maps showed that the highest SOC contents are located along the Lachlan River, and areas at higher elevation (Fig. 6.5). While the spatial patterns of topsoil SOC content were similar at both time points, the change map revealed that there were predicted decreases in SOC content in the southwest of the study area on the alluvial floodplains, and predicted increases in the northeast of the study area. In the subsoil, the spatial patterns of the SOC content maps at both time points were very similar, and there appeared to be no obvious driver of the spatial distribution of SOC content. The change map for the subsoil revealed that very little of the area was predicted to have experienced a shift in SOC content, however, a generally similar trend to the topsoil was observed, with a decrease in the southwest, and an increase in the northeast. It appeared that land use was not a clear driving factor of the spatial patterns of SOC content in both the topsoil and subsoil maps (Fig. 6.5).

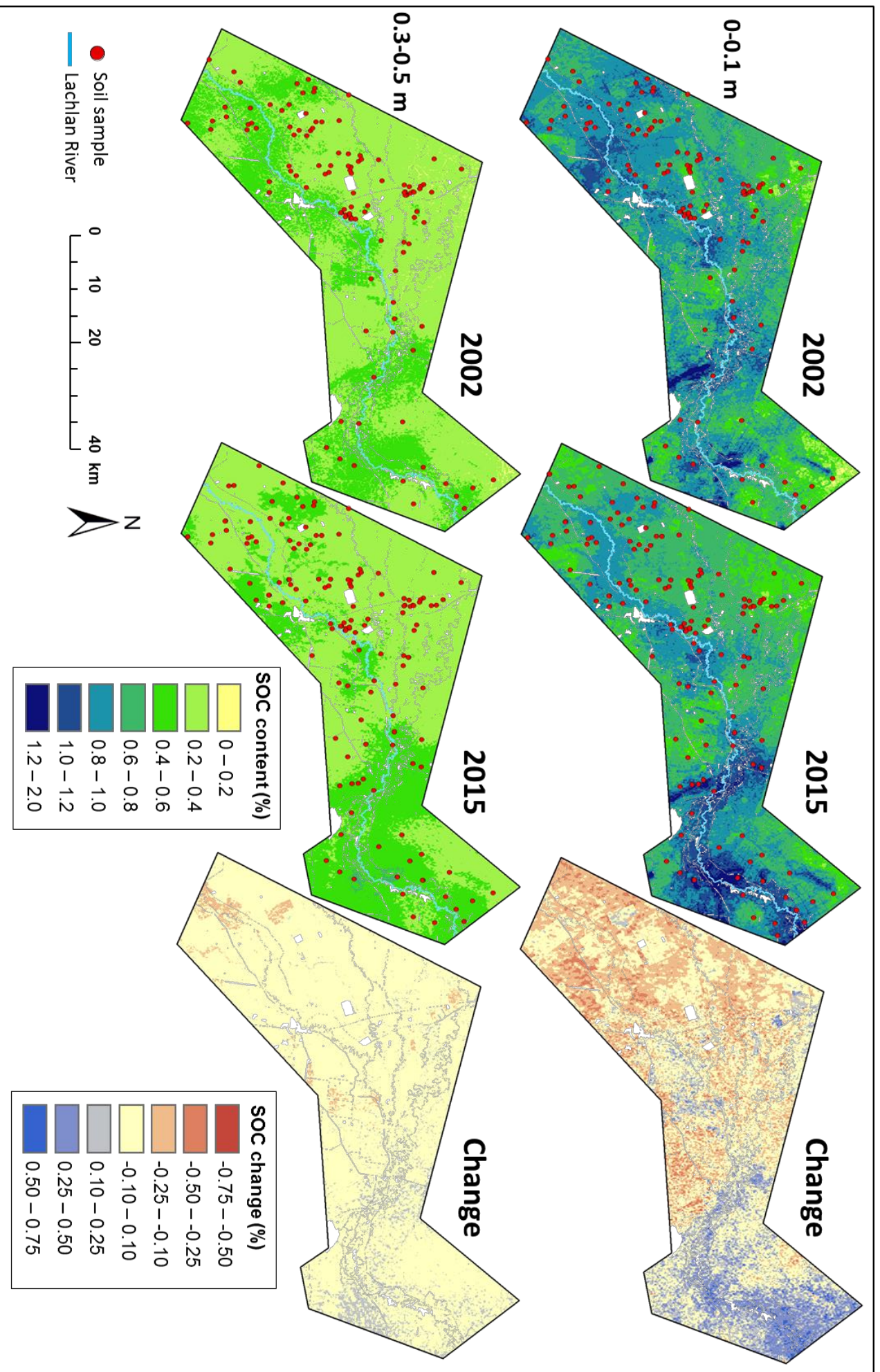


Fig. 6.5 – Soil organic carbon (SOC) content (%) of the 0-0.1 (top) and 0.3-0.5 m (bottom) sampling depths in 2002 and 2015, modelled with a single multivariate linear mixed model, and the change in SOC during this time period (2015-2002)

6.3.4. Statistical significance of soil organic carbon change

While the change map of SOC content at the 0-0.1 m depth increment showed large areas were predicted to have undergone a shift in SOC content (Fig. 6.5), the z-score map revealed that only small areas underwent a statistically significant change (Fig. 6.6). In particular, an area in the northeast of the study area experienced a statistically significant increase (p -value 0.05-0.10) in SOC content during the study period (Fig. 6.6). This particular area of the study region was covered by several land uses (cotton, natural, and dryland), which suggests that the reasons for this increase could be climatically-driven. There were also some small, isolated areas on the alluvial floodplains in the south-west of the study area that underwent a statistically significant decrease in SOC content during the study period (Fig. 6.6). Most of these decreases occurred in areas under irrigated cotton production. In the 0.3-0.5 m sampling depth, almost no statistically significant change in SOC content was detected across the entire study region (Fig. 6.6).

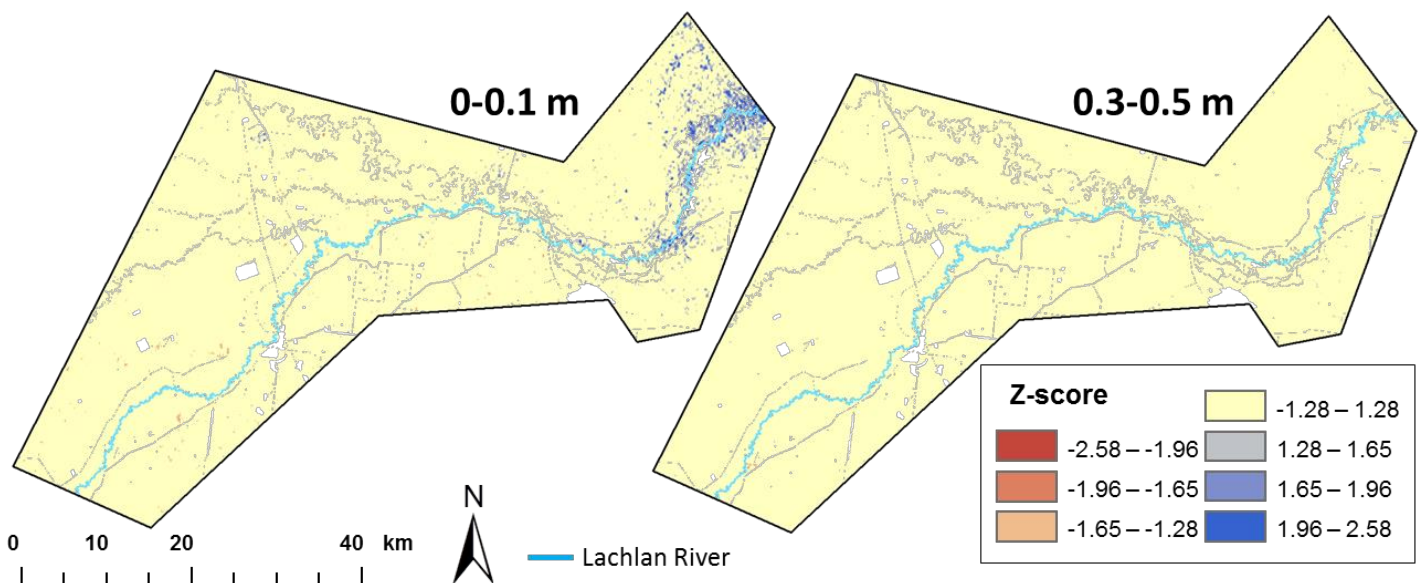


Fig. 6.6 – Z-scores showing the statistical significance of the change from 2002 to 2015 in soil organic carbon content for the 0-0.1 and 0.3-0.5 m sampling depths

6.4. Discussion

6.4.1. Soil organic and inorganic carbon predictions with visible near infrared (VisNIR)

Overall, both SOC and SIC contents of samples from the cotton-growing district of Hillston could be accurately predicted with spectroscopic techniques. It was clear from the results that combining soil pH

and STC content data as predictor variables with VisNIR spectra dramatically improved the accuracy of both SOC and SIC predictions compared to solely using VisNIR spectra.

In particular, SOC content was predicted with very high accuracy by the model that included VisNIR, mid-depth, pH and STC (model D), with an LCCC of 0.94 when predicted on the validation dataset compared to an LCCC of 0.81 for the model that contained only VisNIR and mid-depth (model A). This is logical, as SOC content is highly positively correlated (r) with STC content (0.86), and moderately negatively correlated with pH (-0.36). When predicting on both the calibration and validation datasets with model D, it was apparent that SOC content was predicted to ~100% accuracy for several samples, as can be seen in Fig. 6.1. While these results initially look unbelievable, these very accurate predictions can be logically explained. It is likely that the Cubist model is detecting that there is no inorganic carbon in the sample, and because Cubist models are essentially rule-based decision trees, the model is simply assigning the inputted STC value as the SOC content prediction. There are particular VisNIR wavelengths that are associated with SIC (Gaffey 1986), and if these wavelengths of the scanned sample do not possess the appropriate reflectance, the model is likely determining that there is no SIC present in the sample. While the inclusion of soil pH alone with VisNIR and mid-depth did not improve SOC predictions (model B), when this was included in combination with STC (model D), the predictions were slightly better than model C (VisNIR, mid-depth and STC), suggesting that there is an advantageous interaction occurring with STC and pH in these models.

Studies have reported that the accuracy of predicting SOC and SIC content of samples with VisNIR is generally quite similar (Chang *et al.* 2001; McCarty *et al.* 2002), although this depends on a number of factors. In our study, this was not the case, and the final models (model D) predicted the validation dataset with an LCCC of 0.94 for SOC content, and 0.83 for SIC content (Table 6.2). A possible reason for the poorer predictions of SIC content is due to the nature and distribution of the SIC dataset. The SIC dataset in our study is zero-inflated (contains many zero values), and consequently there are fewer samples that contain some amount of SIC in the training dataset. In addition, SIC content of samples was not directly measured by laboratory methods, and was determined by the difference between measured SOC and STC content, which includes a greater amount of error. While there were multiple occurrences of SOC content being predicted to ~100% accuracy, this was not the case for SIC content. This is logical, as all soil samples in the study contain some amount of SOC, even if it is a very small amount, but not all samples contain SIC. As a result, the model could not simply assign the SIC value as the STC value.

While the inclusion of soil pH as a predictor variable with VisNIR and mid-depth (model B) did not improve the SOC content predictions, it made a significant improvement in SIC content predictions. Soil pH was found to be positively correlated (r) with SIC content (0.39) and it is known that very alkaline pH levels indicate the presence of considerable amounts of carbonate in a soil. Soils that possess a pH of less than 7 (soil: water 1:5 extract) also commonly do not contain SIC (Wang *et al.* 2015). Although the correlation (r) of SIC with STC content was relatively weak (0.35), including STC content as a predictor improved predictions on the validation dataset from an LCCC of 0.35 for model A to 0.82 for model C. While combining pH with spectra improved SIC content predictions, this positive impact seemed to be masked when both pH and STC content were included as predictor variables together. The LCCC of predictions on the validation dataset with VisNIR, mid-depth, pH and STC content (model D) was 0.83, which is only slightly better than VisNIR, mid-depth, and STC content (model C), which was 0.82.

It must be acknowledged that including STC content data with VisNIR spectra to predict the SOC content of a sample is likely impractical and unnecessary for most studies. The prediction of SOC content with spectra alone was of high quality in our study, and this has been the case for many other studies (Viscarra Rossel *et al.* 2006). Our study, however, demonstrates that there is considerable benefit in measuring STC content and including this in model predictions with VisNIR to predict SIC content. Although inorganic carbon is not found in all soils, this approach could be particularly appropriate for areas that typically possess soils with carbonates, such as arid and semi-arid areas. Our results also suggest that there is considerable benefit in including soil pH data with VisNIR spectra to predict SIC content. Soil pH data is also typically more available than STC content data, as it can be rapidly and cheaply measured by traditional laboratory methods. As soil pH is often correlated with different soil attributes such as nutrient availability, this also shows promise for combining soil pH data with spectra to predict other soil properties.

This also opens the discussion as to the possible benefits of combining other cheaply-measured and readily available soil data with VisNIR spectra to predict different soil properties, as there are many soil properties that are highly correlated with each other. For example, soil electrical conductivity (EC) is easily measured by traditional laboratory methods and hence this data is often available. It is known that EC is well correlated with other soil properties that are typically laborious to measure, such as soil particle size, and cation exchange capacity (Grisso *et al.* 2005). While studies often use ancillary soil data combined with pedotransfer functions to estimate the value of a soil property (McBratney *et al.* 2002), there are no studies, to our knowledge, that use ancillary soil data in combination with spectra to

predict another soil property. There are some limitations to adopting this approach, as including additional soil property data with spectra in predictive models requires that both the training dataset and prediction dataset possess a value for that soil property. Despite this, when ancillary soil data is available, it could be very useful in improving the quality of soil spectroscopic predictions.

6.4.2. Spatio-temporal patterns of soil organic carbon

Overall, the levels of SOC in the Hillston district are low, with a mean content of 0.79% in the topsoil (0-0.1 m), and 0.38% in the subsoil (0.3-0.5 m) in 2015. These values are similar to SOC levels reported by other studies in semi-arid areas of eastern Australia (e.g. Chan *et al.* 1995; Conyers *et al.* 2015). While there was very little spatial variation in SOC content in the subsoil, the topsoil maps showed that the highest SOC contents are located on areas at higher elevation, as well as along the Lachlan River (Fig. 6.5). The larger SOC values at higher elevations could be because these areas are not used for agriculture, and there is typically greater ground cover and density of trees compared to the rest of the study area. The higher SOC contents along the Lachlan River are likely due to the relatively more fertile and higher yielding Grey Vertosols being situated in those locations. Interestingly, land use did not appear to have a strong impact on the spatial patterns of SOC content in both the topsoil and subsoil maps.

Of all soil properties, SOC content is the most commonly spatio-temporally monitored over a region (Filippi *et al.* 2016). There has been some contrasting trends observed in such studies, with some detecting trends of increasing SOC content (Minasny *et al.* 2012; Ross *et al.* 2013), while others have detected decreases in SOC over time (Saby *et al.* 2008; Deng *et al.* 2013). In our study, no considerable change in SOC content was detected in the subsoil (0.3-0.5 m), but the change map of SOC content for the topsoil (0-0.1 m) revealed extensive predicted decreases in the southwest of the study area, and widespread predicted increases in the northeast of the study area. Nevertheless, the only part of the study region that underwent a widespread statistically significant change was an area in the northeast, where a statistically significant increase was predicted (p -value 0.05-0.10). There were also some isolated areas in the south-west of the study area that underwent a statistically significant decrease in SOC content during the study period, although this area was relatively small (Fig. 6.6). The results from these z-score maps reinforce the importance of placing a prediction interval around the predicted change to prevent misinterpretation. A study that similarly monitored the change in SOC content at decadal scales in a cotton-growing region in eastern Australia, found a contrasting trend to ours, where decreases in SOC content over time were observed (Karunaratne *et al.* 2014). Although rarely performed

in the literature, Karunaratne *et al.* (2014) also analysed the statistical significance of this change and found that only a very small section of the study area decreased significantly (p -value 0.05-0.10), with the authors attributing this change to a warming climate during the study period.

The particular part in the northeast of the study area that underwent a statistically significant increase in SOC content was not limited to a single land use, with the cotton, natural and dryland land uses covering the area. This suggests that the observed increase in SOC content is not a result of agricultural management practices, but rather is climatically-driven. This increase could be driven by the three consecutive years of high rainfall received in Hillston from 2010 to 2012, as it is known that there is a strong positive correlation with rainfall amount and SOC content (Jobbágy and Jackson 2000). The reasons for this particular part of the study area undergoing an increase in SOC content could be due to the position in the landscape, the soil type, or an interaction of several factors. In addition, this area possessed relatively low topsoil SOC contents in 2002 (Fig. 6.5). Because the soils in this location are starting off at a lower baseline SOC content, there is greater potential for this to increase, as soils with a low in SOC content are often more responsive to factors that drive SOC accumulation, and these changes can occur rapidly (Klemmedson 1989).

