Relationships in soil distribution from digital soil modelling and mapping over eastern Australia under past, present and future conditions

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

Jonathan M Gray
BSc (Hons), MPhil (Physical Geography)

September 2016

Faculty of Agriculture and Environment
The University of Sydney
New South Wales
Australia
Statement of Originality

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing it has been acknowledged. No part of the thesis has been submitted towards any other degree, however Chapters 2 to 6 have been submitted to journals.

J. Gray, 30 September 2016

Dedication

I dedicate this thesis to my late father, Ronald Lawson Gray, who died too young. He helped and encouraged me to pursue my dreams in science.

Acknowledgements

Firstly I wish to thank my primary supervisor, Associate Professor Tom Bishop, for his ongoing support, guidance and friendship throughout the course of my research, which has extended over 6 years (part time). I will miss our stimulating meetings to discuss and further develop my research. My initial associate supervisor, Dr Peter Smith (formerly of NSW Office of Environment and Heritage, OEH) is thanked for his assistance, particularly in the earlier years.

I wish to thank many other staff and fellow students of the Faculty of Agriculture and Environment for their general assistance, support and friendship over the years. I particularly appreciated the support from Senani Karunaratne and Darren Kidd (now both Drs) and also Drs Margaret Barbour and Andrew Merchant during annual progress reviews.

I also thank several of my colleagues at OEH for their technical help and encouragement over the years, particularly Drs Mark Littleboy, Xihua Yang and Hugh Jones. More generally, I acknowledge the assistance of my employer OEH in providing a flexible working arrangement allowing me to undertake my studies. I thank the three examiners for their constructive review comments.

Finally, I wish to thank my entire family for their support and encouragement throughout my studies. My loving wife Melissa has in particular been a wonderful support, and I could not have completed this without her behind me.
Summary

This research project applied digital soil modelling and mapping (DSMM) techniques to elucidate relationships between key soil properties and the main soil-forming factors. It attempted to address several broad research issues relating to quantifying the factors that control soil distribution and identifying how these combine together to control soil distribution and their change due to alteration in land use and climate over New South Wales and eastern Australia.

These broad issues were examined through a number of more specific research issues that were progressively addressed over five chapters, each intended as publishable journal papers. These chapters/journal papers relate to (i) the influence of lithology in soil formation and its application in DSMM (ii) relationships of soil-forming factors to key soil properties and their use in digital soil mapping; (iii) factors controlling the distribution of soil organic carbon stocks (SOC), spatially and with depth; (iv) change in SOC stocks following historic clearing of native vegetation, and (v) change in SOC stocks with projected climate change.

The strong influence of lithology in controlling soil distribution was demonstrated. Following its classification into 12 classes based on mineral and chemical composition, it was shown to have the highest influence of all soil-forming factors for six key soil properties (SOC, pH, cation exchange capacity (CEC), sum-of-bases, total phosphorous and clay content) examined over NSW. Lithology had similar influence at the scale of eastern Australia; however climate variables were of equivalent or slightly stronger influence for SOC and pH. It was shown to have two to five times more influence than the next highest ranked geophysical covariate such as gamma radiometrics in the models. A marked improvement in the statistical quality of digital models and maps was demonstrated when lithology was applied together with other geophysical covariates.

Quantitative relationships that are readily interpreted were developed with eight key properties (those listed above plus sand and silt contents) over eastern Australia. These relationships at least partially solve Jenny’s fundamental soil equation in a manner that is more universally applicable and readily interpreted than appears to have been reported previously. Using these relationships, the quantitative influence of the different factors on each soil property is determined, including the unit change per unit
Summary

variation in the factor, for example a decrease of 0.11 pH units for each 100 mm
increase in annual rainfall for the 0-10 cm interval (other factors remaining constant).
These relationships were applied together with readily available covariate grids to
prepare digital soil maps (DSMs) with 100-m resolution for the eight soil properties
over NSW. The predictive ability demonstrated by the maps was broadly moderate,
with Lin’s concordance generally between 0.4 and 0.7. They compared well with maps
prepared using more sophisticated modelling methods and covariate data. They have the
ability to be readily prepared and interpreted and thus have the potential to serve as a
useful introduction to the more sophisticated DSMM approaches.

Systematic patterns of SOC stock levels were graphically demonstrated over 45
different climate-parent material-vegetation cover regimes for upper soils (0-30 cm) and
lower soils (30-100 cm) over eastern Australia. There are generally uniform trends of
increasing SOC stocks with increasingly moist climate, increasing mafic character of
parent material and increasing vegetation cover. Average SOC stocks in the 0-30 cm
depth interval range from 16.3 Mg ha\(^{-1}\) (t/ha) in dry, highly siliceous parent material and
low vegetation cover environments, up to over 145.0 Mg ha\(^{-1}\) in wet, mafic parent
material and high vegetation cover environments. It was demonstrated that the
proportion of SOC stored in the subsoil (30-100 cm) relative to the top 100 cm varies
systematically from an average of 43% in moist climates to an average of 54% in dry
climates.

Digital soil maps of pre-clearing (pre-European) SOC stocks (100-m resolution)
were prepared over NSW. These maps may be used to provide baseline soil carbon
levels for carbon turnover models and carbon accounting and trading schemes. They
were demonstrated to outperform the existing equivalent maps produced by
conventional soil survey methods, with independent validation RMSE values being 33%
lower. Comparison of these maps with current SOC stock maps allowed an examination
of the change in SOC over NSW following native vegetation clearing. A total SOC loss
of approximately 0.53 Gt (530 million Mg or tonnes), or 12.6% over the entire State
was revealed. It was demonstrated that the change in SOC stocks following clearing
increases (in both absolute and relative terms) with increasingly cool (moist) climate,
more mafic parent material and more intensive land use. In the 56 different climate-
parent material – land use regimes, the loss varied from less than 1 Mg ha\(^{-1}\) (or 4%) in
warmer climates over highly siliceous parent materials under grazing land uses to 44.3
Mg ha\(^{-1}\) (or 50.0\%) in cooler (moist) conditions over mafic parent materials under intensive cropping land use.

Digital soil mapping techniques involving Cubist piecewise linear decision trees, in combination with a space-for-time substitution process (DSM-SFTS), were demonstrated to be effective in mapping the potential change in SOC stocks due to projected climate change over NSW until approximately 2070. Considerable variation in both direction and magnitude of change was demonstrated with application of the 12 different climate change models with their differing climate trajectories. For the mean state-wide change there were some climate models that predicted an increase but others that predicted a decrease over the two depth intervals studied (0-30 and 30-100 cm). Greater consistency between climate change models is required. The predicted SOC changes are primarily controlled by the balance between changing temperatures and rainfall. However, the extent of change is also shown to be dependent on the precise environmental regime, with systematically differing changes demonstrated over 36 current climate-parent material-land use combinations. For example, the projected mean decline of SOC is less than 1 Mg ha\(^{-1}\) for dry-highly siliceous-cropping regimes but over 15 Mg ha\(^{-1}\) for wet-mafic-native vegetation regimes.

The study has provided quantitative data on the influence of the main soil-forming factors. The necessity of considering the combined influence of multiple soil-forming factors to make meaningful quantitative estimates of current and potential future soil properties is demonstrated. Clear patterns of soil property distribution and change under changing land use and climate conditions are identified, particularly for the vital soil property of SOC. The presentation of relationships that are readily interpreted can assist in their application in natural resource planning and management activities and also in other environment modelling programs. They may thus potentially help to address a range of environmental challenges facing eastern Australia and beyond.
Contents

Summary ........................................................................................................................................... ii
Contents .............................................................................................................................................. v
Published peer reviewed publications from Thesis ................................................................. xii

Chapter 1 : Introduction...................................................................................................................... 1
  1.1 General introduction .................................................................................................................. 1
    1.1.1...Soil-environment relationships through space and time .............................................. 1
    1.1.2...SOC distribution through space and time .................................................................... 3
    1.1.3...Dissemination of soils knowledge .............................................................................. 4
    1.1.4...Broad research questions ......................................................................................... 4
  1.2 Specific research issues ........................................................................................................... 5
    1.2.1...Influence of lithology in soil formation and its application in DSMM ....................... 5
    1.2.2...Readily interpreted soil–environment relationships and their potential use in digital soil mapping ........................................................................................................... 6
    1.2.3...The factors controlling the distribution of soil organic carbon stocks, spatially and with depth ........................................................................................................ 7
    1.2.4...Change in SOC stocks with the clearing of native vegetation .................................. 7
    1.2.5...Change in SOC stocks with projected climate change .............................................. 8
  1.3 Organisation of thesis ............................................................................................................. 9
  1.4 References............................................................................................................................... 10

Chapter 2 : Lithology and soil relationships for soil modelling and mapping ...................... 14
  Abstract ........................................................................................................................................ 14
  2.1 Introduction ............................................................................................................................ 15
    2.1.1...Sources of parent material data for soil modelling and mapping ............................... 16
    2.1.2...Aims ........................................................................................................................... 17
  2.2 Classification of parent material for pedological purposes ..................................................... 18
  2.3 Methods .................................................................................................................................. 23
    2.3.1...Overview ..................................................................................................................... 23
    2.3.2...Overview of NSW study area .................................................................................... 24
    2.3.3...The soil dataset .......................................................................................................... 25
    2.3.4...Parent material covariates ....................................................................................... 25
    2.3.5...Other covariates ......................................................................................................... 27
    2.3.6...Model and map development and statistical analysis ............................................... 28
  2.4 Results .................................................................................................................................... 29
    2.4.1...Quantitative influence of lithology in models ............................................................ 29
    2.4.2...Relative influence of different parent material covariates in models ....................... 30
    2.4.3...Model performance with different parent material covariates ................................ 32
    2.4.4...Map validation with different parent material covariates ........................................ 34
  2.5 Discussion .............................................................................................................................. 36
Contents

2.5.1. Relationships between lithology and soil distribution..........................36
2.5.2. Use of lithology and geophysical covariate in DSMM .......................38
2.5.3. Suggestion for use of lithology in DSMM .....................................41

2.6. Conclusion .......................................................................................42
2.7. References .....................................................................................43

Chapter 3 : Pragmatic models for the prediction and digital mapping of soil properties in eastern Australia ............................................51

Abstract ..............................................................................................51

3.1. Introduction ......................................................................................52

3.1.1. Development of quantitative soil-environment models......................52
3.1.2. Use of models for digital soil mapping ...........................................53
3.1.3. Aims ............................................................................................54

3.2. Methods .........................................................................................55

3.2.1. Overview of study area ................................................................55
3.2.2. Soil profile dataset .......................................................................56
3.2.3. Soil properties .............................................................................56
3.2.4. Depth intervals ...........................................................................58
3.2.5. Covariates ..................................................................................58
3.2.6. Statistical analysis and validation of models ..................................63
3.2.7. Preparation of digital soil maps ....................................................64

3.3. Results and validation ......................................................................65

3.3.1. Multiple linear regression model development ...............................65
3.3.2. Influence of covariates .................................................................65
3.3.3. Validation of models ....................................................................69
3.3.4. Comparison with Cubist modelling approach and remotely sensed covariates .............................................................71
3.3.5. Digital soil map preparation ..........................................................72

3.4. Discussion .......................................................................................76

3.4.1. Predictive performance of models and maps ..................................76
3.4.2. Comparison with other modelling approaches ...............................78
3.4.3. Pedologic insights - influence of factors .......................................79
3.4.4. Future climate and land use change .............................................81
3.4.5. Sources of uncertainty .................................................................81

3.5. Conclusion ......................................................................................83
3.6. References .....................................................................................84

Chapter 4 : Factors controlling soil organic carbon stocks with depth over eastern Australia ........................................................................92

Abstract ..............................................................................................92

4.1. Introduction ......................................................................................93
4.2. Methods ..........................................................................................95

4.2.1. The soil dataset ............................................................................95
4.2.2. The covariates ............................................................................96
Chapter 6: Change in soil organic carbon stocks under twelve climate-change projections over New South Wales, Australia

Abstract 155

6.1 Introduction ........................................................................................................... 156

6.2 Assessment methodology .................................................................................... 158

6.2.1 Soil profile dataset ........................................................................................... 159
Chapter 7: Summary and discussion

7.1 Specific research issues

7.1.1 Influence of lithology in soil formation and its use in digital soil mapping

7.1.2 Relationship of soil-forming factors to key soil properties and their use in digital soil mapping

7.1.3 Factors controlling the distribution of soil organic carbon stocks, spatially and with depth

7.1.4 Change in soil organic carbon stocks following clearing of native vegetation

7.1.5 Change in soil organic carbon stocks with projected climate change

7.2 Broad research issues

7.2.1 Can we better elucidate the influence of different factors in controlling the distribution of key soil properties?

7.2.2 How do these factors combine to control the distribution of soil properties?

7.2.3 How does SOC respond to changes in the environment such as altered land use and global climate change? Can readily interpreted relationships and patterns in change be identified?

7.2.4 How effective are pragmatic soil relationships and data products, as derived in the study, in disseminating soil knowledge?

7.3 Future research directions

7.4 Conclusion

7.5 References

Appendix 1: Background to development of soil modelling and digital soil mapping

A1.1 The State Factor model

A1.2 Development of digital soil mapping

A1.3 Use of geophysical covariates to represent soil and parent material in DSM

A1.4 References
Appendix 2: Behaviour of pH and sum-of-bases under projected climate change over NSW

A2.1 The soil properties.................................................................................................................. 225
A2.2 Results........................................................................................................................................ 226
A2.3 Discussion.................................................................................................................................. 234
A2.4 References................................................................................................................................ 235

Appendix 3: Digital soil maps of bulk density for NSW .................................................................. 238

Error! Bookmark not defined.

List of tables
Table 2.1: Broad classification of parent material for pedologic purposes .................................. 20
Table 2.2. Chemical composition of common rock types ................................................................. 21
Table 2.3. Lithology classes and typically associated soils (generalized first approximation only) .................................................................................................................................. 23
Table 2.4. MLR model validation and change in soil properties with 10% and 50% change in silica index ........................................................................................................................................ 30
Table 2.5. Model strength and validation results for key soil properties over New South Wales (5-15 cm) ......................................................................................................................... 33
Table 2.6. Map validation results for key soil properties over NSW (5-15 cm interval, Cubist approach)......................................................................................................................... 35
Table 3.1. Soil properties: laboratory methods and sample numbers .............................................. 57
Table 3.2. Parent material classes and Silica index .......................................................................... 60
Table 3.3. Rules used to define the topo-slope index ....................................................................... 61
Table 3.4. The aspect index ................................................................................................................ 62
Table 3.5. The land disturbance index ............................................................................................... 62
Table 3.6. Multiple regression relationships and associated statistics ............................................. 66
Table 3.7. Standardized regression coefficients of covariates in regression relationships .......... 67
Table 3.8. Influence of covariates on key soil properties ................................................................. 68
Table 3.9. Model validation ............................................................................................................... 69
Table 3.10. Comparison of model approaches (0-10 cm) ................................................................. 71
Table 3.11. Validation of NSW digital soil maps .............................................................................. 74
Table 3.12. Validation of Hunter region digital soil maps (0-10 cm depth) ...................................... 76
Table 4.1. Parent material classes and typically associated soils ...................................................... 98
Table 4.2. MLR models for SOC density (kg m\(^{-3}\)) ...................................................................... 101
Table 4.3. Validation of MLR and Cubist models ............................................................................ 101
Table 4.4. MLR standardised regression coefficients and relative rankings of covariates .......... 103
Table 4.5. Influence and relative ranking of covariates in Cubist models ........................................ 103
Table 4.6. Relative change in SOC density per unit change in covariates ........................................ 104
Table 4.7. Density and stocks of SOC in upper and lower depth intervals to 100 cm ............. 107
Table 5.1. Typical soil types associated with different parent material classes.................128
Table 5.2. Pragmatic MLR models for pre-clearing SOC% over eastern Australia ..........133
Table 5.3. SOC (log%) model validation ........................................................................134
Table 5.4. Standardised regression coefficients of covariates in SOC% regression models.................................................................134
Table 5.5. Frequency of covariate use in the Cubist models...........................................134
Table 5.6. Validation of NSW pre-clearing SOC maps, Banks and McKane map and present day SOC mass map .................................................................135
Table 5.7. Soil organic carbon stocks over NSW and ACT (0-30 cm) – pre-clearing and current levels.................................................................138
Table 5.8. Change since clearing in SOC density over top 30 cm by climate (temperature regime), parent material and land use ..................................................139
Table 6.1. Parent material classes and typically associated soils......................................165
Table 6.2. Validation statistics of soil organic carbon Cubist models ................................166
Table 6.3. Change in SOC stocks and CO₂ equivalent over NSW over 2nd change period (to approximately 2070) .................................................................169

List of figures
Figure 2.1. Lithology class map for NSW ........................................................................26
Figure 2.2. Random Forest variable importance plots (based on the increase in mean square error (MSE %) with virtual omission of the variable) ..................31
Figure 2.3. Change in Lin’s concordance of Cubist models with different parent material covariates........................................................................................................34
Figure 3.1. Location of profile points (shading represents effective modelling area)........57
Figure 3.2. The topo-slope index ..................................................................................61
Figure 3.3. Observed versus predicted values for OC, pHca, sum-of-bases and sand (0-10 cm depth) ..........................................................70
Figure 3.4. Digital soil maps for selected properties (0-10 cm) over NSW ................73
Figure 3.5. Digital soil map of soil pHca over the Hunter region, with underlying covariate data..................................................................................................................75
Figure 4.1. SOC modelling points over eastern Australia (shading represents reliable modelling area) .................................................................96
Figure 4.2. Validation plot for SOC density (5-15 cm, from Cubist model) ..............102
Figure 4.3. SOC stocks over eastern Australia (0-30 cm and 30-100 cm) (Mg ha⁻¹) ....104
Figure 4.4. Variation in SOC stock by climate, parent material and vegetation cover sub-classes (0-30 cm, Mg ha⁻¹, showing mean as dark line and 95% spread of predictions) ...................105
Figure 4.5. Variation in SOC stock by climate, parent material and vegetation cover sub-classes (30-100 cm, Mg ha⁻¹, showing mean as dark line and 95% spread of predictions) ..................................................105
Figure 4.6. Proportions of carbon stock in 30-100 cm depth relative to top 100 cm, by climate, parent material and vegetation cover class over eastern Australia (showing mean as dark line and 95% spread of predictions) ......107
Figure 5.1. Pre-clearing soil organic carbon map of NSW, 0-30 cm (from Banks and McKane 2002) .......................................................... 124
Figure 5.2. Pre-clearing modelling points over eastern Australia (shaded area denotes region that can be effectively represented by the points) ......................... 127
Figure 5.3. Ranges of most commonly used covariates for the eastern Australia pre-clearing point dataset and the NSW grids .............................................. 130
Figure 5.4. Pre-clearing SOC mass for NSW from current model (Mg ha\(^{-1}\), 0-30 cm) ...... 135
Figure 5.5. Observed versus predicted values for pre-clearing SOC mass from NSW digital map (0-30 cm, original scale) .................................................... 136
Figure 5.6. Change in SOC mass since clearing for NSW Mg ha\(^{-1}\), 0-30 cm) ................. 137
Figure 5.7. Absolute change in SOC density (0-30 cm) following clearing by temperature regime and parent material for intensive cropping and grazing land uses (Mg ha\(^{-1}\)) ................................................................. 141
Figure 5.8. Relative change in SOC density (0-30 cm) following clearing by temperature regime and parent material for intensive cropping and grazing land uses (% of original) ........................................... 141
Figure 6.1. Location of SOC modelling profile points and NSW study area ....................... 159
Figure 6.2. Change in SOC stocks from the 12 climate models for both change periods .......................................................................................... 167
Figure 6.3. Average change in SOC stock across NSW for 2nd change period (0-30 cm, Mg ha\(^{-1}\), mean of 12 NARClIM models) ............................................. 168
Figure 6.4. 95% confidence interval and mean change in SOC stocks by physical zone from the 12 NARClIM models (Mg ha\(^{-1}\), 0-30 cm, 2\(^{nd}\) change period) ................................................................... 170
Published peer reviewed publications from Thesis

Chapter 2


doi.org/10.1016/j.catena.2016.07.045


Chapter 3


doi:10.1071/SR13306

Chapter 4


Chapter 5


Chapter 6

Chapter 1: Introduction

1.1 General introduction

Soils are vital for humankind. They are the basis of most of our food and textiles. The preservation and maintenance of the world’s soils is imperative for human civilisation. With the world population projected to grow from its current 7 billion to 9 billion by 2050, the pressure on this vital resource will only increase. Soils also play an integral role in a range of essential ecosystem services in addition to agriculture. They play a key role in ecosystem health, hydrological cycles and in the global cycles of carbon and greenhouse gases influencing climate. As recently stated by the United Nations Secretary-General:

“Sustainable soil management is fundamental to achieving the Sustainable Development Goals – many of which reflect the centrality of soils to sustain life, food, and water” (Ki-moon 2015).

The need for fine resolution spatial soil information is increasing (Grunwald 2005; Hartemink and McBratney 2008; Sanchez et al. 2009; Brevik et al. 2016; Minasny and McBratney 2016). It is an essential input into developing sustainable soil and land management systems that are vital for global soil and food security. It is a key component of many climate change, ecological, hydrological and other modelling systems.

1.1.1 Soil-environment relationships through space and time

In addition to fine resolution maps, there is a need to understand the fundamental relationships describing soil distribution through space and time. What are the factors that control the distribution of key soil properties, and how do these properties respond to changes in the environment such as altered land use and global climate change? Understanding the spatial and temporal distribution of soils is considered essential for the sustainable and effective management of soils and the environment generally (Grunwald 2005)

Soil-environment relationships have been the subject of study since Dokuchaev first put forward his theories on soil formation in the late 1800s (Dokuchaev 1899). He and other early workers all contributed to the State Factor model (clorpt), most clearly...
enunciated by Jenny (1941). In recent decades, more sophisticated relationships have been developed as a component of the emerging digital soil mapping (DSM) methods, typically coming under the *scorpan* framework of McBratney *et al.* (2003). Further background to the development of the State Factor model and digital soil mapping is provided in Appendix 1.

Despite the recent advances in quantitative soil modelling and digital soil mapping, there is still a need to further elucidate important relationships controlling the distribution of various key soil properties. The fundamental soil equation of Dokuchaev and Jenny or the more recent *scorpan* framework have still not been effectively solved with full numeric coefficients, certainly not at any universal level. Few readily interpreted relationships have considered the combined influence of multiple variables, most being the relatively simple pedo-functions promoted by Jenny (1961) using just one or two variables and being restricted to particular regions. As noted by Heuvelink (2005), the pedological literature presents us with quasi-mathematical equations, but details of the function have only been partially revealed. One reason that author gives for this is that we do not have sufficient understanding of some of the mechanisms involved in soil landscape formation and development.

More sophisticated quantitative soil-environmental models, involving the use of remote sensed covariates such as gamma radiometrics and satellite derived spectral imagery, have been prepared for a range of soil properties, particularly in the initial modelling component of digital soil mapping projects (McBratney *et al.* 2003). However, the typical complexity of most of these models and their underlying covariates means that they cannot be readily interpreted to gain pedologic knowledge such as the precise influence of individual soil-forming factors, although Bui *et al.* (2009) and Viscarra Rossel *et al.* (2014) did make some such interpretations in relation to soil organic carbon in Australia.

It appears that most readily interpreted information on the role of the soil-forming factors and processes remains in a qualitative, conceptual form rather than quantitative form, as lamented by Heuvelink (2005). For example, it is widely reported that parent material influences soil character, but there appears almost a complete lack of research on its precise influence in any quantitative sense, such as the systematic variation in SOC or pH between soils derived from parent materials ranging from basalt to granite.
Chapter 1: Introduction

It has been asserted that there can be no universal equation that fits all soil landscapes; the equations must be domain dependent (Grunwald 2005). Phillips (2001) goes further in arguing that intrinsic variability within homogeneous landscape units is more important in determining pedo-diversity than is the extrinsic variability associated with measuring differences in topography, parent material and vegetation/land use. However, these arguments should not be taken as justification to not pursue elucidation of universal soil – environment relationships.

There appears to be a gap in the global research literature on relatively straightforward and easily applied quantitative or semi-quantitative models between key soil properties and the major soil-forming factors. Further investigation is required on the broad patterns of variation in the distribution and change of key soil properties relative to the main soil-forming factors, and combinations of these factors. With respect to the drivers of change in soil properties, major land use change and global climate change are major issues requiring attention. Digital soil modelling and mapping (DSMM) techniques offer an approach to investigate these soil-environment relationships in ways that appear not to have been previously widely explored.

1.1.2 SOC distribution through space and time

Soil organic carbon (SOC) is a soil property that requires particular attention. The long term storage of carbon in our soils offers a potentially important avenue to offset increasing atmospheric carbon levels and thus help mitigate potential climate change (Lal et al. 2007; Smith 2012; IPCC 2014). High SOC levels are also associated with improved soil health and agricultural productivity (Baldock et al. 2009; Sanderman et al. 2010).

A number of specific research issues with respect to SOC have been identified recently. Stockmann et al. (2013) highlighted the “unknowns” of SOC including the need for research on the average net change in soil organic carbon due to environmental conditions or management practices. McBratney et al. (2014) raised four broad challenges, three of which related to (i) the concept of soil SOC saturation and identifying the limiting capacity of soils to accumulate carbon, (ii) the influence of land management; and (iv) modelling SOC dynamics in space and time, including with depth. These “unknowns” and challenges provide useful guidance for ongoing research with respect to SOC.
1.1.3 Dissemination of soils knowledge

The need to facilitate communication of soil data and knowledge to end users is widely recognised, particularly for practical purposes such as policy making, land use planning and land management (Bouma 2014; Brevik et al. 2016; Minasny and McBratney 2016). A “flexible and pragmatic” approach to presenting soil data was called for by Calzolari and Filippi (2016).

The potential to identify patterns in soil distribution and soil change in terms of variables that can be easily interpreted may facilitate application into natural resource and other environmental programs. The ability to identify spatially explicit zones, which can be subjected to particular planning or land management strategies is desirable. This was the concept behind the carbon matrix zones of Murphy et al. (2010) that identify well defined zones of potential carbon storage that can be easily interpreted and implemented by land managers. Aggregating and interpreting spatial data has always been a part of soil classification and cartography, and can now be done more efficiently with DSMM techniques (Brevik et al. 2016).

Further investigation is desirable into means to present soil-environment relationships and soil data in formats that can be readily understood and interpreted by a wide range of end users. Formats that provide both continuous raster data and polygonal class data may be valuable.

1.1.4 Broad research questions

The application of digital soil modelling and mapping (DSMM) techniques opens the door to explore new approaches to examining soil – environment relationships and shed further light on the factors controlling soil distribution and change through space and time. Several innovative applications of DSMM are attempted in this research project. SOC and other key soil properties are modelled and mapped under current, past and future land use and climatic conditions over both eastern Australia and NSW and important patterns of distribution and change identified.
Key broad research questions to be examined are:

- Can we better elucidate the influence of different factors in controlling the distribution of key soil properties?
- How do these factors combine to control distribution of soil properties?
- How does SOC respond to changes in the environment such as altered land use and global climate change? Can readily interpreted relationships and patterns in change be identified?
- How effective are pragmatic soil relationships and data products, as derived in the study, in disseminating soil knowledge?

1.2 Specific research issues

To help address the above broad research questions, five more specific research issues were identified, as outlined below.

1.2.1 Influence of lithology in soil formation and its application in DSMM

The importance of parent material in soil formation has been recognised since the early years of pedology (Dockuchaev 1899; Hilgard 1906; Glinka 1914). It is a fundamental component of the clorpt framework of Jenny (1941) and the scorpan framework of McBratney et al. (2003). However, the precise role and influence of parent material in soil formation and in controlling soil distribution appears to be not fully understood. There appear few meaningful relationships reported in the literature.

Readily available geological and lithological data are not widely used in DSMM studies, probably due in part to its typical categorical data format, but also possibly due to a lack of knowledge on its potential influence and how to most effectively apply it. There is strong reliance on geophysical data (McBratney et al. 2003; Mulder et al. 2011) such as gamma radiometric data; multi- and hyper-spectral data such as visible and near infrared (VNIR) and Landsat Thematic Mapper; electromagnetic induction/electrical conductivity and others, which have the benefit of being continuous data forms. However, their relationships to lithology or to direct soil properties are not always well defined. Lithology covariates may be useful in complementing other geophysical covariates in DSMM programs.
Specific research questions include:

- Can we further elucidate the relationship of lithology to key soil properties?
- How effective is properly organised lithology data relative to other geophysical covariates in digital soil model and map products developed for NSW?
- Can we demonstrate an appropriate methodology for the use of lithology class data in DSMM programs?

1.2.2 Readily interpreted soil – environment relationships and their potential use in digital soil mapping

A better understanding of the main environmental factors that control the formation and distribution of a number of key soil properties is required. These properties include soil organic carbon (SOC), pH, cation exchange capacity (CEC), base content, particle sizes (clay, silt and sand) and total phosphorous. The development of relatively straightforward relationships (or models) relating these properties to the main soil-forming factors, similar to that called for by Heuvelink (2005) would be useful, particularly in helping us to predict and understand their distribution across the landscape. They may allow us to better predict changes due to changing conditions for example, the change in pH with each 100 mm rise in rainfall (with other factors remaining constant), or the changes due to the combined influence of several factors. The use of readily available covariates will assist in their application and interpretation. They may also allow us to predict soil properties at individual sites, using only readily collected field data.

The pragmatic models so developed may have potential for use in preparing digital soil maps in a manner which is more readily understood and transparent than the more advanced DSMM techniques. They may therefore serve as a useful introductory tool to DSMM and thus help to address an apparent reluctance to adopt these techniques by at least some soil scientists (Scull et al. 2003; Hartemink et al. 2008; Hempel et al. 2008; Moore et al. 2010).

Specific research questions include:

- Can we develop effective pragmatic relationships, such as multiple linear regression (MLR) models, for key soil properties over eastern Australia, based on readily available pragmatic variables?
• Do they inform us on the quantitative influence of each soil forming factor?
• What is the feasibility of preparing DSMs using this pragmatic approach over NSW?
• How do the pragmatic models and maps compare with other more advanced DSM techniques?
• What are their main weaknesses and potential benefits?

1.2.3 The factors controlling the distribution of soil organic carbon stocks, spatially and with depth

Further work is required to elucidate the relative levels of influence of important soil-forming factors in controlling SOC stocks. We need to understand and quantify how the influence of these drivers changes with increasing depth in the soil profile. More quantitative data are needed on the combined influence of the key factors and how they work together to produce different SOC stocks in different environmental regimes. These issues particularly address the SOC “unknowns” of Stockmann et al. (2013) and two of the four key SOC research challenges of McBratney et al. (2014). This knowledge will allow us to develop realistic strategies to promote long term increases in soil carbon levels, thus help mitigate potential climate change and promote improved soil health.

Specific research questions include:

• What are the key drivers of soil organic carbon stocks in the soils of eastern Australia and how do they vary with depth?
• Are there systematic patterns of SOC stock levels according to climate – parent material/soil type – groundcover/land use?
• Are their systematic trends in topsoil / subsoil SOC storage ratios?

1.2.4 Change in SOC stocks with the clearing of native vegetation

Data on SOC stocks prior to native vegetation clearing can be an important requirement for carbon accounting systems such as those prescribed by the Intergovernmental Panel on Climate Change (IPCC 2006). Such data is often also applied in the initialisation and validation of soil carbon turnover models such as Roth C (Coleman and Jenkinson 1999), which project SOC behaviour under different climate
and land management regimes. Such data are typically acquired using conventional techniques that apply data from representative areas of essentially undisturbed native vegetation over broad polygonal units. It may be possible to improve on this data by applying DSMM techniques.

Knowledge on SOC change since native vegetation clearing provides important data and understanding on the impacts of land use change on soil carbon stocks. There is a need to understand how these changes relate to the main soil-forming factors. Do different combinations of factors such as climate, parent material and final land use result in systematically different levels of change? There appears to be little attempt in the literature to examine and identify systematic patterns in the level of SOC change following major land use change. Such knowledge may also allow us to assess potential gains in SOC following the conversion of agricultural land back to native vegetation. Ultimately, the knowledge may allow us to better estimate the potential contribution that land use change may play in soil carbon sequestration as a means to mitigate climate change (Lal 2004; Wilson et al. 2011; Baldock et al. 2012).

Specific research questions include:

- *Can we improve on currently available data on pre-clearing SOC stocks by applying DSMM techniques?*
- *Can we use DSMM techniques to determine the loss in SOC over NSW since clearing?*
- *Are there systematic patterns of change relative to climate, parent material and final land use?*

### 1.2.5 Change in SOC stocks with projected climate change

Climate change has the potential to impact on many elements of our world, including our soils. We need to gain knowledge and understanding of how key soil properties such as SOC will change due to projected climate change. Potential changes may impact on soil condition and agricultural productivity and also be important for climate change modelling and mitigation programs. Stockmann et al. (2013) note the uncertainty as to whether soil will act as a source or sink to carbon under future climate change.
To date most modelling of the impacts of climate change on soil properties has been carried out using simulation methods such as RothC (Yurova et al. 2010; Gottschalk et al. 2012; Smith 2012). A trial project relating to SOC carried out by Minasny et al. (2013) is a rare example of use of DSM techniques in this field. The use of DSM techniques in combination with a space-for-time substitution (SFTS) process offers a new approach to examining soil property change under climate change.

Specific research questions include:

- **What is the feasibility of using DSM – SFTS techniques to spatially quantify changes in SOC due to the influence of future climate change over NSW?**
- **How consistent are the predictions of SOC change between different global and regional climate models?**
- **What factors drive the predicted changes in SOC with climate change? Do the changes vary systematically according to environmental conditions based on current climate–parent material (soil type)–land use regimes?**

### 1.3 Organisation of thesis

This thesis includes seven chapters, comprising the current Introduction, which is followed by five chapters dealing with the specific research issues raised in the previous section (representing journal articles that are either published or under review) and finishing with a final Summary and discussion chapter.

**Chapter 1: Introduction**

The project is introduced and the five broad and 17 specific research issues outlined

**Chapter 2: Lithology and soil relationships and their use in digital soil mapping**

**Chapter 3: Pragmatic models for the prediction and digital mapping of soil properties in eastern Australia.**

**Chapter 4: Factors controlling soil organic carbon stocks with depth in eastern Australia.**

**Chapter 5: Digital mapping of pre-European soil carbon stocks and decline since clearing over New South Wales, Australia.**
Chapter 6: Change in soil organic carbon stocks under twelve climate-change projections over New South Wales, Australia.

Chapter 7: Summary and discussion

The main results from the research program are summarised and briefly discussed through the answering of each of the 17 specific, then five broad research questions as raised in the Introduction. Future research directions are proposed before the final conclusion.

1.4 References


Translated from German by CF Marbut in 1927, Edwards Bros., Ann Arbor, Michigan.


Chapter 1: Introduction

Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.


Chapter 1: Introduction


Chapter 2: Lithology and soil relationships for soil modelling and mapping

Abstract

Relationships between parent material and soil are not well understood and generally only reported in qualitative form. We present a classification of parent material for pedologic purposes, which includes twelve lithology classes based on mineralogical and chemical composition. The relationships of these lithology classes with six key soil properties (soil organic carbon, pH, cation exchange capacity, sum-of-bases, total P and clay %) were examined in a case study over New South Wales, Australia. We used multiple linear regression, Random Forest and Cubist tree models based on a soil dataset of over 3200 points. Semi-quantitative estimates are derived of change in these soil properties with a change in lithology class, and an associated silica index, for example, a 22% relative decrease in soil organic carbon with each 10% rise in silica, broadly equivalent to a change from shale to granite, assuming other factors remain constant.

Parent material covariates are vital for the effective modelling and mapping of soil properties. Widely available lithology data have the potential for greater use in digital soil modelling and mapping (DSMM) programs. We compared the performance of the classified lithology data with other continuous, geophysical parent material covariates such as gamma radiometrics in digital soil models and maps over NSW. The lithology covariate was demonstrated to exert the greatest influence on all six soil properties, coming well ahead of all geophysical parent material and other environmental covariates. Validation statistics demonstrated strong improvement in both model and map quality when the lithology covariate was included. For example, Lin’s concordance for the Cubist sum-of-bases model rose from 0.46 with no parent material covariates, to 0.58 with the continuous geophysical covariates, to a high of 0.77 when lithology was also used. The improvement was typically slightly less marked in the final digital maps than for the calibration models, probably due to the lower reliability of the lithology grid derived from broad scale polygonal geological and soil data. A process is suggested for the application of lithology data into DSMM programs. Despite the potential drawbacks of using polygonal data, properly organised categorical lithology data can be
a strong covariate to complement other continuous geophysical data sources in DSMM programs, particularly where reliable and fine scale geological and soil data are available.

**Key words:** lithology; geology; parent material; soil properties; quantitative relationships; digital soil mapping

### 2.1 Introduction

The importance of parent material in soil formation has long been recognised. Soil has been described as a “kind of pathologic condition of the native rock” (von Richthofen 1882) and “the residual product of the physical disintegration and chemical decomposition of rocks” (Hilgard 1906). Parent material was given prominence in the earliest theories of soil formation (Dokuchaev 1899; Glinka 1927; Hilgard 1906; Joffe 1936). It provides the raw starting material of the soil, be it bedrock or other unconsolidated material, upon which soil forming processes will act to create a particular soil. The essential chemical character of the parent material will be imparted into the derivative soil.

Parent material is recognised as a key component of most models of soil formation, and is an integral part of the fundamental soil equation \((clorpt)\) of Jenny (1941). However, there appears to be little rigorous examination of broad universal relationships between parent material or lithology to soil formation and distribution, for example, how soil organic carbon (SOC) or pH systematically vary between soils derived from basalt to granite. Detailed investigation through literature search engines reveals a scarcity of studies on systematic lithology – soil relationships.