While it is generally expected that soils under natural vegetation are in equilibrium, and that organic carbon content of these soils would not change over decadal time scales (West and Post 2002), the results from our study suggest that several years of high rainfall can significantly impact SOC content of soils in a semi-arid landscape. Studies have proposed that there is considerable potential for fluctuating rainfall patterns to significantly alter the levels of organic carbon in a soil (Aanderud *et al.* 2010), but it is generally expected that this would not be significant at such short time scales. Despite this, the climatic conditions experienced during the study period from 2002 to 2015 represent quite an extreme case of fluctuating rainfall patterns and amounts. The three consecutive years of heavy rainfall from 2010 to 2012 was the wettest period that Hillston had experienced on record, and in the eight years prior to this from 2002 to 2009, Hillston experienced the worst drought ever recorded (BOM 2017). It is probable that the lengthy drought caused a decrease in SOC content across the study area, although this is simply speculation without a measurement at the junction of the shift in rainfall patterns in 2010. Nevertheless, it appears that the period of high rainfall had at least negated the impacts of the drought on SOC content, or superseded them.

There has also been considerable discussion in the literature about the potential for irrigated agriculture to increase the accumulation of SOC in semi-arid areas, due to the inherently low SOC levels

in such areas (Entry *et al.* 2002). The topsoil z-score map of SOC change produced in our study indicated that only a small area had experienced a statistically significant increase in SOC content in the topsoil, but the role that irrigation played in this was unclear. The mean SOC content of all sampled sites also indicates that both natural and cotton sites have increased by 0.05% from 2002 to 2015 (Table 6.3), however, the data from sites only sampled at both time points suggests that there was essentially no change in SOC content during the study period (Table 6.4). Consequently, it is very difficult to disentangle the impacts of management and climate from these results. The mean topsoil SOC content for natural sites was 0.07% higher than cotton sites in both the 2002 and 2015 surveys (Table 6.3), which suggests that the cotton land use has a relatively negative impact on SOC content compared to the natural land use. This is also supported by the small, isolated areas on the alluvial floodplains in the south-west of the study area that underwent a statistically significant decrease in SOC content during the study period, as these decreases solely occurred in soils under irrigated cotton production. Similar trends have been found in a study on Vertosols in eastern Australia where soil under cotton production generally displayed lower SOC levels compared to adjacent native soils (Conteh *et al.* 1997). These trends suggest that although the plant biomass entering the soil is much higher under cotton production compared to the natural land use, this is not resulting in overall increases in SOC content at Hillston. It is known that the process of cultivating soil promotes the loss of SOC (Studdert and Echeverria 2000), and it is likely that the negative effects of tillage on SOC levels are counteracting the positive effects of increased biomass inputs. Soils under the dryland land use in Hillston are also typically cultivated, and a similar trend of a decrease in SOC over time was observed, with a mean SOC content of 0.78% in 2002 shifting to 0.69% in 2015 (Table 6.4). It is possible that a net accumulation of SOC could occur under irrigated cotton production if soils were not required to be constantly cultivated.

6.4.3. Multivariate linear mixed models

The SOC content prediction quality of the MLMM in this study is high (0.68 LCCC using LOSOCV), and is generally higher than other studies that have spatially modelled SOC at a similar density, although the results in the published literature vary considerably. Karunaratne *et al.* (2014) found that the quality of SOC predictions ranged from an LCCC of 0.20 to 0.76 after fitting separate univariate LMMs to two different datasets and testing this with leave-one-out cross-validation. A study of a similarly-sized region in Tibet also found that SOC could be predicted with an LCCC of 0.54 using independent validation (Yang *et al.* 2016). While SOC content could be predicted with high accuracy using our MLMM, it is also crucial that our fitted model has appropriate SSPE values, so that we know that the prediction variance reflects

the actual error of the predictions. Our fitted model had a mean SSPE close the expected value of 1.0, but the median SSPE was 0.34, which is slightly lower than the expected value of 0.455, indicating that some outliers may be present. Despite this, we have an acceptable estimate of the reliability of the prediction variance, and this is crucial, as the prediction variance is used to calculate the z-statistic, and subsequently calculate the statistical significance of the estimated change in SOC content.

In the previous research chapters in this thesis we have utilised co-located soil information from different time points to create a single bivariate linear mixed model (BLMM), and thus account for spatial correlation through time. However, in this chapter we have also utilised vertical spatial correlation from other co-located soil depths. The utilisation of this correlation is a likely reason for the high quality of the predictions presented in this study. In addition, simultaneously modelling multiple response variables increases the size of the training dataset. The only published study, to our knowledge, that utilises co-located soil data and the covariance between different sampling depths was undertaken by Orton *et al.* (2014). In their study, the authors used a linear model of co-regionalization (LMCR) to model two soil properties at three depths, resulting in six variables in the LMCR. This meant that the correlation between the two soil properties, and the correlation between the different depths, was utilised. Results showed improved predictions, particularly where missing samples occurred (Orton *et al.* 2014). Another significant advantage of modelling multiple soil depths using this approach is that there is a logical connection between soil maps of the same property at different depth increments. This ultimately results in soil maps that pedologically make sense, as the values of the soil property at a particular depth has been influenced by the value of the soil property at different depth increments above, below, or both above and below the depth of interest. This approach is particularly advantageous when some sampling depths have more samples than others.

While using MLMMs to model SOC content and other soil properties in space and time contains many benefits, there are a few limitations that must be acknowledged. The most notable is the amount of time and computing power required to create these models, as this is a serious limitation to their widespread adoption. In this study, manual backwards elimination of predictor variables was performed and each trend included nine predictor variables (two categorical, and seven numerical), resulting in 36 predictor variables in the model. To reach our final fitted MLMM, we eliminated 23 of the possible 36 predictor variables. Each time a predictor variable was removed, a new model needed to be created with the remaining suite of predictor variables. Using a 3.20 GHz processor with 4.00 GB of RAM per core, it took approximately 28 hours to create a model, meaning that a minimum of ~27 days was

required to reach our final result. This would be even more of a limitation if the dataset was larger (i.e. more soil samples), and/or if more responses were included (e.g. four depths and two time points). While there are ways of streamlining this approach, such as removing redundant covariates by running a simple linear model on each individual response variable and trend, this is nevertheless a serious limitation to the adoption of multivariate linear mixed models.

6.5. Conclusions and future directions

It was clear from the results in this study that the inclusions of soil pH and STC content as predictor variables dramatically improved the prediction of both SOC and SIC content when combined with VisNIR wavelengths of the scanned soil samples. This shows promise for including other readily available soil data with VisNIR to predict different soil properties, particularly when the soil property used as a predictor is correlated with the soil property to be predicted. The overall SOC content of the cotton-growing region of Hillston was found to be low. No considerable change in subsoil (0.3-0.5 m) SOC content was observed from 2002 to 2015, but topsoil (0-0.1 m) maps of SOC content revealed a trend of a statistically significant increase in the northeast of the study area. This increase appeared to be climatically-driven, rather than a function of land management, and is likely due to the three consecutive years of heavy rainfall received in the study area from 2010 to 2012. The results suggested that irrigated cotton production did not increase SOC content from natural levels, and small, isolated areas under irrigated cotton production were predicted to have undergone a statistically significant decrease in SOC content from 2002 to 2015. These results suggest that while irrigated cotton production increases the amount of plant biomass that enters the soil, the negative effects of tillage counteracts this. The MLMM used in this study predicted SOC content with high accuracy, with an LCCC of 0.68 when tested with LOSOCV. These high quality results are likely due to the utilisation of the correlation and covariance between samples extracted at different time points, as well as the correlation and covariance between different depth increments of the same soil core. This study provides evidence of the advantages of using MLMMs to monitor soil properties in space and time. The subsequent increase in the sensitivity of detecting a statistically significant change in soil properties over time that these types of models provide is also demonstrated.

References

- Aanderud, ZT, Richards, JH, Svejcar, T, James, JJ (2010) A shift in seasonal rainfall reduces soil organic carbon storage in a cold desert. *Ecosystems* **13**, 673-682.
- Bishop, TFA, McBratney, AB, Laslett, GM (1999) Modelling soil attribute depth functions with equal-area quadratic smoothing splines. *Geoderma* **91**, 27-45.

- Bureau of Meteorology - BOM (2017) Monthly rainfall – Hillston airport. Available from: <http://www.bom.gov.au/climate/averages/tables/cw_075032_All.shtml> [25 April 2017]
- Cécillon, L, Barthès, BG, Gomez, C, Ertlen, D, Genot, V, Hedde, M, Stevens, A, Brun, JJ (2009) Assessment and monitoring of soil quality using near-infrared reflectance spectroscopy (NIRS). *European Journal of Soil Science* **60**, 770-784.
- Chan, K, Hodgson, A, Bowman, A (1995) Degradation of Australian vertisols after conversion from native grassland (*Astrelba lappacea*) to continuous cropping in a semi-arid subtropical environment. *Tropical Grasslands* **29**, 210-217.
- Chang, C-W, Laird, DA, Mausbach, MJ, Hurburgh, CR (2001) Near-infrared reflectance spectroscopy–principal components regression analyses of soil properties. *Soil Science Society of America Journal* **65**, 480-490.
- Conteh, A, Blair, GJ, Lefroy, RDB, Macleod, DA (1997) Soil organic carbon changes in cracking clay soils under cotton production as studied by carbon fractionation. *Australian Journal of Agricultural Research* **48**, 1049-1058.
- Conyers, M, Liu, DL, Kirkegaard, J, Orgill, S, Oates, A, Li, G, Poile, G, Kirkby, C (2015) A review of organic carbon accumulation in soils within the agricultural context of southern New South Wales, Australia. *Field Crops Research* **184**, 177-182.
- Deng, F, Minasny, B, Knadel, M, McBratney, A, Heckrath, G, Greve, MH (2013) Using Vis-NIR Spectroscopy for Monitoring Temporal Changes in Soil Organic Carbon. *Soil Science* **178**, 389-399.
- Entry, JA, Sojka, RE, Shewmaker, GE (2002) Management of Irrigated Agriculture to Increase Organic Carbon Storage in Soils. *Soil Science Society of America Journal* **66**, 1957-1964.
- Filippi, P, Minasny, B, Cattle, SR, Bishop, TFA (2016) Monitoring and Modeling Soil Change: The Influence of Human Activity and Climatic Shifts on Aspects of Soil Spatiotemporally. *Advances in Agronomy* **139**, 153-214.
- Gaffey, SJ (1986) Spectral reflectance of carbonate minerals in the visible and near infrared (0.35 - 2.55 microns). Part 1: Calcite, aragonite, and dolomite. *American Mineralogist* **71**, 151-162.
- Grisso, RD, Alley, MM, Holshouser, DL, Thomason, WE (2005) Precision Farming Tools. Soil Electrical Conductivity.
- Jobbágy, EG, Jackson, RB (2000) The vertical distribution of soil organic carbon and its relation to climate and vegetation. *Ecological applications* **10**, 423-436.
- Karunaratne, SB, Bishop, TFA, Odeh, IOA, Baldock, JA, Marchant, BP (2014) Estimating change in soil organic carbon using legacy data as the baseline: issues, approaches and lessons to learn. *Soil Research* **52**, 349-365.
- Klemmedson, JO (1989) Soil organic matter in arid and semiarid ecosystems: sources, accumulation, and distribution. *Arid Land Research and Management* **3**, 99-114.
- McBratney, AB, Minasny, B, Cattle, SR, Vervoort, RW (2002) From pedotransfer functions to soil inference systems. *Geoderma* **109**, 41-73.
- McCarty, G, Reeves, J, Reeves, V, Follett, R, Kimble, J (2002) Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. *Soil Science Society of America Journal* **66**, 640-646.
- Miklos, M, Short, MG, McBratney, AB, Minasny, B (2010) Mapping and comparing the distribution of soil carbon under cropping and grazing management practices in Narrabri, north-west New South Wales. *Soil Research* **48**, 248-257.
- Minasny, B, McBratney, AB, Hong, SY, Sulaeman, Y, Kim, MS, Zhang, YS, Kim, YH, Han, KH (2012) Continuous rice cropping has been sequestering carbon in soils in Java and South Korea for the past 30 years. *Global Biogeochemical Cycles* **26**, GB3027.

- Orton, TG, Pringle, MJ, Page, KL, Dalal, RC, Bishop, TFA (2014) Spatial prediction of soil organic carbon stock using a linear model of coregionalisation. *Geoderma* **230–231**, 119-130.
- Ross, CW, Grunwald, S, Myers, DB (2013) Spatiotemporal modeling of soil organic carbon stocks across a subtropical region. *Science of The Total Environment* **461–462**, 149-157.
- Saby, NPA, Arrouays, D, Antoni, V, Lemerrier, B, Follain, S, Walter, C, Schwartz, C (2008) Changes in soil organic carbon in a mountainous French region, 1990–2004. *Soil Use and Management* **24**, 254-262.
- Studdert, GA, Echeverria, HE (2000) Crop Rotations and Nitrogen Fertilization to Manage Soil Organic Carbon Dynamics. *Soil Science Society of America Journal* **64**, 1496-1503.
- Viscarra Rossel, RA, Walvoort, DJJ, McBratney, AB, Janik, LJ, Skjemstad, JO (2006) Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* **131**, 59-75.
- Walkley, A, Black, IA (1934) An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method. *Soil Science* **37**, 29-38.
- Wang, C, Li, W, Yang, Z, Chen, Y, Shao, W, Ji, J (2015) An invisible soil acidification: Critical role of soil carbonate and its impact on heavy metal bioavailability. *Nature* **5**, 12735.
- West, TO, Post, WM (2002) Soil Organic Carbon Sequestration Rates by Tillage and Crop Rotation. *Soil Science Society of America Journal* **66**, 1930-1946.
- Yang, R-M, Zhang, G-L, Yang, F, Zhi, J-J, Yang, F, Liu, F, Zhao, Y-G, Li, D-C (2016) Precise estimation of soil organic carbon stocks in the northeast Tibetan Plateau. *Scientific reports* **6**, 21842.

Chapter 7

7. Modelling subsoil inorganic carbon in space and time with mixture models

Abstract

Soil inorganic carbon (SIC) accounts for approximately 30-40% of global soil carbon stocks, and up to 90% of total carbon stocks in arid and semi-arid regions. Agricultural management, particularly irrigation, has the capacity to alter these stocks of SIC. This study uses soil data from surveys performed in 2002 and 2015 to spatio-temporally monitor subsoil (0.3-0.5 m) inorganic carbon content in a semi-arid irrigated cotton-growing region in the lower Lachlan River valley in southern NSW. Modelling SIC content is often challenging as SIC datasets are often zero-inflated and highly skewed compared to the distributions of other soil properties. This study uses mixture models to overcome this; such an approach involved initially modelling the presence or absence of SIC in the study area with a Random Forest model, and then separately modelling SIC content using values from the dataset that are above-zero only, with a bivariate linear mixed model. The results from these two models are then combined, producing a mixture model. This mixture model approach was deemed successful, suggesting that this method could be adopted by other studies that model SIC content, or to model other environmental properties that have a similarly unusual distribution. This approach was also compared to the prediction of SIC content through the difference between soil total carbon (STC) and soil organic carbon (SOC) content maps produced from multivariate linear mixed models (MLMM). The results from this second approach were poor, and there were many limitations. The study found that there were generally no statistically significant changes in SIC content between 2002 and 2015, although both increases and decreases in SIC content were observed at some paired-sites sampled in both surveys in the upper 0.5 m of the soil profile. These changes did not appear to be a result of land use and management, but rather due to the heavy rainfall experienced during 2010-2012 resulting in dissolution and leaching of carbonates to/from other layers of the soil profile. The high spatial variability of SIC content was identified as an important limitation to the detection of statistically significant changes. Future research should focus on monitoring spatio-temporal aspects of SIC deeper in the subsoil, as well as monitoring several sampling depths in conjunction to analyse the downward movement of SIC within the profile over time.

7.1. Introduction

It is well known that soil is the largest terrestrial pool of carbon, and over the last few decades there has been increased effort into quantifying, mapping, and monitoring the world's soil carbon stocks. The organic component of soil carbon has almost solely been the focus of these studies, despite soil inorganic carbon (SIC) accounting for approximately one third of global soil carbon stocks (Lal 2004). Arid and semi-arid areas possess intrinsically low soil organic carbon (SOC) levels, as well as high SIC levels, with this SIC accounting for up to 90% of total soil carbon stocks in these landscapes (Suarez 2006). While SIC typically refers to calcium carbonate (CaCO_3), sodium and magnesium carbonates can often be present in substantial quantities in the saline and sodic soils of arid and semi-arid regions (Sanderman 2012).

While carbonates are not found in all soils, SOC is a ubiquitous feature that is highly impacted by land management (Sanderman 2012). In contrast, SIC is considered to be more temporally stable, although land management practices can still significantly alter the direction and severity of changes in SIC (Lal and Kimble 2000; Sanderman 2012). These changes in SIC could have an important impact on soil pH buffering capacity and the soil growing environment of crops (Entry *et al.* 2004). In particular, irrigation can significantly influence changes in SIC, and many arid and semi-arid areas have started to be used for irrigated agriculture across the globe. Irrigated cotton production in the semi-arid regions of eastern Australia is widespread and continues to grow, with the bulk of this undertaken on Vertosols that contain significant amounts of SIC (Knowles and Singh 2003; Sanderman 2012). Despite this, there has been very little research examining the effects of irrigation on SIC stocks in these areas.

The impact that irrigation practices have on SIC content is not clear, as many studies often come to contrasting conclusions (Entry *et al.* 2004; Wu *et al.* 2008; Mikhailova *et al.* 2013). It is generally accepted that irrigation decreases the carbonate content of upper parts of soil profiles due to the leaching and dissolution of these materials (Eswaran *et al.* 2000; Suarez 2006). This is logical, as the lack of rainfall in arid and semi-arid regions is the primary reason that these areas possess high amounts of SIC. Irrigation in these landscapes often encourages the redistribution and downward migration of carbonates, as they are dissolved and leached from the upper part of the soil profile and re-precipitated at deeper layers (Khokhlova *et al.* 1997). However, this is dependent on the amount of irrigation applied and how well drained the soils are (Sanderman 2012). Regardless, a redistribution of SIC away from the plant-root zone could have various impacts on crop production and soil pH buffering capacity.

It has also been proposed that the composition of irrigation water is critical in determining changes to SIC content, with studies suggesting that soils irrigated with water with a high electrical conductivity (EC) increases SIC content, whereas using irrigation water with a low EC decreases SIC over time (Wu *et al.* 2008). In addition, irrigation generally increases CO₂ concentrations in the soil in arid and semi-arid landscapes (Sanderman 2012), and this affects both carbonate dissolution and precipitation reactions, although there is debate as to the exact mechanisms (Entry *et al.* 2004; Suarez 2006). There has been much discussion of the role that SIC plays in sequestering CO₂ from the atmosphere, but it is generally accepted that it is relatively insignificant, as an increase in SIC does not automatically result in atmospheric CO₂ sequestration (Eswaran *et al.* 2000; Sanderman 2012).

There are very few studies that monitor spatio-temporal changes in SIC over a region (Wu *et al.* 2009), and this is likely due to a number of reasons. It is difficult to measure and monitor SIC, as unlike SOC that primarily accumulates at the surface, it preferentially accumulates at depth (Sanderman 2012). Furthermore, the spatially variable nature of SIC makes it difficult to model and monitor changes in space and time, as contrasting SOC, SIC can vary considerably over very short distances. Soil inorganic carbon is often found in accumulations, and two soil cores extracted five metres from each other can contain SIC contents that are magnitudes different, with this being even starker in deeper layers of the subsoil. Consequently, the monitoring of SIC and returning to georeferenced sites to observe temporal change is often challenging and confounded by the inherent spatial variability of SIC. As carbonates are not found in all soils, SIC datasets are often zero-inflated (i.e. many zero values) and highly skewed (Grinand *et al.* 2012), which presents challenges in statistical analysis and spatio-temporal modelling that often cannot be overcome by statistically transforming the data. This issue is commonplace with other environmental variables, such as soil contaminants; where sites are either un-contaminated or contaminated with values ranging from low to very high (Johnson *et al.* 2017), and streamflow; where some streams are not flowing, while others are flowing at varying rates (Van Ogtrop *et al.* 2011).

A modelling approach that suitably deals with highly-skewed and zero-inflated datasets is the mixture model method, where two separate models are used to achieve the final result. Initially, a model is created to predict the presence or absence of the response variable, and then a model is created to predict the quantity of the response variable where only values above zero are used in model creation. Subsequently, these two models are combined, which results in a value that is either zero (where the initial model predicted an absence), or the same value as the predicted response variable quantity model (where the initial model predicted a presence). In this study, we use the mixture model

approach to model subsoil inorganic carbon content in space and time in a semi-arid irrigated cotton-growing district in south-west NSW, Australia. This approach is also compared to the prediction of SIC content through the difference between soil total carbon (STC) and SOC content maps produced from multivariate linear mixed models (MLMM), as described in Chapter 6. The aim of the study is to analyse the impact of different land uses and shifting rainfall patterns on changes in SIC content, as well as investigate the appropriateness of different methods of modelling SIC in space and time.

7.2. Materials and methods

7.2.1. Soil datasets and soil property analyses

This study uses data from two soil surveys of a semi-arid area surrounding the township of Hillston, in south-west NSW, where 113 soil cores from a survey in 2002 are used, as well as 160 soil cores from a survey performed in 2015. Many of the same sites were sampled in both surveys ($n = 103$), as the locations were georeferenced. At 20 of the sites in 2015, a soil core was extracted ~30-50 m away and analysed separately to examine soil spatial variation at short distances. This chapter uses soil data at different sampling depth intervals to 0.5 m. The subsampling intervals in the two surveys differed, with the 2002 survey sampled at 0-0.2 and 0.3-0.4 m, and the 2015 survey sampled at 0-0.1, 0.1-0.3 and 0.3-0.5 m. Because of the differing sampling depths between the two surveys, equal-area quadratic smoothing splines were used to standardise the depths of the 2002 survey to the depths of the 2015 survey (0-0.1, 0.1-0.3 and 0.3-0.5 m) (Bishop *et al.* 1999). As described in Chapter 6, SIC content was determined by both traditional laboratory methods, and through the use of visible near infrared (VisNIR) spectroscopy. The concentration of SIC, rather than SIC stock is the focus of this chapter, as the purpose is not that of carbon accounting.

7.2.2. Soil inorganic carbon prediction by mixture models

Mixture models are used in this study to deal with the zero-inflated and highly-skewed nature of the SIC dataset. Firstly, a model is created to predict the presence or absence of SIC in the study area, and then another model is created to predict the amount of SIC, using only values above zero from the dataset. Consequently, these two maps are combined, creating the mixture model. While mixture models are sporadically used in other fields of research, there are no studies to our knowledge that have adopted this approach to model the spatial distribution of SIC. Only models and maps of the 0.3-0.5 m depth are discussed in this study, as this sampling depth contained the most samples with SIC present, and the largest amount of SIC, compared to the 0-0.1 and 0.1-0.3 m sampling depths. The predictor

variables/covariates used for spatial modelling and mapping in this chapter has been discussed in detail in Chapter 3 and 4.

7.2.2.1. The probability model for the presence/absence of SIC

A Random Forest (RF) model was used to predict the presence or absence of SIC in the study area. The samples that contained any SIC value greater than zero were assigned the category of 'presence', and the samples that contained a SIC value equal to zero were assigned the category of 'absence'. In total, 733 samples from the 0-0.1, 0.1-0.3 and 0.3-0.5 m sampling depths in both 2002 and 2015 were used to create one model. In addition to the covariates described in Chapter 3, the mid-depth of the sample (the middle point of the upper and lower sampling depth) and the year that the sample was collected in, were included as predictor variables. The parameters of the RF models were set to 500 trees and three variables at each split. Predictions were assigned based on a probability threshold of 0.5; i.e. if the predicted probability of SIC presence was greater than 0.5, a 'presence' value was assigned, and if the probability of SIC presence was less than 0.5, an 'absence' value was assigned. The created model was then used to predict onto a 100 m grid at each of the two time points at the 0.3-0.5 m sampling depth, with the only difference in the study area grid between these being the 'year' category. After the maps were created, 'absence' predictions were assigned the numerical value of 0, and 'presence' predictions were assigned the numerical value of 1. Receiver operator characteristic (ROC) curves and the calculation of the area under curve (AUC) were used to assess model quality for both in-bag and out-of-bag predictions (Breiman 1996).

7.2.2.2. SIC content model with above zero values only

One bivariate linear mixed model (BLMM) was used to model SIC content at the 0.3-0.5 m sampling depth for 2002 and 2015 concurrently, with the majority of zero values removed from the dataset. The theory and details of the BLMM used here is described in greater detail in Chapter 3 and 4. Co-located sites (sites sampled in both years) that contained no SIC in both years were removed from the dataset, and non-co-located sites that did not contain SIC were also discarded. If a co-located site had SIC present in one year, and no SIC in the other year, both sites were retained for modelling. After this data filtering process, there were 103 individual sites used for modelling and 155 individual measurements. There were 52 co-located sites ($n = 104$ samples), with 4 sites sampled in 2002 only ($n = 4$ samples), and 47 sites sampled in 2015 only ($n = 47$ samples). At these co-located sites, 10 contained a zero value in 2002 and a value above zero in 2015, and 12 contained a zero value in 2015 and a value above zero in 2002. Due to the skewed nature of the remaining dataset, a log-normal transformation was performed

on the SIC content data. The combination of covariates to be included in the final model was determined by backwards elimination, and this is discussed in further detail in Chapter 3. After the final model was selected, predictions of SIC content on the covariate grid of the study area at 100 m resolution were made. The SIC data was subsequently back-transformed and digital soil maps produced.

7.2.2.3. Mixture model

The corresponding SIC presence map and SIC content map for each time point were combined by multiplication. This resulted in an additional map that contained either zero values, where the SIC presence map predicts an absence, or the same value from the SIC content maps, where the SIC presence map predicts a presence (i.e. 1 multiplied by the SIC content).

7.2.3. Prediction of SIC by difference of STC and SOC maps produced with MLMs

In Chapter 6, multivariate linear mixed models (MLMs) were used to predict SOC content, and the same approach was used to model STC content in this chapter. Details of the STC dataset used are described in the preceding chapter. The SOC and STC content maps produced from this were subsequently used to calculate the amount of SIC content in the study area by simply calculating the difference between SOC and STC content for the 0.3-0.5 m sampling depth. The details of this modelling approach can be viewed in Chapter 6. Directly modelling the full SIC content dataset using this approach was not a viable option due to the zero-inflated and skewed nature of the dataset. However, it is possible to do this for SOC and STC content as both can easily conform to the requirements (i.e. normally distributed) of the MLM, after statistical transformation.

7.3. Results

7.3.1. Summary statistics of inorganic carbon

The distribution of the SIC data at all sampling depths was highly skewed (Fig. 7.1) and zero-inflated (Table 7.1). Only a small number of samples in the 0-0.1 m sampling depth contained calcium carbonate, and approximately half of the samples contained some amount of SIC in the 0.3-0.5 m depth, with no clear differences in distribution between land uses or time points (Table 7.1). The quantity of SIC at 0.3-0.5 m depth was also largely similar for the different land uses, apart from dryland sites where SIC levels were considerably higher. The distribution of the above-zero SIC content values ranged from 0.01% to 1.60% (Table 7.1). While there were no clear differences in SIC content between the two time points, both increases and decreases were observed at individual sites that had been sampled in both surveys (Fig. 7.2). Data from co-located sites sampled in both 2002 and 2015 in the upper 0.5 m of the soil

profile showed a trend of a slight decrease of SIC content in the topsoil, and a slight increase in the subsoil under both cotton and natural land uses (Fig. 7.3). This suggests that carbonate dissolution has occurred in the topsoil, and then re-precipitated in the subsoil during the study period.

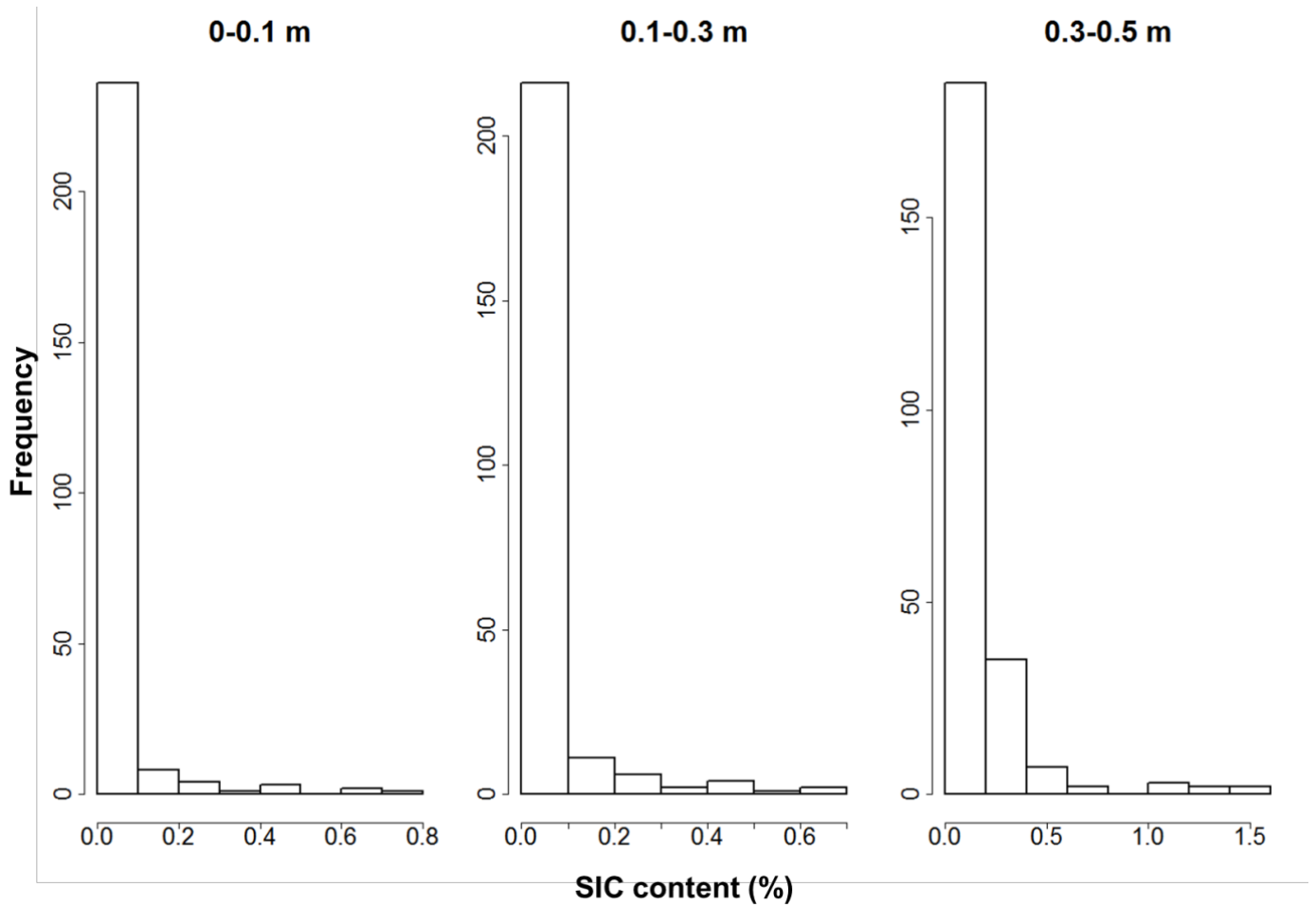


Fig. 7.1 – The distribution of soil inorganic carbon content (%) at the 0-0.1, 0.1-0.3 and 0.3-0.5 m sampling intervals at all sites

Table 7.1 – Soil inorganic carbon content (%) statistics of sites that possessed values above zero, and the number of samples that contained a zero value or a value above zero, by land use, depth and time point

Statistic	Depth	Cotton		Dryland		Natural		Horticulture	
		2002	2015	2002	2015	2002	2015	2002	2015
Mean	0-0.1 m	0.22	0.12	0.28	0.21	0.44	0.26	0.15	0.13
	0.1-0.3 m	0.09	0.11	0.20	0.31	0.12	0.15	0.15	0.08
	0.3-0.5 m	0.16	0.24	0.35	0.39	0.19	0.26	0.15	0.14
Minimum	0-0.1 m	0.04	0.01	0.15	0.14	0.01	0.09	0.04	0.13
	0.1-0.3 m	0.01	0.01	0.02	0.06	0.01	0.04	0.04	0.07
	0.3-0.5 m	0.03	0.03	0.04	0.01	0.02	0.01	0.04	0.01
Maximum	0-0.1 m	0.50	0.33	0.41	0.27	0.69	0.73	0.25	0.13
	0.1-0.3 m	0.42	0.40	0.67	0.49	0.66	0.52	0.25	0.08
	0.3-0.5 m	0.48	1.31	1.01	1.60	0.62	1.51	0.25	0.25
Standard deviation	0-0.1 m	0.192	0.105	0.184	0.092	0.376	0.312	0.148	NA
	0.1-0.3 m	0.102	0.129	0.239	0.222	0.184	0.153	0.148	0.007
	0.3-0.5 m	0.124	0.229	0.367	0.472	0.158	0.367	0.148	0.094
> zero values (n =)	0-0.1 m	6	11	2	2	3	4	2	1
	0.1-0.3 m	25	9	6	3	14	14	2	2
	0.3-0.5 m	25	35	6	12	13	35	2	5
= zero values (n =)	0-0.1 m	37	54	11	19	33	57	1	12
	0.1-0.3 m	18	15	7	3	14	19	1	3
	0.3-0.5 m	16	27	6	7	13	26	1	7