There are many studies that confirm the strong influence of lithology on soil distribution (Bui et al. 2006; Greve et al. 2012; Hengl et al. 2014; Xiong et al. 2014) but they rarely attempt to elucidate the actual relationships. Several studies have examined the differences in various soil properties under specific parent materials over particular regions (Cathcart et al. 2008; Chaplot et al. 2003; Cline 1953; Graham and Franco-Vizcaino 1992; Gruba and Socha 2016; Jaiyeoba 1995; van de Wauw et al. 2008) but results are generally not synthesised to draw out clear universal trends. Existing relationships are at best qualitative; there appears an almost complete lack of any quantitative or semi quantitative relationships, a concern more broadly expressed by
Chapter 2: Lithology – soil relationships and their use in DSMM

Heuvalink (2005). It has been suggested that this deficiency is due to difficulty in quantifying parent material in a meaningful way (Schaetzl and Anderson 2005; Yaalon 1975). This problem has been addressed to some extent through the use of geophysical indicators such as gamma radiometric or spectral imagery data in digital soil modelling and mapping (DSMM) programs, but the relationships so derived are difficult to interpret and translate more universally. The lead author has attempted to investigate parent material – soil relationships previously (Gray and Murphy 1999, Gray et al. 2009, 2014, 2015), but further work is required.

2.1.1 Sources of parent material data for soil modelling and mapping

In addition to its vital use in conventional soil mapping programs, parent material data is widely used in digital soil mapping programs, being an element of the scorpan framework of McBratney et al. (2003). Parent material information is generally readily available in the form of lithology data in geological maps ranging from broad to fine scale, and now usually in digitised format. Lithology refers to the gross physical character of parent material, including its mineral composition, colour and grain size. The data is often collected in soil survey programs, but a more systematic collection and recording of subsolum data in these programs is required, as recently called for by Juilleret et al. (2016).

Despite its availability, lithology data is frequently omitted altogether as a data source in DSMM programs. Where it is used it may be in a simplistic manner where different stratigraphic units (for example, Rylstone Volcanics or Winton Formation) are used as separate classes and not re-classified in any meaningful way. This can be cumbersome when large areas with a large number of different units are involved. In other cases, geological materials are broadly grouped together into very general classes such as coarse igneous, sedimentary or alluvial material that do not sufficiently distinguish key soil forming attributes, for example, grouping diorites with granites or feldspathic sandstones with quartz sandstones. Other approaches to classifying geological data for DSMM purposes have also been trialled (Vaysse and Lagacherie 2015).

Remotely or proximally sensed geophysical and other modelled sources of parent material data are frequently used in digital soil mapping programs, having the benefit of providing continuous datasets down to fine pixel resolutions, such as 30 m or finer.
Chapter 2: Lithology – soil relationships and their use in DSMM

(McBratney et al. 2003; Mulder et al. 2011). Foremost amongst these are gamma radiometric data (Taylor et al. 2002; Wilford 2012; Wilford and Minty 2007); multi- and hyper-spectral data such as visible and near infrared (VNIR) (Lagacherie and Gomez 2014; Viscarra Rossel and Webster 2012) and Landsat Thematic Mapper (Boettinger et al. 2008); and electromagnetic induction/electrical conductivity (Triantafillis et al. 2009, Zhu et al. 2010). Other geophysical data sources such as magnetometry (Jordanova et al. 2008; Ryan et al. 2000) and gravity anomalies (Viscarra Rossel et al. 2015) are occasionally used. However the geophysical signals can be distorted in various ways and their relationships to lithology or direct soil properties are not always strong and well defined. It can be difficult to clearly understand and interpret the role that the data are having in the soil model, meaning there can be a lack of transparency and less opportunity for the gaining of pedological knowledge.

An examination of 265 recent DSMM papers from around the globe as presented in the conference proceedings of Minasny et al. (2012) and Arrouays et al. (2014); and the meta-studies of McBratney et al. (2003), Grunwald (2009) and Minasny et al. (2013) provide an indication of the extent and variety of auxiliary data sources used to represent parent material or direct soil conditions. The application rates for the different sources was as follows: soil maps/data - 40% of studies; geology and lithology maps - 22%; spectral sensing techniques (Landsat, hyper-spectral VNIR, etc) - 21%; gamma radiometrics - 9%, other geophysical sources (electromagnetic induction, electrical conductivity, etc) - 7%; and nil parent material/soil data - 25%. It would appear there is potential for primary geological/lithologic data to be more widely utilised as an auxiliary data source in DSMM projects being carried out around the globe.

2.1.2 Aims

There is a need to elucidate relationships between parent material and key soil properties, so as to improve our knowledge of factors controlling soil formation and distribution. Quantitative or semi-quantitative relationships with lithology would be a useful addition to the generally poorly defined and qualitative relationships that exist at present. Widely available lithology data could provide a strong and easily applied predictor in both conventional and digital soil modelling and mapping programs, to complement other geophysical, continuous parent material data sources. This paper
builds on previous work of the lead author and others to develop lithology – soil relationships and to assess the potential effectiveness of incorporating lithology into DSMM programs. More specifically, the paper aims to:

- present a possible classification scheme of parent material for pedologic purposes based on broad chemical composition with 12 lithology classes
- derive semi-quantitative relationships between lithology and six key soil properties in a case study over New South Wales (NSW), Australia
- demonstrate the effectiveness of lithology as a covariate in DSMM, including comparing its effectiveness relative to other potentially available geophysical parent material covariates.
- suggest a strategy for the inclusion of lithology as a predictor in DSMM programs.

2.2 Classification of parent material for pedological purposes

For pedological purposes, the most important feature of parent material is its lithology, and more specifically its mineralogy and chemical composition. These greatly influence both the chemical and physical properties of derivative regolith material and soils. Key chemical characteristics are the silica (SiO$_2$) content and selected base cation content (Ca, Mg and K), which usually have an inverse relationship with each other. The higher the silica content of a parent material, the generally higher the quartz content and lower the clay and base cation content of derivative soils. Ultimately all key soil properties are greatly influenced by the original parent material.

Other physical characteristics of the parent material such as grain size and macrostructure (layering, fracturing, etc) are generally of lesser significance, although they can be important in some situations. Most major minerals apart from quartz will weather to clay irrespective of whether they were originally coarse or fine grained, thus for example, basalt and its coarse grained equivalent gabbro will normally give rise to similar soils. When quartz is a major component of the parent material, such as in siliceous sedimentary or igneous rocks, its grainsize becomes a more important factor and will determine whether the material classifies as coarse sand, fine sand or silt. In younger soils, such as derived from alluvial deposits, the grainsize of all minerals can
be important. Bedrock structure such as the degree and orientation of fracturing can influence soil hydrology and depth of weathering properties.

The origin of the material, be it igneous plutonic or volcanic, sedimentary, metamorphic, alluvial, aeolian, etc, should not in itself directly influence soil formation. It is only the inherent chemical and to a lesser extent physical properties of the material that are important.

We propose a possible parent material classification with 12 lithological classes based on their chemical composition, as presented in Table 2.1. The first eight categories are based on silica and base cation levels, ranging from extremely siliceous (>85% silica) to ultra-mafic (<45% silica). Each of these first eight classes may be allocated a single “silica index”, being its median silica percentage, which can be useful in semi-quantitative soil modelling activities. The last four categories are defined by other chemical characters, namely calcareous, sesqui-oxide, organic and evaporite materials, and cannot be allocated a meaningful silica index.

The classification places more emphasis on composition and less on origin compared to other schemes such as those presented in FAO (2006) and Juilleret et al. (2016). We consider this scheme will be widely applicable around the world, but variation and addition of new classes may be appropriate in many regions. The scheme can act as a base which can be adapted as required to meet the needs of any regions or users. For example, in Central and Northern Australia and elsewhere in the world, a further sub-division of the sesqui-oxide class may be appropriate. It is the concept of systematically ordering into a manageable number of classes that is the most important.
Table 2.1: Broad classification of parent material for pedologic purposes

<table>
<thead>
<tr>
<th>Lithology class</th>
<th>Silica ((\text{SiO}_2) %^1)</th>
<th>Key base cation oxides (%^2)</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Range</td>
<td>Median (\text{&quot;Silica index&quot;})</td>
<td></td>
</tr>
<tr>
<td>1. Extremely siliceous</td>
<td>&gt;85</td>
<td>88</td>
<td>quartz sands (beach, riverine or aeolian), chert, pure quartzite, jasper, quartz reefs, silicified rocks</td>
</tr>
<tr>
<td>2. Siliceous upper</td>
<td>77-85</td>
<td>80</td>
<td>quartz sandstone, quartz siltstone, unqualified quartzite and alluvial sands</td>
</tr>
<tr>
<td>3. Siliceous mid</td>
<td>70-77</td>
<td>73</td>
<td>granite, rhyolite and siliceous tuff, arkose sandstone, most unqualified sandstone</td>
</tr>
<tr>
<td>4. Siliceous lower</td>
<td>65-70</td>
<td>68</td>
<td>adamellite, granodiorite, dacite, monzogranite, siliceous/intermediate tuff, most greywacke &amp; lithic sandstone, unqualified siltstone</td>
</tr>
<tr>
<td>5. Intermediate upper</td>
<td>60-65</td>
<td>62</td>
<td>syenite, trachyte, most argillaceous rocks (mudstone, claystone, shale, slate, phyllite and schist), alluvial loams and non-cracking clays</td>
</tr>
<tr>
<td>6. Intermediate lower</td>
<td>52-60</td>
<td>57</td>
<td>monzonite, trachy-andesite, diorite, andesite, intermediate tuff, alluvial cracking clays (not black)</td>
</tr>
<tr>
<td>7. Mafic</td>
<td>45-52</td>
<td>49</td>
<td>gabbro, dolerite, basalt, mafic tuff, amphibolite, alluvial black cracking clays</td>
</tr>
<tr>
<td>8. Ultra-mafic</td>
<td>&lt;=45</td>
<td>42</td>
<td>serpentinite, dunite, peridotite, tremolite-chlorite-talc schists</td>
</tr>
<tr>
<td>9. Calcareous</td>
<td>variable</td>
<td>na(^3)</td>
<td>variable</td>
</tr>
<tr>
<td>10. Sesquioxide</td>
<td>variable</td>
<td>na(^3)</td>
<td>variable</td>
</tr>
<tr>
<td>11. Organic</td>
<td>variable</td>
<td>na(^3)</td>
<td>variable</td>
</tr>
<tr>
<td>12. Evaporite</td>
<td>variable</td>
<td>na(^3)</td>
<td>variable</td>
</tr>
</tbody>
</table>

1 Approximate compositions from Best (1982), Duff (1993), Joplin (1965), Mason (1966) and Pettijohn (1963)

2 Calcium (Ca), magnesium (Mg) and potassium (K) oxides, note that in soil science sodium (Na) is normally included as a base cation, but is excluded here as it tends to decrease soil productivity

3 not applicable, the Silica index cannot be applied to these materials.

The average composition of a range of common rocks belonging to most of these lithology classes is shown in Table 2.2. Examination of this table allows insights to be gained on the reason many key properties vary between soils derived from different parent materials under otherwise equivalent conditions (Gray and Murphy 1999; Gray et al. 2014). Key base content, as given by the sum of the oxides of Ca, Mg and K, can be seen to increase from the highly siliceous materials (eg, average granite: 6.6 %) to the mafic materials (eg, average basalt: 17.3%). This should be reflected in the associated soils, with a corresponding increase in sum-of-bases (macro-nutrients) and
Table 2.2. Chemical composition of common rock types

<table>
<thead>
<tr>
<th></th>
<th>Extremely siliceous upper</th>
<th>Siliceous mid</th>
<th>Siliceous lower</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dune sand¹ (s)</td>
<td>Quartz sandstone¹ (s)</td>
<td>Arkose sandstone (p)</td>
</tr>
<tr>
<td>SiO₂</td>
<td>97.62</td>
<td>82.05</td>
<td>77.1</td>
</tr>
<tr>
<td>TiO₂</td>
<td>-</td>
<td>0.35</td>
<td>0.3</td>
</tr>
<tr>
<td>Al₂O₃</td>
<td>1.32</td>
<td>9.20</td>
<td>8.7</td>
</tr>
<tr>
<td>Fe₂O₃</td>
<td>0.4</td>
<td>0.50</td>
<td>1.5</td>
</tr>
<tr>
<td>FeO</td>
<td>-</td>
<td>1.92</td>
<td>0.7</td>
</tr>
<tr>
<td>MnO</td>
<td>-</td>
<td>0.05</td>
<td>0.2</td>
</tr>
<tr>
<td>MgO</td>
<td>0.13</td>
<td>0.39</td>
<td>0.5</td>
</tr>
<tr>
<td>CaO</td>
<td>0.28</td>
<td>0.22</td>
<td>2.7</td>
</tr>
<tr>
<td>Na₂O</td>
<td>-</td>
<td>0.16</td>
<td>1.5</td>
</tr>
<tr>
<td>K₂O</td>
<td>-</td>
<td>1.50</td>
<td>2.8</td>
</tr>
<tr>
<td>H₂O</td>
<td>0.52</td>
<td>2.48</td>
<td>0.9</td>
</tr>
<tr>
<td>CO₂</td>
<td>-</td>
<td>1.34</td>
<td>3.0</td>
</tr>
<tr>
<td>P₂O₅</td>
<td>-</td>
<td>-</td>
<td>0.1</td>
</tr>
<tr>
<td>Total</td>
<td>100.27</td>
<td>100.16</td>
<td>100.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Intermediate lower</th>
<th>Mafic (v)</th>
<th>Ultermatic (p)</th>
<th>Calcareous (s)</th>
<th>Sesquioxide (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shale (s)</td>
<td>Diorite (p)</td>
<td>Andesite (v)</td>
<td>Basalt (v)</td>
<td>Peridotite (p)</td>
<td>Limestone (s)</td>
</tr>
<tr>
<td>SiO₂</td>
<td>64.80</td>
<td>57.48</td>
<td>57.94</td>
<td>49.2</td>
<td>42.26</td>
</tr>
<tr>
<td>TiO₂</td>
<td>0.78</td>
<td>0.95</td>
<td>0.87</td>
<td>1.84</td>
<td>0.63</td>
</tr>
<tr>
<td>Al₂O₃</td>
<td>16.90</td>
<td>16.67</td>
<td>17.02</td>
<td>15.74</td>
<td>4.23</td>
</tr>
<tr>
<td>Fe₂O₃</td>
<td>-</td>
<td>2.50</td>
<td>3.27</td>
<td>3.79</td>
<td>3.61</td>
</tr>
<tr>
<td>FeO</td>
<td>5.70</td>
<td>4.92</td>
<td>4.04</td>
<td>7.13</td>
<td>6.58</td>
</tr>
<tr>
<td>MnO</td>
<td>0.06</td>
<td>0.12</td>
<td>0.14</td>
<td>0.2</td>
<td>0.41</td>
</tr>
<tr>
<td>MgO</td>
<td>2.85</td>
<td>3.71</td>
<td>3.33</td>
<td>6.73</td>
<td>31.24</td>
</tr>
<tr>
<td>CaO</td>
<td>3.56</td>
<td>6.58</td>
<td>6.79</td>
<td>9.47</td>
<td>5.05</td>
</tr>
<tr>
<td>Na₂O</td>
<td>1.15</td>
<td>3.54</td>
<td>3.48</td>
<td>2.91</td>
<td>0.49</td>
</tr>
<tr>
<td>K₂O</td>
<td>3.99</td>
<td>1.76</td>
<td>1.62</td>
<td>1.1</td>
<td>0.34</td>
</tr>
<tr>
<td>H₂O</td>
<td>-</td>
<td>1.36</td>
<td>1.17</td>
<td>1.38</td>
<td>4.22</td>
</tr>
<tr>
<td>CO₂</td>
<td>-</td>
<td>0.10</td>
<td>0.05</td>
<td>0.11</td>
<td>0.3</td>
</tr>
<tr>
<td>P₂O₅</td>
<td>0.11</td>
<td>0.29</td>
<td>0.21</td>
<td>0.35</td>
<td>0.1</td>
</tr>
<tr>
<td>Total</td>
<td>99.9</td>
<td>99.98</td>
<td>99.93</td>
<td>99.95</td>
<td>99.46</td>
</tr>
</tbody>
</table>

Average igneous rocks (Best 1982); average arkose sandstone and greywacke (Pettijohn 1963); shale: North American composite, iron reduced analysis (Gromet et al. 1984), average limestone: (Mason 1966); ¹site samples (Joplin 1965) including: dune sand, Cronulla, NSW; quartz sandstone, Pyrmont, NSW; laterite, Darling Range, WA; bauxite, Wingello, NSW; p: plutonic igneous (coarse grained); v: volcanic igneous (fine grained); s: sedimentary, tr: trace
pH, assuming other factors remain constant. Phosphorous ($P_2O_5$) shows a similar increase, ranging from 0.12% in granites, to 0.22% in shales, to 0.35% in basalt before dropping to 0.1% in peridotite, reflecting the unusual chemistry of ultra-mafic rocks. Soil sodicity problems may be expected to be influenced by the $Na_2O/CaO$ ratio of the parent materials, which shows a steady decrease with increasing mafic character, from 2.0 for average granite down to 0.3 for average basalt.

The broad association of common soil types with each of the 12 lithological classes is presented in Table 2.3. Soil types are given in terms of the World Reference Base for Soil Resources (WRB, IUSS Working Group WRB 2014a), Soil Taxonomy (Soil Survey Staff 2010) and the Australian Soil Classification (ASC, Isbell 2002). These associations should be considered first approximations only; most soil types will extend into adjoining lithology classes and precise soil types are the product of all soil-forming factors. The association of soil types with parent material, climate and topography has been discussed and graphically presented for WRB soils in Gray et al. (2011) and IUSS Working Group WRB (2014b) and for ASC soils in Gray and Murphy (1999).
Chapter 2: Lithology – soil relationships and their use in DSMM

Table 2.3. Lithology classes and typically associated soils\(^1\) (generalized first approximation only)\(^1\)

<table>
<thead>
<tr>
<th>Lithology class</th>
<th>Common WRB soils(^2)</th>
<th>Common Soil Taxonomy soils(^3)</th>
<th>Common Australian Soil Classification soils(^4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Extremely siliceous</td>
<td>Arenosols; Podzols</td>
<td>Spodosols; \textit{Quartzipsammements}</td>
<td>arenic Rudosols; Podosols</td>
</tr>
<tr>
<td>2. Siliceous upper</td>
<td>Arenosols</td>
<td>Entisols; \textit{Psammments}</td>
<td>Kandosols; sandy Tenosols</td>
</tr>
<tr>
<td>3. Siliceous mid</td>
<td>Durisols; Lixisols; Planosol; Solonet; Umbrisols</td>
<td>low fertility Alfisols &amp; Aridosols; Inceptisols</td>
<td>Kandosols; dystrophic (low fertility) Chromosols, Kurosols &amp; Sodosols</td>
</tr>
<tr>
<td>4. Siliceous lower</td>
<td>Acrisols; Alisols; Cambisols; Retisols</td>
<td>moderate fertility Alfisols &amp; Aridosols</td>
<td>mesotrophic (moderate fertility) Chromosols, Kurosols &amp; Sodosols</td>
</tr>
<tr>
<td>5. Intermediate upper</td>
<td>Luvisols; Ferralsols</td>
<td>high fertility Alfisol &amp; Aridosol; Oxisols; \textit{Udults}</td>
<td>Dermosols; eutrophic (high fertility) Chromosols, Kurosols &amp; Sodosols</td>
</tr>
<tr>
<td>6. Intermediate lower</td>
<td>Chernozems; Kastanozems</td>
<td>Mollisols; \textit{Udolls}; \textit{Ustolls}; \textit{Kandi-udults}; \textit{Mollic paleudalfs}</td>
<td>Ferrosols; grey &amp; brown Vertosols</td>
</tr>
<tr>
<td>7. Mafic</td>
<td>Nitisols; Phaeozems; Vertisols</td>
<td>Vertisols</td>
<td>black Vertosols; Ferrosols</td>
</tr>
<tr>
<td>8. Ultra-mafic</td>
<td>Vertisols; Gleysols (high heavy metals(^5))</td>
<td>Vertisols (high heavy metals)</td>
<td>Vertosols (high heavy metals(^5))</td>
</tr>
<tr>
<td>9. Calcareous</td>
<td>Calcisols</td>
<td>Aridosols; \textit{Calcids}; \textit{Argids}; \textit{Calciaargids}</td>
<td>Calcarosols</td>
</tr>
<tr>
<td>10. Organic</td>
<td>Histosols</td>
<td>Histosols; Gellisols; \textit{Histels}</td>
<td>Organosols</td>
</tr>
<tr>
<td>11. Sesquioxide</td>
<td>Ferralsols; Plinthosols</td>
<td>Oxisols; \textit{Plinthaquox}</td>
<td>Ferrosols</td>
</tr>
<tr>
<td>12. Evaporite</td>
<td>Solonchaks</td>
<td>\textit{Salids}; \textit{Aquisalids}; \textit{Haplosalids}</td>
<td>Hydrosols</td>
</tr>
</tbody>
</table>

\(^1\) Most soil types will extend into adjoining lithology classes
\(^2\) World Reference Base for Soil Resources; based on IUSS Working Group WRB (2014a) and Gray \textit{et al.} (2011)
\(^3\) Based on Soil Survey Staff (2010) and Gray \textit{et al.} (2011). Common suborders or sub-groups in italics.
\(^4\) Based on Isbell \textit{et al.} (1997) and Gray and Murphy (1999)
\(^5\) Includes lead, chromium, arsenic, zinc, cadmium, copper, mercury and nickel

2.3 Methods

2.3.1 Overview

To serve as a case study, digital models describing the relationships of six key soil properties to lithology and the other main soil-forming factors were developed over NSW, Australia, using multiple linear regression, Random Forest and Cubist linear
piecewise decision tree techniques. The properties considered were soil organic carbon (SOC), pH (in CaCl$_2$), cation exchange capacity (CEC), sum-of-bases, total phosphorous (P$_{\text{total}}$) and clay content. Only the 5-15 cm depth interval was examined. Semi-quantitative estimates on the influence of lithology were derived, for example, the change in pH with 10% change in silica%.

The relative influence of lithology was compared to other soil-forming factors and to other geophysical parent material related covariates in the models. The statistical performance of the soil models and maps prepared over NSW was determined and compared using the differing combinations of parent material related covariates, ranging from (i) no parent material data; to (v) all sources (lithology and geophysical data).

2.3.2 Overview of NSW study area

The state of NSW in eastern Australia (Figure 2.1) encompasses an area of 810 000 km$^2$, slightly larger than France or Texas, and takes in a wide range of environments. Climate varies from warm temperate in the north, to hot arid in the far western areas, temperate in the south and sub alpine in the highlands of the south east. Mean annual maximum daily temperatures range from 10 to 30 degrees C while rainfall varies from less than 200 mm to over 3000 mm per annum. The physiography is marked by a range of mountains, the Great Dividing Range, which runs down the east coast (generally 100-300 km inland). This mountain range is low by world standards, only reaching a maximum of 2200 m in the south. Heading further west from this range the undulating terrain gives way to flat inland plains.

Surface geology of the region is characterised by Palaeozoic and Mesozoic siliceous and intermediate igneous and sedimentary rocks in the higher relief eastern regions with Tertiary alluvial sands, silts and clays occupying most of the flatter western regions. Remnants of Tertiary age mafic volcanics are widespread throughout much of the higher relief eastern areas. The limited extent of Holocene glaciation, as common in the northern hemisphere, means many landscape surfaces are more mature with at least moderately weathered materials, however active geomorphic processes over most of the State help to bring fresh parent material to the surface. Soils vary from very high to very low fertility types, depending on parent material, climatic and topographic conditions, as first mapped by Jensen (1914) and more recently by OEH.
These variable conditions also give rise to a great diversity of land use, ranging from intensive cropping to nature reserves.

### 2.3.3 The soil dataset

The soil dataset contained soil profiles acquired from the NSW State Government Office of Environment and Heritage (OEH). The P$_{\text{total}}$ data were, however, derived from all the eastern State Government soil agencies due to very low sample numbers in NSW. Sample numbers for each of the six properties examined were as follows: SOC, 1788; pH, 3246; CEC, 2318; sum-of-bases, 2290; P$_{\text{total}}$, 1804; and clay, 2609. Further details on laboratory methods and associated issues are given in Gray et al. (2015a).

The depth interval of 5-15 cm was adopted for the study, following conversion from the originally recorded horizon depths using the spline process of Bishop et al. (1999) and Malone et al. (2009). This depth interval is one of the standard intervals adopted in the GlobalSoilMap.net program (Sanchez et al. 2009; Arrouays et al. 2014), it is representative of upper soil horizons but tends to be less variable than the top surface layer.

### 2.3.4 Parent material covariates

The following covariates representing parent material were applied and their relative performance in the models compared. The latter three covariates (ii to iv) are referred to throughout this study as the “geophysical” parent material covariates.

1. **Lithology** – based on the 12 lithology classes as presented in Table 2.1. Some applications used the associated “Silica index”. For the initial model development, data was derived from parent material descriptions recorded at each site by the soil surveyor. For the final NSW lithology map preparation (Figure 2.1), a grid was prepared from lithology classes that had been applied manually to each of the 4115 geological units identified in the 1:250 000 scale digital geology map of the Geological Survey of NSW (undated). These were based on their lithological descriptions contained in the associated attribute table, using the rules presented in Table 2.1. An initial alphabetic sorting of the lithology column expedited this classification process. For poorly defined Cainozoic unconsolidated material, such as unqualified “alluvium” or “colluvium” for which their broad composition was unknown, lithological classes were allocated.
following reference to an existing NSW soil type map (OEH, 2012). This exploited clear soil type to parent material relationships as presented in Table 2.3.

(ii) Gamma radiometrics – radiometric potassium ($rad_K$), uranium ($rad_U$) thorium ($rad_Th$) and the ratio of K to U ($KU\_ratio$); 90 m grids developed by and sourced from Geoscience Australia.

(iii) NIR clay components – the relative proportions of kaolin, illite and smectite clays and the smectite/kaolin ratio ($S/K\_ratio$) derived from DSMM techniques based on laboratory near infra-red (NIR) spectroscopy (Viscarra Rossel 2011); 90 m grids sourced through the CSIRO Data Access Portal via the Soil and Landscape Grid of Australia (TERN 2014; Grundy et al. 2015).

(iv) Weathering index ($W\_I$) – an index to represent the degree of weathering of parent materials, regolith and soil, based on gamma radiometric data (Wilford 2012); 90 m grids were sourced from Geoscience Australia. This index is also reflective of the age factor in the clorpt and scorpan frameworks.

Figure 2.1. Lithology class map for NSW
2.3.5 Other covariates

Covariates relating to the other soil-forming factors are listed below, with further detail provided in the cited references:

Climate: (i) Mean annual rainfall (mm pa, *Rain*) derived from 2.5 km Australia wide climate grids from the Australian Bureau of Meteorology with interpolation of cell values down to a 100 m grid. The values represent mean values obtained over the 1961-1990 period, which coincides with the period when a large proportion of the soil profiles were collected.

(ii) Mean annual daily maximum temperature (°C, *T*$_{max}$) – as above

Relief: (i) *Topo-slope index (TSI)* – an index that combines topographic position and slope gradient (Gray et al. 2015a)

(ii) *Topographic wetness index (TWI)* – a widely used index that represents potential hydrological conditions based on slope and catchment area, as derived from digital elevation models (DEMs) (Gallant and Austin 2015; TERN 2014)

(iii) *Slope* - slope gradient in percent as derived from field data and a 100 m DEM (Gallant and Austin 2015)

(iv) *Aspect index (Asp)* – an index to represent the amount of solar radiation received by sites, ranging from 1 for flat areas and gentle N or NW facing slopes (high radiation in southern hemisphere) to 10 for steep S and SE slopes (low radiation) (Gray et al. 2015a)

Biota: (i) *Land disturbance index (LDI)* – an index that reflects the intensity of disturbance associated with the land use (Gray et al. 2015a), where 1 denotes natural ecosystems and 6 denotes intensive cropping, based on 1:25 000 scale land use mapping (OEH, 2007)

(ii) *Ground cover (Veg_tot)* – total vegetation cover (photo-synthetic and non-photo-synthetic) derived from CSIRO, 2011 MODIS fractional vegetation data (Guerschman et al. 2009).

Age: *Weathering index (W_I)* – as referred to above.
2.3.6 Model and map development and statistical analysis

Separate datasets were created for each soil property, each containing laboratory data for that property alone, plus the associated site covariate data. Twenty percent of points from each dataset were randomly extracted for validation purposes. Multiple linear regression (MLR), Random Forest (Breiman et al. 2015) and Cubist linear piecewise decision tree models (Kuhn et al. 2014; Quinlan 1992) were fitted on the training data using R statistical software (R Core Team 2015).

Simplified MLR models, using only key variables for each soil forming factor, were prepared for each soil property in order to provide quantitative data on the influence of lithology in the models. In these models, a lithology *silica index* (based on the silica percentage for the first eight lithology classes of Table 2.1) was the only parent material variable applied, apart from the weathering index that also represents the age factor. If all twelve lithology classes had been applied as nominal categorical variables in the MLR models they would act as twelve separate variables rather than just one as for the *silica index*, thus complicating the assessment of lithology influence in the models.

Random Forest models were developed using the entire covariate suite, then variable importance plots prepared to demonstrate the relative influence of the different covariables, including lithology and the geophysical parent material covariables. Lithology was applied as the twelve nominal classes. These plots were based on the percentage increase in mean square error (MSE) when the individual covariables are virtually omitted from the models (that is, replaced with random but still realistic values).

Random Forest and Cubist models were developed for each soil property using different combinations of parent material covariables ranging from (i) no parent material data; (ii) gamma radiometrics only; (iii) several geophysical derived data sources; (iv) lithology data only, then finally, (v) all sources (lithology and geophysical data). Maps were prepared over NSW for each soil property and combination of parent material covariables using the Cubist models together with the NSW covariable grids. Log values for SOC, CEC, sum-of-bases and P$_{\text{total}}$ were back-transformed onto their original scales. Both the models and maps for each soil property/covariable selection were validated using the validation datasets. Lin’s concordance correlation coefficient was used to measure the level of agreement of predicted values with observed values relative to the
Chapter 2: Lithology – soil relationships and their use in DSMM

1:1 line (Lin, 1989). Also determined were the coefficient of determination ($R^2$, for the training dataset only), root mean square error (RMSE) and mean error. These results allowed comparison of the effectiveness of the different parent material covariates in the original models and final maps.

2.4 Results

2.4.1 Quantitative influence of lithology in models

The model performance indicators in the simplified MLR models (5-15 cm) for each soil property from the NSW case study are presented in Table 2.4. In these models lithology (silica index) is the only parent material variable applied, apart from the weathering index as it also represents the age factor. The results demonstrate an overall moderate performance of the simplified MLR models, with the validation Lin’s concordance values varying between 0.45 and 0.73.

First approximations of the changes in each soil property per 10% silica change, and also 50% silica change (representing the near maximum, likely potential change) assuming other variables remain constant are also presented in Table 2.4. These are derived from the coefficients for the silica variable from the MLR models. The change is relative for those properties predicted in the log scale, but absolute for pH and sand. It shows, for example, that for a 10% rise in silica %, roughly equivalent to a change from an upper intermediate material like shale to a mid siliceous material like granite (see Table 2.1) there is a corresponding 22.0% relative decrease in SOC, and a 0.34 unit absolute decrease in pH. With the near maximum likely change of 50% silica, as in a change from a mafic material like basalt to an extremely siliceous material like quartz sand, there is a corresponding 71.1% decrease in SOC and a 1.7 unit decrease in pH. Note that for those properties predicted in the log scale, the changes for each additional 10% silica % change are not linear. Further such results were presented in Gray et al. (2015a) for soil property relationship over all of eastern Australia at 0-10, 10-30, and 30-100 cm depth intervals. Those results reveal that the rates of change remain relatively constant through the different depth intervals.
Table 2.4. MLR model validation and change in soil properties with 10% and 50% change in silica index

<table>
<thead>
<tr>
<th>Soil property (5-15 cm depth)</th>
<th>Model calibration $R^2$</th>
<th>Model validation concordance</th>
<th>Coefficient from model</th>
<th>Change per 10% increase in silica index</th>
<th>Change for 50% increase in silica index (maximum)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOC</td>
<td>0.35</td>
<td>0.53</td>
<td>-0.0248</td>
<td>-22.0% (relative)</td>
<td>-71.1% (relative)</td>
</tr>
<tr>
<td>pH</td>
<td>0.56</td>
<td>0.68</td>
<td>-0.0341</td>
<td>-0.34 pH units (absolute)</td>
<td>-1.71 pH units (absolute)</td>
</tr>
<tr>
<td>CEC</td>
<td>0.52</td>
<td>0.66</td>
<td>-0.0497</td>
<td>-39.2% (relative)</td>
<td>-91.7% (relative)</td>
</tr>
<tr>
<td>Sum-of-bases</td>
<td>0.57</td>
<td>0.73</td>
<td>-0.058</td>
<td>-44.1% (relative)</td>
<td>-94.5% (relative)</td>
</tr>
<tr>
<td>Pttotal</td>
<td>0.31</td>
<td>0.45</td>
<td>-0.0437</td>
<td>-35.4% (relative)</td>
<td>-88.8% (relative)</td>
</tr>
<tr>
<td>Clay</td>
<td>0.49</td>
<td>0.63</td>
<td>-0.0834</td>
<td>-8.3% (absolute)</td>
<td>-41.7% (absolute)</td>
</tr>
</tbody>
</table>

MLR: multiple linear regression

This approach as applied here over NSW demonstrates a possible avenue for deriving meaningful quantitative relationships between soil properties and lithology. As noted in section 2.2, the silica index as applied here has the drawback of not covering all lithology types, as it is invalid to apply it to calcareous, sesqui-oxide, organic and evaporite classes.

2.4.2 Relative influence of different parent material covariates in models

The dominant influence of lithology over other parent material covariates, and also all other environmental covariates for all studied soil properties over NSW, is demonstrated by the Random Forest variable importance plots of Figure 2.2. For all properties, lithology is clearly the most influential in terms of the percentage increase in mean square error (MSE) when each variable is virtually excluded. It is typically double that of the next highest ranked covariate of any type for all properties, with the exception of SOC and pH where it is only approximately 25% and 50% higher respectively than the next highest ranked covariates of maximum temperatures and rainfall.

The other geophysical parent material covariates such as gamma radiometrics or NIR clay components typically have considerably less influence. Lithology is generally 2 to 5 times more dominant than the next highest ranked parent material covariate, based on the percentage increase in MSE. Nevertheless, other parent material covariates still feature in the higher rankings of covariates, occupying two of the five top rankings.
for all properties except SOC and pH, which have only one in the top five. There is no clear pattern with respect to the order and degree of influence of other geophysical parent material covariates, with the next most important following lithology being gamma radiometrics for SOC, CEC and sum-of-bases; one of the NIR clays for pH and $P_{\text{total}}$; and weathering index for clay%.

Figure 2.2. Random Forest variable importance plots (based on the increase in mean square error (MSE %) with virtual omission of the variable)
2.4.3 Model performance with different parent material covariates

A comparison of model strength and validation results from both the Cubist and Random Forest techniques is presented in Table 2.5, showing model calibration $R^2$ and validation Lin’s concordance and RMSE. The results are always weakest when no parent material covariates are used, then typically get progressively stronger with the addition of gamma radiometrics, all geophysical continuous parent material covariates, lithology alone and then finally being strongest with all parent material covariates (all geophysical covariates and lithology). This trend is illustrated by Figure 2.3, which graphically presents the Lin’s concordances of the Cubist models.

There is normally always notable improvement (up to 30% in relative terms) in model performance for all properties when using any of the parent material covariates compared to no parent material covariates at all, for example, a rise in Lin’s concordance from 0.54 to 0.62, or 15%, for the Cubist CEC model. However, for SOC and pH the improvement is only marginal (<5% in relative terms) when only the geophysical covariates (radiometrics, NVIR clays and weathering index) are added, particularly with the Cubist approach.

There is typically further marked improvement in model performance with the use of lithology, even when used alone with no other geophysical parent material covariates. As would be expected, the highest performance is achieved when lithology and the continuous geophysical parent material covariates are applied together. The improvement relative to using only the geophysical covariates typically varies between 15-25% for Lin’s concordance, reaching a high of 35% for sum-of-bases (Cubist approach). The overall improvement relative to no parent material covariates at all was typically 15-45% for Lin’s concordance, reaching a high of 67% for sum-of-bases (Cubist approach). For example, Lin’s concordance for the Cubist sum-of-bases model rose from 0.46 with no parent material covariates, to 0.57 with all geophysical parent material covariates, to a high of 0.77 when lithology was also applied.
Table 2.5. Model strength and validation results for key soil properties over New South Wales (5-15 cm)

<table>
<thead>
<tr>
<th>Soil property</th>
<th>Parent material covariate</th>
<th>Training model $R^2$</th>
<th>Validation Lin’s concordance</th>
<th>RMSE</th>
<th>RF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Cubist</td>
<td>RF</td>
<td>Cubist</td>
<td>RF</td>
</tr>
<tr>
<td>SOC (log %)</td>
<td>No PM covariates</td>
<td>0.16</td>
<td>0.29</td>
<td>0.43</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>Radiometrics</td>
<td>0.16</td>
<td>0.33</td>
<td>0.44</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>All geophysical</td>
<td>0.16</td>
<td>0.32</td>
<td>0.46</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>Lithology</td>
<td>0.22</td>
<td>0.38</td>
<td>0.51</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>All PM covariates</td>
<td>0.24</td>
<td>0.38</td>
<td>0.54</td>
<td>0.55</td>
</tr>
<tr>
<td>pH</td>
<td>No PM covariates</td>
<td>0.48</td>
<td>0.62</td>
<td>0.67</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>Radiometrics</td>
<td>0.49</td>
<td>0.64</td>
<td>0.68</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>All geophysical</td>
<td>0.50</td>
<td>0.64</td>
<td>0.71</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>Lithology</td>
<td>0.56</td>
<td>0.68</td>
<td>0.72</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>All PM covariates</td>
<td>0.55</td>
<td>0.68</td>
<td>0.74</td>
<td>0.76</td>
</tr>
<tr>
<td>CEC (log cmol/kg)</td>
<td>No PM covariates</td>
<td>0.25</td>
<td>0.38</td>
<td>0.46</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>Radiometrics</td>
<td>0.29</td>
<td>0.46</td>
<td>0.53</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>All geophysical</td>
<td>0.38</td>
<td>0.46</td>
<td>0.61</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>Lithology</td>
<td>0.46</td>
<td>0.62</td>
<td>0.71</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>All PM covariates</td>
<td>0.49</td>
<td>0.63</td>
<td>0.72</td>
<td>0.75</td>
</tr>
<tr>
<td>Sum-of-bases (log cmol/kg)</td>
<td>No PM covariates</td>
<td>0.27</td>
<td>0.45</td>
<td>0.46</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>Radiometrics</td>
<td>0.38</td>
<td>0.53</td>
<td>0.56</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>All geophysical</td>
<td>0.38</td>
<td>0.53</td>
<td>0.57</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>Lithology</td>
<td>0.52</td>
<td>0.67</td>
<td>0.76</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>All PM covariates</td>
<td>0.56</td>
<td>0.68</td>
<td>0.77</td>
<td>0.80</td>
</tr>
<tr>
<td>P (total) (log mg/kg)</td>
<td>No PM covariates</td>
<td>0.32</td>
<td>0.37</td>
<td>0.52</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>Radiometrics</td>
<td>0.38</td>
<td>0.47</td>
<td>0.53</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>All geophysical</td>
<td>0.44</td>
<td>0.48</td>
<td>0.50</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>Lithology</td>
<td>0.42</td>
<td>0.54</td>
<td>0.54</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>All PM covariates</td>
<td>0.52</td>
<td>0.57</td>
<td>0.59</td>
<td>0.63</td>
</tr>
<tr>
<td>Clay (%)</td>
<td>No PM covariates</td>
<td>0.22</td>
<td>0.41</td>
<td>0.48</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>Radiometrics</td>
<td>0.31</td>
<td>0.47</td>
<td>0.54</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>All geophysical</td>
<td>0.40</td>
<td>0.47</td>
<td>0.56</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>Lithology</td>
<td>0.38</td>
<td>0.56</td>
<td>0.66</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>All PM covariates</td>
<td>0.46</td>
<td>0.57</td>
<td>0.70</td>
<td>0.73</td>
</tr>
</tbody>
</table>

RF: Random Forest; RMSE: root mean square error; PM: parent material
Chapter 2: Lithology – soil relationships and their use in DSMM

2.4.4 Map validation with different parent material covariates

Validation results of the final digital soil maps for each soil property over NSW (derived using the Cubist models) are presented in Table 2.6. The performance indicators are typically 10-20% weaker than those derived for the model validation, for example, for CEC, Lin’s concordance for model validation was 0.72, but for the map validation it was only 0.60.

The results reveal the expected pattern of being weakest where no parent material covariates were used, then progressively improving with the application of the geophysical continuous parent material covariates, and being strongest when lithology is also applied. The degree of improvement with the addition of lithology data is, however, slightly less in the final maps relative to the models, for all properties with the exception of $P_{\text{total}}$ (for which there were low map sample validation numbers). Notable examples of this were for sum-of-bases and clay, where the improvement in model performance (Lin’s concordance) was 35 and 25% respectively, but the improvement in map performance was only 18 and 9% respectively.

These results for map validation reflect the lower reliability of lithology data when using the broader state-wide geology and soil map grids, compared to the individual site data mostly used in the model training data, as further discussed later. All results
presented in this case study are influenced by the quality and precision of the source geological and other geophysical covariate data, and thus will vary between regions depending on this source data.

Table 2.6. Map validation results for key soil properties over NSW (5-15 cm interval, Cubist approach)

<table>
<thead>
<tr>
<th>Soil property</th>
<th>Parent material covariate</th>
<th>N</th>
<th>Lin’s concordance</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOC (log %)</td>
<td>No PM covariates</td>
<td>316</td>
<td>0.40</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>Radiometrics</td>
<td>315</td>
<td>0.38</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>All geophysical</td>
<td>315</td>
<td>0.39</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>Lithology</td>
<td>316</td>
<td>0.45</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>All PM covariates</td>
<td>316</td>
<td>0.40</td>
<td>0.72</td>
</tr>
<tr>
<td>pH</td>
<td>No PM covariates</td>
<td>610</td>
<td>0.62</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>Radiometrics</td>
<td>595</td>
<td>0.65</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>All geophysical</td>
<td>607</td>
<td>0.66</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>Lithology</td>
<td>631</td>
<td>0.68</td>
<td>0.79</td>
</tr>
<tr>
<td></td>
<td>All PM covariates</td>
<td>616</td>
<td>0.68</td>
<td>0.78</td>
</tr>
<tr>
<td>CEC (log cmol/kg)</td>
<td>No PM covariates</td>
<td>450</td>
<td>0.44</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>Radiometrics</td>
<td>437</td>
<td>0.41</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>All geophysical</td>
<td>435</td>
<td>0.50</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>Lithology</td>
<td>449</td>
<td>0.60</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>All PM covariates</td>
<td>436</td>
<td>0.60</td>
<td>0.73</td>
</tr>
<tr>
<td>Sum-of-bases (log cmol/kg)</td>
<td>No PM covariates</td>
<td>468</td>
<td>0.46</td>
<td>1.03</td>
</tr>
<tr>
<td></td>
<td>Radiometrics</td>
<td>446</td>
<td>0.50</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>All geophysical</td>
<td>439</td>
<td>0.56</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>Lithology</td>
<td>466</td>
<td>0.67</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>All PM covariates</td>
<td>457</td>
<td>0.66</td>
<td>0.80</td>
</tr>
<tr>
<td>P (total) (log mg/kg) *</td>
<td>No PM covariates</td>
<td>44</td>
<td>0.34</td>
<td>2.34</td>
</tr>
<tr>
<td></td>
<td>Radiometrics</td>
<td>44</td>
<td>0.18</td>
<td>1.89</td>
</tr>
<tr>
<td></td>
<td>All geophysical</td>
<td>44</td>
<td>0.58</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td>Lithology</td>
<td>44</td>
<td>0.57</td>
<td>1.19</td>
</tr>
<tr>
<td></td>
<td>All PM covariates</td>
<td>44</td>
<td>0.81</td>
<td>0.87</td>
</tr>
<tr>
<td>Clay (%)</td>
<td>No PM covariates</td>
<td>487</td>
<td>0.44</td>
<td>15.0</td>
</tr>
<tr>
<td></td>
<td>Radiometrics</td>
<td>494</td>
<td>0.42</td>
<td>15.2</td>
</tr>
<tr>
<td></td>
<td>All geophysical</td>
<td>475</td>
<td>0.54</td>
<td>14.0</td>
</tr>
<tr>
<td></td>
<td>Lithology</td>
<td>481</td>
<td>0.60</td>
<td>12.8</td>
</tr>
<tr>
<td></td>
<td>All PM covariates</td>
<td>493</td>
<td>0.59</td>
<td>12.6</td>
</tr>
</tbody>
</table>

RMSE: root mean square error; PM: parent material; * note low sample number
2.5 Discussion

2.5.1 Relationships between lithology and soil distribution

Our results from the NSW case study demonstrate a strong relationship between properly organised lithology and soil properties. Lithology was consistently shown to have the highest influence of all covariates in the models at this scale, coming well ahead of the geophysical parent material covariates and also all other environmental covariates including climate.

We consider that our proposed lithological classification scheme and the results from our NSW case study are, in principle, relevant to most other pedological contexts around the globe. Although relatively minor modification and addition to the lithological classification scheme may be appropriate for many other countries or regions, most of the scheme should be globally applicable. In DSMM programs, specific numerical values such as the precise variable importance ratings will vary depending on local conditions, scale of the study and data quality, but we believe our finding of the strong influence of lithology would be repeated more globally.

Our results over NSW parallel those the lead authors recently presented for models over all of eastern Australia in Gray et al. (2015a). However, at that scale climate was revealed to be the strongest driver for SOC, a result confirmed in Gray et al. (2015b), presumably due to the broader range of climate exhibited in the larger study area. In both those previous studies, the model strength was considerably greater (Lin’s concordance approximately 0.7) than was achieved here for the smaller area of NSW alone. Those previous studies also show climate to be approximately equivalent in influence to lithology for pH.

A semi-quantitative relationship between the silica index (or silica content) that applies to eight of the twelve lithology classes and the six soil properties has been demonstrated. First approximations of the change in level of each soil property per unit increase in silica content were provided, for example, a 0.34 unit absolute decrease in pH with a 10% increase in silica, assuming other factors remain constant. Similar estimates were made based on the maximum normal range in silica of 50%. Further exploration of these relationships is recommended, possibly using more advanced sensitivity analysis techniques.
The results clearly confirm broad trends of increasing mafic character (less siliceous character) being associated with higher SOC, pH, CEC, sum-of-bases, \( P_{\text{total}} \) and clay levels. Such trends are widely known from first principles of soil science and have been frequently reported; for all of these properties (Gray et al. 2009, 2015a) and more specifically for SOC (Badgery et al. 2013; Baldock et al. 2009; Gray et al. 2015b; Heckman et al. 2009), pH (Jaiyeoba 1995; Reuter et al. 2008), CEC and sum-of-bases (Cathcart et al. 2008; Jaiyeoba 1995) and clay (Barshad 1958; Birkeland 1999, Schaetzl and Anderson 2005). However, clear quantitative or semi-quantitative relationships appear to be rarely reported.

The basis of these strong relationships lies in the power of lithology to represent both chemistry and texture of a derivative soil, as outlined in section 2.2. It will provide a strong indication of the clay content, clay type and base cation content, particularly when set in the context of other soil-forming factors.

The *silica index*, as presented for eight of the twelve lithology classes, provides a potentially useful semi-quantitative index to account for relationships in the silica based parent materials. However, its non-applicability to calcareous, sesquioxide, organic and evaporite materials means it is not a universal parent material index and limits its application. It does go at least some way to addressing the problem noted by Yaalon (1975) that unless a numerical coding of parent materials is developed, litho-functions are likely to remain essentially qualitative. Badgery et al. (2013) also applied silica (derived from MIR spectroscopy) as a predictor of SOC in central NSW with at least moderate success, noting that it aggregates the influence of particle size and mineralogy on the protection of SOC better than the other variables examined.

Clear patterns of soil property variation typically emerge when lithology is considered in conjunction with other environmental variables. The lead authors recently demonstrated this in Gray et al. (2015b), where clear trends of increasing SOC stocks over eastern Australia were evident with more mafic (decreasing siliceous) character, in equivalent climatic and vegetation cover regimes. For example, within wet and high vegetation cover regimes, average SOC stocks in the 0-30 cm depth interval systematically varied from 55 Mg ha\(^{-1}\) in soils with highly siliceous parent materials, up to 145 Mg ha\(^{-1}\) in soils with mafic parent materials.
Chapter 2: Lithology – soil relationships and their use in DSMM

It is important to recognise that the influence of parent material diminishes with age of the landscape and weathering system, as the chemistry of the weathered material (saprolite) converges into more uniform material rich in aluminium, ferric iron and silica (Chesworth 1973). Such final equilibrium states, which Chesworth suggests are reached after one to several million years, typically relate to sesqui-oxides materials such as laterites and bauxites, and to siliceous sands, common in the ancient weathered landscapes of central and western Australia, Africa and South America, but they are less common in the younger landscapes of the northern hemisphere.

The influence of scale also needs to be recognised when considering the magnitude and relative influence of the different soil-forming factors (Grunwald 2005; Miller et al. 2015 and Miller and Schaetzl 2016). For example, climate typically has greater influence at coarser scales, while topography has greater influence at finer scales.

2.5.2 Use of lithology and geophysical covariate in DSMM

The results have demonstrated that parent material covariates are vital for the effective modelling and mapping of soil properties. Lithology is a powerful covariate providing it is systematically classified. The inclusion of reliable categorical lithology data as a covariate in our NSW case study, in conjunction with other continuous geophysical parent material covariates, substantially increased model and map performance for all soil properties, based on model calibration $R^2$ values, and validation concordance and RMSE results. For example, Lin’s concordance for the Cubist sum-of-bases model rose from 0.46 with no parent material covariates to 0.57 with the continuous geophysical covariates to a high of 0.77 when lithology was also applied.

We have observed that lithology and other geological data was applied in only 22% of our compilation of 267 DSMM studies derived from Arrouays et al. (2014), Grunwald (2009), McBratney et al. (2003) and Minasny et al. (2012, 2013). However, even in these studies it is likely that the lithology data were not suitably organised into meaningful classes, for example, being used in too broad classes or as simple stratigraphic units with poor relation to parent material composition. Soil type was used in 33% of the studies, but again, these data may not have been organised into meaningful classes to achieve optimal model and map quality. Based on our results, a
more widespread incorporation of lithology, derived from geological and/or soil data, into DSMM programs may be beneficial.

Lithology classes provide a useful means to understand and interpret patterns of variation and change in many soil properties, particularly when combined with other categorical variables, such as climate and land use, as was recently demonstrated for SOC (Gray et al. 2015b, 2016). Such clear patterns of variation are less evident when other geophysical continuous covariates such as gamma radiometrics or spectral imagery are applied.

The main weakness associated with use of lithology in DSMM is the lack of reliable, continuous grid data with which to prepare the final digital maps. Where reliable lithology data at each soil profile has been collected during field survey, which is common but should be further promoted (Juilleret et al. 2016), one can have high confidence in the model training data and thus the actual calibration model. However, for the production of final digital soil maps, one is generally dependent on polygon based geology and soil maps, frequently only at coarse scales, for which there is lower confidence. In Australia, mapping of geology and soils is rarely available at scales finer than 1: 250 000. This contrasts with geophysical parent material covariates, which are entirely continuous and often available in high spatial resolutions (100 m or finer).

In cases where lithology has been incorrectly mapped in the original source data, or is not recorded due to scale limitations, it can give misleading results in the final soil maps. For example, if a band of shale is not recorded within a larger unit of quartz sandstone, the predicted soil properties for that band will reflect the more siliceous sandstone, thus resulting in prediction error. Another potential limitation is the wide variation in composition of some common materials such as shales, which may vary from lower intermediate to mid siliceous, thus can be difficult to reliably classify.

The merit of lithology as a covariate in DSMM may be viewed as a balance between its inherent high predictive ability and its coarser spatial resolution, being typically based on coarse polygons. The problems with coarse resolution are however, at least partly addressed by inclusion of the other continuous geophysical parent material covariates, which can help predict the within-polygon variation in soil properties.
All geophysical techniques are, however, also subject to a number of limitations, these generally being more serious with remote sensed rather than proximally sensed data (McBratney et al. 2003; Mulder et al. 2011; Wilford and Minty 2007). These are at least partly attributable to the complexity of geophysical-soil relationships, even at field scale. Correlations of geophysical signals with soil properties may be weak or non-existent and particular soil properties do not always have unique geophysical signatures. The signals often only relate to surface properties of the top few millimetres. There may be high noise to signal ratios, brought about by coarse resolutions of the acquisition system and partial attenuation and interference of signals by vegetation, soil moisture, the atmosphere and topographic reflections. Differences in observational conditions such as differing intensity and direction of measurement also give rise to signal discrepancies.

It is evident that despite their benefits, all of these geophysical data sources have limitations that may temper their effectiveness in DSMM programs. Ideally, uncertainty data should be attached to the geophysical data surfaces. Mulder et al. (2011) report that the feasibility of using the wide range of spectrometry techniques for soil survey is generally low to medium for remotely sensed applications, but medium to high for proximally sensed applications. Used individually, geophysical covariates may not always be effective in digital soil models and maps, however, when used in combination with other parent material and environmental covariates, they have the potential to be highly effective (Mulder et al. 2011).

In their study on the comparative effectiveness of lesser and greater detailed covariates in the digital soil mapping of three soil properties in Brazil, Samuel-Rosa et al. (2015) demonstrated that the use of more detailed geophysical covariates resulted in only a modest increase in prediction performance and may not outweigh the extra costs of using them. They concluded it may be more efficient to spend extra resources on increasing the detail of only those covariates that have the strongest improvement effect. They suggested this may possibly include a more detailed geological map for two of the three properties.

Establishing clear relationships between geophysical data and the twelve lithology classes may be a valuable research endeavour. The use of the increasingly available hyper-spectral imagery in this way offers some particularly exciting prospects. Such
relationships may prove more robust than relationships with individual soil properties or soil types.

In summary, it is evident that both categorical polygonal lithological data and continuous geophysical data both have their merits and limitations. Best results in soil modelling and mapping programs are generally derived by combining these two data sources.

2.5.3 Suggestion for use of lithology in DSMM

We have demonstrated the strong influence of lithology in controlling soil distribution and that it is a potentially important covariate in DSMM over NSW, Australia. It is typically strongest in model development where reliable lithology data are available with training data, but is less strong for final map production where only coarse scale polygonal data is normally available. The finer the scale and more reliable the lithology data, for both training data and final mapping grid layers, the more effective it will be. It is however necessary to properly organise the available parent material data into suitable classes to achieve its maximum potential.

A possible broad process for applying lithology class data in DSMM programs more broadly is as follows:

i. include parent material and soil type descriptions recorded in field site data in the initial training datasets
ii. acquire the best, finest scale geological and soil maps available over the study area
iii. for both training data and final map grids, classify all parent material data (site data or geological/soil units from a map) into the 12 lithology groups based on Table 2.1, or a modified version as appropriate for the study region.
iv. where a combination of parent materials is present, determine the best overall single class, which may be an average of several classes
v. check the derived lithology class against the available soil site data and maps; ensure lithology and soil type are at least broadly consistent, for example, WRB Vertisols should coincide with mafic or lower intermediate lithology classes, and Arenosols should coincide with highly or extremely siliceous lithology class. Use Tables 2.1 and 2.3 to assist in this process. Where there is
a discrepancy between the two sources, lean towards the product with finer scale and greater reliability.

vi. apply the final lithology data in a digital soil model. This would normally be as nominal categorical covariates if most of the 12 lithology classes are present, but they may be applied as ordinal classes or silica indices if only the silica based classes 1 to 8 are present. Other available parent material related covariates (such as radiometric and spectral imagery data) and other environmental covariates are also applied.

vii. undertake variable selection by assessing the relative performance of the different parent material and other covariates in both the models and resulting maps.

viii. prepare the final digital soil maps using the best models that either include or exclude lithology data.

There are a number of issues to be considered when making decisions on the use of categorical lithology data in DSMM programs. The suitability of available lithology data should be assessed against the proposed scale and purpose of the final map product. Coarse scale lithology data may be suitable where a coarse scale final map product is proposed, for example to guide delineation of broad land use zones. However, for finer scale map products, such as for precision agriculture, fine scale lithology data may be necessary. For example, if reliability at field scale down to 30 m or finer is critical then inclusion of lithology data derived from a 1: 250 000 scale geological map may not be appropriate. The importance of transparency and ability to readily interpret the final model or map results should also be considered. Lithology can be more readily understood and related to known features on the ground than is normally possible with geophysical covariates, which can aid in the interpretation and application of final results.

2.6 Conclusion

This study has demonstrated clear and strong relationships between parent material lithology and six key soil properties, and added some quantitative detail to the current mainly qualitative knowledge in this area of pedology. The organisation of lithology data into twelve classes, based on mineralogical and chemical composition, was instrumental in revealing the relationships in our NSW case study. The proposed
lithology classification scheme is considered to be applicable globally, but minor adaptations may be appropriate for different countries or regions. It has been possible to make quantitative estimates of the change in individual soil properties as the lithology class changes, at least for the eight silica based classes over NSW.

The inclusion of organised lithology data is shown to greatly assist in the digital modelling and mapping of soil properties, providing it is available at a suitable scale and properly organised. It was revealed to exert the greatest influence on six key soil properties in models prepared over NSW, coming well ahead of all other environmental and geophysical parent material covariates. Despite the potential drawbacks of using polygonal data, it is evident that properly organised categorical lithology data can be a strong covariate in all soil modelling and mapping programs and be a valuable complement to other continuous geophysical parent material covariates. Research aimed at establishing clear relationships between geophysical data and the twelve lithology classes would be useful. The inclusion of this readily understood covariate can facilitate the interpretation of soil models and maps and ultimately allow further insights into factors controlling the distribution of soil properties.

2.7 References


Australia with the potential to enhance soil carbon content. CSIRO Sustainable Agriculture Report, CSIRO, Canberra.


Dokuchaev VV, 1899, On the Theory of Natural Zones, St. Petersburg, Russia.


44


Geological Survey of NSW, undated. 1:250 000 scale geology map of NSW. 


Chapter 2: Lithology – soil relationships and their use in DSMM


Jensen HI, 1914. The Soils of New South Wales. NSW Department of Agriculture, Sydney.

Joffe JS, 1936. Pedology. Rutgers University. Press, New Brunswick, New Jersey, USA


AI ’92: Proceedings of the 5th Australian Joint Conference on Artificial 


mapping using European soil profile data: soil pH. In ‘SAGA – Seconds Out’. 
Hamburger Beiträge zur Physischen Geographie und Landschaftsökologie. 
Hamburger Beiträge zur Physischen Geographie und Landschaftsökologie, Vol. 
19. (Eds J Böhner, J Blaschke, L Montanarella) pp. 91–102. (Universität 
Hamburg, Institut für Geographie, Hamburg.

Ryan PJ, McKenzie NJ, O’Connell DO, Loughhead AN, Leppert PM, Jacquier D, 
Ashton L, 2000, Integrating forest soils information across scales: spatial 
prediction of soil properties under Australian forests. Forest Ecology and 
Management 138, 139-157.

environmental covariates deliver more accurate soil maps? Geoderma 243–244, 
214–227.

Sanchez PA, Ahamed S, Carré F, Hartemink A, Hempel J, Huising J, Lagacherie 
P,McBratney A, McKenzie N, de Lourdes Mendonça-Santos M, Minasny B, 
Montanarella L, Okoth P, Palm C, Sachs J, Shepherd K, Vågen T, Vanlauwe 

University Press, New York.

Soil Survey Staff (2010). Keys to Soil Taxonomy, 11th ed. USDA Natural Resources 
Conservation Service. Washington DC.

relationships between soil properties and high resolution radiometrics, central 
eastern Wheat belt, Western Australia. Exploration Geophysics. 33, 95–102.

TERN, 2014. The Soil and Landscape Grid of Australia. Terrestrial Ecosystem 


Chapter 3: **Pragmatic models for the prediction and digital mapping of soil properties in eastern Australia**

**Abstract**

To help meet the increasing need for knowledge and data on the spatial distribution of soils, readily applied multiple linear regression models were developed for key soil properties over eastern Australia. Selected covariates were used to represent the key soil-forming factors of climate (annual precipitation and maximum temperature), parent material (a lithological silica index) topography (new topo-slope and aspect indices) and biota (a modified land disturbance index).

The models are presented at three depth intervals (0-10, 10-30 and 30-100 cm) and are of variable but generally moderate statistical strength, with concordance correlation coefficients in the order of 0.7 for organic carbon (OC) upper depth, pH_{ca}, sum-of-bases, cation exchange capacity (CEC) and sand, but somewhat lower (0.4-0.6) for OC lower depths, total phosphorous, clay and silt. The pragmatic models facilitate soil property predictions at individual sites using only climate and field collected data. They were also moderately effective for deriving digital soil maps over the State of New South Wales and a regional catchment. The models and derived maps compared well in predictive ability to those derived from more sophisticated techniques involving Cubist decision trees with remotely sensed covariates. The readily understood and interpreted nature of these products means they may provide a useful introduction to the more advanced digital soil modelling and mapping (DSMM) techniques. The models provide useful information and broader insights into the factors controlling soil distribution in eastern Australia and beyond, including the change in a soil property with a given unit change in a covariate.

**Keywords:** regression models, digital soil maps, soil formation, organic carbon, pH, bases, particle sizes
3.1 Introduction

The importance of soils and the need for their protection was recently well articulated by the then Australian Prime Minister, Julia Gillard:

*Soil is the very basis of our survival. Clean air and water; food and fibre; and our unique biodiversity all rely on protecting our soil* (Gillard 2012).

Gaining information on the distribution of soils across the landscape is fundamental to the protection of this vital asset. Soil information is becoming increasingly important in Australia and globally for the understanding and modelling of a range of earth and environmental science systems. It is an essential input into developing sustainable soil and land management systems that are vital for global soil and food security. It is key to carbon sequestration issues and associated climate change modelling; and also ecological, hydrological and other modelling systems. The prohibitive time and resource requirements of traditional soil data collection mean alternative modelling and soil mapping strategies must be investigated as a means to collect the required soil information. This has led to the rise of digital soil modelling and mapping (DSMM) techniques over the last two decades, made possible through the enormous advances in computer, database mining and geospatial technologies. In this paper we explore the development of pragmatic, easy to apply and interpret DSMM techniques, which may complement other more widely used and complex techniques.

3.1.1 Development of quantitative soil-environment models

Despite the recent advances in DSMM there are still no clearly articulated, easily interpreted, universal quantitative models to describe the occurrence and distribution of most important soil properties. The fundamental soil equation (Dokuchaev 1899; Jenny 1941) has still not been effectively solved with full numeric coefficients, at least at broad regional, national or universal levels. Up until the last decades of the 1900s most soil-environment relationships were generally relatively simple mono- or bi-variate pedofunctions, such as “climo-functions” (Jenny 1980; Webb *et al.* 1986) or “topo-functions” (Furley 1971). These rarely considered the combined influence of multiple variables and were restricted to particular regions. Few relationships involving parent material were published. Even recently it has been lamented that most information on
the role of the soil-forming factors and processes is of a qualitative, conceptual form rather than quantitative form (Heuvelink 2005).

In recent years, more sophisticated quantitative soil-environmental models, involving the use of remote sensed covariates such as gamma radiometrics and satellite derived vegetation indices, have been prepared for a range of soil properties including OC, pH, EC, exchangeable bases, phosphorous, particle size and others by several workers (Rasmussen et al. 2005; Sumfleth and Duttman 2008; Wang et al. 2008; Kroulik et al. 2010; Guo et al. 2011; Webster et al. 2011; Shi et al. 2012 and Hattar et al. 2010). Such quantitative models have also been developed in the initial modelling component of digital soil mapping (DSM) projects (McBratney et al. 2003). However, the typical complexity of most of these models means that they cannot be readily interpreted to gain pedologic knowledge such as the precise influence of individual soil-forming factors. They are not readily applicable for deriving predictions of soil properties at individual sites using readily available data, such as that collected in the field. There appears to be only few attempts to develop widely applicable and pragmatic quantitative relationships between soil properties and the combined multiple soil-forming factors (Gray et al. 2009; Phachomphon et al. 2010).

3.1.2 Use of models for digital soil mapping

The major use of quantitative soil-environment models is the production of digital soil maps. In Australia and around the world, the application of DSMM techniques appears to be gaining wider acceptance in recent years, with active DSMM projects being particularly driven through the GlobalSoilMap.net project (Sanchez et al. 2009; GlobalSoilMap.net 2011; Hempel et al. 2012) and in Australia through the TERN (Terrestrial Ecosystem Research Network) program (Grundy et al. 2015). Recent projects in Australia, carried out using sophisticated techniques and covariates, include ones at national scale (Henderson et al. 2005; Bui et al. 2009; Viscarra Rossel et al. 2014), at State and regional scale (Wheeler et al. 2011; Holmes et al. 2014; Hopley et al. 2014; Kidd et al. 2014; Liddicoat et al. 2014; Odgers et al. 2015) and at local or field scale (Malone et al. 2009; Triantafilis et al. 2009).

Despite the increasing acceptance of DSMM in Australia and worldwide, some reluctance to adopt the techniques appears to remain amongst at least some soil scientists (Hartemink et al. 2008; Hempel et al. 2008; Moore et al. 2010) perhaps due to
a lack of understanding of the processes behind many of these techniques, and a lack of experienced mentors in the field. The presentation of DSMM techniques and products in a more readily understood and transparent manner may serve as a useful introduction to the more advanced DSMM products and thereby promote their greater acceptance amongst the wider soil science community.

### 3.1.3 Aims

The presentation of such readily applied and understood DSMM products is the major aim of this paper. More specific aims are to:

- develop relatively straightforward and easily applied multiple linear regression models between major soil-forming factors and a number of important soil properties (including organic carbon (OC), pH, base content, cation exchange capacity (CEC), particle sizes and total phosphorous) using a large soil profile dataset covering eastern Australia. Three depth intervals are to be examined: 0-10 cm, 10-30 cm and 30-100 cm. These models are intended to facilitate the prediction of soil properties at specific sites in this province with readily available field data, without relying on complex data sources and computer systems as in most currently applied DSMM strategies.
- use the pragmatic models to produce digital soil maps over the State of New South Wales (NSW) and the Hunter region, which, with the underlying covariate layers, can be readily understood and interpreted by potential users.
- compare the effectiveness of these pragmatic models and maps with more sophisticated models and maps derived from the Cubist piecewise linear decision tree technique and a range of remotely sensed covariates, to help gauge the potential worth of the pragmatic approach.
- use the models to gain quantitative information on and broader insights into the factors controlling soil distribution in eastern Australia and beyond, including the relative influence of each covariate on each soil property and the quantitative change in a soil property with a given unit change in a covariate.
3.2 Methods

The overall strategy was to prepare multiple linear regression models with associated statistics using a large dataset with laboratory data for key soil properties and readily available covariates covering eastern Australia. The models were validated against data points initially withheld from the analysis. The models were then used in conjunction with spatial grids representing each covariate to prepare digital soil maps for each soil property over the State of NSW and then more specifically over the Hunter region, which were also validated using the withheld dataset. Finally, these models and maps were compared with more sophisticated products derived from the Cubist piecewise linear decision tree technique and a range of remotely sensed covariates.

3.2.1 Overview of study area

The eastern Australia province over which the quantitative models are prepared takes in the states of Queensland, NSW, Victoria, South Australia and Tasmania, plus the small Australian Capital Territory (Figure 3.1). It extends over some 3700 km² in the north-south dimension, and encompasses a total area 3.8 million km². Covering such a large area it takes in a wide range of environments.

Climate varies from equatorial in the far north, to hot arid in the western areas, cool temperate in the south and sub alpine in the highlands of NSW and Victoria. Mean annual maximum daily temperatures range from 10 to 35 degrees C, while rainfall varies from less than 200 mm to over 3300 mm per annum. The physiography of the region is marked by a range of mountains, the Great Dividing Range, that runs down the entire mainland east coast (generally 100-300 km inland) before swinging around and fading out in western Victoria. This mountain range is low by world standards, only reaching a maximum of 2200 m in southern NSW. Heading further west from this range the undulating terrain gives way to flat inland plains.

Surface geology of the region is characterised by Paleozoic and Mesozoic siliceous and intermediate igneous and sedimentary rocks in the higher relief eastern regions, with Tertiary alluvial sands, silts and clays occupying most of the flatter western regions. Remnants of tertiary age mafic volcanics are widespread throughout much of the higher relief eastern areas. The limited extent of Holocene glaciation over the continent means most landscapes are relatively old with strongly weathered
materials, particularly in the flat inland regions. There is great diversity of land use over the region, with intensity of use typically following climate, topography and soil fertility gradients. Major land uses range from intensive cropping, horticulture, grazing, plantation forestry, native forestry to environmental protection reserves.

The state of NSW, over which digital soil maps are here prepared, occupies a middle region of eastern Australia and encompasses an area of 800 000 km$^2$. The character and range of environmental conditions is similar to that of the broader region, however, the maximum temperatures and rainfall levels are somewhat lower.

3.2.2 Soil profile dataset

Soil profile datasets over eastern Australia were acquired from the five State government soil resource agencies, based on their 2011 data holdings, plus the Federal Government’s CSIRO data from 2001. These included data collected back to the 1960s and earlier. Together they effectively represent an updated version of the 2001 point dataset of the Australian Soil Resource Information System (ASRIS, eastern states component) (Johnston et al. 2003) and a precursor to part of the Australian soil point dataset established for the TERN program (Searle 2014).

Only those profiles with laboratory data, plus parent material and topographic descriptors that could be reliably classified were used for the final analysis. The final dataset contained profile numbers as follows: Queensland (2283), NSW (3318), Victoria (402), Tasmania (501), South Australia (752) and CSIRO (eastern States, 1147), amounting to a total of 8403 profiles; however sample numbers varied significantly for the different properties. Figure 1 presents the sample location points across the five States.

3.2.3 Soil properties

A total of eight soil properties were examined. These are listed in Table 3.1, together with the laboratory test methods used in their derivation and final sample numbers (with the required environmental covariates). These properties are essential for effective climatic, agricultural, hydrological, ecological and/or other scientific modelling. The properties of organic carbon (OC), pH$_{ca}$, sum-of-bases, cation exchange capacity (CEC) and total phosphorous (P) are indicators of a soil’s chemical condition, its nutrient status and potential to retain nutrients. The storage of carbon in soil is
considered vital in addressing global climate change (Lal 2004). The clay, silt and sand content plus the bulk density of a soil control its texture and physical behaviour, including water holding capacity and permeability, and also influence many chemical characteristics.

Figure 3.1. Location of profile points (shading represents effective modelling area)

Table 3.1. Soil properties: laboratory methods and sample numbers

<table>
<thead>
<tr>
<th>Soil property</th>
<th>Units</th>
<th>Laboratory method with test number from Rayment and Lyons (2011)</th>
<th>Sample number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organic carbon</td>
<td>%, kg/m³</td>
<td>Walkley-Black wet oxidation method (6A1, approx 95%), LECO and other combustion methods.</td>
<td>5825</td>
</tr>
<tr>
<td>pHₐ</td>
<td>pH units</td>
<td>pH of 1:5 soil/0.01M calcium chloride extract (4B1, 4B2). Includes conversions from pH 1:5 soil/water suspension (4A1)</td>
<td>7682</td>
</tr>
<tr>
<td>Sum-of-bases³</td>
<td>cmol/kg</td>
<td>Various methods (15A – 15F)</td>
<td>6277</td>
</tr>
<tr>
<td>CEC</td>
<td>cmol/kg</td>
<td>Various methods (15A – 15F)</td>
<td>4315</td>
</tr>
<tr>
<td>P total</td>
<td>mg/kg</td>
<td>X-ray fluorescence (9A1), sodium carbonate fusion (9A2) and semi-micro kjeldahl, automated colour (9A3)</td>
<td>1837</td>
</tr>
<tr>
<td>Clay</td>
<td>%</td>
<td>Various methods including pipette, hydrometer and plummet balance (P10)</td>
<td>5253</td>
</tr>
<tr>
<td>Sand</td>
<td>%</td>
<td>As above</td>
<td>5215</td>
</tr>
<tr>
<td>Silt</td>
<td>%</td>
<td>As above</td>
<td>5253</td>
</tr>
</tbody>
</table>

¹ Ca, Mg, Na and K
The variation in different laboratory methods for the same soil property, due to the different dates and jurisdictions of the analyses, results in a degree of inconsistency in the test results. This presents a source of potential error in the predictive models. The Walkley-Black method has been reported to underestimate total OC levels (Skjemstad 2000, Bui et al. 2009), but no correction factor was applied for this. Final analysis excluded samples with less than 0.1% OC, as these were considered unreliable and generally not true soil material, and greater than 18% OC, as these are always defined as organic materials in the Australian Soil Classification (Isbell 2002). Organic carbon mass (kg/m³) was also derived, following the determination of bulk density that was derived by applying a variant of the pedo-transfer function reported in Tranter et al. (2007) and Minasny et al. (2013):

\[
BD = \frac{100}{(1.724\times OC\%/224 + (100-1.724\times OC\%)/(1.351 + 0.0045 \times Sand + (Sand - 44.65)^2 \times -0.0000614 + 0.0596 \times \log\text{(depth)}))}
\]

To avoid reporting two separate pH test results, pH\text{w} values were converted into pH\text{ca} values using the correlation tables of Henderson and Bui (2002). The latter mode is preferred in Australia as it more closely represents the ionic soil solutions typically found in the field, and thus gives more consistent results.

3.2.4 Depth intervals

Models were developed over numerous depth intervals, but are only presented for the 0-10, 10-30 and 30-100 cm intervals. These are intervals commonly used in NSW and Australian monitoring evaluation and reporting programs, thus have advantages with respect to ongoing data validation and comparison. Preliminary models for depth intervals as required for the global digital soil mapping project (GlobalSoil.Net 2011) and its Australian TERN component (Grundy et al. 2015), that is, 0-5, 5-15, 15-30, 30-60, 60-100 and 100-200 cm were also derived but are not presented here. Soil property values reported for the original depth interval of each soil horizon were converted into the standard depth intervals using the equal area spline process of Bishop et al. (1999) and Malone et al. (2009).

3.2.5 Covariates

Covariates were selected to effectively represent each of the key soil-forming factors of climate, parent material, relief and biota, as outlined below. A prerequisite
was that they be readily understood and easily applied into the final models and maps by non-statistical experts at individual sites based only on climate and field observations. This facilitates the use of the models by field officers in various natural resource sciences that rely on soil data. The grids for each covariate layer used in final map production are available from the authors.

**Climate**

a) Mean annual precipitation (Precip, mm pa) – values were derived from 5 km Australia wide climate grids obtained from the Australian Bureau of Meteorology (BoM). They represent mean values obtained over the 1961-1990 period. For the final NSW map preparation, cell values were interpolated from the original grid size down to a 100 m grid, based on adjoining cell values, using the ArcGIS interpolation spline tool (i.e., “fine-gridded”).

b) Mean annual maximum daily temperature (T_max, °C) – values were obtained from equivalent BoM grids to precipitation above, and likewise interpolated down to a 100 m grid. Another climate covariate trialled was annual evapotranspiration, along with various combinations of this with Precip and T_max, but it proved less effective and was not adopted.

**Parent material**

The basis of this covariate was the lithology of the parent material, and more specifically the silica content (%) which is applied as a silica index. Silica content provides a meaningful quantitative estimation of the chemical composition of most parent materials. It generally has a direct relationship to quartz content and an inverse relationship with basic cation content (Gray and Murphy 1999; Gray et al. 2014) as shown in Table 3.2. Note that calcareous and some other lithology types cannot be characterized in any meaningful way by their silica content, meaning these materials need to be treated differently in the modelling process.

For the model development, the description of parent material or geologic unit recorded at each site by the soil surveyor was used. For the final NSW and Hunter region map preparation, lithological classes and silica index values were applied manually to each geological formation as identified in the 1:250 000 scale digital geology map of the Geological Survey of NSW (NSW Department of Resources and Energy, undated). These were based on their lithological descriptions, using the rules
presented in Table 3.2. For poorly defined Cainozoic unconsolidated material, such as unqualified “alluvium” or “colluvium” for which their broad composition is unknown, lithological classes were allocated following reference to existing soil type maps. This exploited clear soil type to parent material relationships, such as Black Vertosols (WRB Vertisols) with the mafic class and highly sandy Tenosols or Rudosols (WRB Arenosols) with the upper siliceous class.