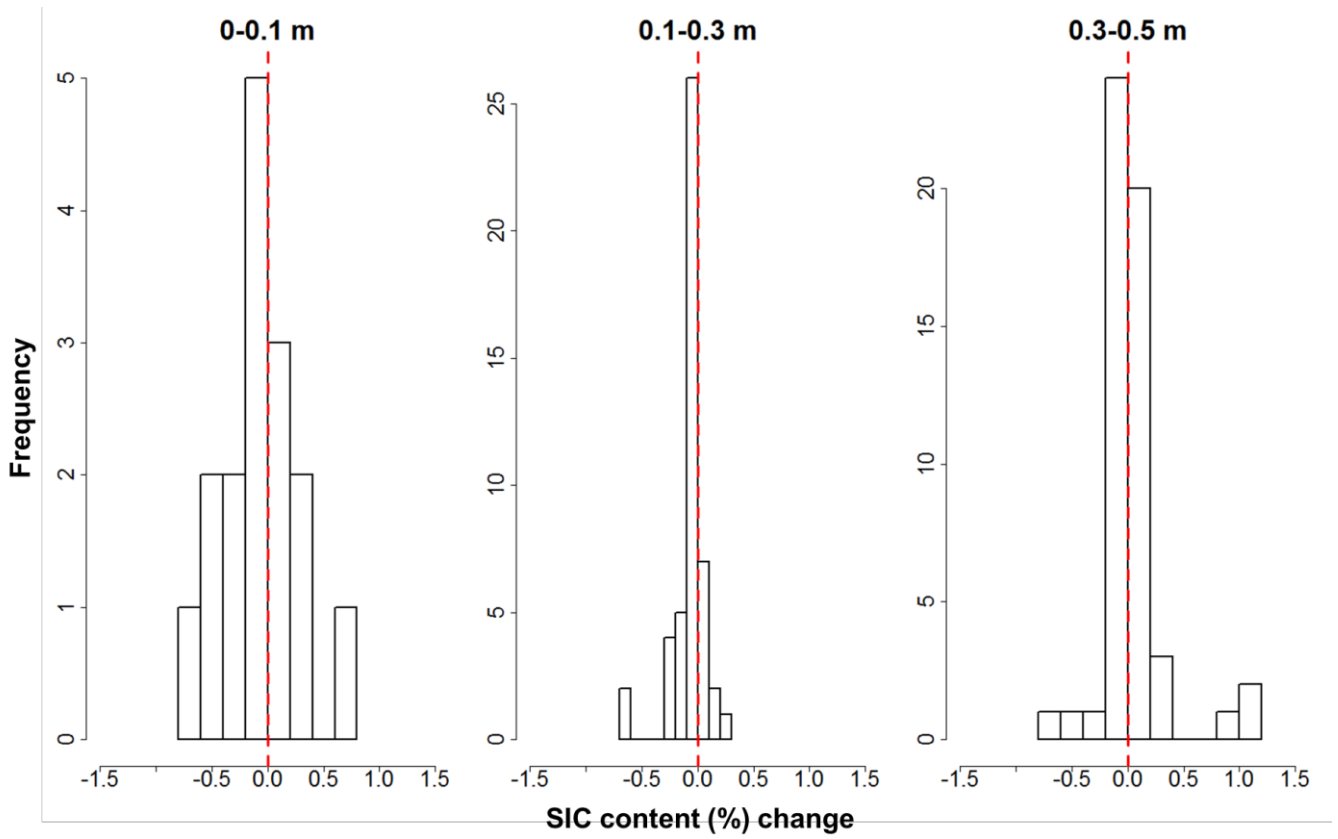


Fig. 7.2 – Change in the soil inorganic carbon content (%) at the 0-0.1, 0.1-0.3 and 0.3-0.5 m sampling intervals at co-located sites that included a presence of carbonates in both 2002 and 2015

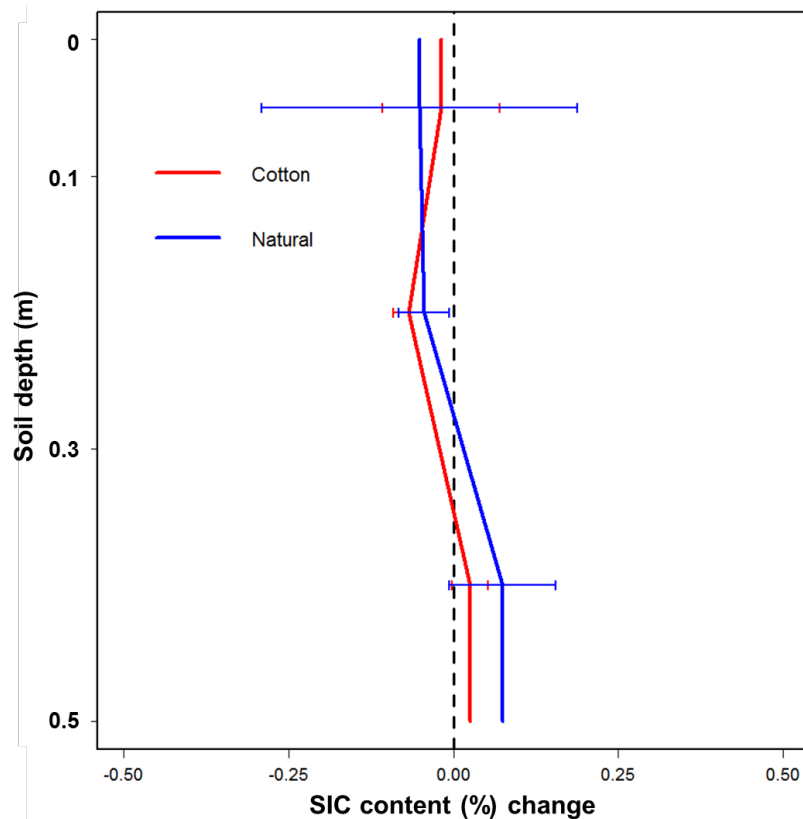


Fig. 7.3 – Mean change in soil inorganic carbon contents (%) from 2002 to 2015 of co-located sites under the cotton and natural land use to the depth of 0.5 metres (with standard error bars)

7.3.2. Inorganic carbon presence/absence predictions with Random Forest

The Random Forest model created to predict the presence or absence of SIC in the study area was highly accurate (Fig. 7.4). The ROC curve summarises the trade-off between true positives and false positives for different probability thresholds used to define when a class is predicted to be present. The true positive rate is the probability of a correct prediction, and the false positive rate is the probability of a false alarm. The closer the fitted line is to the 1:1 line, the poorer the model is. A plot that shows a perfect prediction is where the fitted line runs parallel to the y-axis on the x-axis origin, and parallel to the x-axis near the maximum y-axis value. Another useful measure of model quality is the area under the curve (AUC), which is 1.0 for a perfect model. Our out-of-bag (OOB) predictions showed that the model predicted with high accuracy, with an AUC of 0.88 (Fig. 7.4). The estimation of error rate for OOB predictions was 20.76% overall (Table 7.2). Out-of-bag error is a method of measuring the prediction error of RF models, and is essentially a method of cross-validating models. The model was better at predicting the absence of SIC, only incorrectly predicting this 10.8% of the time, while the presence of

SIC was incorrectly predicted 40.6% of the time (Table 7.2). Overall, this shows promise for using the model to predict the presence of SIC at locations where SIC has not been measured.

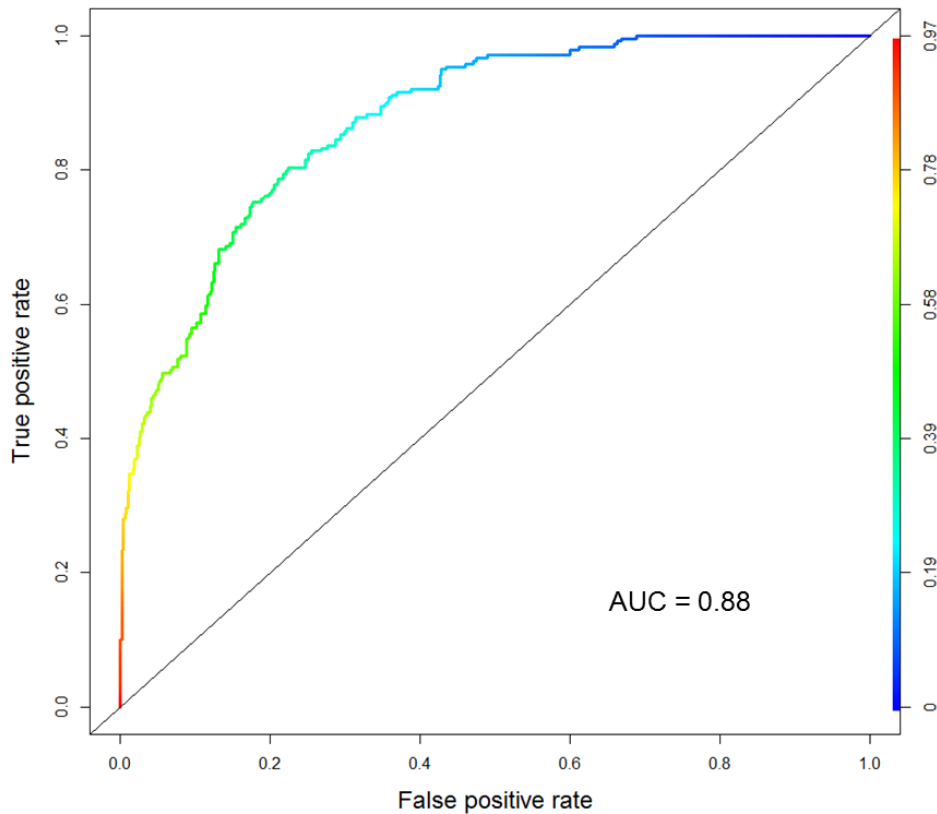


Fig. 7.4 – ROC (receiver operator characteristic) curve for out-of-bag predictions of the presence or absence of soil inorganic carbon with a Random Forest model

Table 7.2 – Confusion matrix – out-of-bag estimate of error rate = 20.76% – based on threshold probability of 0.5

		Observed		
		Absence	Presence	Error rate (%)
Predicted	Absence	423 True negative	51 False negative	10.8
	Presence	97 False positive	142 True positive	40.6

The importance of different predictor variables in the SIC presence models was tested by the mean decrease in accuracy, and the mean decrease in node impurity. Mean decrease in accuracy is calculated on the OOB data, where for each tree, the prediction error (error rate) on the OOB portion of the data is recorded. This is then done after permuting each predictor variable and the difference between the two

are then averaged over all trees, and normalised by the standard deviation of the differences. The mean decrease in accuracy plot essentially shows the impact of removing each predictor variable on the misclassification of observations (Fig. 7.5). Mean decrease in node impurity is measured by the Gini index, where the total decrease in node impurities from splitting on the variable is averaged over all trees (R Core Team 2017). The most important predictor variables were mid-depth, followed by PC2, and eastings (Fig. 7.5). While soil type and year were important in the mean decrease in accuracy, they were less important in the mean decrease in node impurity. Land use was consistently a poor predictor of the presence of SIC (Fig. 7.5).

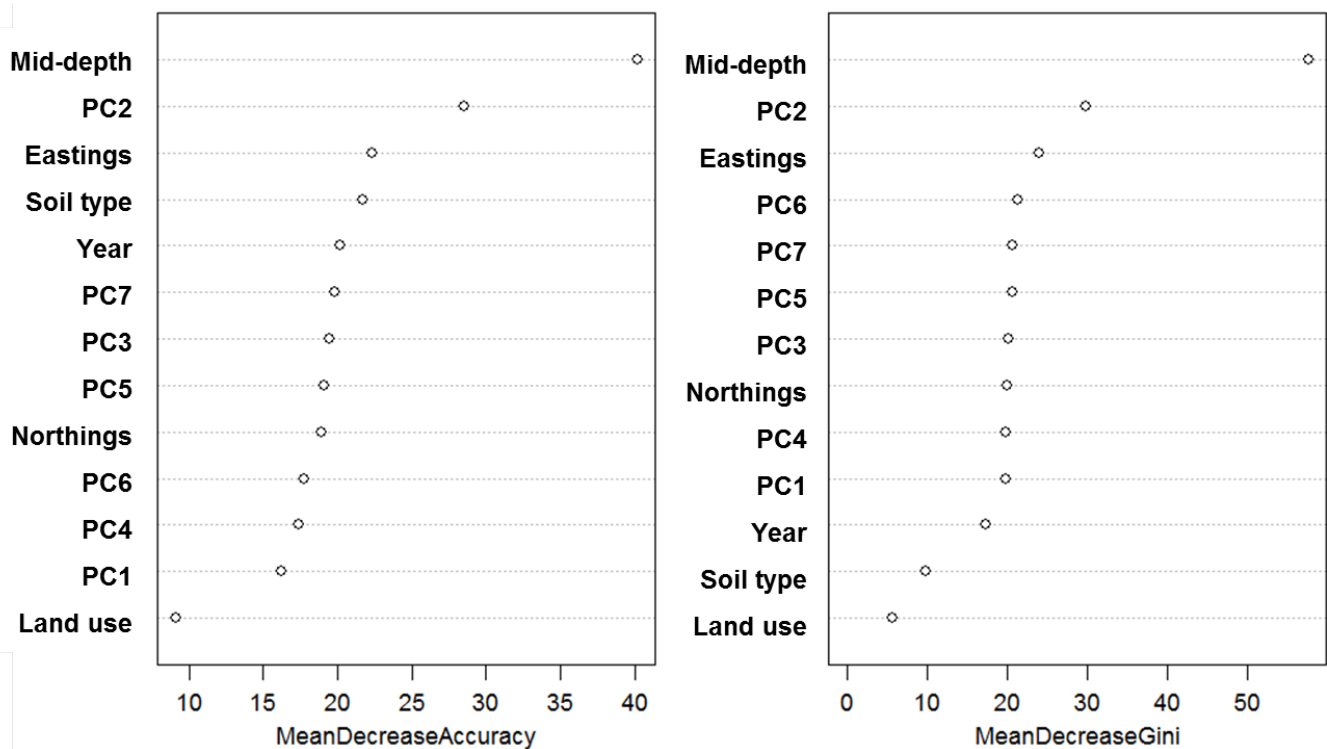


Fig. 7.5 – The importance of each predictor variable in the Random Forest model of the probability SIC with the contribution to mean decrease in accuracy (left) and mean decrease in node impurity (right)

The maps of the presence or absence of SIC, based on a probability threshold of 0.5, showed that SIC was predicted to be present in 47% and 43% of the study area at the 0.3-0.5 m sampling depth in 2002 and 2015, respectively (Fig. 7.6). There was little difference in the spatial distribution of the presence and absence of SIC at the different time points. Areas where SIC was predicted to be present were primarily located on the alluvial floodplains, southeast of the Lachlan river, and northeast of the study area (Fig. 7.6).