Table 3.2. Parent material classes and Silica index

<table>
<thead>
<tr>
<th>Parent material class</th>
<th>Silica index</th>
<th>Bases</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(average</td>
<td>(%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>silica, %)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Siliceous - extreme</td>
<td>86</td>
<td>0.4</td>
<td>quartz sands (alluvial and aeolian), pure quartzite, chert, jasper, quartz reefs</td>
</tr>
<tr>
<td>Siliceous - high</td>
<td>80</td>
<td>7</td>
<td>quartz sandstone, quartz siltstone, unqualified quartzite, unqualified sands</td>
</tr>
<tr>
<td>Siliceous - mid</td>
<td>73</td>
<td>10</td>
<td>granite, rhyolite, siliceous tuff, unqualified sandstone</td>
</tr>
<tr>
<td>Siliceous - lower-mid</td>
<td>69</td>
<td>11</td>
<td>adamellite, monzogranite, unqualified siltstone</td>
</tr>
<tr>
<td>Siliceous/Intermediate transition</td>
<td>66</td>
<td>12</td>
<td>granodiorite, tonalite, quartz diorite, dacite, trachyte, syenite, greywacke, feldspathic or lithic sandstone, mudstone</td>
</tr>
<tr>
<td>Intermediate - mid</td>
<td>62</td>
<td>~13</td>
<td>monzonite, trachy-andesite, argillaceous sediments (mudstone, claystone, shale, slate, phyllite), clay, gneiss, schist, unqualified loess</td>
</tr>
<tr>
<td>Intermediate - lower</td>
<td>57</td>
<td>15</td>
<td>diorite, andesite, low-quartz tuff</td>
</tr>
<tr>
<td>Mafic</td>
<td>48</td>
<td>20</td>
<td>gabbro, dolerite, basalt, amphibolite, alluvial black cracking clay</td>
</tr>
<tr>
<td>Ultramafic</td>
<td>42</td>
<td>37</td>
<td>serpentine, greenschist, dunite, peridotite</td>
</tr>
<tr>
<td>Calcareous</td>
<td>na¹</td>
<td>na</td>
<td>limestone, dolomite, calcareous shale and sands</td>
</tr>
<tr>
<td>Other</td>
<td>na</td>
<td>na</td>
<td>sesqui-oxide (laterite, bauxite), organic material, evaporites, unqualified alluvium</td>
</tr>
</tbody>
</table>

¹ Approximate compositions from Best (1982), Duff (1993) and Joplin (1963,1965)
² Bases: average Ca, Mg, Na and K³ na – variable and not applicable

Relief

a) a new topo-slope index (TSI) that combines topographic position and slope gradient was developed to represent this factor. It is similar in concept to the compound topographic index (CTI) and topographic wetness index (TWI). The simple 1 to 6 index attempts to represent the extent to which a site is subject to either depletion or accumulation of water, soil particles and chemical materials. It can be readily derived from limited physiographic information in the field or office, with or without DEM data, using the rule set outlined in Table 3.3. A broad visual representation is given in
Chapter 3: Pragmatic models for soil prediction in eastern Australia

Figure 3.2. The topographic position and slope gradient data used to derive the index for the initial model development were derived from site descriptions recorded by the soil surveyor for each profile. For final NSW and Hunter region map preparation the index was derived from a 100 m digital elevation model (DEM), which was used to determine the slope% and Topographic Position Index, an existing index that identifies position in the landscape (Jenness 2006).

Table 3.3. Rules used to define the topo-slope index

<table>
<thead>
<tr>
<th>Topographic position</th>
<th>Slope gradient (%)</th>
<th>Topo-slope index</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crest, upper and mid slopes</td>
<td>&gt;10</td>
<td>1</td>
<td>Highly depletive</td>
</tr>
<tr>
<td></td>
<td>3-10</td>
<td>2</td>
<td>Moderately depletive</td>
</tr>
<tr>
<td></td>
<td>&lt;=3</td>
<td>3</td>
<td>Weakly depletive</td>
</tr>
<tr>
<td></td>
<td>unknown</td>
<td>2</td>
<td>Moderately depletive</td>
</tr>
<tr>
<td>Lower and foot slopes</td>
<td>&gt;20</td>
<td>1</td>
<td>Highly depletive</td>
</tr>
<tr>
<td></td>
<td>10-20</td>
<td>2</td>
<td>Moderately depletive</td>
</tr>
<tr>
<td></td>
<td>3-10</td>
<td>4</td>
<td>Weakly accumulative</td>
</tr>
<tr>
<td></td>
<td>&lt;=3</td>
<td>5</td>
<td>Moderately accumulative</td>
</tr>
<tr>
<td></td>
<td>unknown</td>
<td>4</td>
<td>Weakly accumulative</td>
</tr>
<tr>
<td>Valley floor, plain</td>
<td>&gt;20</td>
<td>1</td>
<td>Highly depletive</td>
</tr>
<tr>
<td></td>
<td>10-20</td>
<td>3</td>
<td>Weakly depletive</td>
</tr>
<tr>
<td></td>
<td>3-10</td>
<td>5</td>
<td>Moderately accumulative</td>
</tr>
<tr>
<td></td>
<td>&lt;=3</td>
<td>6</td>
<td>Strongly accumulative</td>
</tr>
<tr>
<td></td>
<td>unknown</td>
<td>5</td>
<td>Moderately accumulative</td>
</tr>
<tr>
<td>Unknown position</td>
<td>&gt;10</td>
<td>1</td>
<td>Highly depletive</td>
</tr>
<tr>
<td></td>
<td>3-10</td>
<td>3</td>
<td>Weakly depletive</td>
</tr>
<tr>
<td></td>
<td>&lt;=3</td>
<td>5</td>
<td>Moderately accumulative</td>
</tr>
</tbody>
</table>

Figure 3.2. The *topo-slope index*

b) Aspect - a new 1 to 10 *aspect index* (*Asp*) was developed to represent the influence of aspect as shown in Table 3.4. Sites that receive high solar radiation such as on gentle slopes and those facing north and north-west have low indices, whilst sites
that receive low solar radiation such as those with steep south and south-easterly facing slopes have high indices. In the model development stage, aspect and slope gradient data recorded by the soil surveyor for each site was used to derive the index, however, for final map preparation it was derived from a 100m DEM.

Other relief variables were trialled including slope%, log\(_e\)(slope%) and an index representing relative position on the catena, but they proved less effective and were not adopted.

Table 3.4. The aspect index

<table>
<thead>
<tr>
<th>Slope %</th>
<th>N</th>
<th>NE</th>
<th>E</th>
<th>SE</th>
<th>S</th>
<th>SW</th>
<th>W</th>
<th>NW</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;10</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10-30</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>&gt;30</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>10</td>
<td>8</td>
<td>6</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

Biota

This factor is represented by a land disturbance index (LDI) that reflects the intensity of disturbance associated with the land use at a site. The 1 to 6 index as shown in Table 3.5 is a slight modification of the site disturbance table presented in the *Australian Soil and Land Survey Field Handbook* (NCST 2009). For model development, site land use was taken from profile descriptions, or where this was not recorded, from a 2006 Australia wide digital land use map (1 km grid), downloaded from the Australian government ACLUMP (2010) website. For the final NSW and Hunter region map, the LDI was derived from 1:25 000 scale polygonal land use mapping (OEH 2007).

Table 3.5. The land disturbance index

<table>
<thead>
<tr>
<th>Land disturbance index</th>
<th>Description and typical land uses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No effective disturbance, eg, national park, nature reserve</td>
</tr>
<tr>
<td>2</td>
<td>Limited disturbance, minor native vegetation clearing, eg, selective</td>
</tr>
<tr>
<td></td>
<td>logging, production forestry</td>
</tr>
<tr>
<td>3</td>
<td>Moderate disturbance, moderate native vegetation clearing, light grazing in woodland, hardwood plantation</td>
</tr>
<tr>
<td>4</td>
<td>High disturbance, complete native vegetation clearing, eg, native and improved pasture, softwood plantation</td>
</tr>
<tr>
<td>5</td>
<td>Very high disturbance, eg, improved pasture with moderate cropping, orchards, viticulture</td>
</tr>
<tr>
<td>6</td>
<td>Extreme disturbance, predominant cropping (rain fed or irrigated)</td>
</tr>
</tbody>
</table>
3.2.6 Statistical analysis and validation of models

Separate datasets were created for each soil property, each containing laboratory data for that property alone, plus the associated site covariate data. Fifteen percent of points from each dataset were extracted for validation purposes, following stratification by State, this being a relatively low proportion so as to maximize numbers in the remaining training data. Using the training data, multiple linear regression (MLR) models were fitted for the soil property based on the above mentioned covariates using R statistical software (R Core Team 2013) and a standard suite of associated statistics derived. Covariates were rejected from the relationship where the p value rose above 0.075. A natural log transformation was applied to many of the properties to address the observed skewness in model predictions.

Validation using the 15% of withheld data points involved the determination of root mean square error (RMSE), mean error and median absolute error. Plots of observed versus predicted values, with associated statistics, were prepared. Lin’s concordance correlation coefficient (CCC) was used to measure the level of agreement of predicted values with observed values, i.e., a 1:1 relationship (Lin 1989).

Soils derived from calcareous parent materials and those listed as “other” in Table 3.2 require separate treatment from those derived from the silica-based parent materials as they cannot be well characterized by the silica index, so a different suite of models was derived for each of these groups. As these materials are of restricted occurrence in eastern Australia these models are not presented here, but are available from the authors on request.

Regression models, their coefficients and accompanying statistics such as standardised regression coefficients and t values were examined to derive quantitative data on the relative roles and behaviour of each covariate in the model, with the aim of gaining pedologic insights into soil formation.

A comparison of these pragmatic multiple linear regression models was undertaken with models derived using the Cubist piecewise linear decision tree approach (Quinlan 1992) combined with more sophisticated covariates obtained from the TERN covariate dataset. These include potassium, uranium and thorium gamma radiometrics, kaolin, illite and smectite clays from visible near infra red (VNIR) spectroscopy (Viscarra Rossel and Webster 2012), topographic wetness index, MODIS mean photosynthetic
vegetation fractional cover (Guerschman et al. 2012) and the weathering index (Wilford 2012). The same training and validation sets were used for each modelling approach.

3.2.7 Preparation of digital soil maps

Trial digital soil maps representing each of the key soil properties at the three selected depth intervals were prepared over the State of NSW and the Hunter region, as outlined below:

Compile GIS grids for each covariate for the State as described in the earlier section on covariates, with pixel size of 100 m.

Produce digital soil map layers for each soil property and depth interval over NSW by applying the derived multiple linear regression models to each 100-m pixel. The soil property estimate and all covariate data may be viewed for each pixel with a GIS tool for identifying individual values.

Prepare preliminary maps of upper and lower 95% confidence levels where we assume that the RMSE calculated from the validation samples represents the standard deviation of the distribution of our predictions at sites, which if we assume follows a normal distribution means that the 95% confidence interval is defined as +/- 1.96 x RMSE. One weakness of this approach is that it assumes the variance of our predictions is uniform across the region which is unlikely but nevertheless it provides a first approximation. As an example, a pH prediction of 6.0 may have a 95% confidence interval between 4.7 and 7.8.

Prepare maps over the Hunter region using (a) the pragmatic approach outlined in this paper, and (b) the more sophisticated approach using the Cubist system with advanced remotely sensed covariates from the TERN covariate dataset.

Validate maps for each soil property against the NSW and Hunter region sites contained within the original validation sets. Validation statistics were as described for the model development.
3.3 Results and validation

3.3.1 Multiple linear regression model development

The multiple linear regression models derived between the eight soil properties and the six covariates at each of the three depth intervals, together with associated statistics are presented in Table 3.6.

The relationships are of generally moderate statistical strength ($R^2 = 0.48$ to 0.56) for OC (0-10 cm depth interval), pH$_{ca}$, sum-of-bases, CEC and sand but are of only low to moderate strength ($R^2 = 0.20$ to 0.40) for OC (lower depth intervals), total P, clay and silt. Although the relationship strengths reduce markedly with depth for OC, for other properties the strengths remain generally steady or actually increase, eg, $R^2$ rises for pH$_{ca}$ from 0.51 to 0.55 and for sum-of-bases from 0.48 to 0.56.

3.3.2 Influence of covariates

Most of the models include five or all six of the covariates, meaning most covariates had acceptable levels of statistical confidence ($p < 0.075$) but some do have only four or as low as three of the covariates. The standardised regression coefficients for each covariate (Table 3.7) provide useful indications of the relative influence each covariate has in the models. For example, it reveals that for organic carbon in the upper depth interval, $T_{max}$ appears to have the greatest influence, followed equally by $Precip$ and $Silica$, with $TSI$, $LDI$ and $Asp$ having only a relatively small influence. At the deepest depth interval $Silica$ appears to have the greatest influence on OC%, followed by $Precip$, with $T_{max}$ having much weaker influence. For most properties at most depths, $Silica$ appears to be the covariate exerting the greatest influence. For example, for CEC upper depth interval the standardized coefficient for $Silica$ is 0.62, with the next highest being $T_{max}$ at 0.18. The knowledge and insights into soil formation that can be gained from this table is discussed later.

The quantitative influence of each covariate on each soil property can also be estimated directly from their partial regression coefficients. A full listing of the influence per unit change of the six covariates on the key soil properties examined, assuming other factors are held constant, is presented in Table 3.8. These results also
Chapter 3: Pragmatic models for soil prediction in eastern Australia

Table 3.6. Multiple regression relationships and associated statistics

<table>
<thead>
<tr>
<th>Soil property</th>
<th>Depth (cm)</th>
<th>Regression relationship</th>
<th>N</th>
<th>R²</th>
<th>F</th>
<th>Resid. SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Org C</td>
<td>0-10</td>
<td>OC = exp(3.270 + 0.000622Precip - 0.0722Tmax - 0.0194Silica - 0.0468TSI - 0.0473LDI + 0.0174Asp)</td>
<td>3942</td>
<td>0.50</td>
<td>646</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>10-30</td>
<td>OC = exp(2.315 + 0.000701Precip - 0.0370Tmax - 0.0226Silica - 0.0378TSI - 0.0812LDI)</td>
<td>3384</td>
<td>0.40</td>
<td>443</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>OC = exp(0.951 + 0.00593Precip - 0.0416Tmax - 0.0242Silica + 0.0382TSI - 0.0372LDI)</td>
<td>2456</td>
<td>0.20</td>
<td>122</td>
<td>0.81</td>
</tr>
<tr>
<td>pHca</td>
<td>0-10</td>
<td>pHca = 6.437-0.00109Precip + 0.0780Tmax - 0.0383Silica + 0.1117TSI + 0.0629LDI + 0.0591Asp</td>
<td>4961</td>
<td>0.54</td>
<td>976</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>10-30</td>
<td>pHca = 6.595 - 0.00132Precip + 0.0790Tmax - 0.0383Silica + 0.1317TSI + 0.0842LDI + 0.0604Asp</td>
<td>4501</td>
<td>0.57</td>
<td>997</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>pHca = 6.935 - 0.00169Precip + 0.0687Tmax - 0.0352Silica + 0.137TSI + 0.159LDI + 0.0373Asp</td>
<td>4102</td>
<td>0.57</td>
<td>886</td>
<td>0.99</td>
</tr>
<tr>
<td>Sum-of-bases</td>
<td>0-10</td>
<td>SoB = exp(4.892 - 0.000454Precip + 0.0411Tmax - 0.0566Silica + 0.0361TSI + 0.0659LDI + 0.0495Asp)</td>
<td>4193</td>
<td>0.48</td>
<td>651</td>
<td>0.81</td>
</tr>
<tr>
<td>(cmol/kg)</td>
<td>10-30</td>
<td>SoB = exp(4.641 - 0.000595Precip + 0.0552Tmax - 0.0583Silica + 0.0300TSI + 0.110LDI + 0.0392Asp)</td>
<td>3963</td>
<td>0.52</td>
<td>706</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>SoB = exp(4.610 - 0.00103Precip + 0.0663Tmax - 0.0551Silica + 0.161LDI)</td>
<td>3848</td>
<td>0.56</td>
<td>1201</td>
<td>0.87</td>
</tr>
<tr>
<td>CEC</td>
<td>0-10</td>
<td>CEC = exp(4.945 - 0.000196Precip + 0.0415Tmax - 0.0512Silica + 0.0399TSI + 0.0187Asp)</td>
<td>3173</td>
<td>0.48</td>
<td>594</td>
<td>0.69</td>
</tr>
<tr>
<td>(cmol/kg)</td>
<td>10-30</td>
<td>CEC = exp(4.917 - 0.000277Precip + 0.0404Tmax - 0.0516Silica + 0.0269TSI + 0.0702LDI)</td>
<td>3122</td>
<td>0.52</td>
<td>674</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>CEC = exp(4.951 - 0.000397Precip + 0.0446Tmax - 0.0513Silica + 0.0963LDI)</td>
<td>2878</td>
<td>0.51</td>
<td>759</td>
<td>0.71</td>
</tr>
<tr>
<td>P total (mg/kg)</td>
<td>0-10</td>
<td>Ptot = exp(6.328 + 0.000864Precip + 0.0345Tmax - 0.0389Silica - 0.0702LDI)</td>
<td>1416</td>
<td>0.28</td>
<td>136</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>10-30</td>
<td>Ptot = exp(6.329 + 0.000871Precip + 0.0305Tmax - 0.0398Silica + 0.0599LDI)</td>
<td>1253</td>
<td>0.27</td>
<td>116</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>Ptot = exp(6.642 + 0.000671Precip + 0.0343Tmax - 0.0431Silica)</td>
<td>1133</td>
<td>0.26</td>
<td>130</td>
<td>0.91</td>
</tr>
<tr>
<td>Clay (%)</td>
<td>0-10</td>
<td>Clay = exp(4.711 - 0.000226Precip + 0.0293Tmax - 0.0419Silica + 0.0203TSI + 0.105LDI + 0.0243Asp)</td>
<td>3431</td>
<td>0.39</td>
<td>364</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>10-30</td>
<td>Clay = exp(5.181 - 0.000320Precip + 0.0248Tmax - 0.0425Silica + 0.109LDI + 0.0219Asp)</td>
<td>3289</td>
<td>0.39</td>
<td>419</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>Clay = exp(5.682 - 0.000385Precip - 0.0348Silica - 0.0323TSI - 0.126LDI)</td>
<td>3177</td>
<td>0.31</td>
<td>360</td>
<td>0.76</td>
</tr>
<tr>
<td>Sand (%)</td>
<td>0-10</td>
<td>Sand = 1.768 - 0.00620Precip - 0.485Tmax + 1.279Silica - 1.590TSI - 1.225LDI - 0.619Asp</td>
<td>3362</td>
<td>0.50</td>
<td>560</td>
<td>16.2</td>
</tr>
<tr>
<td></td>
<td>10-30</td>
<td>Sand = -8.599 - 0.00443Precip - 0.478Tmax + 1.328Silica - 1.039TSI - 1.429LDI - 0.503Asp</td>
<td>3208</td>
<td>0.51</td>
<td>546</td>
<td>16.5</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>Sand = -17.11 - 0.00258Precip + 1.157Silica - 0.459TSI - 2.223LDI</td>
<td>3087</td>
<td>0.39</td>
<td>493</td>
<td>18.4</td>
</tr>
<tr>
<td>Silt (%)</td>
<td>0-10</td>
<td>Silt = exp(5.142 + 0.000151Precip - 0.0362Tmax - 0.0318Silica + 0.0516LDI + 0.0403Asp)</td>
<td>3443</td>
<td>0.27</td>
<td>259</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>10-30</td>
<td>Silt = exp(5.105 + 0.000167Precip - 0.0374Tmax - 0.0310Silica + 0.0386LDI + 0.0494Asp)</td>
<td>3291</td>
<td>0.27</td>
<td>245</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>Silt = exp(4.790 + 0.000135Precip - 0.0307Tmax - 0.0293Silica + 0.0324LDI + 0.0599Asp)</td>
<td>2973</td>
<td>0.22</td>
<td>171</td>
<td>0.72</td>
</tr>
</tbody>
</table>

Precip = annual precipitation (mm pa); Tmax = mean annual daily max temperature (°C); Silica = silica index; TSI = topo slope index; LDI = land disturbance index; Asp = aspect index
Table 3.7. Standardized regression coefficients of covariates in regression relationships

<table>
<thead>
<tr>
<th>Soil property</th>
<th>Depth (cm)</th>
<th>Precip</th>
<th>Tmax</th>
<th>Silica</th>
<th>TSI</th>
<th>LDI</th>
<th>Asp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Org C (log %)</td>
<td>0-10</td>
<td>0.27</td>
<td>-0.40</td>
<td>-0.27</td>
<td>-0.11</td>
<td>-0.07</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>10-30</td>
<td>0.33</td>
<td>-0.19</td>
<td>-0.31</td>
<td>-0.09</td>
<td>-0.13</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>0.25</td>
<td>-0.15</td>
<td>-0.29</td>
<td>-0.08</td>
<td>-0.06</td>
<td>-</td>
</tr>
<tr>
<td>pHca</td>
<td>0-10</td>
<td>-0.32</td>
<td>0.28</td>
<td>-0.35</td>
<td>0.17</td>
<td>0.06</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>10-30</td>
<td>-0.37</td>
<td>0.25</td>
<td>-0.33</td>
<td>0.19</td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>-0.42</td>
<td>0.19</td>
<td>-0.26</td>
<td>0.17</td>
<td>0.14</td>
<td>0.04</td>
</tr>
<tr>
<td>Sum-of-bases</td>
<td>0-10</td>
<td>-0.15</td>
<td>0.16</td>
<td>-0.58</td>
<td>0.06</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>(log cmol/kg)</td>
<td>10-30</td>
<td>-0.19</td>
<td>0.20</td>
<td>-0.55</td>
<td>0.05</td>
<td>0.12</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>-0.31</td>
<td>0.23</td>
<td>-0.48</td>
<td>-</td>
<td>0.16</td>
<td>-</td>
</tr>
<tr>
<td>CEC</td>
<td>0-10</td>
<td>-0.08</td>
<td>0.18</td>
<td>-0.62</td>
<td>0.08</td>
<td>-</td>
<td>0.03</td>
</tr>
<tr>
<td>(log cmol/kg)</td>
<td>10-30</td>
<td>-0.11</td>
<td>0.17</td>
<td>-0.61</td>
<td>0.05</td>
<td>0.09</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>-0.15</td>
<td>0.18</td>
<td>-0.59</td>
<td>-</td>
<td>0.12</td>
<td>-</td>
</tr>
<tr>
<td>P total (ppm)</td>
<td>0-10</td>
<td>0.32</td>
<td>0.21</td>
<td>-0.44</td>
<td>-</td>
<td>0.08</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>10-30</td>
<td>0.32</td>
<td>0.18</td>
<td>-0.44</td>
<td>-</td>
<td>0.07</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>0.25</td>
<td>0.20</td>
<td>-0.47</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Clay (log %)</td>
<td>0-10</td>
<td>-0.09</td>
<td>0.13</td>
<td>-0.51</td>
<td>0.04</td>
<td>0.13</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>10-30</td>
<td>-0.12</td>
<td>0.11</td>
<td>-0.52</td>
<td>-</td>
<td>0.14</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>-0.14</td>
<td>-</td>
<td>-0.46</td>
<td>-</td>
<td>0.15</td>
<td>-</td>
</tr>
<tr>
<td>Sand (%)</td>
<td>0-10</td>
<td>-0.10</td>
<td>-0.09</td>
<td>0.66</td>
<td>-0.13</td>
<td>-0.06</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td>10-30</td>
<td>-0.07</td>
<td>-0.09</td>
<td>0.67</td>
<td>-0.08</td>
<td>-0.07</td>
<td>-0.04</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>-0.04</td>
<td>-</td>
<td>0.58</td>
<td>-0.04</td>
<td>-0.11</td>
<td>-</td>
</tr>
<tr>
<td>Silt (log %)</td>
<td>0-10</td>
<td>0.07</td>
<td>-0.20</td>
<td>-0.47</td>
<td>-</td>
<td>0.08</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>10-30</td>
<td>0.08</td>
<td>-0.20</td>
<td>-0.46</td>
<td>-</td>
<td>0.06</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>0.09</td>
<td>-</td>
<td>-0.42</td>
<td>-</td>
<td>0.02</td>
<td>0.14</td>
</tr>
</tbody>
</table>

assume linear relationships, which may not always be the case. It is suggested, for example, that over the 0-10 cm depth interval, for each 100 mm increase in annual precipitation there is an 0.062 log e % (6.4% proportional) increase in OC% and 0.11 unit decrease in pHca; and that with each 1 degree Celsius rise in annual maximum temperature there is a corresponding 0.072 log e % (7.0% proportional) decrease in OC% and 0.08 unit rise in pHca. A synthesis of the results from this table and other study results is presented in the Discussion.
Table 3.8. Influence of covariates on key soil properties

<table>
<thead>
<tr>
<th>Soil property</th>
<th>Covariate (units of change)</th>
<th>Change per covariate unit&lt;sup&gt;1&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0-10 cm</td>
</tr>
<tr>
<td>OC</td>
<td>Precip (% per 100 mm pa)</td>
<td>6.4</td>
</tr>
<tr>
<td></td>
<td>Tmax (% per deg C)</td>
<td>-7.0</td>
</tr>
<tr>
<td></td>
<td>Silica (% per 10% silica)</td>
<td>-17.6</td>
</tr>
<tr>
<td></td>
<td>TSI (% per TSI unit)</td>
<td>-4.6</td>
</tr>
<tr>
<td></td>
<td>LDI (% per LDI unit)</td>
<td>-4.6</td>
</tr>
<tr>
<td></td>
<td>Asp (% per Asp unit)</td>
<td>1.8</td>
</tr>
<tr>
<td>pHca</td>
<td>Precip (pH units per 100 mm pa)</td>
<td>-0.11</td>
</tr>
<tr>
<td></td>
<td>Tmax (pH units per deg C)</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>Silica (pH units per 10% silica)</td>
<td>-0.38</td>
</tr>
<tr>
<td></td>
<td>TSI (pH units per TSI unit)</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>LDI (pH units per LDI unit)</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>Asp (pH units per Asp unit)</td>
<td>0.06</td>
</tr>
<tr>
<td>SoB</td>
<td>Precip (% per 100 mm pa)</td>
<td>-4.4</td>
</tr>
<tr>
<td></td>
<td>Tmax (% per deg C)</td>
<td>4.2</td>
</tr>
<tr>
<td></td>
<td>Silica (% per 10% silica)</td>
<td>-43.2</td>
</tr>
<tr>
<td></td>
<td>TSI (% per TSI unit)</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td>LDI (% per LDI unit)</td>
<td>6.8</td>
</tr>
<tr>
<td></td>
<td>Asp (% per Asp unit)</td>
<td>5.1</td>
</tr>
<tr>
<td>CEC</td>
<td>Precip (% per 100 mm pa)</td>
<td>-1.9</td>
</tr>
<tr>
<td></td>
<td>Tmax (% per deg C)</td>
<td>4.2</td>
</tr>
<tr>
<td></td>
<td>Silica (% per 10% silica)</td>
<td>-40.1</td>
</tr>
<tr>
<td></td>
<td>TSI (% per TSI unit)</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>LDI (% per LDI unit)</td>
<td>7.3</td>
</tr>
<tr>
<td></td>
<td>Asp (% per Asp unit)</td>
<td>1.9</td>
</tr>
<tr>
<td>Ptot</td>
<td>Precip (% per 100 mm pa)</td>
<td>9.0</td>
</tr>
<tr>
<td></td>
<td>Tmax (% per deg C)</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>Silica (% per 10% silica)</td>
<td>-32.2</td>
</tr>
<tr>
<td></td>
<td>TSI (% per TSI unit)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>LDI (% per LDI unit)</td>
<td>7.3</td>
</tr>
<tr>
<td></td>
<td>Asp (% per Asp unit)</td>
<td>-</td>
</tr>
<tr>
<td>Clay</td>
<td>Precip (% per 100 mm pa)</td>
<td>-2.2</td>
</tr>
<tr>
<td></td>
<td>Tmax (% per deg C)</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>Silica (% per 10% silica)</td>
<td>-34.2</td>
</tr>
<tr>
<td></td>
<td>TSI (% per TSI unit)</td>
<td>2.1</td>
</tr>
<tr>
<td></td>
<td>LDI (% per LDI unit)</td>
<td>11.1</td>
</tr>
<tr>
<td></td>
<td>Asp (% per Asp unit)</td>
<td>2.5</td>
</tr>
<tr>
<td>Sand</td>
<td>Precip (absol % per 100 mm pa)</td>
<td>-0.6</td>
</tr>
<tr>
<td></td>
<td>Tmax (absolute % per deg C)</td>
<td>-0.5</td>
</tr>
<tr>
<td></td>
<td>Silica (absolute % per 10% silica)</td>
<td>12.8</td>
</tr>
<tr>
<td></td>
<td>TSI (absolute % per TSI unit)</td>
<td>-1.6</td>
</tr>
<tr>
<td></td>
<td>LDI (absolute % per LDI unit)</td>
<td>-1.2</td>
</tr>
<tr>
<td></td>
<td>Asp (absolute % per Asp unit)</td>
<td>-0.6</td>
</tr>
<tr>
<td>Silt</td>
<td>Precip (% per 100 mm pa)</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>Tmax (% per deg C)</td>
<td>-3.6</td>
</tr>
<tr>
<td></td>
<td>Silica (% per 10% silica)</td>
<td>-27.2</td>
</tr>
<tr>
<td></td>
<td>TSI (% per TSI unit)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>LDI (% per LDI unit)</td>
<td>5.3</td>
</tr>
<tr>
<td></td>
<td>Asp (% per Asp unit)</td>
<td>4.1</td>
</tr>
</tbody>
</table>

<sup>1</sup> Change is relative for most properties, but absolute for pH and sand
3.3.3 Validation of models

Results of the validation process using the 15% of withheld profiles are presented in Table 3.9, with validation plots for four of the properties shown in Figure 3.3. These applied the derived regression models to the recorded covariate data at each site. Validation results are shown on the scale at which the models were developed, meaning several of the properties are presented on the natural log scale.

### Table 3.9. Model validation

<table>
<thead>
<tr>
<th>Soil property</th>
<th>Depth</th>
<th>Lin’s CCC</th>
<th>R²</th>
<th>N</th>
<th>RMSE</th>
<th>Mean error</th>
<th>Median absolute error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Org C (log %)</td>
<td>0-10</td>
<td>0.70</td>
<td>0.54</td>
<td>685</td>
<td>0.58</td>
<td>0.02</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>10-30</td>
<td>0.59</td>
<td>0.43</td>
<td>608</td>
<td>0.65</td>
<td>0.00</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>0.35</td>
<td>0.20</td>
<td>460</td>
<td>0.83</td>
<td>0.02</td>
<td>0.46</td>
</tr>
<tr>
<td>pHca</td>
<td>0-10</td>
<td>0.69</td>
<td>0.51</td>
<td>956</td>
<td>0.85</td>
<td>0.04</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>10-30</td>
<td>0.71</td>
<td>0.54</td>
<td>892</td>
<td>0.88</td>
<td>0.04</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>0.73</td>
<td>0.57</td>
<td>809</td>
<td>0.97</td>
<td>0.04</td>
<td>0.66</td>
</tr>
<tr>
<td>Sum-of-bases (log cmol/kg)</td>
<td>0-10</td>
<td>0.66</td>
<td>0.49</td>
<td>750</td>
<td>0.77</td>
<td>0.01</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>10-30</td>
<td>0.70</td>
<td>0.53</td>
<td>721</td>
<td>0.78</td>
<td>-0.01</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>0.73</td>
<td>0.56</td>
<td>673</td>
<td>0.81</td>
<td>-0.01</td>
<td>0.50</td>
</tr>
<tr>
<td>CEC (log cmol/kg)</td>
<td>0-10</td>
<td>0.66</td>
<td>0.50</td>
<td>769</td>
<td>0.67</td>
<td>0.04</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>10-30</td>
<td>0.69</td>
<td>0.52</td>
<td>560</td>
<td>0.66</td>
<td>0.02</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>0.68</td>
<td>0.50</td>
<td>504</td>
<td>0.69</td>
<td>0.02</td>
<td>0.40</td>
</tr>
<tr>
<td>P total (log mg/kg)</td>
<td>0-10</td>
<td>0.40</td>
<td>0.25</td>
<td>238</td>
<td>0.89</td>
<td>0.00</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>10-30</td>
<td>0.41</td>
<td>0.26</td>
<td>217</td>
<td>0.91</td>
<td>-0.03</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>0.36</td>
<td>0.22</td>
<td>200</td>
<td>0.97</td>
<td>-0.02</td>
<td>0.62</td>
</tr>
<tr>
<td>Clay (log %)</td>
<td>0-10</td>
<td>0.58</td>
<td>0.39</td>
<td>630</td>
<td>0.70</td>
<td>0.03</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>10-30</td>
<td>0.59</td>
<td>0.38</td>
<td>616</td>
<td>0.68</td>
<td>0.02</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>0.49</td>
<td>0.29</td>
<td>660</td>
<td>0.69</td>
<td>0.01</td>
<td>0.36</td>
</tr>
<tr>
<td>Sand (%)¹</td>
<td>0-10</td>
<td>0.66</td>
<td>0.48</td>
<td>627</td>
<td>16.4</td>
<td>-0.3</td>
<td>10.9</td>
</tr>
<tr>
<td></td>
<td>10-30</td>
<td>0.68</td>
<td>0.50</td>
<td>614</td>
<td>16.3</td>
<td>-0.6</td>
<td>11.1</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>0.58</td>
<td>0.41</td>
<td>573</td>
<td>18.0</td>
<td>-0.2</td>
<td>12.7</td>
</tr>
<tr>
<td>Silt (log %)</td>
<td>0-10</td>
<td>0.43</td>
<td>0.27</td>
<td>632</td>
<td>0.70</td>
<td>0.02</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>10-30</td>
<td>0.42</td>
<td>0.26</td>
<td>623</td>
<td>0.72</td>
<td>0.03</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>0.37</td>
<td>0.20</td>
<td>553</td>
<td>0.69</td>
<td>-0.07</td>
<td>0.39</td>
</tr>
</tbody>
</table>

CCC: concordance correlation coefficient, RMSE: root mean square error, Mean error = mean of predicted – observed values
¹ note the high RMSE resulting from use of normal rather than a log scale

Lin’s concordance correlation coefficients reaches maximums of 0.73 for both pHca and sum-of-bases, and is above 0.50 for most other properties indicating at least moderate predictive performance for these properties. The models for total P, silt and the lower depth intervals for OC have somewhat lower predictive quality. The RMSE
values suggest moderate errors in the predictions; particularly evident where a normal (non-log) scale is used such as for sand. The median absolute differences are, as expected, lower than the RMSE as this measure is less influenced by outliers, and generally indicate satisfactory performance of the models.

A general weakness in the predictive pattern of many of the models is revealed by the plots of Figure 3.3. There is a general tendency to under predict at the high levels, particularly evident for $\text{pH}_{\text{ca}}$ and sum-of-bases, and to over predict at lower levels, particularly evident for OC and sand. This may suggest, at least in some cases, that one or more influencing factors are not being addressed in the models, for example, the presence of carbonates for $\text{pH}$.

![Graphs showing observed versus predicted values for OC, $\text{pH}_{\text{ca}}$, sum-of-bases, and sand](image)

**Figure 3.3.** Observed versus predicted values for OC, $\text{pH}_{\text{ca}}$, sum-of-bases and sand (0-10 cm depth)
### 3.3.4 Comparison with Cubist modelling approach and remotely sensed covariates

A comparison of these pragmatic multiple linear regression models was undertaken with models derived using the more advanced Cubist piecewise linear decision tree technique, with and without the replacement of the pragmatic covariates with more sophisticated covariates such as radiometrics, VNIR clay values, topographic wetness index, photosynthetic vegetation fractional cover and weathering index. The same validation dataset was applied. The key validation statistics for Lin’s CCC and RMSE over the 0-10 cm interval are presented in Table 3.10.

<table>
<thead>
<tr>
<th>Soil property</th>
<th>MLR Pragmatic covariates</th>
<th>Cubist Pragmatic covariates</th>
<th>Cubist TERN covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CCC</td>
<td>RMSE</td>
<td>CCC</td>
</tr>
<tr>
<td>Org C (log %)</td>
<td>0.70</td>
<td>0.58</td>
<td>0.65</td>
</tr>
<tr>
<td>pHca</td>
<td>0.69</td>
<td>0.85</td>
<td>0.70</td>
</tr>
<tr>
<td>Sum bases (log cmol/kg)</td>
<td>0.66</td>
<td>0.77</td>
<td>0.39</td>
</tr>
<tr>
<td>CEC (log cmol/kg)</td>
<td>0.66</td>
<td>0.67</td>
<td>0.25</td>
</tr>
<tr>
<td>P total (log mg/kg)</td>
<td>0.43</td>
<td>0.91</td>
<td>0.33</td>
</tr>
<tr>
<td>Clay (log %)</td>
<td>0.58</td>
<td>0.70</td>
<td>0.32</td>
</tr>
<tr>
<td>Sand (%)³</td>
<td>0.66</td>
<td>16.4</td>
<td>0.23</td>
</tr>
<tr>
<td>Silt (log %)</td>
<td>0.43</td>
<td>0.70</td>
<td>0.21</td>
</tr>
</tbody>
</table>

CCC: Lin’s concordance correlation coefficient, RMSE: root mean square error
³ note the high RMSE resulting from use of normal rather than a log scale

The table broadly indicates that with the replacement of the standard covariates with sophisticated covariates in the MLR model there is typically a significant decrease in statistical strength of the models. This appears to be largely attributable to the removal of the lithology (*silica index*) covariate.

With the application of the Cubist modelling approach to the standard set of covariates a significant improvement in validation statistics is observed for several properties (pHca, Ptot, clay and silt) and marginal improvements for the other properties (OC, sum-of-bases, CEC and sand). The largest improvement is demonstrated with Ptot,
where Lin’s CCC rises from 0.43 to 0.63 and RMSE drops from 0.91 to 0.79 log ppm units. With the replacement of the standard covariates with sophisticated covariates in the Cubist model, there is again a marked decrease in performance. Similar trends are observed for the other depth intervals (not presented here).

An analysis of the validation plots suggests that the Cubist modelling approach tends to give improved predictions at the upper ranges of the soil property values. For example, the MLR approach did not predict any pH\textsubscript{ca} (0-10 cm) values higher than 7.5, whereas the Cubist approach gave predictions up to 8.2. Similarly, the MLR approach did not predict clay (0-10 cm) values above 65%, whereas the Cubist approach predicted up to 75%. A similar pattern was observed for most soil properties, with the notable exception of OC (0-10 and 10-30 cm).

Overall, it is evident that the MLR modelling tool is slightly less effective in its predictive performance to the Cubist tool. On the other hand, the standard pragmatic covariates used in this study appear to out-perform the more sophisticated remotely sensed TERN covariates, mainly due to the strong influence of the lithology (silica index) covariate, which is not applied in the latter covariate set. The best overall modelling approach would appear to be the use of Cubist tool with a mixture of sophisticated TERN covariates and reliable pragmatic covariates, particularly fine scale lithology data.

### 3.3.5 Digital soil map preparation

Digital maps were prepared over the State of NSW for all of the eight soil properties for each depth interval using the pragmatic MLR models. Additionally, maps of bulk density and soil organic carbon mass down to 30 cm were prepared. Maps for OC mass, pH\textsubscript{ca}, sum-of-bases, total P, clay and sand content for the 0-10 cm interval are presented in Figure 3.4. Preliminary maps presenting the 95% upper and lower confidence levels for each soil property were also prepared. With the covariate grids and the MLR models ready, these maps were able to be prepared easily and quickly (in the order of minutes) in a GIS framework, even with a low powered computer.
Validation of these maps was again undertaken using the validation dataset, with results presented in Table 3.11. The predictive ability demonstrated by the maps is generally only moderate, with Lin’s CCC generally between 0.4 and 0.7. The application of the models to spatial coverages introduced new potential errors, for example, the assumptions of the parent material class (silica index) over broad areas.
based on the broad 1: 250 000 scale geological layer. The poor definition of Cainozoic alluvials in the geology layer, which necessitated reference to NSW soil mapping data, was another weakness. As a consequence, the accuracy of predictions from the maps are generally significantly lower than for the predictions directly from the models, which used more reliable covariate data collected at individual sites by soil surveyors, for example, parent material (silica index) from the broad scale map rather than the soil surveyor site description.

Table 3.11. Validation of NSW digital soil maps

<table>
<thead>
<tr>
<th>Soil property</th>
<th>Depth</th>
<th>Lin’s CCC</th>
<th>R²</th>
<th>N</th>
<th>RMSE</th>
<th>Mean error</th>
<th>Median absolute error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Org C (log %)</td>
<td>0-10</td>
<td>0.52</td>
<td>0.39</td>
<td>294</td>
<td>0.65</td>
<td>0.02</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>10-30</td>
<td>0.46</td>
<td>0.33</td>
<td>286</td>
<td>0.71</td>
<td>0.01</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>0.21</td>
<td>0.10</td>
<td>238</td>
<td>0.93</td>
<td>0.11</td>
<td>0.52</td>
</tr>
<tr>
<td>pHca</td>
<td>0-10</td>
<td>0.64</td>
<td>0.45</td>
<td>507</td>
<td>0.81</td>
<td>0.12</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>10-30</td>
<td>0.67</td>
<td>0.47</td>
<td>498</td>
<td>0.86</td>
<td>0.15</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>0.69</td>
<td>0.52</td>
<td>437</td>
<td>0.98</td>
<td>0.19</td>
<td>0.65</td>
</tr>
<tr>
<td>Sum-of-bases</td>
<td>0-10</td>
<td>0.53</td>
<td>0.37</td>
<td>363</td>
<td>0.89</td>
<td>-0.07</td>
<td>0.47</td>
</tr>
<tr>
<td>(log cmol/kg)</td>
<td>10-30</td>
<td>0.58</td>
<td>0.42</td>
<td>358</td>
<td>0.91</td>
<td>-0.14</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>0.61</td>
<td>0.46</td>
<td>310</td>
<td>0.97</td>
<td>-0.23</td>
<td>0.62</td>
</tr>
<tr>
<td>CEC</td>
<td>0-10</td>
<td>0.51</td>
<td>0.32</td>
<td>364</td>
<td>0.76</td>
<td>0.04</td>
<td>0.47</td>
</tr>
<tr>
<td>(log cmol/kg)</td>
<td>10-30</td>
<td>0.48</td>
<td>0.34</td>
<td>358</td>
<td>0.96</td>
<td>-0.54</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>0.53</td>
<td>0.37</td>
<td>314</td>
<td>0.82</td>
<td>-0.10</td>
<td>0.50</td>
</tr>
<tr>
<td>P total</td>
<td>0-10</td>
<td>0.36</td>
<td>0.54</td>
<td>41</td>
<td>1.16</td>
<td>0.01</td>
<td>0.83</td>
</tr>
<tr>
<td>(log mg/kg)</td>
<td>10-30</td>
<td>0.36</td>
<td>0.56</td>
<td>41</td>
<td>1.20</td>
<td>0.03</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>0.33</td>
<td>0.51</td>
<td>41</td>
<td>1.24</td>
<td>0.10</td>
<td>0.82</td>
</tr>
<tr>
<td>Clay (%)</td>
<td>0-10</td>
<td>0.48</td>
<td>0.31</td>
<td>429</td>
<td>14.51</td>
<td>-3.79</td>
<td>7.13</td>
</tr>
<tr>
<td></td>
<td>10-30</td>
<td>0.51</td>
<td>0.32</td>
<td>419</td>
<td>15.59</td>
<td>-4.54</td>
<td>8.54</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>0.39</td>
<td>0.20</td>
<td>362</td>
<td>19.89</td>
<td>-6.92</td>
<td>12.33</td>
</tr>
<tr>
<td>Sand (%)</td>
<td>0-10</td>
<td>0.49</td>
<td>0.31</td>
<td>480</td>
<td>18.28</td>
<td>0.07</td>
<td>12.63</td>
</tr>
<tr>
<td></td>
<td>10-30</td>
<td>0.51</td>
<td>0.33</td>
<td>469</td>
<td>18.28</td>
<td>0.30</td>
<td>12.90</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>0.39</td>
<td>0.24</td>
<td>413</td>
<td>20.58</td>
<td>2.23</td>
<td>14.82</td>
</tr>
<tr>
<td>Silt (%)</td>
<td>0-10</td>
<td>0.13</td>
<td>0.04</td>
<td>514</td>
<td>13.71</td>
<td>-5.70</td>
<td>6.55</td>
</tr>
<tr>
<td></td>
<td>10-30</td>
<td>0.15</td>
<td>0.06</td>
<td>506</td>
<td>13.17</td>
<td>-5.64</td>
<td>6.29</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>0.14</td>
<td>0.05</td>
<td>447</td>
<td>12.10</td>
<td>-5.31</td>
<td>6.01</td>
</tr>
</tbody>
</table>

CCC: concordance correlation coefficient, RMSE: root mean square error

1 note the high RMSE resulting from use of normal rather than a log scale

A map of pHca over the Hunter region of NSW is presented in Figure 3.5. This figure demonstrates the presentation of the underlying data on covariates applying to given pixels that is available in GIS mode, using a tool to identify individual values. This information allows the derivation of predictions for individual sites to be
understood and promotes transparency in all predictions. Thus, for example, it can be seen that Site B has a higher pH than site A, which can be explained by the lower rainfall, higher temperatures, more mafic parent material and a lower topographic position.

Figure 3.5. Digital soil map of soil pHca over the Hunter region, with underlying covariate data

For comparison purposes, a digital soil map over the Hunter region was also prepared using the Cubist tool with TERN remotely sensed covariates. Validation results from an external dataset of approximately 300 points are presented in Table 3.12. This reveals that for most properties the Cubist tool with TERN covariates approach gives slightly higher concordance and lower RMSE values than the pragmatic MLR approach, with the notable exceptions of organic carbon and pHca. The advanced remotely sensed covariates are likely to be more reliable in representing each pixel than the pragmatic covariates, which are based on broader polygonal data, for the purposes of preparing digital soil maps. The application of residual kriging (Odeh et al. 1995) may further improve the reliability of the more sophisticated mapping approach, but this was not implemented in this comparison exercise.
Table 3.12. Validation of Hunter region digital soil maps (0-10 cm depth)

<table>
<thead>
<tr>
<th>Soil property</th>
<th>Pragmatic MLR</th>
<th>Cubist with TERN covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lin’s CCC</td>
<td>RMSE</td>
</tr>
<tr>
<td>Org C (log %)</td>
<td>0.32</td>
<td>0.64</td>
</tr>
<tr>
<td>pHca</td>
<td>0.60</td>
<td>0.71</td>
</tr>
<tr>
<td>Sum-of-bases (log cmol/kg)</td>
<td>0.57</td>
<td>0.90</td>
</tr>
<tr>
<td>CEC (log cmol/kg)</td>
<td>0.59</td>
<td>0.70</td>
</tr>
<tr>
<td>Clay (log %)</td>
<td>0.41</td>
<td>0.83</td>
</tr>
<tr>
<td>Sand (%)</td>
<td>0.45</td>
<td>18.2</td>
</tr>
<tr>
<td>Silt (log %)</td>
<td>0.27</td>
<td>0.84</td>
</tr>
</tbody>
</table>

Ptot - no observed values available for the region

### 3.4 Discussion

#### 3.4.1 Predictive performance of models and maps

The performance of the pragmatic regression models and resulting predictive maps varies between the different properties and the three depth intervals. Many of the models demonstrated at least moderate strength, particularly those for OC (upper depth), pHca, sum-of-bases, CEC and sand, however, the models for OC (mid and lower depths), total P, clay and silt were generally only of low to moderate strength.

The models provide a useful first approximation of soil property values at a given site where only climate and field data representing each of the covariates is available. They may therefore be readily adopted by various natural resource scientists that require soil data for specific locations, particularly if no other suitable source of soil information is available. Soil scientists, agronomists, ecologists, hydrologists and other scientists may all periodically have a need to derive soil property estimates over specific survey sites, using as little as a hand held scientific calculator if need be. The field collected covariate data may often be more reliable than regional scale data (such as regional geology or land use maps), leading to more accurate final soil property estimates for individual sites. Comparison of the model derived estimates with reliable laboratory data for a particular site could provide a tentative indication of whether the soil properties are significantly above or below what might be considered normal for
that soil and set of environmental conditions, thus informing on land management conditions at the site.

The models appear to be moderately effective in the development of digital soil maps over large regions, as has been carried out over NSW and the Hunter region in this study. The time and computer resources required to produce the final digital maps in a GIS framework is very low. It would be relatively straightforward to prepare similar maps more extensively over the whole of eastern Australia. The broad scale of the digital soil map development needs to be considered when assessing map predictions at individual points.

From these maps the underlying information for each cell can be easily viewed and interpreted, allowing identification of the controlling factors at any particular site, a feature not usually provided in most DSM products. This is demonstrated by Figure 3.5, which shows that the different pH at the two points A and B can be the result of differences in their respective covariates; as was noted above. This ability to examine the underlying covariates opens up a capability to modify soil property predictions at individual sites based on local knowledge. For example, where one knows that the parent material or land use at a particular location is different to that recorded from the digital data source, then the regression models could be re-run over these sites with the corrected covariate values. As with all digital soil maps, they could be relatively easily upgraded as new, more reliable and detailed covariate layers become available.

Most models appear to explain a significant proportion (typically 35-55%) of the total variation in these properties. This performance is satisfactory given the high natural variation in these properties. In the case of OC, contents are known to vary widely, even under apparently uniform environmental conditions over small areas. For example, Wilson et al. (2010) report high variance in soil carbon over 25 m x 25 m monitoring plots in woodland, pasture and cropland of northern NSW; Pringle et al. (2011) report similarly high sampling requirements for carbon stock measurements to overcome local variations in the rangelands of north Queensland. Variation appears to occur down to the pedon scale (metres) and can thus be very difficult to accurately model at the regional or broader scale. Significant “intrinsic” factors rather than purely externally controlled “extrinsic” factors (Phillips 2001) may be important for many of these soil properties. The need to consider “chaotic” behaviour in soil modelling is also raised by Addiscott and Mirza (1998).
3.4.2 Comparison with other modelling approaches

It was revealed that the pragmatic approach adopted here was generally comparable in predictive performance to modelling using the Cubist piecewise linear decision tree approach with more sophisticated covariates. Overall, it is evident that the Cubist modelling approach is generally slightly superior in its predictive performance to the pragmatic MLR approach. The addition of the more sophisticated covariates to either modelling approach results in a further slight improvement, however if the pragmatic silica index (lithology) covariate is removed, there is generally a marked decrease in model performance. Where these models are applied to generate digital soil maps, there may be relatively greater improvement by using the more sophisticated approach because of the greater reliability of the covariate layer down to individual pixel level.

There appear to be few other straightforward quantitative models using readily interpreted covariates on key soil properties carried out in Australia or even globally with which to compare the current results. The current models improve on the regression models for OC, pH\textsubscript{w}, sum-of-bases and clay reported by Gray et al. (2009) based on their study of the ISRIC WISE Global database, which had R\textsuperscript{2} values between 0.1 and 0.4. The pragmatic regression model for OC in Laos prepared by Phachomphon et al. (2010) had an R\textsuperscript{2} value of 0.36, but when co-kriging techniques were applied this rose to 0.42.

Recent digital soil maps for OC over Australia prepared by Viscarra Rossel et al. (2014) using the Cubist modelling technique with residual kriging yielded validation statistics including a mean Lin’s concordance of 0.81. Other models developed for OC DSM projects in Australia and overseas for areas greater than 500 km\textsuperscript{2} generally had R\textsuperscript{2} values between 0.2 and 0.6 (Minasny et al. 2013). The national scale digital soil maps over Australian agricultural zones developed by Henderson et al. (2005), using Cubist techniques on over 40 auxiliary variables, demonstrated validation R\textsuperscript{2} values for topsoil up to 0.67 for pH\textsubscript{ca}, 0.41 for OC, 0.44 for clay and 0.62 for total P. The pH\textsubscript{ca} model of Reuter et al. (2008) over Europe using regression kriging with 54 auxiliary variables had an R\textsuperscript{2} of 0.43. Gonzalez et al. (2008) in Honduras gained R\textsuperscript{2} values of 0.45 for pH, 0.17 for clay and 0.24 for sand using Gaussian process models with 32 variables.
Chapter 3: Pragmatic models for soil prediction in eastern Australia

The strength and effectiveness of the models and the resulting digital soil maps generated in this study appear to be comparable with most of these previous studies. The resulting digital soil maps may be generally slightly less reliable than when using the most sophisticated techniques and covariate assemblage, but the difference is not large. This study has shown that when reliable lithology data is applied, models and resulting digital maps can be even more reliable than when using remotely sensed parent material data such as gamma radiometrics and NVIR spectroscopy alone. Furthermore, models and maps derived from the current pragmatic approach have the benefit of being user friendly, i.e., readily applied, understood and interpreted, particularly by non-DSM expert users. They may thus provide a potentially useful introduction to many soil scientists to the whole digital soil modelling and mapping arena. The need to provide DSM products that are easily discovered, understood, accessed and used is stressed by Wilson and Bleys (2008) and Wilson et al. (2012).

3.4.3 Pedologic insights - influence of factors

Useful pedologic insights into the factors controlling the levels and distribution of these soil properties can be gained from the derived regression models, particularly the partial regression coefficients as presented in Table 3.8. The standardised regression coefficients for each covariate as presented in Table 3.7, and the t and p values from the regression models (available from authors), also inform on the relative influence of each covariate in the models, although the interpretation of these coefficients and values is not always straightforward. From the data it is possible to gain useful insights into the relative roles of the various soil-forming factors in controlling the levels and distributions of many key soil properties at the eastern Australian scale:

- Organic carbon (%) – appears to be predominantly controlled at near surface depths by precipitation and maximum temperatures followed by parent material, with topography, land use and aspect being of lesser significance. At deeper depth intervals, maximum temperature appears to have a lower relative influence and parent material a higher relative influence. The apparent relatively low potential influence of land use at this scale is noteworthy, bearing in mind this is the only factor that humans can easily alter in order to promote soil organic carbon. This
mirrors results of Badgery et al. (2013) who found only weak significance of land management relative to other environmental factors in Lachlan and Macquarie catchments of central NSW.

- $\text{pH}_{\text{ca}}$ – appears to be predominantly controlled by precipitation followed by parent material and maximum temperatures, then of moderate influence is topography, with only a weak influence from land use and aspect. Again, the relative influence of maximum temperature drops off at deeper depth intervals. Note that pH, and other properties reflecting soil fertility, may be a controlling factor for land use rather than the reverse, with intensive agricultural uses being associated with higher pH soils. Thus, the apparent trends as revealed by the regression relationships should be treated with caution when considering pH behaviour under changing land use.

- Sum-of-bases and CEC – these appear to be very strongly controlled by parent material, then precipitation and maximum temperatures having moderate influence, with aspect, land use and topography being of lesser significance. As for pH, the apparent trends in base character with land use may reflect the association of intensive agricultural activity with naturally fertile soils, rather than a change in base content or CEC as a result of land use change.

- P total – appears to be predominantly controlled by parent material (also reported by Walker and Syers 1976) and precipitation, followed by maximum temperatures and land use, with topography and aspect not being significant at this scale.

- Clay, sand and silt – these all appear to be very strongly controlled by parent material, with climate factors having a moderate influence and land use, topography and aspect having lesser influence. At this scale, topography has only a slight influence but when regional climate influences are removed at a more local scale it does have a strong influence. Clay levels increase and sand levels decrease in lower positions in the landscape.

   It is important to appreciate that these results apply at this broad eastern Australian scale and their apparent influence will likely differ at local scales. The apparent lesser influence of topography is a consequence of this scale factor. As noted above, at this near national scale, the influence of topography appears to be subsumed by the influence of climate, to the point where it comes out as only a weakly significant
variable for many soil properties. This scale issue may also impact on the apparent influence of other covariates such as land use \((\text{land disturbance index})\), which may be much more significant when considered at a local scale. It also should be recognised that the models typically explain only half or less of the total variation, so other unidentified factors may also be of significant influence.

### 3.4.4 Future climate and land use change

The models provide a broad quantitative estimate of potential change in soil property levels given differing climatic and land use scenarios that may be present in the future. For example, if one considers a hypothetical situation some 50 years in the future where a locality undergoes a 2 degree rise in annual maximum temperature and a 200 mm annual precipitation drop. If the soil \((0-10 \, \text{cm})\) currently had an \(\text{OC}\) level of 4.0\%, a \(\text{pH}_{ca}\) level of 6.0, and a total base content of 10 cmol\(_c/\)kg, by applying the models we might expect the soil under the future climate conditions to have \text{OC} decreased by 1.0\%, \(\text{pH}_{ca}\) to have increased by 0.4 units and base content to have increased by 1.8 cmol\(_c/\)kg, assuming re-equilibration with the new environmental conditions. In addition, if the land use at the locality was to change from undisturbed woodland to crop \((5 \, LDI \, \text{unit increase})\), the \text{OC} levels may drop by a further 1.2\% down to approximately 1.8\%. These predictions are based on an assumption that equilibrium with the prevailing climate and land use conditions are reached, which may be unrealistic over short time frames.

Such predictions may be important in understanding how eastern Australian soils are likely to behave under altered future conditions, given potential climate change and land use pressures. They may have particular bearing on our understanding of the potential of these soils to sequester carbon.

### 3.4.5 Sources of uncertainty

Apart from the wide naturally occurring variation in soil properties as noted earlier, there are several inherent weaknesses in the analytical process used here, which serve to reduce the strength and predictive ability of the models and resulting digital soil maps. These give rise to the wide variations in predictions demonstrated by the upper and lower 95\% confidence level maps. Primary sources of uncertainty include:
Chapter 3: Pragmatic models for soil prediction in eastern Australia

- Simplification in covariate categories: the conversion of parent material descriptors to average silica% values in the silica index is a simplification, eg, all shales assumed to be 62%, whereas in reality they vary from 50 to 75%. The land disturbance index (LDI) is a very coarse indicator of land use, land management and biotic conditions at a site. The topo-slope index similarly involves grouping a large range of topographic conditions into a single index value. There was no effective inclusion of the soil age factor.

- Weaknesses in model development data: these include the lack of representation of some environments, possible errors in grid coordinates (affecting climate values) and inaccurate site descriptions. Potential errors in parent material identification are a particular concern, in many cases the recorded geology in a profile description may not be the true parent material. The LDI does not consider the period of time a new land use regime has been in operation, thus soil property imprints from a previous land use may still be evident in the analysed soil. Where the 2006 digital land use map was relied upon, errors due to scale may apply and the current recorded land use may differ from the use at the time of profile collection.

- Weaknesses in the spatial data grids for DSM development: errors will occur due to the scale (i.e., pixel or polygon size), particularly with coarse scale datasets. There may be considerable lithological variation within individual geological units and the associated map polygons. Likewise, land use may vary from that identified in the land use grid.

- Laboratory analysis errors: including: sample collection and handling errors, differences due to different laboratory techniques used and, for OC, possible under-estimation of values by the Walkley-Black method.

The presence of polygenetic soils may also introduce further weaknesses into the models and derivative maps. This issue arises when soils are old enough to have been influenced by previous climatic conditions and they therefore carry an imprint from those conditions, rather than being entirely influenced by current climate conditions. However, despite these weaknesses, the validation results are promising and suggest the models are useful in providing first approximations of soil properties across the landscape in eastern Australia.
Chapter 3: Pragmatic models for soil prediction in eastern Australia

3.5 Conclusion

Readily applied quantitative models to predict a range of key soil properties have been developed for eastern Australia. The multiple linear regression models are of generally moderate statistical strength (Lin’s concordance approximately 0.7) for OC% (0-10 cm depth), pHca, sum-of-bases, CEC and sand%, but only of low-moderate strength (Lin’s concordance 0.4-0.6) for OC% (>10 cm depths), total P, clay% and silt%. Overall, they provide a pragmatic tool for modelling and first approximation prediction of soil properties in this part of the continent. The models may have particular application in facilitating soil property predictions at individual sites, using only field collected and climate data, without the need for complex remotely sensed data sources and computer systems. They can also be used to prepare digital soil maps that appear to be at least moderately effective, as undertaken over NSW and a large regional catchment in this study. The models and derived maps have the potential for ready application into a range of agricultural, ecological, climatic and other environmental modelling systems.

The models and derived maps appear to compare favourably with those derived from more sophisticated approaches such as the Cubist piecewise linear decision tree technique using advanced remotely sensed covariates. This study has demonstrated there is overall only a slight improvement in validation statistics with some more sophisticated techniques. More specifically, it was observed that:

- the MLR technique is slightly less effective than the Cubist technique
- for the development of models, the pragmatic covariates from site specific data used in this study appear to be equal to or outperform the remotely sensed covariates used
- for the preparation of digital soil maps, remotely sensed covariates appear slightly more robust than broader polygon based pragmatic covariates, as they provide more accurate data at individual pixel level.
- lithology (silica index) emerges as a powerful covariate, at least for this study area. Where reliable fine scale lithology data is available it has the potential to significantly strengthen digital soil modelling and mapping products.
The pragmatic digital soil models and maps presented here have the benefit that the logic and covariate data underlying them can be readily understood, accessed and interpreted, particularly by non-DSMM experts. This at least partly compensates for their apparent slight loss of performance relative to more advanced products. This approach may therefore serve as a useful introduction to the more sophisticated DSMM approaches for many soil scientists and encourage greater acceptance of DSMM products and strategies generally.

The models also provide useful broad scale pedologic insights into the factors governing soil property levels and their variability in eastern Australia and beyond. Quantitative estimates on the influence of the main soil-forming factors on different soil properties have been derived, such as the change in a soil property per degree change of temperature. The models can be used to provide specific estimates of soil property change under future altered climate and land use scenarios. Ongoing research will attempt to improve the effectiveness of these pragmatic models, through for example the trialling of other readily available covariates.

Acknowledgements

The authors acknowledge the provision of soil datasets by CSIRO and the Queensland, NSW, Victorian, Tasmanian and South Australian state government natural resource agencies. Peter Smith, Hugh Jones and Senani Karunaratne provided important assistance and statistical advice. Useful comments and suggestions were provided by anonymous reviewers.

3.6 References


Badger W, Simmons AT, Murphy BW, Rawson A, Andersson KO, Lonergan VE, van de Ven R, 2013. Relationship between environmental and land-use variables
on soil carbon levels at the regional scale in central New South Wales, Australia. *Soil Research* 51, 645-656.


Chapter 3: Pragmatic models for soil prediction in eastern Australia


Holmes KW, Odgers NP, Griffin EA, van Gool D, 2014. Spatial disaggregation of conventional soil mapping across Western Australia using DSMART In: Arrouays
Chapter 3: Pragmatic models for soil prediction in eastern Australia


Chapter 3: Pragmatic models for soil prediction in eastern Australia


Chapter 4: Factors controlling soil organic carbon stocks with depth over eastern Australia

Abstract

Understanding the potential of soil to store organic carbon (SOC) is important for potential climate change mitigation strategies and assessing soil health issues. We examined the factors controlling SOC storage in eastern Australian soils and how these vary with depth. Models were developed using a set of readily interpreted covariates to represent key soil-forming factors together with multiple linear regression and Cubist piecewise decision tree techniques. Independent validation demonstrated concordance correlation coefficients up to 0.68 for SOC density in near surface layers but progressively decreasing with depth.

The results demonstrate the key role of climate (rainfall and maximum temperatures) in controlling SOC stocks, with parent material (lithology) and vegetation cover also being key drivers, whilst topography and aspect are of lesser influence, at least at this sub-continental scale. The relative influence of temperature and land use/vegetation cover decreases with depth, while that of parent material increases. The necessity of considering a combination of factors when deriving meaningful estimates of current or projected SOC storage is demonstrated with quantitative estimates of SOC stocks in 45 different climate-parent material-vegetation cover regimes. Average SOC stocks in the 0-30 cm depth interval range from 16.3 Mg ha$^{-1}$ (tonnes/ha) in dry, highly siliceous parent material environments with low vegetation cover, up to over 145.0 Mg ha$^{-1}$ in wet, mafic parent material environments with high vegetation cover. Results suggest that the proportion of SOC stock in the 30-100 cm interval as a proportion of the top 100 cm varies from a low of 41% in wet climates up to a high of 59% in dry climates. Climate appears to be the dominant controller of subsoil SOC storage proportion, with parent material and vegetation cover also having restricted influence.

Keywords: carbon sequestration, soil geography, distribution, digital soil mapping, climate change
Chapter 4: Factors controlling SOC with depth in eastern Australia

4.1 Introduction

Soil organic carbon is a major component of the global carbon budget, with the estimated 1500 Gt representing more than the combined stocks of the atmosphere and biosphere (Eswaran et al. 2000; Lal 2004a). There is now widespread recognition that maximising storage of carbon in our soils offers a potentially crucial avenue to offset increasing atmospheric carbon levels and thereby help mitigate the effects of human induced climate change (Lal et al. 2007; Smith 2012; IPCC 2014). The association of soil organic carbon (SOC) with soil health and agricultural productivity provides an added incentive to promote soil carbon levels (Baldock et al. 2009a; Sanderman et al. 2010). The potential importance of soil carbon to humankind was expressed by Lal (2004b):

“The close link between soil C sequestration and world food security on the one hand and climate change on the other can neither be overemphasized nor ignored”.

Uncertainty remains regarding the relative importance and interplay of the various driving factors and mechanisms that control SOC storage. There is a need to clearly understand and quantify potential SOC levels under different environmental and land use combinations, this being imperative if carbon trading schemes are to be effectively implemented as a means to help address climate change (Sanderman et al. 2010; Wilson et al. 2011; Cotching 2012; Badgery et al. 2013; Viscarra Rossel et al. 2014; IPCC 2014). Difficulties achieving this have been attributed to the inherent variability of SOC under uniform soil types (Batjes 1996; Cotching 2012) even under localised areas with apparently uniform environmental conditions (Cerri et al. 2000; Wilson et al. 2010), and difficulty in reliably estimating bulk densities (Wilson et al. 2011).

The key driver of soil organic carbon is widely reported to be climate, broadly comprising precipitation and temperature (Jenny 1980; Bui et al. 2009; Minasny et al. 2013; Viscarra Rossel et al. 2014; Hobley et al. 2015). There is however less consistency in the literature regarding the relative influence attributed to other factors, such as those relating to land use/management (including agricultural intensity and vegetation cover levels), parent material (including lithology and clay content), and topography (including slope position and aspect). The relative influence of some factors varies with scale. For example, land use and topographic factors, appear to increase in
importance at finer, more localised scales (Minasny et al. 2013). There is widespread recognition that it is the interaction of these drivers that determines final SOC levels, and that individual factors cannot be considered in isolation (Baldock et al. 2009b; Murphy et al. 2010; Powers et al. 2011; Mishra et al. 2012; Xiong et al. 2014; Mayes et al. 2014).

Although most SOC research to date has focused on surface layers (generally down to 30 cm) it is also being increasingly recognised that subsurface soils play an important role in SOC storage (Rumpel and Kögel-Knabner 2011; Lou et al. 2010; Cotching 2012), particularly considering the higher total volumes and bulk densities of these soils and the greater stability and longevity of SOC than that in surface soils (Fontaine et al. 2007; Sanderman et al. 2010; Wilson and Lonergan 2013). However, much uncertainty remains about how the factors controlling SOC levels vary in subsurface relative to surface soils (Rumpel and Kögel-Knabner 2011; Hobley et al. 2015).

Further work is required to elucidate the relative levels of influence of a range of important factors in controlling SOC stocks. We need to understand and quantify how the influence of these drivers change with increasing depth in the soil profile. More quantitative data are needed on the combined influence of the key factors and how they work together to produce different SOC stocks in different environmental regimes. Only by understanding these mechanisms can we hope to develop realistic strategies to promote long term increases in soil carbon levels. This study attempts to address these issues, through a digital soil modelling and mapping program undertaken over eastern Australia, covering an area of 3.8 million km². Our strategy was to:

- prepare digital soil models of SOC density (kg m⁻³) over the eastern states of Australia at five depth intervals to 100 cm.
- derive quantitative estimates of the influence of several key factors at the different depth intervals, using standardised regression coefficients from MLR models and frequency of use in Cubist models
- from the resulting digital soil maps, derive estimates of SOC stock (Mg ha⁻¹) over 45 different climate-parent material-vegetation cover sub-classes at 0-30 cm and 30-100 cm intervals.
- Derive relative proportions of the SOC stock over these two depth intervals for each sub-class.
Chapter 4: Factors controlling SOC with depth in eastern Australia

- Identify and discuss key trends in the results.

4.2 Methods

In overview, a legacy dataset of soil profiles with associated environmental covariates over the eastern states of Australia was used to develop multiple linear regression (MLR) models and Cubist piecewise linear decision tree models that described the relationship of SOC density relative to key soil-forming factors. The models, prepared at a number of depth intervals down to 1 m, were examined to determine the relative influence of each factor and how these vary with increasing depth in the profile. Digital soil maps were prepared, which were then partitioned into 45 subclasses based on climate, parent material and vegetation cover, for which average SOC stock levels were determined and key trends examined. A broad overview of the region is provided in Gray et al. (2015).

4.2.1 The soil dataset

A dataset of 5188 soil profiles containing SOC laboratory results and extensive site data was compiled over eastern Australia, a subset of that reported in Gray et al. (2015) (Figure 4.1) and mostly collected during the years 1970 to 2010. Breakdown by jurisdiction was as follows: NSW: 1778, Queensland: 1499, CSIRO (eastern states) 757, South Australia: 586, Tasmania: 345 and Victoria 223. Soil organic carbon values reported for each soil horizon over the entire original dataset were converted into five standard depth intervals: 0-5, 5-15, 15-30, 30-60, and 60-100 cm using the equal-area spline process of Bishop et al. (1999) and Malone et al. (2009). These intervals conform with those adopted in the Soil and Landscape Grid of Australia (TERN 2014; Viscarra Rossel et al. 2015) (www.csiro.au/soil-and-landscape-grid) and GlobalSoilMap.net (Sanchez et al. 2009) down to the 100 cm level.

Excluded from the analyses were organic soils including Organosols under the Australian Soil Classification (Isbell 2002) or Peats under the Great Soil Group Classification (Stace et al. 1968), which are equivalent to Histosols from Soil Taxonomy (Soil Survey Staff 2010) as these are not common in the region (eg, only 0.06% of NSW) and are difficult to incorporate into models, because their relatively extreme SOC levels tend to distort normal trends in SOC distribution. Additionally, all
sites with less than 0.1% SOC in the uppermost depth interval were excluded as many of these were considered unreliable. For most profiles (approx 95%), the Walkley-Black wet oxidation method had been used to derive the SOC values, with LECO and other combustion methods used for the remainder (Rayment and Lyons 2011). No correction factors were applied to account for possible under-estimations of SOC values by the Walkley-Black method used in earlier decades (Skjemstad et al. 2000), as there is uncertainty regarding the most appropriate correction factor and whether or not it had already been applied (Conyers et al. 2011; Bui et al. 2009).

SOC density (kg m\(^{-3}\)) was added to the dataset by applying bulk density estimates from the Soil and Landscape Grid of Australia bulk density layers, derived by digital soil mapping at 3 arcsecond grid spacing (Viscarra Rossel et al. 2014). The following simple relationship was applied:

\[
\text{SOC} (\text{kg m}^{-3}) = \text{SOC}\% \times \text{BD} (\text{Mg m}^{-3}) \times 10
\]

Figure 4.1. SOC modelling points over eastern Australia (shading represents reliable modelling area)

### 4.2.2 The covariates

A relatively small number of readily interpreted covariates were selected to represent the key soil-forming factors of climate, parent material, relief and biota as outlined below. The choice and small number of covariates was designed to reduce the
extent of correlation between them and facilitate clear interpretation on the relative influence of each soil forming factor. The range and variability of these covariates was presented in Gray et al. (2016).

Climate

- *Mean annual precipitation* (mm pa, Precip) – derived from 2.5 km Australia wide climate grids from the Australian Bureau of Meteorology with interpolation of cell values down to a 100 m grid, using the ArcGIS Interpolation Spline tool. The dataset represents mean values obtained over the 1961-1990 period
- *Mean annual daily maximum temperature* (°C, T\text{max}) – as above
- *Precipitation/max temperature ratio* (P/Tm ratio) – the ratio of the above two covariates was derived to provide a single climatic index, used only for post-modelling interpretation purposes. This was classified into three classes: dry: <30; moist: 30-60; wet: >60

Parent material

- *Silica index (lithology)* – an index denoting the silica (SiO\text{2}) content to represent the lithological character of the parent material, estimated using documented average chemical composition of the materials (Gray et al. 2014; Gray et al. 2015). For example, granite with moderately siliceous lithology has a silica content of approximately 73%, while basalt with mafic lithology has a silica content of approximately 48%. Higher silica content parent materials typically give rise to soils with more quartzose, coarser, sandier textures with lower chemical fertility. Parent material descriptors recorded at each site were used to derive silica indices for model development; the 1:250 000 scale NSW Geological Survey polygonal geology map and 1:1 million scale Geoscience Australia geology map were used for the final digital soil maps. For post-modelling interpretation purposes, these were grouped into five classes as shown in Table 4.1, which also presents typically associated soil types.


Table 4.1. Parent material classes and typically associated soils

<table>
<thead>
<tr>
<th>Parent material class</th>
<th>Silica (approx. %)</th>
<th>Base cations (approx. %)</th>
<th>Common examples</th>
<th>Typical ASC soils</th>
<th>Typical Soil Taxonomy soils</th>
<th>Typical WRB soils</th>
</tr>
</thead>
<tbody>
<tr>
<td>Siliceous - extreme</td>
<td>&gt;82</td>
<td>&lt;=4</td>
<td>quartz sands (alluvial and aeolian), pure quartzite, chert, quartz sandstone &amp; siltstone, granite, rhylolite, siliceous tuff</td>
<td>arenic</td>
<td>Spodosols; (Quartzipsamments)</td>
<td>Podzols</td>
</tr>
<tr>
<td>Siliceous - high</td>
<td>70-82</td>
<td>4-10</td>
<td>Kandosols; Tenosols</td>
<td>low clay activity</td>
<td>Alfisol, Aridosols &amp; Oxisols; Inceptisols; (Psamments)</td>
<td>Arenosols; Durisols; Lixisols; Planosols; Solonetz</td>
</tr>
<tr>
<td>Intermediate /siliceous transitional</td>
<td>60-70</td>
<td>10-13</td>
<td>adammellite, granodiorite, tonalite, dacite, syenite, monzonite, greywacke, feldspathic/lithic sandstone, mudstone, argillites (shale, slate, etc)</td>
<td>Chromosols, Kurosols &amp; Sodosols</td>
<td>high clay activity</td>
<td>Alfisol, Aridosols &amp; Oxisols; (Udults)</td>
</tr>
<tr>
<td>Intermediate - lower</td>
<td>52-60</td>
<td>13-20</td>
<td>diorite, andesite, alluvial grey &amp; brown clays; most calcareous materials (eg, limestone)</td>
<td>Dermosols; Ferrosols; grey &amp; brown Vertosols</td>
<td>Mollisols</td>
<td>Chernozems; Kastanozems</td>
</tr>
<tr>
<td>Mafic</td>
<td>&lt;=52</td>
<td>&gt;20</td>
<td>gabbro, dolerite, basalt, amphibolite, alluvial black cracking clay</td>
<td>black Vertosols, Ferrosols</td>
<td>Mollisols</td>
<td>Nitisols; Phaeozems; Vertisols</td>
</tr>
</tbody>
</table>

1 Typical soil types are first approximations only; most extend into adjoining parent material classes.
2 Cations comprise average Ca, Mg, Na and K oxides
3 Australian Soil Classification (Isbell, 2002)
4 Common suborders or sub-groups given in italics (Soil Survey Staff, 2010)
5 World Reference Base for Soil Resources (IUSS Working Group WRB, 2014)

Other sources: Gray and Murphy (1999); Gray et al. (2011, 2014, 2015); Isbell et al. (1997)
Chapter 4: Factors controlling SOC with depth in eastern Australia

Relief

- **Topo-slope index (TSI)** – an index that can be derived from field observations that combines topographic position and slope gradient. It represents the degree to which a site is subject to depletion (1) or accumulation (6) of water, soil particles and chemical materials (Gray et al. 2015). Model development relied on soil surveyor site data for individual sites; map development used a 100 m DEM (resampled from Gallant et al. 2010) to derive the Topographic Position Index (Jenness 2006) and slope%.