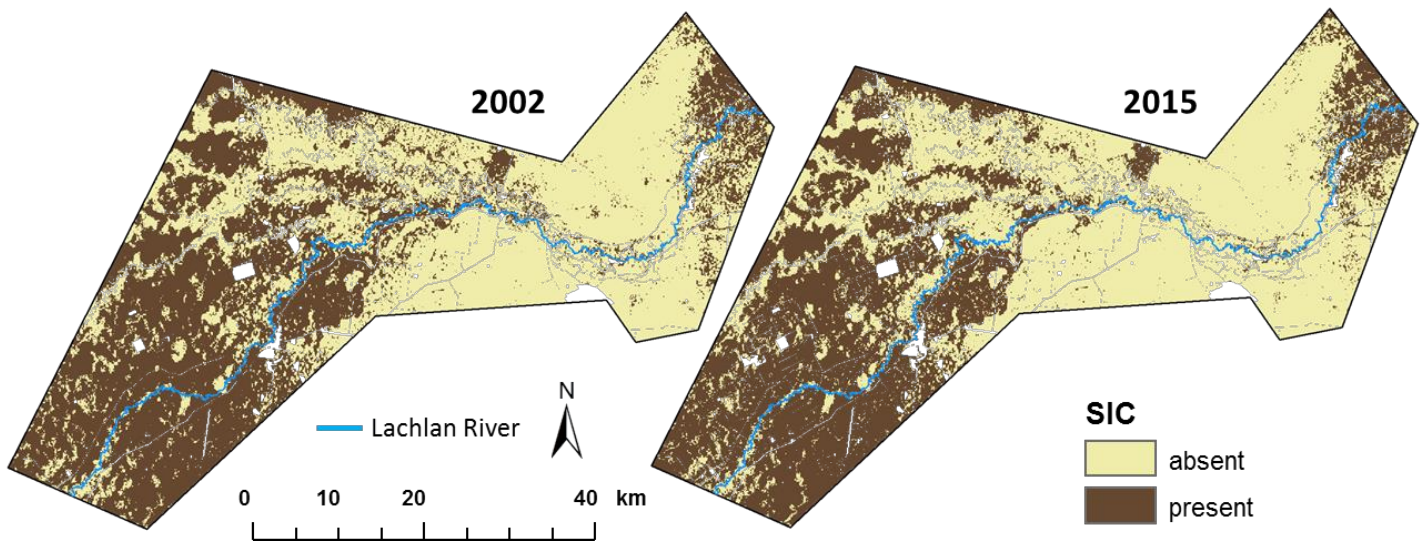


Fig. 7.6 – The predicted distribution of the presence or absence of soil inorganic carbon for year 2002 (left) and 2015 (right) for the 0.3-0.5 m sampling depth

7.3.3. Soil inorganic carbon content predictions with bivariate linear mixed models

The maps produced by BLMMs and modelled with sites that only contained SIC showed similar spatial trends at both time points (Fig. 7.7). Similarly to the SIC presence model (Fig. 7.6), PC2 was an important predictor variable in both trends (Table 7.3), and this PC is largely driven by radiometrics and hence geology, as can be seen in Chapter 4. In addition, soil type was an important predictor (Table 7.3), and these covariates are clearly reflected in the spatial patterns of the maps (Fig. 7.7). Similarly to the SIC presence model, the impact of land use was insignificant in the SIC content BLMM. The quality of the BLMM was generally low, with an LCCC (Lin's concordance correlation coefficient) of 0.25 (Table 7.4). The fitted model had mean and median standardised squared prediction errors (SSPE) that were close to the expected values of 1.0 and 0.455, respectively (Table 7.4), which means that while the predictions are not of high quality, the estimates of the uncertainty are quite good. While the SIC presence/absence maps showed that SIC was predicted to be primarily located west and southeast of the Lachlan River (Fig. 7.6), it can be seen that the SIC content BLMM is predicting the highest SIC values at locations where SIC was predicted to be absent (Fig. 7.6; 7.7). The reason for these clearly incorrect predictions is most likely due to the removal of samples from these locations for modelling, as they contained no SIC. The model is thus extrapolating SIC content values in this area; however, this ultimately does not affect our final maps, as these areas were predicted to contain no SIC and will thus be dealt with in the mixture modelling process (Fig. 7.8).

Table 7.3 – Included covariates in the final BLMM of SIC content

Trend	Covariates
2002	Soil type + PC2 + PC3 + PC6 + PC7
2015	Soil type + PC1 + PC2 + PC7

Table 7.4 – Quality statistics of the SIC content BLMM modelled with above-zero values only

Sampling depth	LCCC	Bias	Mean (SSPE)	Median (SSPE)
0.3-0.5 m	0.25	-0.02	0.97	0.38

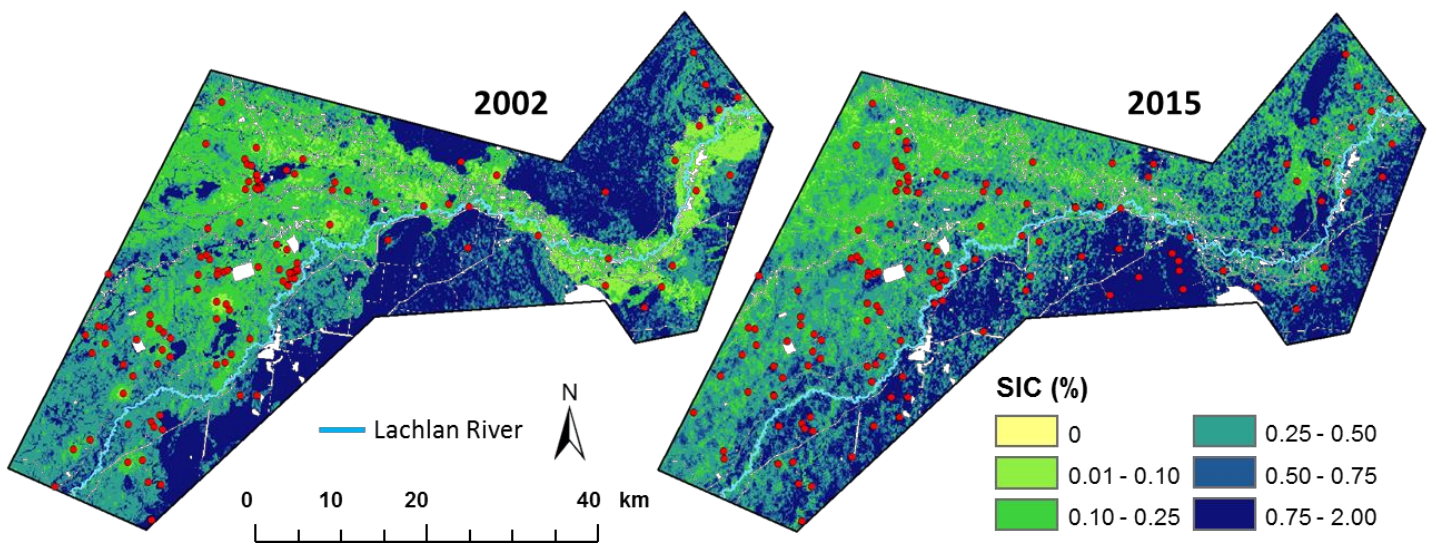


Fig. 7.7 – Soil inorganic carbon content (%) in 2002 (left) and 2015 (right) for the 0.3-0.5 m sampling depth produced from a bivariate linear mixed model using above-zero values only

7.3.4. Inorganic carbon maps from mixture models

Upon combining the corresponding SIC presence maps (Fig. 7.6) and SIC content maps (Fig. 7.7) for each time point, mixture model maps were produced (Fig. 7.8). There are some slightly differing patterns in SIC content between the two time points, but the spatial trends are largely similar (Fig. 7.8). Land use visually appears to have little impact on the distribution of SIC, and the spatial patterns seem to be largely driven by soil type. These maps show that there is broadly a low to moderate amount (0.01-0.50%) of SIC found at 0.3-0.5 m depth on the alluvial floodplains west of the Lachlan River where most of the irrigated cotton production is situated. The locations southeast of the study area and east of the Lachlan River contain the largest accumulations of SIC at 0.3-0.5 m depth, with values typically greater than 0.75% (Fig. 7.8). The change map between 2002 and 2015 for the 0.3-0.5 m depth shows that SIC

content was predicted to have increased over time in some areas, and decreased over time in others (Fig. 7.9).

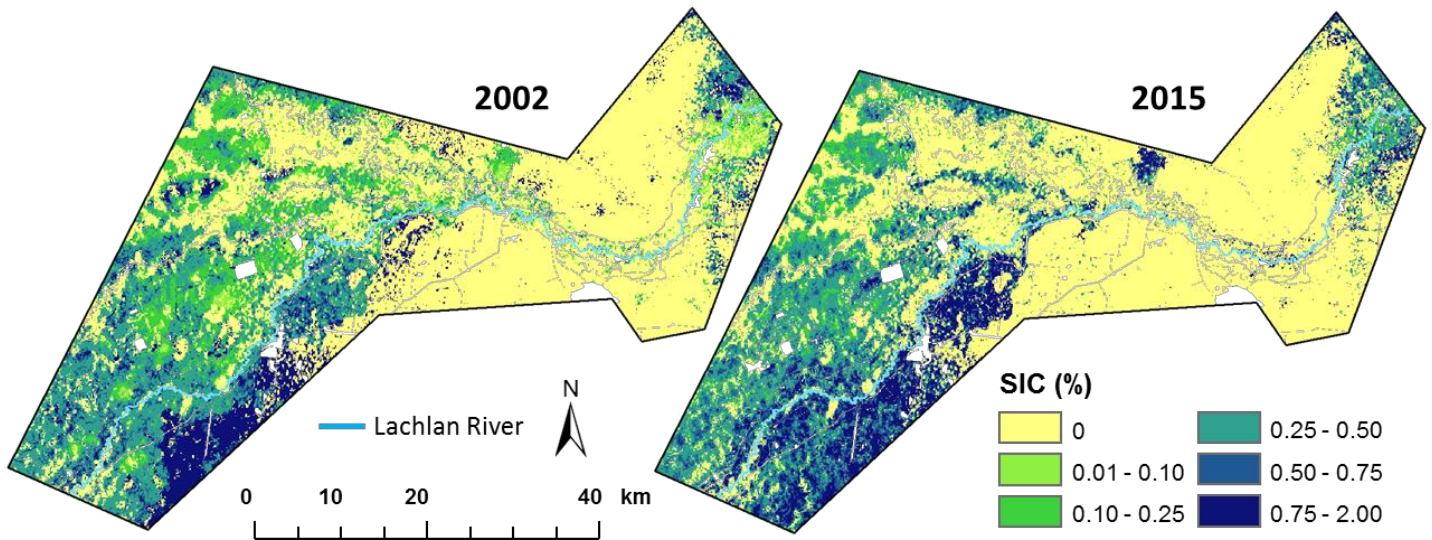


Fig. 7.8 – Soil inorganic carbon content (%) for the 0.3-0.5 m depth produced from mixture models

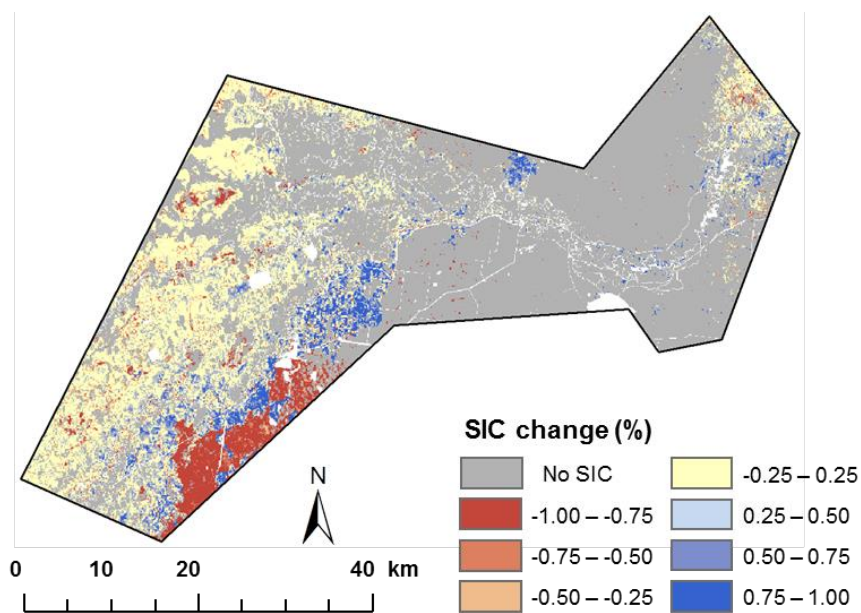


Fig. 7.9 – The change in soil inorganic carbon content (%) from 2002 to 2015 for the 0.3-0.5 m depth

7.3.5. Statistical significance of change in inorganic carbon

The mixture model maps for the two time points show some areas where there appears to be a change in SIC over time (Fig. 7.9). When the statistical significance of this change was investigated, it was apparent that only very small, isolated fragments were predicted to have experienced a statistically

significant change in SIC in the 13-year study period (Fig. 7.10). The area covered by the grey colour is where there was a predicted absence of SIC (Fig. 7.10).

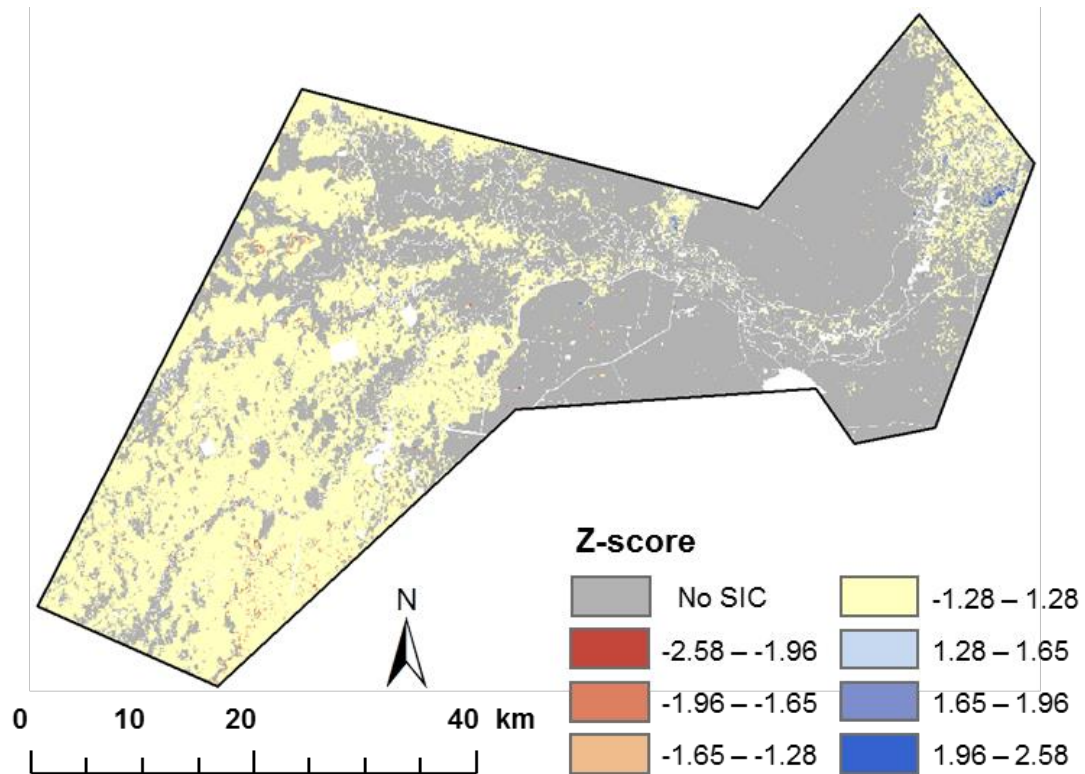


Fig. 7.10 – Z-scores showing the statistical significance of the change in soil inorganic carbon from 2002 to 2015 for the 0.3-0.5 m depth

7.3.6. Inorganic carbon maps from difference of total carbon and organic carbon

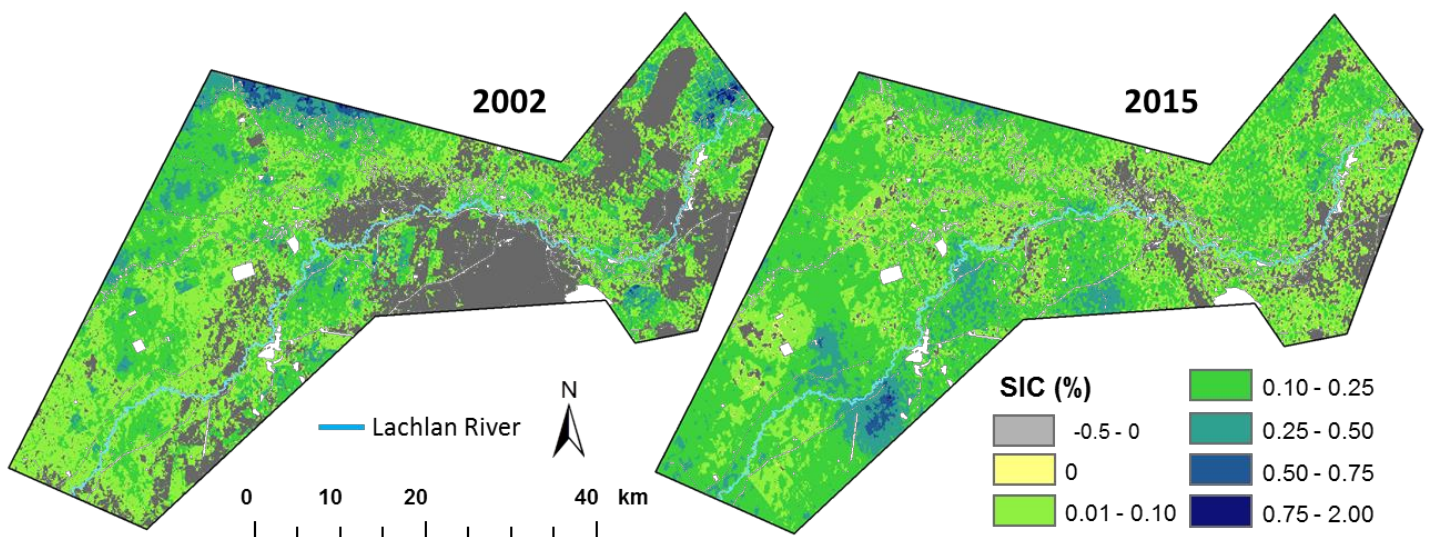
The maps and results from the SOC content MLMM are shown and discussed in Chapter 6, and the quality of the model predictions from this was high, with an LCCC of 0.68 using LOSOCV. The results from the STC content MLMM were also quite good, with an LCCC of 0.55 using LOSOCV (Table 7.5). The covariates included in the final fitted STC content MLMM varied for each individual response variable, although land use, PC1, and PC2 were important predictors for three of the four response variables (Table 7.6). Despite the high quality of the individual SOC and STC content MLMMs, the SIC content map produced by the difference of these maps (STC – SOC) showed unusual spatial patterns (Fig. 7.11) that had little relation to the SIC content maps produced by mixture models (Fig. 7.8). Negative values of SIC content were also predicted in a significant portion of the study area using this approach (Fig. 7.11).