- **Aspect index (Asp)** – an index to represent the amount of solar radiation received by sites, ranging from 1 for gentle N or NW facing slopes to 10 for steep S and SE slopes (Gray et al. 2015). The required data was derived from field collected site data or a 100 m DEM.

Biota

- **Land disturbance index (LDI)** – an index that reflects the intensity of disturbance associated with the land use (1: natural ecosystem to 6: intensive cropping) (Gray et al. 2014, modified from NCST, 2009). For model development, site land use was taken from field profile descriptions; for the final digital SOC map it was derived from 1:25 000 scale polygonal land use mapping (OEH, 2007).

- **Vegetation cover (Veg_cov)** – total vegetation cover (photo-synthetic and non photosynthetic) derived from 2011 MODIS fractional vegetation data, 90 m grids (Geurschman et al. 2009). For post modelling interpretation purposes, these were grouped into three classes: Low <= 60%, moderate 60-80%, high >80%). Generally, vegetation cover decreases with increasing levels of disturbance (higher LDI), thus these two covariates would display at least some collinearity. Although they are treated separately in the statistical analysis, they are considered jointly in the later discussions.

4.2.3 Developing models and statistical analysis

Analysis was carried out using R statistical software (R Core Team 2013). The soil dataset was apportioned 80% as training data and 20% as validation data, with modelling by multiple linear regression (MLR) and Cubist linear piecewise decision tree models (Quinlan 1992) using the Cubist package of Kuhn et al. (2014). A natural
log transformation was applied to the SOC values to address the observed skewness in the response. Models were prepared for both SOC concentration (%) and density (kg m$^{-3}$), but as density units allow SOC stock calculations they were considered the more useful, and only these models are presented.

The models for each depth interval were validated using the validation datasets. Lin’s concordance correlation coefficient (CCC) was used to measure the level of agreement of predicted values with observed values, relative to the 1:1 line (Lin 1989). Also determined were root mean square error (RMSE), mean error and standardised RMSE (being RMSE/mean estimate). Standardised regression coefficients of covariates in the MLR models and data on the frequency of use in the Cubist models were derived to inform on the relative influence of each covariate in the models.

Maps of SOC density at the five depths were prepared using the Cubist models. The layers were combined to give just two maps for the depth intervals 0-30 and 30-100 cm. Note that this simple amalgamation may have introduced additional prediction errors and in future it would be preferable to create a new dataset covering these specific depth intervals. SOC density at these two new intervals were converted to SOC stocks (Mg ha$^{-1}$) and then partitioned into 45 sub-classes according to (i) climate regime ($P/Tm\text{\_ratio}$), (ii) lithology (silica) class and (iii) vegetation cover class. Mean values of each sub-class and their standard deviations (SD) were recorded from their GIS layer information. The mean and 95% spread of predictions (based on 1.96 x SD) were plotted to assist in interpretation. The ratios of SOC over the 30-100 cm interval relative to the total 0-100 cm depth interval for each of these 45 sub-classes were also calculated and plotted in a similar manner to above.

4.3 Results

4.3.1 The models and validation

The MLR models for SOC density (kg m$^{-3}$) for the five depth intervals are presented in Table 4.2. It is evident that the models are strongest in the near surface layers, with $R^2$ reaching a maximum of 0.45 in the 5-15 cm interval model, but they decline with depth with $R^2$ values dropping to less than 0.15 in the 60-100 cm interval. Major declines in the F values are also observed with depth. The Cubist models also revealed a similar decrease in model strength with depth. The models for SOC
Chapter 4: Factors controlling SOC with depth in eastern Australia

concentration (%) were considerably stronger than those for SOC density, with $R^2$ values generally some 15-20% higher, probably mainly due to their non-reliance on bulk density estimates, but these models are not presented here.

Table 4.2. MLR models for SOC density (kg m\(^{-3}\))

<table>
<thead>
<tr>
<th>Depth (cm)</th>
<th>Regression relationship</th>
<th>N</th>
<th>$R^2$</th>
<th>F</th>
<th>Resid. SE(^1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-5</td>
<td>OC0_5d = \exp(3.727+0.000158<em>Rain-0.0471</em>Tmax-0.0178<em>Silica-0.0294</em>TSI+0.0144<em>Asp+0.0180</em>Veg_cov)</td>
<td>3429</td>
<td>0.42</td>
<td>347</td>
<td>0.59</td>
</tr>
<tr>
<td>5-15</td>
<td>OC5_15d = \exp(3.433+0.000308<em>Rain-0.0390</em>Tmax-0.0175<em>Silica-0.0328</em>TSI+0.0155<em>Asp+0.0175</em>Veg_cov)</td>
<td>3222</td>
<td>0.45</td>
<td>434</td>
<td>0.56</td>
</tr>
<tr>
<td>15-30</td>
<td>OC15_30d = \exp(3.125+0.000487<em>Rain-0.0179</em>Tmax-0.0216<em>Silica+0.0221</em>Asp-0.0362<em>LDI+0.0191</em>Veg_cov)</td>
<td>2593</td>
<td>0.31</td>
<td>197</td>
<td>0.65</td>
</tr>
<tr>
<td>30-60</td>
<td>OC30_60d = \exp(3.074+0.000415<em>Rain-0.0231</em>Tmax-0.0224<em>Silica-0.0399</em>LDI+0.00793*Veg_cov)</td>
<td>1899</td>
<td>0.24</td>
<td>119</td>
<td>0.67</td>
</tr>
<tr>
<td>60-100</td>
<td>OC60_100d = \exp(1.767+0.000233<em>Rain-0.0179</em>Silica-0.0205<em>Asp+0.00934</em>Veg_cov)</td>
<td>1341</td>
<td>0.14</td>
<td>54</td>
<td>0.64</td>
</tr>
</tbody>
</table>

\(^1\) residual standard error

Results of validation of the SOC density models, using the initially withheld validation dataset, are presented in Table 4.3. It can be observed that the highest validation performance is achieved for the 5-15 cm depth intervals using the Cubist approach, where the concordance values reach 0.68 (Figure 4.2). Again, a major decline in performance with depth is clearly evident, as also shown by the standardised RMSE. The Cubist models consistently outperformed the MLR models, with concordance values being 5-15% higher.

Table 4.3. Validation of MLR and Cubist models

<table>
<thead>
<tr>
<th>Depth (cm)</th>
<th>Model type (all log scale)</th>
<th>N</th>
<th>$R^2$</th>
<th>Lin’s CCC</th>
<th>RMSE</th>
<th>ME(^1)</th>
<th>Std RMSE(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-5</td>
<td>MLR</td>
<td>850</td>
<td>0.46</td>
<td>0.62</td>
<td>0.56</td>
<td>0.003</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>Cubist</td>
<td>1019</td>
<td>0.50</td>
<td>0.66</td>
<td>0.53</td>
<td>-0.0004</td>
<td>0.037</td>
</tr>
<tr>
<td>5-15</td>
<td>MLR</td>
<td>833</td>
<td>0.46</td>
<td>0.64</td>
<td>0.55</td>
<td>-0.006</td>
<td>0.041</td>
</tr>
<tr>
<td></td>
<td>Cubist</td>
<td>977</td>
<td>0.51</td>
<td>0.68</td>
<td>0.53</td>
<td>-0.02</td>
<td>0.040</td>
</tr>
<tr>
<td>15-30</td>
<td>MLR</td>
<td>631</td>
<td>0.34</td>
<td>0.50</td>
<td>0.62</td>
<td>-0.02</td>
<td>0.060</td>
</tr>
<tr>
<td></td>
<td>Cubist</td>
<td>780</td>
<td>0.42</td>
<td>0.58</td>
<td>0.58</td>
<td>-0.02</td>
<td>0.056</td>
</tr>
<tr>
<td>30-60</td>
<td>MLR</td>
<td>477</td>
<td>0.24</td>
<td>0.40</td>
<td>0.67</td>
<td>0.01</td>
<td>0.108</td>
</tr>
<tr>
<td></td>
<td>Cubist</td>
<td>588</td>
<td>0.29</td>
<td>0.45</td>
<td>0.64</td>
<td>-0.003</td>
<td>0.103</td>
</tr>
<tr>
<td>60-100</td>
<td>MLR</td>
<td>329</td>
<td>0.14</td>
<td>0.26</td>
<td>0.64</td>
<td>0.05</td>
<td>0.153</td>
</tr>
<tr>
<td></td>
<td>Cubist</td>
<td>416</td>
<td>0.12</td>
<td>0.27</td>
<td>0.65</td>
<td>0.04</td>
<td>0.155</td>
</tr>
</tbody>
</table>

\(^1\) Mean error, calculated as the mean predicted – mean observed values.

\(^2\) Standardised root mean square error, calculated as RMSE/mean of estimate.
4.3.2 Influence of covariates

The standardised regression coefficients for each covariate used in the MLR SOC density models are provided in Table 4.4. These provide useful indications of the relative influence of each covariate in the models and thus in driving the SOC content in soils over the various depth intervals. It reveals that in the upper depth intervals (0-5 and 5-15 cm) $T_{max}$, Silica and Veg cover appear to exert the dominant influence, Precip has moderate influence while TSI, Asp and LDI have only a relatively small influence. At the deeper intervals Silica becomes the clearly dominant influence with Precip and Veg_cov also being important. $T_{max}$ progressively declines in influence down to negligible levels, as does TSI also. It was noted that when Veg_cov was omitted from the MLR models, the significance of LDI generally slightly increased, suggesting these two factors are at least partially correlated.

Examination of the frequency of use of the different covariates in the Cubist models (Table 4.5) also provides a broad indication of their relative influence. It suggests that in the upper depth intervals (0-5 and 5-15 cm) $T_{max}$, Precip and Silica are dominant with Veg_cov also important, while TSI, Asp and LDI have only moderate influence. At the deeper levels Silica and Precip are clearly dominant, $T_{max}$ and TSI have moderate influence, whilst the other covariates have generally only minor
influence. Results are broadly similar to those derived from the MLR standard regression coefficients, but they suggest a greater influence of topography and lesser influence of Veg_cov and LDI at the deeper intervals.

Table 4.4. MLR standardised regression coefficients and relative rankings of covariates

<table>
<thead>
<tr>
<th>Depth (cm)</th>
<th>Precip</th>
<th>Tmax</th>
<th>Silica</th>
<th>TSI</th>
<th>Asp</th>
<th>LDI</th>
<th>Veg_cov</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-5</td>
<td>0.07 (4)</td>
<td>-0.29 (1)</td>
<td>-0.27 (3)</td>
<td>-0.07 (5)</td>
<td>0.03 (6)</td>
<td>-2 (7)</td>
<td>0.27 (2)</td>
</tr>
<tr>
<td>5-15</td>
<td>0.15 (4)</td>
<td>-0.24 (3)</td>
<td>-0.26 (1)</td>
<td>-0.08 (5)</td>
<td>0.03 (6)</td>
<td>- (7)</td>
<td>0.27 (1)</td>
</tr>
<tr>
<td>15-30</td>
<td>0.23 (2)</td>
<td>-0.10 (4)</td>
<td>-0.31 (1)</td>
<td>- (7)</td>
<td>0.05 (6)</td>
<td>-0.06 (5)</td>
<td>0.17 (3)</td>
</tr>
<tr>
<td>30-60</td>
<td>0.20 (2)</td>
<td>-0.11 (4)</td>
<td>-0.32 (1)</td>
<td>- (6)</td>
<td>- (6)</td>
<td>-0.06 (5)</td>
<td>0.11 (3)</td>
</tr>
<tr>
<td>60-100</td>
<td>0.12 (3)</td>
<td>- (5)</td>
<td>-0.27 (1)</td>
<td>- (5)</td>
<td>0.05 (4)</td>
<td>- (5)</td>
<td>0.15 (2)</td>
</tr>
</tbody>
</table>

1 ranking relative to other covariates in brackets
2 - no value indicates the covariate was not significant in the MLR model (P >0.1)

Table 4.5. Influence and relative ranking of covariates in Cubist models

<table>
<thead>
<tr>
<th>Depth (cm)</th>
<th>Precip</th>
<th>Tmax</th>
<th>Silica</th>
<th>TSI</th>
<th>Asp</th>
<th>LDI</th>
<th>Veg_cov</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-5</td>
<td>93 (2)</td>
<td>100 (1)</td>
<td>50 (4)</td>
<td>50 (4)</td>
<td>35 (6)</td>
<td>21 (7)</td>
<td>65 (3)</td>
</tr>
<tr>
<td>5-15</td>
<td>97 (2)</td>
<td>100 (1)</td>
<td>69 (3)</td>
<td>47 (4)</td>
<td>36 (6)</td>
<td>16 (7)</td>
<td>47 (4)</td>
</tr>
<tr>
<td>15-30</td>
<td>98 (1)</td>
<td>74 (3)</td>
<td>87 (2)</td>
<td>12 (6)</td>
<td>23 (4)</td>
<td>13 (7)</td>
<td>17 (5)</td>
</tr>
<tr>
<td>30-60</td>
<td>94 (2)</td>
<td>49 (3)</td>
<td>100 (1)</td>
<td>38 (4)</td>
<td>14 (5)</td>
<td>0 (7)</td>
<td>12 (6)</td>
</tr>
<tr>
<td>60-100</td>
<td>65 (2)</td>
<td>42 (3)</td>
<td>71 (1)</td>
<td>29 (4)</td>
<td>29 (4)</td>
<td>0 (7)</td>
<td>19 (6)</td>
</tr>
</tbody>
</table>

1 represents average percentage use in Cubist rule conditions and terminal linear models, as derived using varImp function in the “caret” package
2 ranking relative to other covariates in brackets

Table 4.6 lists the influence per unit change of the seven covariates, assuming other factors are held constant, over the 0-30 and 30-100 cm depth intervals. Values were derived using weighted averages of the partial regression coefficients from the MLR models. It is revealed for example that over the upper depth interval, for each 100 mm increase in annual precipitation there is 3.8% proportional increase in SOC density, and that with each degree C rise in annual maximum temperature there is a corresponding 2.9% proportional decrease in SOC density. The influence of both these climatic factors, particularly temperature, decreases in the lower depth intervals, as do most other factors apart from parent material which slightly increases.
4.3.3 Variation in carbon stocks with climate, parent material and vegetation cover

The derived digital soil maps of SOC stocks for the 0-30 cm and 30-100 cm depth intervals are presented in Figure 4.3. Using these maps with stratification by the three climate classes, five parent material classes and three vegetation cover classes, estimates of the mean SOC mass, and spread of predictions at 95% level, over each of these 45 sub-classes for both intervals were determined (Figures 4.4 and 4.5).
Chapter 4: Factors controlling SOC with depth in eastern Australia

Figure 4.4. Variation in SOC stock by climate, parent material and vegetation cover sub-classes (0-30 cm, Mg ha$^{-1}$, showing mean as dark line and 95% spread of predictions)

Figure 4.5. Variation in SOC stock by climate, parent material and vegetation cover sub-classes (30-100 cm, Mg ha$^{-1}$, showing mean as dark line and 95% spread of predictions)
The plots demonstrate generally uniform trends of increasing SOC stock with increasingly moist climate, increasing mafic character of parent material and increasing vegetative cover. SOC stocks in the upper interval vary from 16.3 Mg ha\(^{-1}\) (t/ha) in dry, highly siliceous parent material environments with low vegetation cover, up to 145.0 Mg ha\(^{-1}\) in wet, mafic parent material environments with high vegetation cover. It can be seen that the increase in SOC density that occurs when moving from an equivalent climate-vegetation cover environment is more pronounced over mafic parent material soils than it is over siliceous parent material soils, at least in absolute terms. In the lower 30-100 cm interval the SOC stocks likewise vary under different environmental conditions, ranging from 18.7 to 106.1 Mg ha\(^{-1}\). The differences are, however, somewhat less pronounced, giving the histogram a broadly flatter structure, mainly due the lesser relative influence of climate and vegetation cover.

4.3.4 Carbon stocks at different depths

Figure 4.6 presents the proportion of SOC stock in the subsoil (30-100 cm) relative to the top 100 cm calculated for each of the 45 climate-parent material-vegetation cover classes, giving the mean values and the 95% spread of values for each class. The wide spread in the upper and lower estimates is indicative of the high degree of uncertainty, as also borne out by the validation statistics, particularly for the lower depth. Nevertheless, it can be seen that mean values vary from a high of 59% in dry climates to a low of 41% in wet climates, although a large range of values is evident for most classes. The results indicate that in dry climates, the majority of carbon stock in the top metre is stored in the subsoil (30-100 cm), whereas in wet climates the majority is stored in the upper soil (0-30 cm). Climate appears to be the key factor driving these proportions. The main difference appears to occur between the dry and moist climate zones, with only a slight decrease between the moist and wet zones, as also borne out by the summary Table 4.7. Parent material and vegetation cover classes do not show strong trends, however relatively higher mean subsoil SOC storage proportions are observed with the extremely siliceous parent materials in the wet and moist climates, and with lower vegetation classes in the dry climates.

The relative increase in SOC stocks in subsoils in the drier inland areas is also demonstrated by the maps of Figure 4.3, which show a less pronounced decline for the lower soils than the upper soils. It is interesting to note from Table 4.7 that over the
whole of eastern Australia the ratio of SOC stock in the top 30 cm relative to the top 100 cm is almost exactly 50%, which is indicative of the relative spread of climate zones in this province. Although carbon stocks (Mg ha\(^{-1}\)) are approximately equivalent in both the upper and lower depth intervals, Table 4.7 reminds us that the carbon densities (kg m\(^{-3}\)) are significantly lower in the subsoils, but this is compensated for by the greater depth interval they cover (70 cm compared to 30 cm). Additional carbon stocks would of course be present below 100 cm, but at increasingly lower densities.

### Table 4.7. Density and stocks of SOC in upper and lower depth intervals to 100 cm

<table>
<thead>
<tr>
<th>Climate zone</th>
<th>Depth interval (cm)</th>
<th>Mean density (kg m(^{-3}))</th>
<th>SOC stock (Mg ha(^{-1}))</th>
<th>Proportion of stock 30-100 cm relative to top 100 cm (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dry</td>
<td>0-30</td>
<td>9.99</td>
<td>29.98</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>4.94</td>
<td>34.56</td>
<td></td>
</tr>
<tr>
<td>Moist</td>
<td>0-30</td>
<td>17.17</td>
<td>51.52</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>6.11</td>
<td>42.79</td>
<td></td>
</tr>
<tr>
<td>Wet</td>
<td>0-30</td>
<td>27.28</td>
<td>81.82</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>8.91</td>
<td>62.36</td>
<td></td>
</tr>
<tr>
<td>All eastern Australia</td>
<td>0-30</td>
<td>11.99</td>
<td>35.97</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>5.04</td>
<td>35.29</td>
<td></td>
</tr>
</tbody>
</table>

Figure 4.6. Proportions of carbon stock in 30-100 cm depth relative to top 100 cm, by climate, parent material and vegetation cover class over eastern Australia (showing mean as dark line and 95% spread of predictions)
4.4 Discussion

4.4.1 Strength of models

This study has developed models of SOC density based on seven readily interpretable covariates to provide insights into factors driving this soil property at varying depths. Validation results suggest the models are of at least moderate reliability, with Lin’s CCC values for SOC density approaching 0.70 for the near surface depth intervals but decreasing with depth. There is high uncertainty associated with results from the lower depth intervals. Validation results were somewhat higher for SOC concentration (%), possibly because they did not rely on previously modelled bulk density data, with its inherent uncertainties.

The strength and effectiveness of the digital SOC models generated in this study appear broadly comparable with previous studies. A digital SOC density map for 0-30 cm interval over Australia prepared by Viscarra Rossel et al. (2014) yielded strong validation statistics including a mean Lin’s concordance of 0.81. The SOC concentration maps over Australian agricultural zones developed by Bui et al. (2009) demonstrated validation $R^2$ values of 0.49 for topsoil and 0.36 for subsoil. Over the same region, Henderson et al. (2005) achieved model validation $R^2$ values of 0.41 for topsoil and 0.24 for subsoil. Hobley et al. (2015) achieved model $R^2$ values of 0.76 in SOC modelling to 30 cm over NSW. Other SOC models and maps developed for DSM projects in Australia and overseas for areas greater than 500 km$^2$ generally had $R^2$ values between 0.2 and 0.6 (Minasny et al. 2013).

The decline in model strength that occurs with depth, as observed in this study (Tables 4.2 and 4.3) and by other workers such as those referred to above, may be at least partly attributable to the decline in absolute SOC values that occurs with depth. With lower mean absolute values, but residual standard errors remaining constant, the model strength will decline. The lower influence of the climatic factors, particularly temperature, with depth, as discussed in the following section, also results in weaker models with depth.

The models developed in this study would appear sufficiently strong to draw useful conclusions on factors controlling SOC stocks. Although the strength of the models could potentially be improved through use of additional, more sophisticated
variables, such as remotely sensed data and advanced climatic and topographic
covariates, this may be at the expense of ease of interpretation. There is likely to be
significantly greater collinearity between covariates relating to similar soil-forming
factors, thereby reducing our ability to draw useful conclusions on the relative influence
of each factor.

4.4.2 Influence of individual factors driving SOC and variation with depth

The results of this study confirm that climate, parent material and vegetation
cover are the key driving factors for SOC density at the eastern Australian sub-
continental scale. Topography and the associated aspect appear of less influence at this
scale. The relative influence of the various factors does however change with depth
(Jobbágy and Jackson 2000) suggesting the mechanisms involved in the stabilization
and dynamics of SOC may be different at different depths (Rumpel and Kögel-Knabner
2011). The combined influence of these factors ultimately controls the SOC densities.

The dominant influence of climate, particularly at broad continental scales, is
almost universally recognised. It controls the production of organic matter and the
extent of its mineralisation and resulting loss from the soil with SOC levels highest
under cool moist conditions (Jenny 1980; Lal 2004a; Sanderman et al. 2010; Lou et al.
2010; Cotching 2012; Badgery et al. 2013). This current study shows that in near
surface layers (0-15 cm) temperature is the dominant driving force, being more
important than precipitation, also reported by Bui (2012) and Wang et al. (2014),
however at deeper levels the influence of temperature reduces and precipitation
becomes more dominant. Both the MLR and Cubist modelling processes in this study
reveal a steady decrease in the influence of temperature with depth, whereas
precipitation appears to increase to a maximum influence at mid depths (15-30 cm) then
decline again at deeper levels (>30 cm).

Parent material composition has been shown by this study to have a strong
influence on SOC content. More mafic, less siliceous materials are associated with
higher SOC levels. It becomes the clearly dominant influence in the mid and lower
layers. The importance of parent material in controlling SOC has also been reported by
several recent workers (Mishra et al. 2012; Chaplot et al. 2010; Vasques et al. 2010;
Powers et al. 2011; Wilson et al. 2011; Cotching 2012; Badgery et al. 2013; Viscarra
Rossel et al. 2014). Its influence is primarily due to its control of soil texture and clay
content, which serves to protect SOC from mineralisation (Oades 1988; Paustian et al. 1997; Baldock and Skjemstead 1999; Baldock et al. 2009b; Heckman et al. 2009), and its control of soil fertility and nutrient supply to promote production of organic material (Badgery et al. 2013). Parent material has a close association with soil type, which is almost universally recognised as having a major influence on SOC content (Batjes 1996; Eswaran et al. 2000; Lal 2004a; Cotching 2012; Xiong et al. 2014). The different parent material classes identified in this study have close relationship with groupings of soil types, as presented in Table 4.1.

The extent of vegetation cover is clearly another dominant factor controlling SOC. The results suggest the high importance of this factor in near surface layers, and then a decreasing influence with depth. However its relative ranking compared to other factors appears variable. Land use, as represented by LDI is shown to have a relatively low influence on SOC levels at this sub continental scale, but it may be at least partly masked by the vegetation cover factor. Its influence appears to decline further with depth, similarly reported by others (Lou et al. 2010; Wilson et al. 2010; 2011; Allen et al. 2013; Badgery et al. 2013), Nevertheless the factor is still considered important in subsoils (Guo and Gifford 2002; Wright et al. 2007; Meersmans et al. 2009; Follett et al. 2009; Vasques et al. 2010; Rumpel and Kögel-Knabner 2011). It has been suggested that SOC in deeper soil layers might reflect historic rather than current land use (Schulp and Veldkamp 2008; Wilson and Lonergan 2013). This suggests the SOC may not be in equilibrium with current environmental conditions.

Topography as represented by TSI and the associated aspect (Asp) are not revealed as strong controlling factors at this sub-continen tal scale, a pattern also noted by Minasny et al. (2013) and Hobley et al. (2015). At more local scales, where climate is more uniform and therefore less of an influence, topography can be shown to be a significant driving factor (McKenzie and Ryan 1999; Allen et al. 2013; Davy and Koen 2013).

Based on results from the two modelling approaches in this study, the ranking of relative importance of environmental and land use factors over the upper depth intervals (0-30 cm) would appear to be temperature > precipitation > parent material > vegetation cover/land use ~= topography/aspect. At the deeper levels (30-100 cm) the order would appear to be parent material > precipitation > vegetation cover/land use ~= temperature ~= topography/aspect. This ranking sequence differs from that
presented by Baldock and Skjemstad (1999) for soils as a whole, which presents land management as the most influential factor and soil mineral composition as the least influential.

A change in the relative influence of the various soil-forming factors with depth has been revealed by this study. The reduction in the significance of climate, particularly temperature, with depth and an increase in the significance of parent material parallels results reported by Jobbágy and Jackson (2000), Albaladejo et al. (2013), Wang et al. (2014) and Hobley et al. (2015). Similarly, Wilson and Lonergan (2013) reported the declining influence of land use with depth.

4.4.3 Combination of factors control SOC stocks

This study has demonstrated that in order to understand and predict SOC storage levels in any soil, the combined influence of the key soil-forming factors must be considered. Each factor has a different broad level of influence, being significant at differing scales, and they combine together to control final SOC stocks. Minor topographic influences are superimposed on the moderate land use/ground cover influences, which in turn are superimposed on the large parent material (soil type) influences which are ultimately superimposed on the very large climatic influences. Figures 4.4 and 4.5 reveals generally uniform trends of increasing SOC density with increasingly moist climate, increasing mafic character of parent material and increasing vegetation cover. The use of only two of these three key factors would clearly result in unreliable estimations. For example, SOC density over the 0-30 cm interval in a wet, high vegetation cover regime varies from 56.0 Mg ha\(^{-1}\) in soils from extremely siliceous parent material up to 145.0 Mg ha\(^{-1}\) for soils from mafic parent material, a 2.6 fold increase.

The necessity of considering a combination of factors when deriving meaningful estimates of potential SOC storage has been similarly reported by other workers (Heckman et al. 2009; Powers et al. 2011; Mishra et al. 2012; Mayes et al. 2014; Viscarra Rossel et al. 2014; Xiong et al. 2014). The differing potential of regions to store SOC according to different climate, soil types and land use forms the basis of the “carbon zone” concept of Murphy et al. (2010) and the “potential capability index” for additional SOC storage of Baldock et al. (2009b). The importance of considering combined multiple factors was also demonstrated by Gray et al. (2016) who examined
the declines in SOC over the top 30 cm following a change from native vegetation to regular cropping in NSW, Australia. They reported declines ranging from just 3 Mg ha\(^{-1}\) or 8% for highly siliceous parent materials in warmer climates up to 44.3 Mg ha\(^{-1}\) or 50.0% loss over mafic parent materials in cooler (moist) conditions. Knowledge on the combined influence of multiple factors can guide the identification of soil-environment regimes/locations that are priorities in carbon sequestration programs. For example, Figures 4.4 and 4.5 would suggest that greater potential gains in SOC density could be made by focusing on soils from mafic rather than siliceous parent materials. Such knowledge may be particularly important for the effective establishment and operation of carbon trading schemes as a means of addressing climate change.

Many workers have referred to and documented the differing SOC storage potential of different soil types, both in Australia (Cotching 2012) and internationally (Batjes 1996; Eswaran \textit{et al.} 2000; Lal 2004a). Large variations within each soil type are reported, for example, Batjes (1996) reported coefficients of variation generally between 50-100\% for SOC contents to 100 cm in world FAO-UNESCO soil groups and Albaladejo \textit{et al.} (2013) reported similar variation in Spain. It is clear that for such estimates to be meaningful they must similarly be stratified according to other environmental factors, particularly climate and vegetation cover or land use attributes, also noted by Jobbágy and Jackson (2000). For many soil orders it may be necessary to define them down to lower classification levels, due to significant variation within the primary Order level. The potential complexity of such an exercise is reason to favour the relative simplicity of parent material composition as a basis of the soil stratification process as in this current study (refer to Table 4.1).

4.4.4 Relative SOC storage in subsoils

The results from this study support the widely held contention that subsoils contribute a substantial proportion of total SOC stocks despite their lower SOC concentration (Batjes 1996; Jobbágy and Jackson 2000; Rumpel and Kögel-Knabner 2011; Cotching 2012). Despite high uncertainty levels of the results, the 30-100 cm interval has been shown here to contribute approximately half of the SOC stocks down to 1 metre. However, the precise contributions of upper and subsoil levels have been shown to vary depending on climatic influences. We have demonstrated that in the dry climate zones of eastern Australia the majority of carbon stock in the top metre is stored
in the subsoil (30-100 cm) with an average 54%, whereas in wet climates a lower proportion is in the subsoil (average 43%). Climate appears to be the dominant driving influence of this ratio. Although no strong trends are evident in relation to parent material and vegetation cover classes, there is an indication that relative SOC storage in the subsoil is greatest over extremely siliceous parent materials in wet and moist climates, and in lower vegetation cover classes in dry climates.

These results compare with the estimates for world soils of the proportion of soil organic matter stored in the first metre below 30 cm depth that range between 46 and 63%, for almost all soil types (Batjes 1996; Rumpel and Kögel-Knabner 2011). An examination of results reported in Cotching (2012) also suggest a higher relative proportion of SOC storage in subsoils in the relatively drier states of eastern Australia than in the more moist States for equivalent soil types. He also found land use to be an important driver of these relative proportions, which contrasts with our results here which indicate no significant influence from vegetation cover levels. A relative proportion of 35% SOC over the same interval was reported in Laos, a notably very wet climate (Chaplot et al. 2010). Jobbágy and Jackson (2000) found the percentage of SOC below 20 cm, relative to the first metre, averaged 67%, 58%, and 50% for shrublands, grasslands and forests, respectively in their global study. They found these proportions varied from 71% in cold arid shrublands to 43% in cold humid forests. The above studies all support our findings of higher relative SOC storage levels of subsoils in dry climates compared to moist climates.

4.5 Conclusion

This study has provided quantitative data to help us understand the factors controlling the storage of organic carbon in the soils of eastern Australia. The results can provide guidance on the physical locations of soils with high and low SOC storage potentials. It has been revealed that SOC stocks over this province are primarily controlled by climate, parent material and vegetation cover. Other factors including topography and the associated aspect tend only to be significant when the over-riding influence of climate in particular is removed, as in more localised scale studies. The relative influence of the different factors has been shown to change with depth, with climate (particularly temperature) and vegetation cover/land use decreasing in influence and parent material increasing in influence.
It has been demonstrated that a combination of factors, particularly climate, parent material (or soil type) and vegetation cover (or land management) are required to understand and make meaningful estimates of SOC storage levels. Without a full consideration of the key controlling factors together, any estimates of current or projected SOC stocks will be unreliable. The study provides further evidence on the importance of SOC subsoil storage, and has demonstrated that SOC storage in subsoils actually exceeds that in upper soils in drier climates (dependent on the defined depth intervals). The results suggest that the proportion of SOC stored in the subsoil appears to be primarily controlled by climate, generally increasing with drier climates, but also possibly influenced by parent material (soil type) and vegetation cover in more complex trends.

The incorporation of knowledge on factors controlling organic carbon stocks in our soils, such as gained from this study, is essential for designing effective strategies of soil carbon sequestration that can help to combat projected climate change. Simultaneous improvements in soil health across our agricultural lands may be another important outcome.

4.6 References


Australia with the potential to enhance soil carbon content. CSIRO Sustainable Agriculture Report, CSIRO, Canberra.


Chapter 4: Factors controlling SOC with depth in eastern Australia


Chapter 4: Factors controlling SOC with depth in eastern Australia


Chapter 4: Factors controlling SOC with depth in eastern Australia


Chapter 4: Factors controlling SOC with depth in eastern Australia


Chapter 5: Digital mapping of pre-European soil carbon stocks and decline since clearing over New South Wales, Australia

Abstract

Digital soil models and maps have been developed for pre-European (pre-clearing) levels of soil organic carbon (SOC) over New South Wales, Australia. These provide a useful first estimate of natural unaltered soil conditions prior to agricultural development, which are potentially important for many carbon accounting schemes such as those prescribed by the Intergovernmental Panel on Climate Change, carbon turnover models such as RothC and for soil condition monitoring programs. The modelling approach adopted included multiple linear regression and Cubist piecewise linear decision trees. It used 1690 soil profiles from undisturbed or only lightly disturbed native vegetation sites across all of eastern Australia, together with a range of covariates representing key soil-forming factors.

The digital soil maps of pre-clearing SOC (% and mass) over NSW provide a more sophisticated alternative to currently available equivalent maps. Independent validation of the SOC mass predictions over the top 30 cm revealed a concordance correlation coefficient of 0.76, which was 13% higher than the currently used map. Total pre-clearing SOC stocks amount to 4.21 Gt in the top 30 cm which compared to a current stock estimate of 3.68 Gt, suggesting a total SOC loss of approximately 0.53 Gt over the entire State. The extent of SOC decline in both absolute and relative terms was found to be highly dependent on the climate-parent material-land use regime, reaching a maximum decline of 44.3 Mg ha\(^{-1}\) or 50.0% relative loss in cooler (moist) conditions over mafic parent materials under regular cropping land use. The models also provide valuable pedological insights into the factors controlling SOC levels under natural conditions.

Keywords: digital soil mapping, carbon modelling, carbon accounting, native vegetation, land use change
Chapter 5: Pre-European soil carbon stocks and change since clearing, NSW, Australia

5.1 Introduction

Australian landscapes, ecosystems and soil conditions have changed significantly since European settlement in the late 1700s with the associated large scale clearing and introduction of agricultural systems. A decline in soil organic carbon (SOC) levels upon conversion from native vegetation to agriculture has been widely reported in Australia (Murphy et al. 2003; Dalal et al. 2005; Baldock et al. 2009a; Lou et al. 2010; Wilson et al. 2011) and globally (Paustian et al. 1997; Post and Kwon 2000; Guo and Gifford 2002; Lal 2004; Lal and Follet 2009).

Knowledge on SOC levels prior to vegetation clearing can be an important requirement for carbon accounting systems such as those prescribed by the Intergovernmental Panel on Climate Change (IPCC 2006). They provide the initial levels for undisturbed soil, which can then be used in other models to derive baseline levels at particular reference dates such as the Kyoto Agreement baseline year of 1990 (IPCC 2006). They thus facilitate the estimation in a verifiable manner of the change in soil carbon stocks under different land use and management, which is necessary for the implementation of carbon trading schemes.

Pre-clearing SOC levels are often applied in the initialisation and validation of soil carbon turnover models such as Roth C, Century or others (Parton and Rasmussen 1994; Smith et al. 1997; Coleman and Jenkinson 1999), which aim to model the behaviour of SOC under different climate and land management regimes. These feed into broader climate change models (Houghton et al. 2009; Scholes et al. 2009; Smith and Fang 2010).

Estimates of SOC change since clearing provide important data and understanding on the impacts of land use change on soil carbon levels. They allow us to better estimate the potential contribution that land use change may play in soil carbon sequestration as a means to mitigate against rising atmospheric carbon levels and associated climate change (Lal 2004; Wilson et al. 2011; Baldock et al. 2012). Soil carbon has been at the centre of soil monitoring programs in Australia, not only for its role in addressing climate change but also as it is considered a key indicator of soil health (Grealish et al. 2011; Baldock et al. 2009b). Information on changes in this soil property is valuable for guiding sustainable land management (McKenzie and Dixon 2006; Campbell 2008).
5.1.1 Approaches to derive pre-clearing SOC levels and SOC change following clearing

The most common approach of deriving original pre-clearing (or pre-European) SOC values for various modelling or monitoring purposes is to apply estimates over broad geographic units based on data from representative areas of essentially undisturbed native vegetation. These often rely on limited data and uncertain assumptions regarding the representativeness of the existing vegetated areas, which may be associated with less productive soils than in adjoining agricultural lands. This is the approach adopted by the IPCC Guidelines for National Greenhouse Inventories (IPCC 2006), where mean reference carbon stock values for different zones of climate and soil type (WRB or Soil Taxonomy) are derived based on the simple averaging of available soil data from native vegetation sites. The approach forms the basis of default tables of reference carbon stock values that are used in the IPCC Tier 1 method where countries or regions are data poor (Batjes 2011) and for the Tier 2 method where country or region specific reference data are available. Tier 3 methods involve more sophisticated process based modelling, but to date have been used by only a few countries (Ogle et al. 2010; Mishra et al. 2012).

This simple approach to deriving pre-clearing SOC estimates has been applied in Australia, as presented in the report *Pre-clearing soil carbon levels in Australia* (Webb 2002). The associated map for New South Wales (NSW) by Banks and McKane (2002) is presented in Figure 5.1. Note however, that for the purposes of Australia’s National Carbon Accounting Scheme (NCAS, Richards & Evans 2004) a recently prepared map representing current OC stocks (Viscarra Rossel et al. 2014) is now primarily used, on the assumption that it effectively represents 1990 conditions. However, for areas where recent (post 1990) vegetation clearance or forestry operations have occurred, pre-clearing estimates such as those from Webb (2002) are still preferred. The simple approach is often applied where pre-clearing SOC data are incorporated in soil carbon turnover models such as RothC or Century (Bortolon et al. 2011). It is also applied in some soil condition monitoring programs, such as a recent SOC monitoring program in NSW (Chapman et al. 2011) and in studies to estimate decline in SOC levels since pre-European times such as by Owens et al. (1999) in Michigan USA.
There appear very few cases of applying more sophisticated modelling approaches to derive pre-clearing SOC data. One example was recently presented by Mishra et al. (2012) for seven states in north east USA, which involved regression kriging with a number of environmental variables. They reported improved predictions in final estimates of change in carbon stock compared to those from the standard IPCC approach with simple averaging. In Australia, digital soil modelling and mapping (DSMM) techniques have been widely used to map current SOC levels (Henderson et al. 2005; Bui et al. 2009; Bui 2012; Grundy et al. 2012; Viscarra Rossel et al. 2014; Gray et al. 2015a). Other Australian and international studies are described in Minasny et al. (2013). These DSMM techniques, however, appear not to have been widely applied to pre-clearing, original SOC levels in Australia or elsewhere to date.

![Pre-clearing soil organic carbon map of NSW](image)

**Figure 5.1. Pre-clearing soil organic carbon map of NSW, Mg ha\(^{-1}\), 0-30 cm (from Banks and McKane 2002)**

Estimates of carbon loss following conversion of natural ecosystems to agriculture have been presented by a number of authors at global scale (Guo and Gifford 2002; Lal and Follett 2009) and in Australia (Murphy et al. 2003; Lou et al. 2010; Wilson et al. 2011; Sanderman et al. 2011; Wang et al. 2013). Most of these are based on direct field investigations, reviews of existing studies or by simulation modelling. There appears to be few detailed attempts to demonstrate variation in SOC change following land use change according to key soil-forming factors, although the influence of climate and soil
types on the level of SOC change has been recognised in a number of studies (Wilson et al. 2004; Baldock et al. 2009a; Murphy et al. 2010; Powers et al. 2011;Allen et al. 2013; Page et al. 2013). There appears to be considerable potential to apply digital soil modelling and mapping techniques to improve estimates of pre-clearing SOC stocks and SOC change since clearing in Australia and internationally.

5.1.2 Aims

This study attempts to

(i) produce quantitative models for the prediction of SOC levels (concentration and mass) under pre-clearing (pre-European) conditions over eastern Australia, using both pragmatic multiple linear regression models and more complex Cubist decision tree models.
(ii) produce digital soil maps of pre-clearing SOC from these models over the State of New South Wales (NSW)
(iii) compare the predictive ability of the NSW map against the currently available map of Banks and McKane (2002)
(iv) compare pre-clearing carbon stocks against current levels to derive estimates on the extent of change in SOC since clearing, with further breakdown by land use, broad climate and parent material and class
(v) gain pedological knowledge on factors controlling SOC in natural environments.

5.2 Data and methods

The methodology involved the initial establishment of a dataset of soil profiles with SOC and numerous environmental covariates across eastern Australia. A subset of undisturbed or only lightly disturbed vegetated sites was identified, over which multiple linear regression (MLR) and Cubist piecewise linear decision tree models for SOC concentration were developed. These facilitated the preparation of a digital map of pre-clearing SOC stock over the State of NSW, including the Australian Capital Territory (ACT). This was compared against the existing equivalent map of Banks and McKane (2002), with validation using an independent dataset. Another map of present day SOC stock was prepared over the same area, using the Cubist approach. By comparison of
the present day map against the pre-clearing map, estimates were derived of the mean decline in SOC density (Mg ha\(^{-1}\), or tonnes/ha), the total SOC loss (tonnes) since clearing. Also derived were the absolute and relative losses according to land use class, broad climate regime and parent material class.

### 5.2.1 The soil dataset

A dataset of 5210 soil profiles containing SOC laboratory results and adequate site data was compiled over eastern Australia as reported in Gray et al. (2015b). From this a subset of 1687 profiles from undisturbed or only lightly disturbed vegetated sites over eastern Australia was identified (Figure 5.2). These sites included national parks, nature reserves, lightly logged native forests and lightly grazed native grasslands on alluvial clay plains and were considered to effectively represent pre-clearing conditions over the region. The breakdown by jurisdiction was as follows: NSW 669, Queensland 575, CSIRO (eastern states) 308, Tasmania 111, Victoria 14 and South Australia 10.

Soil organic carbon values reported for each soil horizon depth interval over the entire original dataset were converted into the standard depth intervals of 0-10 cm, 10-30 cm and 30-100 cm using the equal area spline process of Bishop et al. (1999) and Malone et al. (2009). The upper two layers were also combined to give a weighted average for the 0-30 cm interval (but in future a new specific dataset for this interval should be created to avoid adding unnecessary prediction errors).

Excluded from the analyses were organic soils including Organosols under the Australian Soil Classification (Isbell 2002) or Peats under the Great Soil Classification (Stace et al. 1968). Additionally, all sites with less than 0.1% SOC were excluded as these were considered unreliable. For most profiles (approx 95%), the Walkley-Black wet oxidation method was used to derive the SOC values, with LECO and other combustion methods used for the remainder. Although the Walkley-Black method is known to underestimate SOC values (Skjemstad et al. 2000) no correction values were applied here, as there is much uncertainty regarding the most appropriate correction factor, with SOC recovery rates varying between different substrates (Conyers et al. 2011) and improving in more recent years and also whether or not a correction factor has already been applied to the reported values (Bui et al. 2009). In addition to SOC percentage, estimates of SOC mass (Mg ha\(^{-1}\), tonnes/ha) were also derived for NSW following the determination of bulk density, as described later in this section.
Figure 5.2. Pre-clearing modelling points over eastern Australia (shaded area denotes region that can be effectively represented by the points)

5.2.2 Covariates

Covariates were selected to represent the key soil-forming factors of climate, parent material, relief, age and biota as outlined below. The covariates relating to biota were only considered for the present day SOC map, and not the pre-clearing SOC map where undisturbed vegetation cover was an assumed constant.

**Climate**

- *Mean annual precipitation* (mm pa, Precip) – derived from 2.5 km Australia wide climate grids from the Australian Bureau of Meteorology, interpolation of cell values down to a 100 m grid. They represent mean values obtained over the 1961-1990 period. Data were not available for climate at and prior to clearing.
- *Mean annual daily maximum temperature* (°C, T\(_{\text{max}}\)) – derived as above. For later stratification and interpretation purposes, two broad classes were defined: <=23°C = cool; >23°C = warm.

**Parent material**

- *Lithology class (silica index)* – an index representing the lithological character of the parent material (Gray *et al.* 2014, 2015b). For example, granite belongs to the moderately siliceous lithological class, with an approximate silica content
and index of 73%, while basalt belongs to the mafic class with an approximate silica content and index of 48%. Parent material descriptors recorded at each site were used for model development but the 1:250 000 NSW Geological Survey polygonal geology map was used for final maps.

For later stratification and interpretation purposes, seven broad classes were identified, ranging from mafic to extremely siliceous as presented in Table 5.1, each with commonly associated soil types from the Australian Soil Classification (ASC) scheme (Isbell 2002) and the World Reference Base for Soil Resources (WRB) scheme (IUSS Working Group WRB 2014).

- **Radiometrics** – gamma radiometric potassium (K), uranium (U) thorium (Th) and the ratio of Th to K; 90 m grids were developed by, and sourced from, Geoscience Australia.

- **Clay components** – relative proportions of kaolin, illite and smectite clays and the smectite/kaolin ratio (S/K ratio) derived from modelling with visible near infra-red (VNIR) spectroscopy (Viscarra Rossel 2011).

Table 5.1. Typical soil types associated with different parent material classes

<table>
<thead>
<tr>
<th>Parent material class</th>
<th>Typical ASC soils</th>
<th>Typical WRB soils</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mafic (45-52% silica)</td>
<td>black Vertosols, Ferrosols</td>
<td>Nitisols; Phaeozems;</td>
</tr>
<tr>
<td>Intermediate lower (52-60%</td>
<td>Dermosols; Ferrosols; other</td>
<td>Chernozeoms; Kastanozems</td>
</tr>
<tr>
<td>silica)</td>
<td>Vertosols</td>
<td></td>
</tr>
<tr>
<td>Intermediate upper (60-65%</td>
<td>Dermosols; eutrophic (high base)</td>
<td>Luvisols</td>
</tr>
<tr>
<td>silica)</td>
<td>Chromosols &amp; Sodosols</td>
<td></td>
</tr>
<tr>
<td>Siliceous lower (65-70% silica)</td>
<td>mesotrophic (moderate base)</td>
<td>Acrisols; Alisols; Cambisols; Ferralsols; Retisols; Umbrisols</td>
</tr>
<tr>
<td></td>
<td>Chromosols, Kurosols &amp;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sodosols</td>
<td></td>
</tr>
<tr>
<td>Siliceous mid (70-77% silica)</td>
<td>dystrophic (low base)</td>
<td>Durisols; Lixisols; Planosols; Solonetz</td>
</tr>
<tr>
<td></td>
<td>Chromosols, Sodosols &amp;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kurosols; Kandosols</td>
<td></td>
</tr>
<tr>
<td>Siliceous upper (77-82% silica)</td>
<td>Tenosols</td>
<td>Arenosols</td>
</tr>
<tr>
<td>Siliceous extreme (&gt;82% silica)</td>
<td>arenic Rudosols; Podosols</td>
<td>Podzols</td>
</tr>
</tbody>
</table>

1 First approximations for common soil types only; most soil types will extend into adjoining parent material classes
2 Based on Isbell *et al.* (1997); Gray and Murphy (1999)
3 Based on IUSS Working Group WRB (2014); Gray *et al.* (2011)
Relief

- **Topo-slope index (TSI)** – an index that can be derived from field observations that combines topographic position and slope gradient. It represents the degree to which a site is subject to depletion (1) or accumulation (6) of water, soil particles, and chemical materials (Gray *et al.* 2015). Model development relied on site data collected by soil surveyors for individual sites; map development used a 100 m DEM to derive the Topographic Position Index (Jenness 2006) and slope%.

- **Topographic wetness index (TWI)** – a widely used index that represents potential hydrological conditions based on slope and catchment area, as derived from DEMs. A 90 m grid was used (Gallant and Austin 2015).

- **Aspect index (Asp)** – an index to represent the amount of solar radiation received by sites, ranging from 1 for gentle N or NW facing slopes to 10 for steep S and SE slopes (Gray *et al.* 2015). Data derived from site data collected in the field or a 100 m DEM.

Age

- **Weathering Index (W index)** – an index to represent the degree of weathering of parent materials (Wilford 2012). A 90 m grid was accessed from Geoscience Australia.

Biota (for current SOC map only)

- **Land disturbance index (LDI)** – an index that reflects the intensity of disturbance associated with the land use (1: natural ecosystem to 6: intensive cropping) (Gray *et al.* 2015, modified from NCST 2009). For model development, site land use was taken from field profile descriptions; for the NSW current SOC map it was derived from 1:25 000 scale polygonal land use mapping (OEH 2007).

- **Ground cover** – total vegetation cover (photosynthetic and non-photosynthetic) derived from CSIRO 2011 MODIS fractional vegetation data (90 m grids) (Guerschman *et al.* 2009).

A number of these covariates can be relatively easily acquired from field survey or readily available maps including precipitation, maximum temperature, lithology (silica index), topo-slope index, aspect index and land disturbance index. These are
applied in the pragmatic MLR modelling approach, which facilitate readily made field based predictions for specific sites. The remainder are more sophisticated, complex covariates that rely on remotely sensed data or computation from DEMs, as applied in more sophisticated desktop predictive methods.

The ranges of the most commonly used covariates from the soil point dataset covering the eastern states of Australia, and the grids covering the State of NSW are shown in the box plots of Figure 5.3. This demonstrates that the point dataset does effectively represent the range of values for most covariates across our prediction grid, with the possible exception of kaolin and smectite.

Figure 5.3. Ranges of most commonly used covariates for the eastern Australia pre-clearing point dataset and the NSW grids

5.2.3 Model development, map generation and validation

The pre-clearing sub-dataset (1687 points) and the entire eastern Australian dataset (5210 points) were both apportioned into 80% as training data and 20% as validation data. Two different modelling processes were adopted; (i) pragmatic multiple
linear regression (MLR) models using readily available (generally field based) covariates only; and (ii) Cubist linear piecewise decision tree models (Quinlan 1992) using all readily available and complex covariates. Both approaches used R statistical software (R Core Team 2013). A natural log transformation was applied to the SOC% to meet the assumptions of normality.

The models were validated using the validation datasets, ie, the initially excluded 20% of points. Lin’s concordance correlation coefficient (CCC) was used to measure the level of agreement of predicted values with observed values relative to the 1:1 line (Lin 1989). Also determined were root mean square error (RMSE), mean error (predicted – observed) and median absolute error. Standardised regression coefficients of covariates in the MLR models and data on the frequency of use in the Cubist models were derived to inform on the relative influence of each covariate in the models.

Digital soil maps over NSW (100 m raster) for both the pre-clearing and current SOC levels were derived using a Cubist tool developed for ESRI ArcGIS software (Reuters 2014). The pre-clearing SOC map used the Cubist models derived from the pre-clearing sub-dataset (ie, essentially native vegetation sites) while the present day SOC map used Cubist models derived from the entire eastern Australian point dataset (ie, native vegetation, agricultural and other sites), as described in Chapter 4.

A digital version of the currently applied polygonal pre-clearing SOC map for NSW (Banks and McKane 2002) was obtained from the Commonwealth Department of Environment (formerly Australian Greenhouse Office) (Figure 5.1). Validation of the new digitally modelled pre-clearing map and that of Banks and McKane (2002) used an entirely independent validation dataset, comprising 103 undisturbed or only lightly disturbed sites from the NSW monitoring evaluation and reporting (MER) program (Chapman et al. 2011). The present day SOC map was validated with the full NSW MER dataset (780 profiles), in addition to the validation process described in Chapter 4. This map is not the subject of this chapter so only the new MER validation results are presented here (refer back to Chapter 4 for more details and images, which cover all of eastern Australia).

A comparison of the present day SOC map with the pre-clearing SOC map allowed estimates to be derived of total change in carbon stock (million Mg or tonnes), together with estimates of mean decline in density in absolute terms (Mg ha\(^{-1}\)) and
relative terms (%). The decline since clearing by broad land use (LDI), climate regime ($T_{max}$ less than or greater than 23 °C) and parent material (seven lithology classes and associated soil types) was also determined. Approximations of the upper and lower 95% confidence intervals for total SOC stock and density estimates across the state were derived by adding or subtracting 1.96 x RMSE (from the map validation) to the mean values.

Bulk density, as required for SOC mass predictions, were derived over the 0-30 cm depth interval from Cubist decision tree models using data from the NSW MER program (888 points) with the full suite of covariates listed above. Further details of this map are presented in Appendix 3. For the 30-100 cm depth interval, estimates were derived using a variant of the pedo-transfer functions reported in Tranter et al. (2007) and Minasny et al. (2013) (Equation 1):

\[
BD = \frac{100}{(1.724*OC%/224+(100-1.724*OC%)/(1.351 + 0.0045 * Sand% + (Sand% - 44.65)^2 * -0.0000614 + 0.0596* \log(depth cm)))}
\]

Equation 1

The same bulk density values were applied to both pre-clearing and current SOC stock maps. As agricultural soils typically have a higher bulk density than pre-clearing soils, this may result in relatively higher SOC stocks in the present day maps, resulting in a possible under-estimation of the SOC change since clearing. This problem could have been addressed by comparing the SOC stocks on an equivalent soil mass basis, a concept that recognises the different volumes and depths of equivalent masses of soil material (Wendt and Hauser 2013; Murphy et al. 2003; Ellert and Battany 1995), but for the purposes of simplicity, we reported changes in SOC according to the uniform depth intervals.

Other limitations of the modelling process also need to be borne in mind. There is an assumption that the soils under current natural conditions are representative of pre-clearing conditions, and that equilibrium with environmental conditions have been reached in both periods. The climate of the 1960-1990 period was assumed to be representative of the pre-clearing conditions, as no more specific climatic details were available. There were no details of the specific dates of clearance; with dates varying from over 150 years ago up to only a decade or two ago. This introduced further uncertainty into final estimated SOC stock changes and precluded analysis of actual rates of SOC change, ie, change per year.
5.3 Results

5.3.1 Models for pre-clearing SOC over eastern Australia

The pragmatic multiple linear regression (MLR) models for pre-clearing soil organic carbon (SOC%) at the three depth intervals, together with associated statistics are presented in Table 5.2.

The SOC model is moderately strong in the upper 10 cm depth interval but then drops off in strength at lower depths, as indicated by the decreasing R^2 values, consistent with other SOC digital soil mapping studies (Henderson et al. 2005, Bui et al. 2009, Viscarra Rossel et al. 2014). The inherently low SOC levels at depth display only a weak relationship to the various environmental factors.

Table 5.2. Pragmatic MLR models for pre-clearing SOC% over eastern Australia

<table>
<thead>
<tr>
<th>Depth (cm)</th>
<th>MLR model</th>
<th>N</th>
<th>R^2</th>
<th>F</th>
<th>Residual SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-10</td>
<td>SOC0_10 = exp(3.913 + 0.000480<em>Rain - 0.0970</em>Tmax - 0.0205<em>Silica - 0.0518</em>TSI + 0.0238*Asp)</td>
<td>1252</td>
<td>0.62</td>
<td>403</td>
<td>0.56</td>
</tr>
<tr>
<td>10-30</td>
<td>SOC10_30 = exp(2.993 + 0.000461<em>Rain - 0.0740</em>Tmax - 0.0212<em>Silica - 0.0245</em>TSI + 0.0242*Asp)</td>
<td>968</td>
<td>0.42</td>
<td>139</td>
<td>0.62</td>
</tr>
<tr>
<td>30-100</td>
<td>SOC30_100 = exp(1.316 + 0.000494<em>Rain - 0.0560</em>Tmax - 0.0260<em>Silica + 0.0359</em>TSI + 0.0385*Asp)</td>
<td>664</td>
<td>0.20</td>
<td>32</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Model statistics based on log scale

Results of model validation based on the data splitting from both the MLR (with pragmatic covariates) and Cubist (with all covariates) approaches are presented in Table 5.3. The two approaches yield similar validation statistics, but the Cubist approach is marginally stronger. Lin’s CCC reached a maximum of 0.78 and RMSE a minimum of 0.52 log% using the Cubist approach over the upper interval but with weaker values at lower depth intervals, particularly over the 30-100 cm depth interval.
Chapter 5: Pre-European soil carbon stocks and change since clearing, NSW, Australia

### Table 5.3. SOC (log%) model validation

<table>
<thead>
<tr>
<th>Model tool</th>
<th>Depth (cm)</th>
<th>Lin’s CCC</th>
<th>N</th>
<th>RMSE Mean error</th>
<th>Median absolute error</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLR</td>
<td>0-10</td>
<td>0.75</td>
<td>316</td>
<td>0.55</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>10-30</td>
<td>0.49</td>
<td>236</td>
<td>0.63</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>0.20</td>
<td>161</td>
<td>0.86</td>
<td>-0.02</td>
</tr>
<tr>
<td>Cubist</td>
<td>0-10</td>
<td>0.78</td>
<td>329</td>
<td>0.52</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>10-30</td>
<td>0.55</td>
<td>246</td>
<td>0.61</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>0.22</td>
<td>167</td>
<td>0.85</td>
<td>-0.01</td>
</tr>
</tbody>
</table>

Model statistics based on log scale

#### 5.3.2 Influence of covariates

Useful indications of the relative influence that each covariate has on the variation of soil organic carbon can be gained from an examination of results from both the MLR and Cubist approaches. Table 5.4 presents the standardised regression coefficients for each of the pragmatic covariates from the MLR models. Table 5.5 presents the frequency of use of the top seven covariates used in the Cubist models. Maximum annual temperatures, annual rainfall and parent material indices all come out as covariates of major influence, as is explored later in the Discussion.

### Table 5.4. Standardised regression coefficients of covariates in SOC% regression models

<table>
<thead>
<tr>
<th>Depth (cm)</th>
<th>Precip</th>
<th>Tmax</th>
<th>Silica</th>
<th>TSI</th>
<th>Asp</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-10</td>
<td>0.22</td>
<td>-0.59</td>
<td>-0.25</td>
<td>-0.12</td>
<td>0.05</td>
</tr>
<tr>
<td>10-30</td>
<td>0.23</td>
<td>-0.46</td>
<td>-0.29</td>
<td>-0.06</td>
<td>0.07</td>
</tr>
<tr>
<td>30-100</td>
<td>0.20</td>
<td>-0.25</td>
<td>-0.32</td>
<td>0.07</td>
<td>0.10</td>
</tr>
</tbody>
</table>

#### Table 5.5. Frequency of covariate use in the Cubist models

<table>
<thead>
<tr>
<th>Covariate</th>
<th>0-10 cm</th>
<th>10-30 cm</th>
<th>30-100 cm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tmax</td>
<td>100/100</td>
<td>Rain</td>
<td>Rain</td>
</tr>
<tr>
<td>Rain</td>
<td>89/100</td>
<td>Tmax</td>
<td>Tmax</td>
</tr>
<tr>
<td>TWI</td>
<td>0/100</td>
<td>Silica</td>
<td>Silica</td>
</tr>
<tr>
<td>S/K ratio</td>
<td>0/100</td>
<td>S/K ratio</td>
<td>Asp</td>
</tr>
<tr>
<td>W index</td>
<td>0/100</td>
<td>Kaolin</td>
<td>Smectite</td>
</tr>
<tr>
<td>Silica</td>
<td>0/89</td>
<td>W index</td>
<td>Illite</td>
</tr>
<tr>
<td>Smectite</td>
<td>0/89</td>
<td>Smectite</td>
<td>S/K ratio</td>
</tr>
</tbody>
</table>

1 Percentage use in Cubist modelling: decision tree rules / final regression models
2 There was only one Rule, ie, one regression equation for this depth
5.3.3 Maps of pre-clearing SOC over NSW

Digital soil maps for pre-clearing SOC concentration (%) and mass (Mg ha\(^{-1}\)) were prepared for the three depth intervals plus the 0-30 cm interval over NSW, using both modelling approaches. Figure 5.4 presents the 0-30 cm mass map derived from using the Cubist approach, which performed slightly better than the MLR approach. The results of external validation of the Cubist derived maps using the 103 essentially undisturbed sites from the NSW MER program is shown in Table 5.6 and Figure 5.5.

![Pre-clearing SOC mass for NSW from current model](image)

Figure 5.4. Pre-clearing SOC mass for NSW from current model (Mg ha\(^{-1}\), 0-30 cm)

<table>
<thead>
<tr>
<th>Depth (cm)</th>
<th>Unit</th>
<th>Lin’s CCC</th>
<th>N</th>
<th>RMSE</th>
<th>Mean error</th>
<th>Median absolute error</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-10</td>
<td>%</td>
<td>0.73</td>
<td>103</td>
<td>0.56</td>
<td>-0.28</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>Mg ha(^{-1})</td>
<td>0.68</td>
<td>103</td>
<td>0.53</td>
<td>-0.25</td>
<td>0.38</td>
</tr>
<tr>
<td>10-30</td>
<td>%</td>
<td>0.78</td>
<td>103</td>
<td>0.34</td>
<td>-0.05</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>Mg ha(^{-1})</td>
<td>0.73</td>
<td>103</td>
<td>0.33</td>
<td>-0.04</td>
<td>0.21</td>
</tr>
<tr>
<td>0-30</td>
<td>%</td>
<td>0.80</td>
<td>103</td>
<td>0.38</td>
<td>-0.15</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>Mg ha(^{-1})</td>
<td>0.76</td>
<td>103</td>
<td>0.37</td>
<td>-0.12</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>Mg ha(^{-1}) (Banks and McKane 2002)</td>
<td>0.67</td>
<td>102</td>
<td>0.50</td>
<td>-0.12</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>Mg ha(^{-1}) (our present day SOC map)</td>
<td>0.58</td>
<td>777</td>
<td>0.44</td>
<td>-0.03</td>
<td>0.29</td>
</tr>
</tbody>
</table>

Table 5.6. Validation of NSW pre-clearing SOC maps, Banks and McKane map and present day SOC mass map

statistics based on log scale
Figure 5.5. Observed versus predicted values for pre-clearing SOC mass from NSW digital map (0-30 cm, original scale)

The SOC percentage models and maps appear stronger and more reliable than the SOC mass (Mg ha\(^{-1}\)) models and maps, as evidenced by comparison of Lin’s CCC values at each depth interval in Table 5.6. The predictive map for SOC mass appears to under predict values less than 50 Mg ha\(^{-1}\) and over 120 Mg ha\(^{-1}\), suggesting pre-clearing SOC stocks may be somewhat higher than our maps suggest at the upper and lower stock ranges. One contributing factor to this shortcoming may be the reliance on the bulk density estimates, which introduces another source of uncertainty.

Comparison of the current digital model derived map (Figure 5.4) against the previous map of Banks and McKane (Figure 5.1) demonstrate the greater level of detail in the new maps with continuous values down to each 100-m pixel. They reveal broadly similar patterns, but some significant differences are apparent in some regions, such as in the north-east sector of the State, which warrant further investigation. In terms of their statistical predictive ability, the Cubist derived map from this study moderately outperformed the earlier map with Lin’s concordance being 13% higher and RMSE and median absolute error being approximately 33 and 50% lower respectively. On the original (non-log) scale the concordance value is 20% higher in the current Cubist derived map.
5.3.4 Loss of carbon stock since clearing

Comparison of the modelled pre-clearing SOC mass against modelled present day mass provides data on SOC change that has occurred across NSW (and the ACT) since clearing took place. As expected, a decrease is demonstrated over most of the state where clearing has occurred (Figure 5.6). Although slight increases are suggested in the far west and other localised areas of the state these are within the uncertainty ranges of the models, for example within the model RMSE values. As shown by Table 5.7, total pre-clearing SOC stocks under pre-clearing conditions amount to 4.21 Gt in the top 30 cm which compared to our estimate of current stocks of 3.68 Gt. This suggests a total SOC loss of approximately 0.53 Gt (530 million tonnes or Mg) over the entire State since clearing, which equates to a proportional loss of 12.6% for the entire State from original pre-clearing (pre-European) conditions. The decline in SOC density over the cleared areas only was 7.43 Mg ha\(^{-1}\), which equates to an overall average decline of 16.3% over these lands. The high range between the upper and lower 95% confidence limits reflects the high level of uncertainty associated with the results, and the need to treat them with caution.

![Map of SOC change since clearing in NSW](image)

Figure 5.6. Change in SOC mass since clearing for NSW (Mg ha\(^{-1}\), 0-30 cm)
Table 5.7. Soil organic carbon stocks over NSW and ACT (0-30 cm) – pre-clearing and current levels

<table>
<thead>
<tr>
<th></th>
<th>Stock density (Mg ha(^{-1}))</th>
<th>Total stock (Gt)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Lower</td>
</tr>
<tr>
<td></td>
<td></td>
<td>confidence limit(^1)</td>
</tr>
<tr>
<td>Pre-clearing SOC stock</td>
<td>52.44</td>
<td>25.39</td>
</tr>
<tr>
<td>Current (this study)</td>
<td>45.84</td>
<td>19.35</td>
</tr>
<tr>
<td>Change (total state)</td>
<td>6.60</td>
<td>0.53</td>
</tr>
<tr>
<td>Change (cleared area only)</td>
<td>7.43</td>
<td>0.47</td>
</tr>
</tbody>
</table>

\(^1\) Confidence limits at 95% level

It is noted that our estimate for mean stock density over NSW (and the ACT) is some 8% higher than the value of 42.46 Mg ha\(^{-1}\) derived by Viscarra Rossel \textit{et al.} (2014), however our estimate for current total SOC is very similar (1% lower than their 3.72 Gt). Although those workers had significantly superior validation statistics, with narrower confidence limits, the similarity of the mean results, particularly for total stocks, adds confidence to our results in this study.

The change in SOC content following clearing was examined with respect to different land use and environmental conditions. At the broad State level, breakdown by major land uses reveals a relative decline in SOC density over the top 30 cm for regular grazing (native and improved pasture) of 14.8%, periodic cropping/grazing use of 18.2% and regular copping of 25.9%. However, this is just the broad State average and the actual decline at the local level varies greatly depending on the precise environmental regime. The change in SOC density by parent material class, broad climatic temperature regime and broad land use category over the 0-30 cm depth is presented in Table 5.8, and partly shown graphically in Figure 5.7 (absolute change) and Figure 5.8 (relative change). It can be seen that the change in density varies according to the following trends:

- increasing decline, in absolute and relative terms, in cooler (moister) regimes. A similar, although less consistent, pattern is observed when breakdown is done according to annual rainfall
- increasing decline, in absolute and relative terms, with more mafic (less siliceous) parent material, with substantial decline in mafic materials but generally
negligible change with extremely siliceous materials (refer to Table 5.1 for broadly associated soils types)

- increasing decline, in absolute and relative terms, from light grazing, through regular grazing, light cropping to intensive cropping.

The greatest decline in SOC over the top 30 cm is demonstrated in cooler (moist) conditions over mafic parent materials under intensive cropping land use, with a 44.3 Mg ha\(^{-1}\) or 50.0% loss. Losses in warmer climates over highly or extremely siliceous parent materials under grazing land uses are less than 1 Mg ha\(^{-1}\) or 4%. Note some actual slight increases in SOC content are revealed for some environment-land use combinations, usually with a very small spatial extent (see area column), and are well within the uncertainty ranges of the models. These changes reflect a pattern that the higher the initial SOC storage levels, the greater are both the absolute and relative loss in SOC upon vegetation clearance. The modelling procedure did not consider the period of time since clearing, which represents a source of uncertainty, and precludes the derivation of estimates of annual rates of change.

Table 5.8. Change since clearing in SOC density over top 30 cm by climate (temperature regime), parent material and land use

<table>
<thead>
<tr>
<th>Climate (temperature regime)</th>
<th>Parent material class</th>
<th>Land use</th>
<th>Absolute SOC change (mean Mg ha(^{-1}))</th>
<th>SD</th>
<th>Relative SOC change (mean %)</th>
<th>Area (sq km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cool</td>
<td>Mafic</td>
<td>Intensive crop</td>
<td>-44.25</td>
<td>9.88</td>
<td>-50.01</td>
<td>1114</td>
</tr>
<tr>
<td></td>
<td>Light crop</td>
<td></td>
<td>-29.87</td>
<td>20.64</td>
<td>-29.19</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td>Grazing</td>
<td></td>
<td>-37.88</td>
<td>15.42</td>
<td>-35.09</td>
<td>13890</td>
</tr>
<tr>
<td></td>
<td>Light grazing</td>
<td></td>
<td>-4.85</td>
<td>12.54</td>
<td>-6.1</td>
<td>5620</td>
</tr>
<tr>
<td>Intermediate</td>
<td>Lower</td>
<td>Intensive crop</td>
<td>-25.31</td>
<td>7.72</td>
<td>-50.96</td>
<td>2049</td>
</tr>
<tr>
<td></td>
<td>Light crop</td>
<td></td>
<td>-22.29</td>
<td>11.01</td>
<td>-38.71</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>Grazing</td>
<td></td>
<td>-23.72</td>
<td>9.71</td>
<td>-43.41</td>
<td>5857</td>
</tr>
<tr>
<td></td>
<td>Light grazing</td>
<td></td>
<td>-22.33</td>
<td>11.25</td>
<td>-41.35</td>
<td>289</td>
</tr>
<tr>
<td></td>
<td>Light crop</td>
<td></td>
<td>-5.32</td>
<td>12.64</td>
<td>-7.35</td>
<td>624</td>
</tr>
<tr>
<td></td>
<td>Grazing</td>
<td></td>
<td>-18.24</td>
<td>12.36</td>
<td>-27.3</td>
<td>18474</td>
</tr>
<tr>
<td></td>
<td>Light grazing</td>
<td></td>
<td>-16.62</td>
<td>13.75</td>
<td>-22.76</td>
<td>1206</td>
</tr>
<tr>
<td>Siliceous lower</td>
<td>Intensive crop</td>
<td></td>
<td>-14.70</td>
<td>8.98</td>
<td>-21.96</td>
<td>3487</td>
</tr>
<tr>
<td></td>
<td>Light crop</td>
<td></td>
<td>-14.81</td>
<td>10.39</td>
<td>-18.56</td>
<td>399</td>
</tr>
<tr>
<td></td>
<td>Grazing</td>
<td></td>
<td>-11.67</td>
<td>10.68</td>
<td>-13.68</td>
<td>36734</td>
</tr>
<tr>
<td></td>
<td>Light grazing</td>
<td></td>
<td>-10.52</td>
<td>11.63</td>
<td>-11.85</td>
<td>3473</td>
</tr>
<tr>
<td>Climate (temperature regime)</td>
<td>Parent material class</td>
<td>Land use</td>
<td>Absolute SOC change (mean Mg ha(^{-1}))</td>
<td>SD</td>
<td>Relative SOC change (mean %)</td>
<td>Area (sq km)</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----------------------</td>
<td>-------------------------</td>
<td>-------------------------------------------</td>
<td>-----</td>
<td>-------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Intensive crop</td>
<td>-13.17</td>
<td>9.16</td>
<td>-22.27</td>
<td>4814</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Light crop</td>
<td>-14.28</td>
<td>10.34</td>
<td>-17.63</td>
<td>777</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grazing</td>
<td>-8.22</td>
<td>10.28</td>
<td>-10.58</td>
<td>40981</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Light grazing</td>
<td>-4.85</td>
<td>12.54</td>
<td>-6.1</td>
<td>5620</td>
</tr>
<tr>
<td>Siliceous mid</td>
<td></td>
<td>Intensive crop</td>
<td>-10.40</td>
<td>5.86</td>
<td>-22.77</td>
<td>296</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Light crop</td>
<td>-6.42</td>
<td>12.93</td>
<td>-10.23</td>
<td>129</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grazing</td>
<td>-4.30</td>
<td>11.54</td>
<td>-6.67</td>
<td>3105</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Light grazing</td>
<td>+0.81</td>
<td>15.63</td>
<td>1.18</td>
<td>453</td>
</tr>
<tr>
<td>Siliceous extreme</td>
<td></td>
<td>Intensive crop</td>
<td>-1.04</td>
<td>5.56</td>
<td>-1.57</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Light crop</td>
<td>+9.51</td>
<td>5.51</td>
<td>22.39</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grazing</td>
<td>+3.50</td>
<td>9.02</td>
<td>4.91</td>
<td>289</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Light grazing</td>
<td>+9.09</td>
<td>9.14</td>
<td>12.21</td>
<td>93</td>
</tr>
<tr>
<td>Warm</td>
<td>Mafic</td>
<td>Intensive crop</td>
<td>-32.00</td>
<td>18.00</td>
<td>-46.85</td>
<td>5269</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Light crop</td>
<td>-17.50</td>
<td>18.77</td>
<td>-19.37</td>
<td>158</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grazing</td>
<td>-30.31</td>
<td>27.08</td>
<td>-37.51</td>
<td>7551</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Light grazing</td>
<td>+0.16</td>
<td>10.83</td>
<td>0.33</td>
<td>4324</td>
</tr>
<tr>
<td>Intermediate</td>
<td>lower</td>
<td>Intensive crop</td>
<td>-11.28</td>
<td>10.28</td>
<td>-27.96</td>
<td>27490</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Light crop</td>
<td>-12.28</td>
<td>8.03</td>
<td>-33.8</td>
<td>577</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grazing</td>
<td>-10.64</td>
<td>8.31</td>
<td>-32.9</td>
<td>78470</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Light grazing</td>
<td>-7.82</td>
<td>5.00</td>
<td>-23.37</td>
<td>30460</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Light crop</td>
<td>-4.27</td>
<td>11.69</td>
<td>-7.84</td>
<td>226</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grazing</td>
<td>-4.14</td>
<td>5.72</td>
<td>-13.12</td>
<td>82999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Light grazing</td>
<td>-3.54</td>
<td>6.86</td>
<td>-9.47</td>
<td>6478</td>
</tr>
<tr>
<td>Siliceous</td>
<td>lower</td>
<td>Intensive crop</td>
<td>+0.40</td>
<td>8.22</td>
<td>0.79</td>
<td>4866</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Light crop</td>
<td>+0.44</td>
<td>13.66</td>
<td>0.69</td>
<td>342</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grazing</td>
<td>+0.87</td>
<td>6.13</td>
<td>2.72</td>
<td>80791</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Light grazing</td>
<td>+3.51</td>
<td>13.52</td>
<td>6.27</td>
<td>3100</td>
</tr>
<tr>
<td>Siliceous mid</td>
<td></td>
<td>Intensive crop</td>
<td>-1.54</td>
<td>7.92</td>
<td>-3.03</td>
<td>6265</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Light crop</td>
<td>-3.14</td>
<td>12.83</td>
<td>-5.67</td>
<td>168</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grazing</td>
<td>-0.03</td>
<td>7.09</td>
<td>-0.08</td>
<td>62841</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Light grazing</td>
<td>+0.16</td>
<td>10.83</td>
<td>0.33</td>
<td>4324</td>
</tr>
<tr>
<td>Siliceous upper</td>
<td></td>
<td>Intensive crop</td>
<td>-2.94</td>
<td>6.68</td>
<td>-7.38</td>
<td>1184</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Light crop</td>
<td>-9.89</td>
<td>15.98</td>
<td>-18.85</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grazing</td>
<td>-0.50</td>
<td>3.88</td>
<td>-2.7</td>
<td>1630</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Light grazing</td>
<td>+0.02</td>
<td>9.33</td>
<td>0.07</td>
<td>1437</td>
</tr>
<tr>
<td>Siliceous extreme</td>
<td></td>
<td>Intensive crop</td>
<td>+1.27</td>
<td>2.01</td>
<td>6.39</td>
<td>4.60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Light crop</td>
<td>+3.34</td>
<td>6.87</td>
<td>10.43</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Grazing</td>
<td>-0.64</td>
<td>1.22</td>
<td>-3.46</td>
<td>1630</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Light grazing</td>
<td>+1.07</td>
<td>4.51</td>
<td>3.43</td>
<td>7.10</td>
</tr>
</tbody>
</table>

1 mean annual daily maximum temperatures <=23°C = cool; >23°C = warm
2 see Table 5.1 for silica ranges and broadly associated soil types
3 mean SOC change / pre-clearing SOC x 100 for each parent material, climate and land use class
5.4 Discussion

5.4.1 Effectiveness of models and maps

Models and derived digital soil maps of pre-clearing and current levels of soil organic carbon have been produced. The models, which cover all of eastern Australia, and the maps, which cover NSW only, have at least moderate statistical strength and
predictive ability, despite some inherent weaknesses. This study appears to be one of only a few attempts reported in the literature to digitally model and map pre-clearing SOC levels. Mishra et al. (2012) utilised a regression kriging approach in north eastern USA to model reference SOC stocks, as a component of SOC change modelling. They did not present data on the actual performance of the reference SOC stock modelling but did report superior results than those obtained using the standard IPCC (2006) approach.