Table 7.5 – Quality statistics of the STC content MLMM using LOSOCV

Sampling depth	LCCC	Bias	Mean (SSPE)	Median (SSPE)
0-0.1 & 0.3-0.5 m	0.55	-0.01	1.04	0.35

Table 7.6 – Included covariates in the final STC content MLMM

	Trend	Covariates
0-0.1 m	2002	PC1 + PC7
	2015	Land use + soil type + PC1 + PC2 + PC6
0.3-0.5 m	2002	Land use + soil type + PC1 + PC2 + PC4 + PC5 + PC7
	2015	Land use + PC2

**Fig. 7.11** – Soil inorganic carbon content (%) for the 0.3-0.5 m depth produced from the difference between maps of soil total carbon and maps of soil organic carbon

7.4. Discussion

7.4.1. The dynamics of soil inorganic carbon in the study area

Due to the comparatively more reliable results of the mixture model approach, the maps produced from this method are used to discuss SIC dynamics in the study area in this section. Approximately half of the soils in the cotton-growing region of Hillston contain calcium carbonate at the 0.3-0.5 m depth increment, with the quantity of SIC ranging from 0.01% to 1.60% in these soil samples (Table 7.1). Similar quantities of SIC at comparable depths have been found in Vertosols under irrigated cotton

production in northern NSW, Australia, where the SIC content ranged from 0.01% to 1.03% (Knowles and Singh 2003). In our study region, irrigated cotton production is primarily situated on the alluvial floodplains west of the Lachlan River, and the SIC maps produced show that there is a high correlation between these areas and the presence of SIC. This is likely because the clayey soils that are well-suited for cotton production naturally contain calcium carbonate. The lower hydraulic conductivity of these clayey soils typically results in slower leaching of carbonates out of the soil profile. The spatial distribution of the SIC presence maps were similar in both 2002 and 2015 (Fig. 7.6), and this is logical, as while the amount of inorganic carbon may change over time, it is less likely that a soil will completely lose all traces of carbonates.

In general, no change in SIC content was detected over the 13-year study period; apart from very small, isolated fragments that were predicted to have experienced either a statistically significant increase or decrease in SIC content (Fig. 7.10). Despite this, the standalone change map of SIC content (Fig. 7.9) indicated that some areas were predicted to have both increased and decreased in SIC content over time. These areas also coincided with the highest accumulations of SIC in the study area, and it is logical that the largest changes in SIC content occur here, as there is more carbonate to lose, and there is likely more to gain from depths above. The increases in SIC content in those areas may be due to migration and re-precipitation of carbonates from other depths within the soil profile, and the decreases may be due to the dissolution and leaching of carbonates to deeper soil layers. Data from paired-sites sampled in both 2002 and 2015 also showed a trend of a downwards movement of carbonates, where a slight mean decrease in SIC content was observed in the topsoil and a slight mean increase was observed in the subsoil (0.3-0.5 m) under both cotton and natural land uses (Fig. 7.3). This suggests that carbonate dissolution has occurred in the topsoil, and then re-precipitated in the subsoil during the study period. Similar trends have also been observed by other studies in irrigated regions (e.g. Khokhlova *et al.* 1997; Knowles and Singh 2003).

The impact of land use on the changes in SIC content observed in our study was not clear, and it is difficult to pinpoint the exact drivers of SIC change. The areas predicted to have undergone change in the standalone change map (Fig. 7.9) were covered by a mixture of cotton, dryland, natural, and the horticulture land uses, and the trend of the downward movement of carbonates at paired-sites was observed under both cotton and natural land uses (Fig. 7.3). Literature and field studies likewise often come to conflicting conclusions as to the influence of agricultural management and irrigation on SIC dynamics (Suarez 2006). Net decreases in calcium carbonate content in soils irrigated for four decades

as compared to non-irrigated sites have been found by a study in Israel (Magaritz and Amiel 1981), and similar results have been found in California, where a net loss in carbonates due to irrigation was observed over an 8-year period (Amundson and Smith 1988). In contrast, studies have observed no changes in net inorganic carbon storage after 90 years of irrigation (Suarez 1998). A more comparable study in an irrigated cotton-growing region by Knowles and Singh (2003) also found little difference in SIC content to 0.9 m in Vertisols after 20 years of irrigation compared with a site under remnant native vegetation. The results from our study suggest that the observed change in SIC content is not solely driven by land management, and it is more likely to be climatically-driven. It is possible that the sequential years of high rainfall received in Hillston from 2010 to 2012 has had some influence SIC contents.

Overall, the data from individual paired-sites that contained SIC measured at both time points showed that the majority of sites had experienced little change in SIC content over time (Fig. 7.2). Despite this, the change at certain sites ranged from a SIC content loss of 0.69% up to a gain of 1.19%, although these values could likely be influenced by the highly spatially variable nature of SIC, and not solely due to actual losses or gains of SIC content in those soils. The detection of actual changes in SIC content over time is further complicated in areas under flood-irrigated cotton production (about two thirds of the total area under cotton production in Hillston), due to the required implementation of landforming practices. Cut-and-fill operations move topsoil from one part of a field to another, to achieve a smooth gradient to allow water flow, and this consequently exposes subsoil to the surface, which alters the apparent depth of calcium carbonate (Cay and Cattle 2005).

This study detected some changes in SIC content over time, but very small amounts of this were considered to be statistically significant. While the driving factors of this change were not clear, it was apparent that the shifts in SIC content were more likely driven by climatic fluctuations than human activity. Our study focuses on the 0.3-0.5 m sampling depth, but future work could examine deeper layers of the soil profile, as there are typically larger accumulations of SIC. It would also be beneficial to analyse multiple sampling depths in conjunction, to determine whether there has been a migration of SIC down the profile across the study area.

7.4.2. Modelling approaches

7.4.2.1. Mixture models

The soil inorganic carbon content maps produced from the mixture models gave realistic spatio-temporal patterns for the study area. While studies have spatially modelled SIC at a range of scales, from farm (Miklos *et al.* 2010), to whole countries (Marchant *et al.* 2015), there are no published studies that have used mixture models to map SIC, with most opting for more traditional methods of digital soil mapping (DSM). Despite this, mixture models have been used in other fields of environmental science where the datasets were skewed and zero-inflated. Examples include a study that modelled the depth and duration of soil heating during fire (Stoof *et al.* 2011), and a study that modelled the flow of streams (Van Ogtrop *et al.* 2011).

There are some quite contrasting spatial patterns of SIC from the Random Forest SIC presence maps (Fig. 7.5) and the BLMM SIC content maps (Fig. 7.6). The SIC content maps show that the highest predictions of SIC are at locations where SIC was predicted to be absent by the SIC presence model. These erroneous predictions are most likely due to the removal of samples from these locations for modelling, as they contained no SIC. The SIC content model is thus extrapolating very high values in these areas where there are no representative samples included in the model. While this may appear to be a significant flaw, this ultimately does not affect our final maps, as these areas were predicted to contain no SIC and are dealt with in the mixture modelling process. If the SIC content maps were not combined with the corresponding SIC presence maps, the results would be considerably biased and incorrect.

In our study, the quality of the RF model of SIC presence was high, whereas the BLMM of SIC content was lower. The poor quality of the SIC content BLMM could be due to a number of reasons, such as the inherent high spatial variability of SIC content or the reduced size of the training dataset (as most samples with zero values were removed). In addition to this, the skewed nature of the dataset (even with zero values removed) is a likely limitation. One way to overcome these limitations is to use machine learning approaches like RF to model/map SIC content rather than the BLMM we used, as these do not require a normal distribution of data and can deal with zero-inflated datasets. However, a major pitfall of using machine learning approaches in spatio-temporal monitoring studies is that there is a lack of a spatial component (variogram), and the correlation of the response variables at co-located sites is ignored. While RF models were used to predict the presence of SIC in this study, this was deemed acceptable as whether a soil contains or does not contain carbonates is expected to remain relatively

constant over a 13-year time period. In contrast, it could be expected that the quantity of SIC at a location could fluctuate over time. Simply including the year the sample was collected as a covariate in a RF model, as was done with the SIC presence models, will not appropriately deal with the changes in SIC content at a location over time, as some sites would have experienced an increase in SIC, and others a decrease, but the mean may remain the same. For these reasons, it was decided that a BLMM was more appropriate to model SIC content in space and time.

There are examples of studies in the literature that have modelled/mapped SIC with datasets that contain zero values (Marchant *et al.* 2015; Sreenivas *et al.* 2016), and others that contain no zero values (Yang *et al.* 2010). While some of these studies have dealt with the challenges of modelling SIC suitably, others have not done so. In the study by Yang *et al.* (2010), it is unclear whether all zero values of SIC have been removed, or whether each sample in the dataset contained some amount of SIC, although this is unlikely due to the large size of the study region (~2,000,000 km²). If the zero values of SIC were removed from the dataset, the results of the study would have been heavily biased and compromised, and is comparable to simply using the SIC maps from Fig. 7.7 in our study as the final result. Some studies have also used soil test databases or similar arbitrarily-collected soil data to model and map SIC (Guo *et al.* 2006; Marchant *et al.* 2015), but caution must be taken when using these datasets. As SIC content is only measured if a sample is suspected of containing some amount of carbonates, there may be insufficient samples in these datasets that are known to contain no SIC. Marchant *et al.* (2015) used a large database of soil test results to model and map SIC, and overcame this particular issue by assigning any sample that had a pH less than 6.5 a zero value for SIC. It is essential that studies do not disregard samples that do not contain any SIC during the modelling process, as this can create significant amounts of bias and ultimately result in SIC maps that are not truly representative.

Regardless if zero values are present or not, SIC datasets are still typically highly skewed, and studies have dealt with this in various ways. As the dataset used to create the SIC content BLMM in our study was still skewed after the removal of most zero values, we applied a log-normal transformation. Similarly, Marchant *et al.* (2015) deemed that their SIC dataset was highly skewed and did not conform to standard statistical models and thus they normalised SIC observations by the rank order transform before using a linear mixed model on the transformed properties. In contrast, Sreenivas *et al.* (2016) judged that statistical transformation was not required as Random Forest models were used to model SIC, and it is commonly accepted that these do not depend on data distribution.

Overall, the use of mixture models to model and map SIC content in this study was a success, and this shows promise for using this approach to model SIC in other study regions, or to model similarly distributed soil properties, such as soil contaminants. The mixture model approach is easy to adopt, and would not be difficult to incorporate into future studies. The approach that researchers select to model SIC is largely dependent on the statistical distribution of the dataset, but our study suggests that using mixture models is advantageous where the dataset is both zero-inflated and highly-skewed.

7.4.2.2. SIC difference from STC and SOC maps

Most studies that map SIC content calculate the difference between STC and SOC content to determine SIC content at the observation point (e.g. soil sample), and then model this to create a SIC content map of a study region. However, because of the zero-inflated and highly-skewed nature of the SIC content dataset, it was not possible to effectively model this using more traditional DSM approaches directly in our study. Instead, we have separately modelled STC and SOC content, and then used the maps of STC and SOC content to calculate SIC content across the study area. It was clear that the results from this approach possessed many limitations. Unlike the mixture model approach, it was very difficult to identify soils that contain no SIC with this method. While those parts of the study area that contain no SIC would have the same SOC and STC contents, the separate modelling processes will not always predict the exact same values in these areas. As a result, when the SOC content map is taken away from the STC content map, areas that truly contain no SIC will likely be predicted to have negative or positive values.

It is also apparent that modelling soil total carbon is somewhat illogical and contradictory. While STC is the sum of organic and inorganic carbon of a soil, these two components of carbon are very different in several aspects. The way that SOC and SIC form is different, and often antagonistic, for example areas of high rainfall generally have high SOC and low SIC contents, while areas of low rainfall generally have low SOC and high SIC content. The relationship that SOC and SIC have with the spatial covariates used for modelling (terrain attributes, gamma radiometrics, soil type and land use) would also differ considerably, and thus modelling these together could be deemed negligent.

7.5. Conclusions and future directions

Mixture models were used to successfully model and map soil inorganic carbon at the 0.3-0.5 m depth in a semi-arid cotton-growing region in south-west NSW, Australia. In contrast, modelling and mapping SIC content by taking the difference of STC content and SOC content maps produced from MLMs was unsuccessful and possessed many limitations. The results from this study suggests that mixture models

should be adopted for modelling SIC content in space and time, rather than using more traditional digital soil mapping techniques, particularly when datasets are highly-skewed and zero-inflated. This approach could also be utilised to model other variables that are similarly distributed, such as soil contaminants. The study found that there was generally no statistically significant change in SIC content from 2002 to 2015 in the 0.3-0.5 m depth, although increases and decrease in SIC content were observed at some paired-sites sampled in both surveys in the upper 0.5 m of the soil profile. Land use and management did not seem to have a clear impact on these shifts in SIC content, and it appeared that the increased water percolation by heavy rainfall experienced during 2010-2012 was a more likely reason for these changes. The high spatial variability of SIC content was identified as an important limitation in the detection of statistically significant changes. Future research should focus on monitoring spatio-temporal aspects of SIC deeper in the subsoil in the study area, as it is known that there are larger accumulations at depth. In addition to this, multiple sampling depths should be analysed in conjunction to determine whether there has been a migration and redistribution of SIC down the soil profile.

References

- Amundson RG, Smith VS (1988) Effects of irrigation on the chemical properties of a soil in the western San Joaquin valley, California. *Arid Soil Research Rehabilitation* **2**, 1-17.
- Bishop, TFA, McBratney, AB, Laslett, GM (1999) Modelling soil attribute depth functions with equal-area quadratic smoothing splines. *Geoderma* **91**, 27-45.
- Bishop T.F.A, Daniel R., D.I. Guest, M.A. Nelson, C. Chang (2012) A digital soil map of *Phytophthora cinnamomi* in the Gondwana Rainforests of eastern Australia. In: B. Minasny, B.P. Malone, A.B. McBratney (Eds.), *Digital Soil Assessments and Beyond* (pp. 239-244). London: CRC Press.
- Breiman, L (1996) Out-Of-Bag Estimation. The University of California
- Cay, ER, Cattle, SR (2005) The effects of landforming on soil profile characteristics of an irrigated cotton-producing area of southeastern Australia. *Soil and Tillage Research* **84**, 76-86.
- Entry, JA, Sojka, RE, Shewmaker, GE (2004) Irrigation Increases Inorganic Carbon in Agricultural Soils. *Environmental Management* **33**, 309-317.
- Eswaran, H., Reich, P.F., Kimble, J.M., Beinroth, F.H., Padmanabhan, E., Moncharoen, P., 2000. Global carbon stocks. In: Lal, R., Kimble, J.M., Eswaran, H., Stewart, B.A. (Eds.), *Global Climate Change and Pedogenic Carbonates*. , pp. 15–25.
- Grinand, C, Barthès, BG, Brunet, D, Kouakoua, E, Arrouays, D, Jolivet, C, Caria, G, Bernoux, M (2012) Prediction of soil organic and inorganic carbon contents at a national scale (France) using mid-infrared reflectance spectroscopy (MIRS). *European Journal of Soil Science* **63**, 141-151.
- Guo, Y, Amundson, R, Gong, P, Yu, Q (2006) Quantity and spatial variability of soil carbon in the conterminous United States. *Soil Science Society of America Journal* **70**, 590-600.
- Johnson, LE, Bishop, TFA, Birch, GF (2017) Modelling drivers and distribution of lead and zinc concentrations in soils of an urban catchment (Sydney estuary, Australia). *Science of the Total Environment* **598**, 168-178.

- Khokhlova, OS, Arlashina, EA, Kovalevskaya, IS (1997) The effect of irrigation on the carbonate status of Chernozems of Central Precaucasus (Russia). *Soil Technology* **11**, 171-184.
- Knowles, T, Singh, B (2003) Carbon storage in cotton soils of northern New South Wales. *Soil Research* **41**, 889-903.
- Lal, R (2004) Soil Carbon Sequestration Impacts on Global Climate Change and Food Security. *Science* **304**, 1623-1627.
- Magaritz M, Amiel A (1981) Influence of intensive cultivation and irrigation on soil properties in the Jordan Valley, Israel: recrystallization of carbonate materials. *Soil Science Society of America Journal* **45**, 1201-1205.
- Marchant, BP, Villanneau, EJ, Arrouays, D, Saby, NPA, Rawlins, BG (2015) Quantifying and mapping topsoil inorganic carbon concentrations and stocks: approaches tested in France. *Soil Use and Management* **31**, 29-38.
- Mikhailova, E, Post, C, Cihacek, L, Ulmer, M (2013) Soil Inorganic Carbon Sequestration as a Result of Cultivation in the Mollisols. In 'Carbon Sequestration and Its Role in the Global Carbon Cycle.' pp. 129-133. (American Geophysical Union:
- Miklos, M, Short, MG, McBratney, AB, Minasny, B (2010) Mapping and comparing the distribution of soil carbon under cropping and grazing management practices in Narrabri, north-west New South Wales. *Soil Research* **48**, 248-257.
- R Development Core Team (2017) R: A language and environment for statistical computing, R version 3.2.1. R Foundation for Statistical Computing, Vienna, Austria.
- Sanderman, J (2012) Can management induced changes in the carbonate system drive soil carbon sequestration? A review with particular focus on Australia. *Agriculture, Ecosystems & Environment* **155**, 70-77.
- Sreenivas, K, Dadhwal, VK, Kumar, S, Harsha, GS, Mitran, T, Sujatha, G, Suresh, GJR, Fyze, MA, Ravisankar, T (2016) Digital mapping of soil organic and inorganic carbon status in India. *Geoderma* **269**, 160-173.
- Stoof, CR, De Kort, A, Bishop, TFA, Moore, D, Wesseling, JG, Ritsema, CJ (2011) How Rock Fragments and Moisture Affect Soil Temperatures during Fire. *Soil Science Society of America Journal* **75**, 1133-1143.
- Suarez DL (1998) Impact of agriculture on soil inorganic carbon. *Agronomy Abstract*, Madison, WI, 258-259.
- Van Ogtrop, F, Vervoort, R, Heller, G, Stasinopoulos, D, Rigby, R (2011) Long-range forecasting of intermittent streamflow. *Hydrology and Earth System Sciences* **15**, 3343-3354.
- Wu, H, Guo, Z, Gao, Q, Peng, C (2009) Distribution of soil inorganic carbon storage and its changes due to agricultural land use activity in China. *Agriculture, Ecosystems & Environment* **129**, 413-421.
- Wu, L, Wood, Y, Jiang, P, Li, L, Pan, G, Lu, J, Chang, AC, Enloe, HA (2008) Carbon Sequestration and Dynamics of Two Irrigated Agricultural Soils in California. *Soil Science Society of America Journal* **72**, 808-814.
- Yang, Y, Fang, J, Ji, C, Ma, W, Su, S, Tang, Z (2010) Soil inorganic carbon stock in the Tibetan alpine grasslands. *Global Biogeochemical Cycles* **24**, GB4022.

Chapter 8

8. Concluding remarks

8.1. Overview

In this thesis, the status and changes of various important soil properties – pH, electrical conductivity (EC), exchangeable sodium percentage (ESP), soil inorganic carbon (SIC) content, and soil organic carbon (SOC) content – were monitored, modelled, and mapped in the semi-arid irrigated cotton-growing district of Hillston in the lower Lachlan River valley catchment in south-west, NSW, Australia. The primary aim of this body of work was to assess the impact that different land uses and recent fluctuations in rainfall patterns have had on soil condition to 1.5 m depth from 2002 to 2015. Another aim of this work was to utilise statistical modelling approaches that are advantageous for soil monitoring purposes, and that allow greater sensitivity in predicting a statistically significant change in soil properties over time. Visible near infrared (VisNIR) spectroscopy was used in combination with available soil data to predict difficult-to-measure soil properties to fill in the gaps of the comprehensive soil datasets. Rather than discussing the conclusions of each research chapter in isolation, the original aims identified in Chapter 1 will be addressed:

1. *To explore the status and change of multiple soil properties over a 13-year period in a semi-arid irrigated cotton-growing region using legacy and newly-collected soil survey data*
 - a. *What is the status of soil condition at Hillston?*

There are a number of soil constraints to agricultural production in the semi-arid region of Hillston, although most of these are inherent characteristics of the study region. Sodicity could be considered the biggest soil constraint, with most of the upper soil profile found to be marginally-sodic to sodic (0-0.5 m), and the majority of the lower soil profile (0.5-1.5 m) considered to be highly sodic. The very high ESP values found in the subsoil would likely constrain the rooting depth of crops, and soil structural integrity could be compromised under the application of clean irrigation water, and during rainfall events. The next major constraint is the high degree of alkalinity found in soil across much of the study area. The subsoil was found to possess particularly high pH levels, which could significantly impact the availability of certain nutrients for crops. Soil salinity is generally perceived to be one of the major soil constraints in semi-arid regions, but this was found not to be the case in the Hillston district. Despite this, higher salinity in the deeper subsoil found at some locations could have a potentially negative impact on crop growth. Soil organic carbon contents of the Hillston district were found to be very low in both the topsoil (0-0.1 m) and subsoil (0.3-0.5 m), although this is common for semi-arid and arid regions. Semi-arid areas are often characterised by possessing appreciable soil inorganic carbon, and about half of the soils in the study area contained some amount of SIC in the 0.3-0.5 m sampling depth. While it was not

quantified, SIC was observed to be more widespread and abundant in the deeper layers (0.5-1.5 m) of the subsoil. The Vertosols (Grey, Brown and Red) of the alluvial floodplains typically possessed higher alkalinity, sodicity, salinity and a greater amount of carbonates than the non-Vertosols in the study region.

b. Has soil condition changed at the various sampling depths to 1.5 m since the original 2002 survey?

After the analysis and spatio-temporal modelling of the soil property data from the comprehensive 2002 and 2015 surveys, it was apparent that several aspects of the soil had changed in the cotton-growing district of Hillston during the 13-year period, although soil at many specific locations went unchanged. Most soil monitoring studies solely focus on the topsoil due to the difficulty of sampling soil at depth, and there is a general perception that the subsoil is more resistant to change. Despite this, some of the most substantial changes in soil properties during the study period occurred in the subsoil. While different parts of the study area often experienced contrasting, or various degrees of change, there were a few broad temporal trends. An overall acidification trend through time was observed throughout the profile in many soils at Hillston, with this trend being more widespread and significant in the deeper layers of the subsoil. The non-Vertosols generally underwent a greater amount of acidification compared to the Vertosols, likely due to an intrinsically lower pH buffering capacity. Trends of both decreasing and increasing soil salinity throughout the profile were observed in parts of the study area. The largest and most widespread changes in salinity occurred in the deeper subsoil layers, as well as at the surface soil layers. A trend of increasing soil sodicity through time was observed in both the topsoil and subsoil in several isolated sections of the study region. No considerable change in subsoil (0.3-0.5 m) SOC content was observed, but topsoil (0-0.1 m) maps of SOC content revealed that some areas had experienced a statistically significant increase. Only the 0.3-0.5 m sampling depth was thoroughly analysed for SIC, and there seemed to be no clear trend in temporal change of SIC content over the study period, but there was some evidence of both increases and decreases at certain locations.

c. How have different land uses and land management impacted the change in soil properties?

It was apparent that land use change and agricultural management practices were of considerable importance in driving changes in the various soil properties during the study period, although the relative importance of these was not always clear. Irrigated cotton production is considered to be an

intensive land use, and while this land use was observed to have had an impact on changing soil condition at Hillston, it was apparent that this was not always negative. The driving factors of the observed acidification trend were not distinct, although this was determined to at least partly be due to agricultural management practices. It is likely that the use of acidifying fertilisers commonly applied in areas under cotton production could have contributed to this decline in soil pH, as well as the displacement of basic cations through leaching or removal in the harvesting of plant biomass.

The observed shifts in soil salinity were perhaps the clearest influence of land management on changes in soil condition. Most areas under the natural land use underwent insignificant changes in salinity over time at the various sampling depths. Overall, a trend of a decrease in soil salinity throughout the profile was observed in areas under irrigated cotton production, with this attributed to the leaching of salts out of the profile through continuous irrigation, and the use of higher quality water. In contrast, several irrigated horticultural farms experienced an increase in soil EC at various sampling depths, with this linked to the continual use of salt-containing fertilisers, as well as the addition of salt through irrigation water. These salinity fluctuations in soils under irrigated agriculture is logical, as considerably larger amounts of water percolates through the soil in these land uses compared to non-irrigated land uses (Table 3.1). Although the quality of irrigation water was not quantified, anecdotal evidence from farmers in Hillston suggests that the quality of water used for irrigation in horticultural production is lower than that used for cotton. Overall, the results indicated that irrigation water quality and quantity applied during the study period had a distinct impact on the trajectory and direction of changes in soil salinity. The impact of land use on changes in soil ESP was less clear than that of EC, although most of the statistically significant increases in ESP occurred in areas under irrigated cotton and irrigated horticultural production. This suggests that these irrigated land uses are encouraging sodium ions to displace ions such as calcium and magnesium at exchange sites on clay particles, which is likely driven by the continual addition of sodium to the soil through irrigation or fertilisation.

Some studies have highlighted the potential for irrigated agriculture to encourage SOC accumulation in semi-arid regions, which is believed to be driven by the increased amounts of plant biomass that enters the soil compared to natural levels. Despite this, irrigated cotton production did not increase SOC content from natural levels, and small, isolated areas under irrigated cotton production were predicted to have undergone a statistically significant decrease in SOC content from 2002 to 2015. This trend is likely due to the required continuous cultivation of soil in areas under irrigated cotton production, which would counteract the positive impacts of increased plant biomass entering the soil.

While it is generally believed that irrigation decreases carbonate levels in soil, land use and management did not seem to have a clear impact on these observed shifts in SIC content, and these changes appeared to be more climatically-driven.

d. How has fluctuating rainfall patterns influenced the change in soil properties?

It is well known that land management can influence soil change over relatively short periods, but there is much more uncertainty as to how an increasingly variable climate could influence this. Nevertheless, as evidenced from co-located natural sites, the fluctuating patterns of rainfall received at Hillston during the study period appeared to have a considerable effect on several soil properties. It is important to highlight that the rainfall patterns experienced during the study period at Hillston, were not slightly divergent from the average; they were the most contrasting that Hillston had experienced for more than a century. Since the rainfall records began at Hillston in 1881, the driest period ever recorded was the 2002 to 2009 drought, and the wettest were the years from 2010 to 2012. Consequently, it is understandable that these extreme climatic events have had a significant impact on changing several important soil properties during the study period. With the benefit of hindsight, an additional soil sampling campaign at the end of 2009 would have been very useful for the evaluation of the effects of this prolonged drought on soil condition.

While it appeared that land management encouraged the acidification of soil to some degree, the fluctuating rainfall patterns seemed to be a greater influence on this. The deeper subsoil layers (0.5-1.5 m) in the Hillston region are characterised by large accumulations of carbonates, and upon significant wetting and leaching, there is potential for the dissolution of these carbonates, which ultimately results in a decline in soil pH. It is believed that the three very wet years experienced in Hillston from 2010 to 2012 encouraged this, rather than land management, as areas under irrigated agriculture did not appear to undergo significant acidification in these subsoil layers. This is likely because carbonates have undergone some amount of dissolution when irrigation practices were introduced in preceding years, whereas areas under non-irrigated land uses have only recently been exposed to large amounts of water percolating through the soil. Only the upper 0.5 m of the soil profile was analysed for SIC content, so this hypothesis could not be tested. While there were minimal statistically significant changes in SIC content between 0.3 and 0.5 m depth observed, there were some observations of increases and decreases of SIC content at certain locations, and there appeared to be little influence of land management on this. Additionally, natural paired-sites sampled in both 2002 and 2015 in the upper 0.5 m showed a trend of a slight decrease in SIC content in the topsoil, and a slight

increase in the subsoil, which suggests that the large amounts of rainfall received from 2010 to 2012 has encouraged the dissolution of carbonates in the topsoil, and then re-precipitation in the subsoil.

The parts of the study area that underwent statistically significant increases of SOC content in the topsoil were covered by several land uses, and these increases appeared to be more climatically-driven. This was primarily attributed to the three consecutive years of heavy rainfall received in the study area from 2010 to 2012, and it is known that rainfall and SOC content have a strong positive correlation. It was also acknowledged that the particular parts of the study area that experienced an increase had relatively low SOC contents in the 2002 survey, and because the area is starting off at a lower base SOC level, there is greater potential for this to increase when stimulating factors such as rainfall increase. An impact of fluctuating rainfall patterns on soil salinity and sodicity was not evident, as the changes in these properties seemed to be more clearly driven by land management.

e. Is any detected change in soil properties desirable or undesirable? Has soil condition improved or declined?

Overall, the temporal trends of the different soil properties observed from 2002 to 2015 in the Hillston region can be considered positive in some aspects, and negative in others. Although acidification of soil is generally perceived as an undesirable process, this is not necessarily the case in this situation, due to the high levels of inherent alkalinity of the soils in the study region. The increases in salinity under irrigated horticulture are not necessarily a concern for agriculture, as the largest changes occurred deep in the subsoil, which would have less of an impact on plant productivity. Perhaps the most concerning trend is the decreasing EC and increasing ESP at some locations under irrigated agriculture. The relative values of sodicity and salinity are crucial in determining soil structural integrity during irrigation and rainfall events, and the changes observed in this study could lead to heightened waterlogging issues and losses in plant productivity. The statistically significant increase in SOC content in parts of the study area is considered to be a positive change in soil condition. Overall, it could be inferred from these observed trends that soil condition in the Hillston district has marginally improved from an agricultural viewpoint since the 2002 survey.

2. *To use bivariate linear mixed models (BLMMs) and multivariate linear mixed models (MLMMs) to monitor, model and map soil properties in space and time using co-located soil data*

a. *Can these models be used to accurately predict different soil properties in space and time?*

The use of BLMMs and MLMMs generally resulted in accurate predictions for the various soil properties when tested with leave-one-site-out cross-validation (LOSOCV). While the accuracy of predictions is very important, it is also important that the prediction variances of the model represent the actual errors. To test this, the mean and median standardised squared prediction errors (SSPE) of the fitted models should be close to the values of 1.0 and 0.455, respectively. The various final models generally reflected these values well, and this indicates that even in situations where the models do not predict with high accuracy, we know that the estimates of the uncertainty are quite good, meaning that they can be used with an appropriate level of caution. For example, rather than relying on examining the change in the predicted values, we also used formal hypothesis testing, which incorporates the prediction uncertainty.

b. *How does the inclusion of temporal correlation improve soil maps?*

Bivariate linear mixed models (BLMMs) were used to model two response variables (one soil property, at one depth, at two time points) for soil pH, EC, ESP and SIC content. In this approach, the temporal correlation with co-located samples extracted from the same site at the two time points is used, which is particularly useful when one survey possesses less samples than the other. This approach resulted in maps that displayed a logical connection through time, as the spatial patterns of the maps are influenced by samples from the corresponding time point. A simple example of this was demonstrated in Chapter 5, where BLMMs had clear advantages of modelling topsoil EC compared to using two separate univariate linear mixed models.

c. *How does the inclusion of both temporal and vertical spatial correlation improve soil maps?*

For SOC and STC content, multivariate linear mixed models (MLMMs) were used to model four response variables (one soil property, at two depths, at two time points). In this approach the correlation through time was utilised, as described above, but vertical correlation between two different sampling depths (0-0.1, 0.3-0.5 m) was also included. In this MLMM approach, the same advantages of the BLMM approach are included, but there is also the advantage of a logical connection between the different sampling depths, which is especially advantageous when one sampling depth has more samples than the other.

Having one model for multiple depths creates soil maps that pedologically make sense, as information from depths above and below the depth of interest are utilised.

d. Are these approaches advantageous in detecting statistically significant temporal changes in soil properties?

Another advantage of using BLMMs and MLMMs to monitor the change in soil properties is that the covariance between different response variables is calculated. This covariance term is then used to calculate the contrast variance, which is subsequently used to analyse the statistical significance of any observed changes in soil properties over time. Including the covariance in these calculations ultimately leads to greater sensitivity in detecting statistically significant change, and thus provides a more realistic representation of actual changes in soil condition. The statistical significance of any changes detected by monitoring studies is rarely shown, and most studies simply show standalone change maps, but this can lead to substantial misinterpretation about the changes that have actually occurred in soil condition. In this study, statistically significant changes in several soil properties were detected. In soil monitoring studies where a test statistic has been performed, often no statistically significant change in soil properties is detected, which could be due to an insufficient amount of soil data, but is also likely because the cross-correlation is not modelled and the covariance term is not included.

3. To use visible near infrared (VisNIR) spectroscopy in combination with traditional laboratory methods to predict properties

The use of VisNIR spectra and Cubist models to predict soil properties (ESP, SOC content and SIC content) was an overall success. Not only did this show promise for using VisNIR to analyse difficult-to-measure soil properties in the future, but it was also a rapid and cost-effective method of filling the gaps in the comprehensive 2002 and 2015 soil survey datasets where time and cost was a limiting factor.

a. How accurately can soil exchangeable sodium percentage be predicted using these spectroscopic techniques?

Despite other studies publishing typically poor predictions of soil ESP with spectroscopic techniques, ESP could be predicted with high accuracy (LCCC 0.79 independently-validated) in this study. The high quality predictions of ESP were attributed to several factors, such as the large and comprehensive dataset of the study area, the resampling of the same sites, and the relationship with different levels of ESP and the colour of different soil types in the study region.

- b. How does the inclusion of readily-available, ancillary, laboratory-measured soil data such as soil total carbon content and pH improve the prediction of soil organic and inorganic carbon content?*

When carbonates are present in soil, the determination of SOC content by traditional laboratory approaches such as the combustion method is much more challenging. Likewise, SIC content itself is laborious to analyse by traditional laboratory methods. For these reasons, spectroscopic techniques were used to predict both SOC and SIC content of a portion of samples within the 2002 and 2015 datasets. The results showed that both SOC and SIC content could be accurately predicted, although SOC content was predicted to a higher accuracy than SIC content. Soil pH and soil total carbon (STC) content data were also included as predictor variables with the VisNIR spectra, as this was available for all samples, and it is known that these properties can be highly correlated with both SOC and SIC content. There were significant advantages of including this ancillary soil data with VisNIR spectra. Predictions with VisNIR spectra only, and VisNIR, pH and STC content on the validation datasets improved from 0.81 to 0.94 LCCC for SOC content, and 0.35 LCCC to 0.83 LCCC for SIC content, respectively.

- 4. To determine/develop a spatio-temporal modelling/mapping technique that can sufficiently deal with the zero-inflated and highly-skewed datasets of soil inorganic carbon (SIC) content*
- a. Can the 'mixture model' approach be used to do this?*

The mixture model approach involved initially modelling the presence or absence of SIC in the study area with a Random Forest model, and then separately modelling SIC content with a BLMM using values from the dataset that are greater than zero only. The presence or absence of SIC in the study area could be predicted with high accuracy, but the accuracy of predicting SIC content was relatively poor, which was attributed to the high spatial variability of SIC content. When the corresponding maps from these approaches were combined, it was apparent that the final maps of SIC content produced by mixture models had a logical spatial pattern that could likely not be achieved by other modelling methods. There are no published studies that have used this described approach to model SIC content, but the use of this mixture model approach to map SIC content was deemed a success.

b. Can maps of SIC content accurately be created by taking the difference of soil total carbon and soil organic carbon content maps?

Soil total carbon (STC) and SOC content data can easily be appropriated to suit the requirements of statistical modelling approaches, and thus MLMMs were used to separately model this data. The maps of these soil properties were then used to calculate the SIC content (STC – SOC content) across the study area. It was clear from the results that this approach possessed many limitations, and unlike the mixture model approach, it was very difficult to predict soils that contain no SIC with this method. While those parts of the study area that contain no SIC would have the same SOC and STC contents, the separate modelling processes rarely predicted the same values in these areas. As a result, when the SOC content map is taken away from the STC content map, areas that truly contain no SIC were predicted to be either negative or positive values. In addition, it is apparent that modelling STC content is somewhat illogical. While STC is the sum of SOC and SIC, these two components of carbon are very different in many aspects. The processes of SOC and SIC formation are very different, and the relationship that SOC and SIC have with the spatial covariates used for modelling also differs considerably.

8.2. Future research questions

8.2.1. Monitoring temporal changes in soil

Some interesting shifts in soil condition were observed between 2002 and 2015, and these results have led to other possible avenues of research to conduct in the future for both Hillston, and other study areas. While the soil under cotton land use has been extensively studied in the more traditional cotton-growing regions in northern NSW and southern QLD, there has been little work in the growing regions of southern NSW. Nevertheless, this is the only comprehensive study that has monitored a suite of important soil properties throughout the profile in any of the cotton-growing regions of Australia. The soils of northern and southern NSW are quite different due to differences in up-catchment geology, and the results from this thesis show that there is considerable benefit of monitoring of a suite of important soil properties in these areas.

While the observed acidification trend was deemed to be generally beneficial due to the inherently high soil pH levels, any further changes in the future should be monitored, particularly in those soils with a lower pH buffering capacity, such as the non-Vertosols and Red Vertosols. This acidification was primarily attributed to the dissolution of carbonates, but this was not quantified in the deeper subsoil layers due to time constraints. The very high pH levels in the subsoil indicate that not only calcium carbonate, but also sodium carbonate is present in these soils. Sodium carbonates are

much more soluble than calcium carbonate, and it is likely the dissolution of these materials is the cause of the observed decrease in pH in the subsoil. Future work should analyse archived soil samples from the 2002 and 2015 surveys to quantify the amounts of these different carbonate materials to determine if there has been any changes over time. This should then be linked back to the observed temporal shifts in soil pH, which will ultimately provide greater insight as to the possible causes of any observed changes.

It was determined that land management, and in particular the quality and quantity of irrigation water applied during the study period had an important influence on the dynamics of soil salinity. Despite this, the available information regarding irrigation water quality and quantity applied by each irrigated farm was quite general, and it would be invaluable to collect more detailed information in any future studies to better understand the impact that contrasting irrigation management has on the status and change of soil salinity. An interesting trend was that areas under irrigated horticulture, particularly almond farms, showed increases in soil salinity over time throughout much of the soil profile. The area under irrigated almonds is expanding at Hillston, and further monitoring of soil under this land use should be performed to better understand the possible reasons for these observed increases. While most of the significant changes in salinity occurred in the subsoil, it was apparent that land management still had an impact on increasing this in the topsoil, and monitoring of salinity should be continued to ensure that salinity constraints to plant productivity do not develop in the future.

It was apparent that soil sodicity is the most limiting soil constraint at Hillston, and almost the whole study area was deemed to be sodic in the subsoil. The observed trend of increasing sodicity raises concern, as methods of amelioration for subsoil sodicity are typically costly and ineffective. The research in this thesis was carried out at a relatively broad scale, and more intensive finer-scale research on soil sodicity at Hillston would be beneficial in better understanding the drivers of this increasing sodification of soil. Cotton growers in the Hillston district are becoming more aware of the constraints of soil sodicity on crop growth, and future work should be carried out to help these growers implement management practices that minimise the impact of sodicity, as well as prevent further sodification of soil.

Soil organic carbon levels at Hillston are very low, but this suggests there is potential for significant accumulation of SOC. An interesting finding in this thesis was the statistically significant increase in SOC content that was deemed to be a consequence of the high amounts of rainfall received from 2010 to 2012. It would be interesting to analyse if other arid or semi-arid landscapes that have experienced a similar period of high rainfall have undergone any changes in SOC content over time. The

results from this thesis suggested that irrigated cotton production had a relatively negative impact on SOC content, and this is likely due to the continual cultivation of soil under this land use. In areas under irrigated cotton production, significantly higher amounts of plant biomass enter the soil, and it is hypothesised that if growers were able to adopt minimum or no-till cultivation practices, considerable amounts of SOC would be able to be accumulated under this land use. This theory should be tested in the field to determine if cultivation is in fact the reason for this lack of SOC accumulation, which could lead to an adaptation of management practices.

While the subsoil is generally perceived to be more resistant to change, the results from this thesis show that soil condition can significantly change in the subsoil over a relatively short time period of 13 years. There is a general lack of studies that monitor the subsoil condition of soil properties, and it is clear that more understanding on this is required, particularly as subsoil constraints are often the most limiting in agricultural areas. It is probable that important trends in soil change are overlooked when only the topsoil is focused upon.

There has been interest surrounding the impact that a warming climate and fluctuating rainfall patterns have on soil condition, however, the majority of this attention has been on topsoil SOC. The results from this thesis suggest that highly fluctuating rainfall patterns also have the potential to alter a suite of important soil properties, such as pH, EC, ESP, and SIC content. While this thesis focused on a semi-arid area in eastern Australia, future work should consider analysing the impacts of extremely dry and/or wet periods on soil properties in a range of other landscapes.

8.2.2. Spatio-temporal modelling

The mixture model approach used for modelling SIC was deemed successful, suggesting that this approach should be adopted by other studies where SIC content is modelled across a study region. In the future, SIC content should be modelled in the deeper subsoil layers, as there are generally much larger accumulations of carbonates and less spatial variability at these depths, which would likely lead to fewer challenges in the modelling process. In addition, this mixture model approach could be used to model other environmental variables that are typically difficult to model due to their statistical distributions, such as soil contaminants.

Overall, there were many clear benefits of using bivariate and multivariate linear mixed models to model soil properties in space and time. These approaches are rarely considered by researchers to utilise the correlation between response variables to monitor soil properties, with most opting for

univariate approaches. While soil information at two time points was used in this thesis, future studies could consider incorporating more time points into these models, as this would likely lead to improved predictions. Likewise, numerous sampling depths could be simultaneously modelled with this approach, which would consequently result in soil maps that have a stronger relationship throughout the profile. The inclusion of vertical spatial correlation resulted in highly accurate predictions of SOC content from the MLMM, however, SOC accumulates at the surface, and there is likely more benefit to gain from this approach when modelling soil properties that are more uniform throughout the soil profile, such as pH or texture. The values at different depths of these more uniform properties likely have a stronger relationship with each other, and MLMMs could take advantage of this.

While there is potential in the future to increase the size and complexity of MLMMs by including more response variables to take advantage of the correlation between different co-located soil information, there are some significant limitations to this approach. The most notable is the amount of time and computing power required to create these models, which is a serious challenge in their widespread adoption. There should be increased focus on streamlining this approach in the future, which may increase their adoption to use as a tool to model and monitor soil properties in space and time.

The soil maps were created on a 100 m grid and predicted on point support, but there may be advantages of predicting on a block support. Predicting on a block support results in smaller prediction variances and more statistical power for hypothesis testing, which would be beneficial for monitoring studies. Very few studies implement this, but given these advantages, it would seem wise to consider the correct prediction support of maps of soil change, especially as predictions on block supports are probably more useful for guiding management decisions. For example, a grower may be more interested in the average change in their fields, rather than a point on a 100 m grid.

In the soil modelling/mapping approaches, the laboratory-measured and the spectroscopic-predicted values of soil properties were treated as equal. It would be logical to give greater weighting to the laboratory measured data as it is a more accurate representation of actual soil properties. The different uncertainty in the measurement sources could be treated more formally in the spatial modelling, and should be considered in future work.

8.2.3. VisNIR predictions of soil properties

The advantages of including soil pH and STC data as predictor variables in conjunction with VisNIR spectra in Cubist models to predict SOC and SIC content were clear. While STC is often unavailable, soil pH data is usually easily accessible, as it is rapidly and inexpensively measured by traditional laboratory methods. Soil pH is also often correlated with other soil attributes, which suggests that this could be incorporated with spectra to predict other soil properties in the future. This also opens the discussion as to the possible benefits of combining other easily-measured and readily available soil data with VisNIR spectra to predict different soil properties, as there are many soil properties that are highly correlated with each other. For example, soil electrical conductivity (EC) is easily measured by traditional laboratory methods and hence this data is often available. It is known that EC is well correlated with other soil properties that are typically laborious to measure, such as particle size, and cation exchange capacity. Future research should focus on including other readily-available ancillary soil data with spectra to improve the quality of soil spectroscopic predictions.

8.3. Closing statements

To conclude, soil data from traditional laboratory and spectroscopic methods were used with bivariate and multivariate linear mixed models to predict the change in multiple soil properties at various depth increments to 1.5 m in the semi-arid district of Hillston in south-west NSW. While large parts of the study area did not undergo changes from the 2002 to the 2015 soil survey, a temporal change was detected for several of the soil properties monitored. Overall, there was a general soil acidification trend, both increases and decreases in salinity, increases in sodicity, and increases in SOC content, and it was clear that both land management and fluctuating rainfall patterns influenced these changes. Future soil monitoring studies should consider the approaches used and developed in this thesis.

Appendix

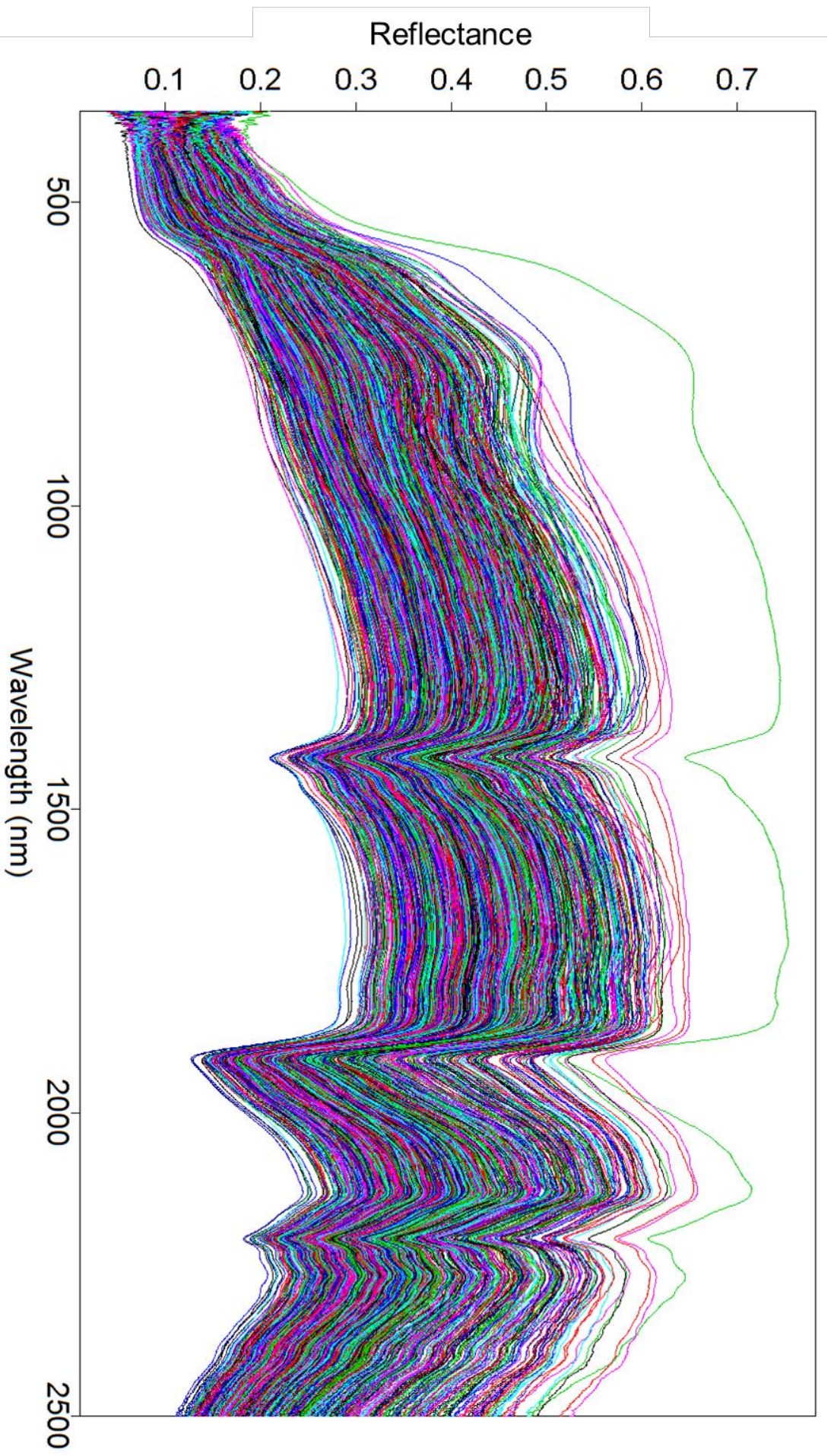


Fig. A1 – Visible near infrared (VisNIR) spectra of all archived soil samples ($n = 1291$) from the 2002 ($n = 385$) and 2015 ($n = 906$) soil surveys