The model and map strengths for both the pre-clearing and present day SOC levels compare reasonably well with most other studies carried out on SOC under current conditions. Viscarra Rossel et al. (2014) achieved concordance correlation coefficients of 0.81 in validation of SOC mass (0-30 cm) maps over Australia using a Cubist approach with residual kriging. Bui (2012) gained validation $R^2$ values of 0.65 for SOC mass under eucalypt forest; Bui et al. (2009) demonstrated model $R^2$ values for SOC content of 0.49 for topsoil and 0.36 for subsoil over Australia’s agro-ecological zone. Other SOC digital soil mapping projects in Australia and overseas for areas greater than 500 km$^2$ generally had model $R^2$ values between 0.2 and 0.6 (Minasny et al. 2013). The inclusion of a pre-European vegetation community map, for example that of Carnahan (1976) in the modelling and mapping process is one area of potential future improvement of our models and maps, as suggested by recent work demonstrating the correlation of SOC with native vegetation patterns/communities (Bui and Henderson 2013; De Vos et al. 2015).

The modelling and digital mapping approach presented in this study offers an alternative and more sophisticated approach to the estimation and mapping of pre-clearing soil organic carbon that is currently adopted throughout Australia. The maps presented for each State in Webb (2002), such as the Banks and McKane map for NSW shown in Figure 5.1, contain only single estimates of soil carbon mass to 30 cm for broad soil types or soil-landscape units within recognised biogeographic regions. There is no direct consideration of environmental covariates, apart from their role in identifying broad soil-landscape units, and they are typically supported by relatively sparse laboratory data. Nevertheless, the Banks and McKane map still appears to be effective in providing useful pre-clearing SOC estimates. The current digital approach does however appear to provide more accurate predictions based on the higher concordance values and lower RMSE and median absolute errors from the independent validation.
The models and resulting maps present potentially useful data on SOC levels prior to clearing. They should improve input data for initial (reference) SOC levels as often applied in carbon models such as RothC and Century, and in carbon accounting methodologies such as outlined in IPCC (2006). They could potentially improve the reliability of this input data into Australia’s National Carbon Accounting System, the Full Carbon Accounting Model (FullCAM, Richards & Evans 2004; Richards 2001), at least for areas subject to clearing or forestry operations since 1990, where the current SOC stock map of Viscarra Rossel et al. (2014) may be less applicable. They may be useful for natural resource management purposes, particularly by being incorporated into soil condition monitoring programs (Grealish et al. 2011; Ballock et al. 2009b; McKenzie et al. 2002). They provide quantitative information on the level of impact of the current and/or historic land management regimes on this soil property, and on soil condition more generally. The models and maps may provide estimates of the soil carbon levels that could reasonably be expected to return if under optimal management.

The pragmatic MLR models using readily available covariates have the potential for easy application, and may assist in the gaining of pre-clearing SOC estimates at individual sites, using only field collected and climate data (Gray et al. 2015b). They do not need the complex remotely sensed data sources and software required by the more sophisticated modelling techniques. They are also easy to interpret and facilitate the gaining of pedological knowledge as discussed later.

5.4.2 Change in carbon stock since clearing

Comparison of the modelled pre-clearing SOC levels against modelled current levels, reveals a substantial decrease in SOC mass since clearing (Figure 5.6). An overall decrease in SOC mass over the top 30 cm for the partially or wholly cleared lands is estimated at approximately 530 million Mg (tonnes), or 1930 million Mg of CO₂ equivalent over the State. The results reveal a significant decline in SOC following conversion of native vegetation to agricultural land. The drop in SOC is larger with an increasing degree of disturbance, that is, with increasing degrees of vegetation clearing and a move towards intensive grazing and cropping regimes. These findings are in accord with widespread reporting from Australia (Murphy et al. 2003; Dalal et al. 2005; Wilson et al. 2008; 2011; Lou et al. 2010; Sanderman et al. 2011; Wang et al. 2013) and internationally (Post and Kwon 2000; Guo and Gifford 2002; Lal and Follet 2009).
They reflect the lower replenishment of organic materials, lower soil moistures, increased organic matter decomposition with subsequent loss of carbon dioxide to the atmosphere, and loss from soil erosion associated with intensive disturbance of soils (Lal 2004; Baldock et al. 2009a).

A meta-analysis of 74 international studies by Guo and Gifford (2002) revealed a decline of 42% of soil carbon stocks following conversion of native forest to cropland. They also indicate an actual increase of 8% following conversion of native forest to pasture but Wilson et al. (2011) suggest the studied areas were not typical of Australia, and that a decline of between 22 to 38% is more likely for northern NSW. This conclusion is supported by Murphy et al. (2003) who report declines in the order of 30-40% for the NSW wheat belt (top 30 cm). Lou et al. (2010) report a loss of 51% in SOC in the top 10 cm following conversion of native ecosystems to cropland in Australia. Wang et al. (2013) estimate a mean decline of 48 million Mg, or 40% over the top 30 cm due to land use change over the wheat belt of NSW between 1960 and 2010. At the global level a loss of between 42 and 78 Gt SOC following land use conversion to agriculture has been reported (Lal and Follet 2009) relative to a current estimate of 1500 Gt over the top metre (Eswaran et al. 1993; Sanderman et al. 2011), suggesting an overall 3-5% decline in total world soil carbon stocks.

This study has demonstrated that the decline in SOC following vegetation clearance is not uniform over broad regions or land uses, but is very dependent on the precise environmental regime in addition to the intensity of the agricultural activity. The climate and parent material combination is particularly important, the latter reflecting soil type as shown in Table 5.1. An increasing decline in SOC density is demonstrated, in both absolute and relative terms, in cooler (moister) climates with more mafic (less siliceous) parent material. The greatest decline in SOC over the top 30 cm, involving a change to regular cropping, is demonstrated in cooler climates over mafic parent materials, with a 44.3 Mg ha\(^{-1}\) or 50.0% loss, whereas the change in warmer climates with highly siliceous parent materials for the same land use change is less than 3 Mg ha\(^{-1}\) or 8%.

The demonstrated changes reflect a pattern that the higher the initial SOC storage level, the greater is the loss in SOC upon vegetation clearance for equivalent land use changes, in both absolute and relative terms. Powers et al. (2011) demonstrated the importance of clay mineral composition, in addition to precipitation, in controlling SOC
levels following land use conversion in the tropics, with greater rates of loss (in relative terms) associated with high activity clays than the low activity clays. Baldock et al. (2009a) similarly report the highest rates of SOC loss following clearing in clay rich soils and minimal change in clay poor, sandy soils. The differing potential of regions to gain or lose SOC with land use change according to different climate and soil types forms the basis of the carbon zone concept of Murphy et al. (2010) and the carbon sequestration predictor of Wilson et al. (2004).

It is noteworthy that the pattern of SOC decline shown by the current study occurs in both absolute and relative terms, as also suggested by Baldock et al. (2009) and Powers et al. (2011). Whilst it might be expected that soils with initially higher SOC levels should undergo the greatest absolute change, it is somewhat surprising that they should also undergo the greatest relative decline. Perhaps these more fertile and productive soils are used relatively more intensively for the same land use, for example grazing with higher stocking rates or cropping at higher frequencies. These soils have the inherent potential to lose a greater proportion of their SOC and still be productive, thus are often worked harder until a high proportion of their initial SOC is lost.

These results, together with the previous studies, point to the potential for land use and management change to influence soil carbon levels, and provide opportunities to mitigate against rising greenhouse gas levels and associated climate change. They suggest the magnitude of the potential change is highly dependent on the land use, climatic and parent material (or soil type) combination.

5.4.3 Factors controlling SOC content in natural conditions

The standardised regression coefficients (SRC) for the MLR models given in Table 5.4 provide a valuable indication of the relative influence of each covariate in the models. The frequency of use of covariates in the Cubist models as given in Table 5.5 is also helpful.

At the eastern Australian scale, it appears that temperature and rainfall are the dominant controlling influences for SOC, having the generally highest SRC values and the most frequent use in the Cubist models. These factors are primary controllers of organic decomposition rates and soil moisture regimes. Parent material composition appears to have the next highest influence. Its relative influence increases at the deeper intervals. The more mafic materials are clearly giving rise to higher fertility soils that is
being reflected in higher SOC levels. Aspect plays a smaller but still significant role, particularly at the deeper level, as moist south-facing slopes have higher SOC levels than drier north-facing slopes; the increase being more pronounced in steeper terrain. The influence of topography at this scale is somewhat confused, as against expectations the results suggest SOC levels decrease in lower parts of the landscape for the upper 30 cm, however they do suggest the expected increase in 30-100 cm interval. These results may reflect a lack of representation of low lying undisturbed areas in the initial dataset. The low $R^2$ values of the models, particularly at depth, suggests a large amount of variability is unexplained, and that there may other be other unaccounted for influencing factors at play.

The ranking of relative importance of soil-forming factors for undisturbed soils based on these SRC values and frequency of use in the Cubist models may be summarised as: temperature $>$ precipitation $>$ parent material $>$ aspect $>$ topography. This compares with the ranking presented by Baldock and Skjemstad (1999) for all soils (including disturbed ones): management $>$ climate $>$ biota (vegetation and soil organisms) $>$ topography $=$ soil mineral composition. Our study suggests that parent material/soil mineral composition has a significant influence on SOC levels, also reported by several other workers (Jenny 1968; Bui et al. 2009; Cotching 2012; Badgery et al. 2013; Page et al. 2013; Minasny et al. 2013; Orgill et al. 2013; Viscarra Rossel et al. 2014) but not widely emphasised. It is also evident that the relative influence of each factor varies at different depths, with temperature becoming less dominant and parent material becoming more dominant with increasing depth (Gray et al. 2015a). Overall, it is apparent that SOC stocks are controlled by a combination of environmental and land management factors, and that individual factors cannot be considered in isolation.

5.5 Conclusion

Models and maps for pre-clearing levels of SOC, of at least moderate predictive strength, have been developed for eastern Australia. There appear to be few other comparable digital soil modelling and mapping studies relating to pre-clearing (pre-European) soil conditions undertaken in Australia or globally reported in the literature. The DSMM approach adopted here may improve on existing methods applied in Australia and internationally that typically involve relatively crude single estimates of
reference SOC over broad soil geographic units. The approach may have potential use in carbon accounting models such as proposed by the IPCC and in carbon turnover models such as RothC.

The comparison of the pre-clearing SOC levels against current levels across NSW revealed a significant decrease in SOC stock since clearing, with an overall 16.3% decline over the cleared lands, however, the high level of uncertainty needs to be recognised. It has been shown that the extent of decline varies greatly over different climate, parent material (soil type) and land use regimes, with increasing decline demonstrated for cooler (more moist) climates over more mafic parent materials with higher intensity agricultural land uses. The results enhance our knowledge on the impacts of land use change on soil carbon storage levels, and the potential opportunities to use land management change as a means to mitigate rising greenhouse gas levels and climate change. They are also potentially important for soil monitoring programs and broader natural resource management purposes, and may help us to better understand and manage changes in soil condition and related ecosystem condition arising from vegetation clearance.

Acknowledgements

Rachel Burgess (Australian Department of Environment) provided useful assistance and advice.

5.6 References

Australia with the potential to enhance soil carbon content. CSIRO Sustainable Agriculture Report.


Chapter 5: Pre-European soil carbon stocks and change since clearing, NSW, Australia


McKenzie NJ, Dixon J (Eds), 2006. *Monitoring Soil Condition Across Australia—Recommendations from the Expert Panels*, prepared on behalf of the National Committee on Soil and Terrain, National Land and Water Resources Audit, Canberra, ACT.


Chapter 5: Pre-European soil carbon stocks and change since clearing, NSW, Australia

sampling for soil carbon estimation – New South Wales. Technical Report No 34,
Australian Greenhouse Office, Canberra, Australia.

matrix for central west NSW. In: Soil Solutions for a Changing World, 19th World
Congress of Soil Science, 1-6 August, Brisbane, Australia.

Committee on Soil and Terrain, CSIRO Publishing, Melbourne.


Ogle SM, Breidt FJ, Easter M, Williams S, Killian K, Paustian K, 2010. Scale and
uncertainty in modelled soil organic carbon stock changes for US croplands using a

Sensitivity of soil carbon to management and environmental factors within
Australian perennial pasture systems. Geoderma 214/215, 70 – 79.

comparison of pre-European settlement and current soil carbon content of a
forested landscape in upper Michigan. Forest Ecology and Management 113 179-
189.

carbon stocks in cropping soils of Queensland, Australia, as affected by tillage
management, climate, and soil characteristics. Soil Research 51, 596-607.

Parton, WJ, Rasmussen, PE, 1994. Long-term effects of crop management in wheat-
58, 530-536.

M, Woomer PL, 1997. Agricultural soils as a sink to mitigate CO2 emissions. Soil
Use and Management 13, 230-244.

Change Biology 6, 317-327.

Powers JS, Corre MD, Twine TE, Veldkamp E, 2011. Geographic bias of field observa-
tions of soil carbon stocks with tropical land-use changes precludes spatial
extrapolation. Proceedings of the National Academy of Sciences USA 108, 6318–
6322.
AI’92: Proceedings of the 5th Australian Joint Conference on Artificial 

R Core Team, 2013. R: A language and environment for statistical computing. R 
Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-
project.org/ (accessed 22.3.2016).

Reuters HI, 2011. JMP and CUBIST Prediction Toolbox. URL: 
http://www.globalsoilmap.net/content/globalsoilmapnet-prediction-toolbox
(accessed 29.9.2016)

Calibration and Implementation for the National Carbon Accounting System, 

(FullICAM Vers. 1.0) for the Australian continent, Australian Forestry 67, No. 4, 
277–283.

review for Australian agriculture. CSIRO Land and Water, A report prepared for Department of Climate Change and Energy Efficiency.


Smith P, Smith JU, Powellson DS, McGill WB, Arah JRM, Chertov OG, Coleman K, 
Franko U, Frolking S, Jenkinson DS, Jensen LS, Kelly RH, Klein-Gunnewiek, 


Stace HCT, Hubble GD, Brewer R, Northcote KH, Sleeman JR, Mulcahy MJ, 


Chapter 6: Change in soil organic carbon stocks under twelve climate-change projections over New South Wales, Australia

Note: Supplementary results relating to pH and sum-of-bases (macro-nutrients) are presented in Appendix 2.

Abstract

Digital soil mapping (DSM) techniques, in combination with a space-for-time substitution (SFTS) process, were used to map (at 100-m resolution) and examine soil organic carbon (SOC) changes due to projected climate change over New South Wales, Australia until approximately 2070. Twelve climate-change projections were applied, derived from four global climate models downscaled with three regional climate models. A marked variation in both direction and magnitude of SOC change was demonstrated with application of the different climate projections. Mean state-wide predictions (0-30 cm depth) ranged between 2.9 Mg ha\(^{-1}\) gain and 8.7 Mg ha\(^{-1}\) SOC loss. Greater consistency between climate-change projections is required before we can confidently predict future changes of SOC and other soil properties.

Nevertheless, using averaged results from the twelve climate projections, broad trends were revealed for the change in SOC over two broad depth intervals (0-30 and 30-100 cm). A mean rate of loss of 2.0 Mg ha\(^{-1}\) for the upper depth interval is demonstrated, and a total loss of 737 million Mg (tonnes) of carbon dioxide equivalent for the entire depth to 100 cm, but there is a wide 95% confidence interval. Although changes are primarily controlled by the balance between changing temperatures and rainfall, the extent of change is also dependent on the precise environmental regime, with differing changes demonstrated over 36 current climate-parent material-land use combinations. For example, the projected mean decline of SOC is less than 1 Mg ha\(^{-1}\) for dry-highly siliceous-cropping regimes but over 15 Mg ha\(^{-1}\) for wet-mafic-native vegetation regimes. DSM-SFTS techniques offer a viable alternative approach to dynamic simulation techniques to predict and identify patterns in the change of soil properties due to climate change.

Keywords: climate change, soil organic carbon, digital soil mapping, space-for-time substitution
Chapter 6: Change in soil carbon stocks under twelve climate-change projections, NSW, Australia

6.1 Introduction

Climate change will impact on many aspects of the global environment and civilization generally, with changes in our soil resources being one important, yet not widely understood, consequence. Changes in key soil properties such as soil organic carbon (SOC) content will influence agricultural productivity and our ability to feed and support the growing world population. Native ecosystems will be impacted, with modifications to species distribution and abundance at local, regional and national scales. Changes in the potential of soils to sequester carbon or release it to the atmosphere is crucial for climate change modelling and mitigation strategies (Lal et al. 2007; Baldock et al. 2012).

Although there has been widespread work on the relationship of climate to soils, dating back to the pioneering work of Dokuchaev (1899) and Jenny (1941) up to more recent studies such as Fantappie et al. (2011), relatively few studies have developed spatial predictions of changes in soil properties under formal climate-change projections, especially at fine regional scales. Studies that have been reported are primarily focused on changes to SOC, reflecting its importance to climate change modelling and mitigation strategies and also agricultural productivity and the ecosystem health more broadly.

The majority of these studies have applied dynamic simulation modelling approaches such as the carbon dynamics simulation models of RothC (Coleman and Jenkinson, 1999) and SOCRATES (Grace et al. 2006). Global simulation studies on SOC change have been undertaken by several workers and different conclusions are drawn depending on the Global Climate Model (GCM), Regional Climate Model (RCM), IPCC emission scenario (IPCC 2000) and carbon dynamics model selected. Overall global increases in SOC by 2100 have been reported by Lucht et al. (2006), Yurova et al. (2010) and Gottschalk et al. (2012), with the magnitude of increase dependent on the GCM and emission scenario. In contrast, however, Jones et al. (2005) reported a decrease in global SOC by the end of the century as did Smith (2012) and Ito (2005) for at least some GCM and emission scenarios. All studies demonstrate large variations in both the direction and extent of SOC change over different regions of the globe, even from the same climate model. Lal (2004) and Gottschalk et al. (2012) report
higher latitude regions undergoing overall losses with tropical regions undergoing overall gains.

At the country or regional level, the change in SOC content under climate change has been spatially simulated in a number of recent studies, ranging from 250 m to over 50 km resolution, including in North America (Smith et al. 2009; Follet et al. 2012; Dib et al. 2014; Zhong and Xu 2014; Byrd et al. 2015; Orem et al. 2015), in Asia (Hashimoto et al. 2012; Zhao et al. 2013, 2015) and in Europe and the Mediterranean region (Smith et al. 2005, 2006; Álvaro-Fuentes and Paustian 2011; Álvaro-Fuentes et al. 2012).

In contrast to dynamic simulation modelling, there appears to be very little previous use of digital soil mapping (DSM) approaches to the prediction of change in SOC and other soil properties under climate change. DSM approaches use data mining and statistical techniques to predict soil properties using a range of environmental covariates (McBratney et al. 2003). The use of DSM approaches for this purpose was proposed by Minasny et al. (2013), who described it as the “static – empirical” modelling approach, being an alternative to the “dynamic – mechanistic simulation” approach. They demonstrated its application with a 500 km² study predicting SOC change in southern New South Wales, Australia, with 250 m grid spacing. Yigini and Panagos (2016) recently adopted a similar approach in their modelling of SOC stock change (1 km resolution) in Europe to 2050 under climate and land use change.

These DSM approaches could be described as “space-for-time substitution (SFTS)”, a process used to infer future trajectories of natural systems from contemporary spatial patterns. It assumes that the drivers of the spatial patterns also drive temporal changes (Pickett 1989, Blois et al. 2013), for example, the change in a soil property over time due to new given climate conditions can be demonstrated by examining soils from different sites with that given climate but otherwise similar soil forming conditions. Changes in climate over space substitute for changes over time. It has been widely used in biodiversity modelling and increasingly for climate change driven biodiversity studies, although there is debate about its effectiveness (Blois et al. 2013). It has to date, however, rarely been reported in pedological studies, although Barraclough et al. (2015) recently used it to test whether climate change had had an impact on soil carbon contents in England and Wales.
In this study we used DSM techniques combined with SFTS to examine the potential change in SOC from projected climate change over the entire State of NSW in the coming decades. Change in SOC is considered a priority for Australian soil monitoring programs (McKenzie and Dixon 2006). Twelve climate-change projections were applied, derived from four global climate models, each downscaled with three regional climate models, sourced from the NSW and ACT Regional Climate Modelling (NARClIM) project (Evans et al. 2014; OEH 2014). The primary aims of our study were to:

- demonstrate the viability of using a DSM-SFTS technique to spatially quantify changes in soil properties due to the influence of future climate change, with reference to SOC
- assess the consistency of predictions of SOC change between different global and regional climate models
- assess whether the changes in SOC content vary systematically according to environmental conditions, based on current climate–parent material (soil type)–land use regimes.

### 6.2 Assessment methodology

In overview, the process commenced with the compilation of the required SOC datasets and environmental grids representing soil-forming factors over eastern Australia and NSW that included current climate data. Statistical models were then developed over eastern Australia, these effectively representing SOC content under the baseline climate. The broader eastern Australian province was used for model development as it encompassed broader climate ranges that may be encountered under the projected climate change, and because data points were scarce in western NSW (see Figure 6.1). Twelve projected climate grids over three time periods (1990-2009, 2020-2039 and 2060-2079) were then substituted in for the baseline climate data in these models to prepare predictive digital maps of the soil properties over NSW under these future climate conditions. By comparing the predicted future soil property maps with current maps, the extent of change in SOC was derived for each 100-m pixel. Further analysis of the change was undertaken with breakdown according to current climate–parent material–land use sub-classes.
6.2.1 Soil profile dataset

Soil profile datasets over eastern Australia were acquired from the five State government soil resource agencies, based on their 2011 data holdings, plus the Federal Government’s CSIRO data from 2001. These included data collected back to the 1960s and earlier. This combined dataset predates the recently compiled National Soil Site Collation (Searle 2014). Only those profiles with SOC laboratory data, plus parent material descriptors that could be reliably classified were used. The final dataset contained profile numbers as follows: Queensland (1504), NSW (1788), Victoria (224), Tasmania (350), South Australia (585) and CSIRO (eastern States, 760), amounting to a total of 5211 profiles (Figure 6.1). Values reported for each soil horizon over the entire original dataset were converted into five standard depth intervals: 0-5, 5-15, 15-30, 30-60 and 60-100 cm using the equal area spline process of Bishop et al. (1999) and Malone et al. (2009). These intervals conform to those adopted in the Soil and Landscape Grid of Australia (TERN 2014; Grundy et al. 2015) and GlobalSoilMap.net (Sanchez et al. 2009; Arrouays et al. 2014) down to the 100 cm level. Models developed from this large eastern Australian dataset were used to generate the digital maps for NSW alone.

Figure 6.1. Location of SOC modelling profile points and NSW study area
6.2.2 SOC data

The great majority of laboratory analysis of SOC from the above dataset were undertaken with the Walkley-Black wet oxidation method, however approximately 5% used LECO and other combustion methods, as described by Rayment and Lyons (2011). The variation in different laboratory methods, due to the different dates and jurisdictions of the analyses, results in a degree of inconsistency in the test results and potential error in the predictive models. The Walkley-Black method has been reported to underestimate total SOC levels (Skjemstad et al. 2000), but no correction was applied as there is much uncertainty regarding the most appropriate correction factor (Conyers et al. 2011; Bui et al. 2009). Final analysis excluded samples with less than 0.1% SOC, as these were considered unreliable, and greater than 18% SOC, as these are always defined as organic materials in the Australian Soil Classification (Isbell 2002). Such soils are not common in the region, are difficult to model, and the extreme SOC values can distort modelling relationships. Organic carbon mass (Mg ha⁻¹, equivalent to tonnes/hectare) was derived in addition to concentration (%), using the bulk density grids produced through DSM techniques from the Soil and Landscape Grid of Australia (TERN 2014, URL: doi.org/10.4225/08/546EE212B0048).

6.2.3 The covariates

Covariates were selected to represent the key soil-forming factors of climate, parent material, relief, biota and age as outlined below.

Climate

For current climate used in the preparation of initial Cubist models: mean annual rainfall (mm pa, Rain) and mean annual daily maximum temperature (°C, Tmax) – derived from 2.5 km Australia wide climate grids from the Australian Bureau of Meteorology (BoM) with interpolation of cell values down to a 100 m grid, using the ArcGIS Interpolation Spline tool. The values represent mean values obtained over the 1961-1990 period, which coincides with the period when a large proportion of the soil profiles were collected, meaning these climate grids are appropriate for this soil modelling exercise.
For projected climate used in the preparation of output grids under climate change: 12 climate models derived from the NSW and ACT Regional Climate Modelling (NARCliM) program (Evans et al. 2014; OEH 2014) were used for *mean annual rainfall* and *mean annual daily maximum temperature* grids across NSW averaged over the three periods: 1990-2009 (representing “current” conditions), 2020-2039 (“near future”) and 2060-2079 (“moderately far future”). These NARCliM grids were derived from modelled outputs from numerous finer time frames over each time period. The models comprised four global climate models:

- **CSIRO_MK30** (henceforth abbreviated as CSIRO), developed by the Commonwealth Scientific and Industrial Research Organisation and the Bureau of Meteorology in Australia
- **CCCMA31** (CCCMA), developed by the Canadian Centre for Climate Modelling and Analysis
- **ECHAM5** (ECHAM), developed by the Max Planck Institute for Meteorology, Hamburg, Germany
- **MIROC32** (MIROC), Model for Interdisciplinary Research on Climate, developed in Japan by several institutions, each downscaled with three regional climate models, R1, R2 and R3 as described in Evans et al. (2014). The global and regional climate models were selected by the NARCliM program on the basis that they could adequately simulate the NARCliM domain climate but also reflect the full range of projected future climate outcomes. The intermediate A2 emission scenario of IPCC was adopted (IPCC 2000). The initial grids were 0.1 degrees (or 10 km) but were interpolated down to 100 m rasters, using the above-mentioned interpolation spline tool. Checks were undertaken to ensure the modelling process was not attempting to predict significantly beyond the bounds of the 1961-1990 BoM climate, as used in the training data.

The twelve different climate models each presented a different future scenario. Although all models tend to reveal increasing temperatures, there is considerable variation in trends with respect to annual rainfall, with the MIROC and CCCMA models projecting more moist conditions over NSW while the CSIRO and to a lesser extent the ECHAM models projecting drier conditions. The different model projections for maximum temperatures and precipitation over NSW and the whole
NARClM domain are presented in Olson et al. (2014) and OEH (2015, Figures 9.15 and 9.16, URL in References).

Other covariates

Covariates relating to other soil-forming factors are listed below. For the purposes of this work these were assumed to be constant over the modelling period.

Parent material: (i) Silica index (lithology class) – an index representing the composition of the parent material estimated using documented average chemical composition (Gray et al. 2014; Gray et al. 2015a). For example, granite is moderately siliceous with approximately 73% silica, while basalt is mafic material with only approximately 48% silica. Higher silica content parent materials typically give rise to soils with more quartzose sandier textures with lower chemical fertility. For model development, parent material descriptors recorded at each site were used to derive silica indices; while for the final digital soil maps, the 1:250 000 scale NSW Geological Survey polygonal geology map were used. For post modelling interpretation purposes, these were grouped into four classes as shown in Table 6.1, which also presents typically associated soil types. (ii) Radiometrics – gamma radiometric potassium (K), uranium (U) and thorium (Th), derived from airborne surveys from Geoscience Australia.

Relief: (i) Topo-slope index (TSI) – an index that combines topographic position and slope gradient, representing the degree to which a site is subject to depletion (1) or accumulation (6) of water, soil particles and chemical materials. Data was derived from field data and 100 m digital elevation model (DEM) (Gray et al. 2015a); (ii) Topographic wetness index (TWI) – a widely used index that represents potential hydrological conditions based on slope and catchment area, as derived from DEMs (TERN 2014; Gallant and Austin 2015): (iii) Slope - slope gradient in percent as derived from a 100 m DEM: (iv) Aspect index – an index to represent the amount of solar radiation received by sites, ranging from 1 for flat areas and gentle N or NW facing slopes to 10 for steep S and SE slopes (in the southern hemisphere, Gray et al. 2015a), derived from a 100 m DEM.

Age: (i) Weathering Index – an index to represent the degree of weathering of parent materials, based on radiometric data (Wilford 2012). A 90 m grid was accessed from Geoscience Australia
Biota: (i) **Land disturbance index (LDI)** – an index that reflects the intensity of disturbance associated with the land use, where 1 denotes natural ecosystems and 6 denotes intensive cropping, based on 1:25 000 scale land use mapping (OEH 2007; Gray et al. 2015a); (ii) **Ground cover** – total vegetation cover (photo-synthetic and non-photo-synthetic) derived from CSIRO 2011 MODIS fractional vegetation data (Guerschman et al. 2009). These variables are held constant into the future.

### 6.2.4 Developing models and statistical analysis

Analysis was carried out using R statistical software (R Core Team, 2013). The soil dataset was apportioned 80% as training data and 20% as validation data (following stratification by State) with modelling by Cubist linear piecewise decision tree models (Quinlan 1992) using the Cubist package of Kuhn et al. (2014). Natural log transformations were applied to the SOC values to address the observed skewness in the response.

The models for each depth interval were validated using the validation datasets. Lin’s concordance correlation coefficient (CCC) was used to measure the level of agreement of predicted values with observed values, relative to the 1:1 line (Lin 1989). Also determined were the coefficient of determination ($R^2$), root mean square error (RMSE), standardised RMSE (being RMSE/mean estimate) and mean error.

The Cubist models were applied to the NSW covariate grids with the NARCliM projected climate layers to prepare maps over NSW for SOC for each of the three time periods in this space for time substitution (SFTS) process (Pickett 1989, Blois et al. 2013).

Log values were back-transformed into natural scales. The maps were replicated using the 12 NARCliM climate model grids; thus there were maps for 5 depths, 3 time periods and 12 climate model grids, giving 180 maps in total. The five depth intervals were then amalgamated into two broad depth intervals: upper soil (0-30 cm), and lower soil (30-100 cm), using a depth weighted averaging process (but in future a new specific dataset covering these depth intervals should be created to avoid adding unnecessary prediction errors). SOC changes over the two change periods (1990-2009 to 2020-2039 and 1990-2009 to 2060-2079) for each climate model were calculated for each pixel. It was necessary to use the climate model projections for the baseline period rather than
the 1961-1990 BoM climate data to gain meaningful levels of change predicted by each model.

The state-wide change in SOC for both depth intervals and change periods for each of the 12 climate models were presented in column plots. These also show the mean change and 95% confidence interval of change (based on 1.96 times the standard error from the 12 model predictions). Uncertainty analysis of the final map layers was not undertaken, partly because no validation points representing future conditions are available, but further research will endeavour to incorporate cross validation techniques such as those demonstrated by Malone et al. (2014) and Kidd et al. (2015).

6.2.5 Partition into environmental sub-classes

In addition to presenting mean results of change over the entire State, the results from the twelve models were also partitioned according to current climate–parent material–land use sub-classes to observe the degree of response in different environmental regimes. The 36 sub-classes were based on the grouping of the three covariates as follows:

- current climate: three broad classes based on the ratio of the annual rainfall and maximum temperature variables \((\text{Rain}/T_{\text{max}})\): dry \(<25\); moist 25-50; wet \(>50\).
- parent material: four classes as shown in Table 6.1, which also presents typically associated soil types.
- land use classes: three broad classes: native vegetation, grazing and cropping.

Plots were prepared showing the mean and 95% confidence interval (being 1.96 x standard error) for each sub-class based on the 12 climate-change projections, derived from the GIS results.
Table 6.1. Parent material classes and typically associated soils

<table>
<thead>
<tr>
<th>Parent material class</th>
<th>Common examples</th>
<th>Typical Australian Soil Classification soils</th>
<th>Typical WRB soils</th>
<th>Typical Soil Taxonomy soils</th>
</tr>
</thead>
<tbody>
<tr>
<td>Siliceous - upper</td>
<td>quartz sands (alluvial/ aeolian), pure quartzite, chert, quartz sandstone &amp; siltstone</td>
<td>quartzose Rudosols &amp; Tenosols; Podosols</td>
<td>Arenosols; Podzols</td>
<td>Spodosols; (Quartzipsamments), Inceptisols; (Psamments)</td>
</tr>
<tr>
<td>Siliceous - lower</td>
<td>granite, rhyolite, adamellite, granodiorite, dacite, greywacke, feldspathic/lithic sandstone, trachyte, syenite, monzonite, diorite, andesite, mudstone, argillaceous sediments (shale, etc), alluvial grey &amp; brown clays</td>
<td>Kandosols, base poor (low fertility) Chromosols, Kurosols &amp; Sodosols</td>
<td>Acrisols; Alisols; Cambisols; Ferralsols; Retisols; Umbrisols; Durisols; Lixisols; Planosols; Solonetz</td>
<td>Alfisol, Aridisols &amp; Oxisols; (Udults)</td>
</tr>
<tr>
<td>Intermediate (52-65% silica)</td>
<td>gabbro, dolerite, basalt, alluvial black cracking clay</td>
<td>Dermosols; Ferrosols; grey &amp; brown Vertosols; base rich (fertile) Chromosols, Kurosols &amp; Sodosols</td>
<td>Chernozems; Kastanozems; Luvisols</td>
<td>Mollisols (Udolls; Ustolls; Kandiuults; Mollic paleudalf)</td>
</tr>
<tr>
<td>Mafic (&lt;52% silica)</td>
<td>black Vertosols, Ferrosols</td>
<td>Nitisols; Phaeozems; Vertisols</td>
<td>Vertisols</td>
<td></td>
</tr>
</tbody>
</table>

1 First approximations for common soil types only; most soil types will extend into adjoining parent material classes
2 Based on Isbell et al. (1997), Gray and Murphy (1999) and Gray et al. (2014)
3 World Reference Base for Soil Resources; based on IUSS Working Group WRB (2014) and Gray et al. (2011)
4 Based on Soil Survey Staff (2010) and Gray et al. (2011). Orders given in normal font, suborders and sub groups in italics

6.3 Results

Validation results for the initial Cubist models developed using the baseline (1961-1990) climate data are presented in Table 6.2. The models are shown to be of moderate strength with Lin’s concordance values up to 0.73 in the surface layers, but decreasing in strength in lower depth intervals, as mirrored by the rise in the standardised RMSE values.
### Table 6.2. Validation statistics of soil organic carbon Cubist models

<table>
<thead>
<tr>
<th>Depth interval (cm)</th>
<th>N</th>
<th>Lin’s CCC</th>
<th>RMSE (log %)</th>
<th>Std RMSE (log %)</th>
<th>ME (log %)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-5</td>
<td>1029</td>
<td>0.73</td>
<td>0.55</td>
<td>0.66</td>
<td>0.018</td>
</tr>
<tr>
<td>5-15</td>
<td>1007</td>
<td>0.73</td>
<td>0.53</td>
<td>0.73</td>
<td>0.015</td>
</tr>
<tr>
<td>15-30</td>
<td>792</td>
<td>0.62</td>
<td>0.62</td>
<td>1.20</td>
<td>0.020</td>
</tr>
<tr>
<td>30-60</td>
<td>588</td>
<td>0.52</td>
<td>0.65</td>
<td>2.20</td>
<td>0.086</td>
</tr>
<tr>
<td>60-100</td>
<td>426</td>
<td>0.37</td>
<td>0.66</td>
<td>3.80</td>
<td>0.073</td>
</tr>
</tbody>
</table>

CCC: concordance correlation coefficient; RMSE: Root mean square error; Std RMSE: Standardised RMSE (RMSE/mean of estimate); ME: Mean error (prediction – observed)

Results from the five depth intervals were combined into just two depth intervals: upper soil (0-30 cm), and lower soil (30-100 cm). Primary focus was given to the results for the second change period (1990-2009 to 2060-2079). More detailed results and full digital maps (100-m resolution) are available for public download through the Adapt NSW website (OEH 2014). The change is reported on a state-wide basis and then by physical zones including climate–parent material (soil type)–land use regime.

#### 6.3.1 State-wide change

The absolute change in SOC stocks across NSW for both depth intervals and change periods predicted from each of the 12 climate models and their mean is presented in Figure 6.2.

The predicted changes vary substantially with the different climate models and their differing projections. The 95% confidence interval encompasses the zero change in all four columns of Figure 6.2, albeit generally only narrowly. The MIROC and CCCMA models, the wetter models, almost all suggest an increase in SOC stocks over both depth intervals and change periods, up to over 2.9 Mg ha⁻¹ (tonnes/ha) increase being predicted by the MIROC3 model for the upper depth interval, 2nd change period. By contrast the CSIRO models, the driest models, suggest notable decreases of up to 8.7 Mg ha⁻¹ for CSIRO1 for the same depth and change period. The ECHAM models all reveal a slight decrease in SOC stocks.
Figure 6.2. Change in SOC stocks from the 12 climate models for both change periods
Notes: yellow for change period 1 (1990-2009 to 2020-2039); orange for change period 2 (1990-2009 to 2060-2079); thin black lines: the mean change across NSW for each climate model; thick red lines: mean from all 12 climate models.

Nevertheless, using averaged results from the twelve climate models, broad trends in change are revealed. The results suggest an overall decline in SOC stocks across NSW, with the extent of change becoming less pronounced at lower intervals and more pronounced over the 2nd change period. From the mean of the 12 models, in the upper depth interval (0-30 cm), there is an average 1.5 Mg ha\(^{-1}\) (tonnes/ha) decrease to the 2030 period and 2.0 Mg ha\(^{-1}\) to the 2070 period. There is only a slight mean decline in the lower depth intervals (30-100 cm), with only a 0.7 and 0.4 Mg ha\(^{-1}\) decline to the 2030 and 2070 periods respectively. Close examination of these results reveals that the overall rate of SOC decline per year is less for the second change period than for the first change period, which is attributable to the increasingly higher rainfall projected in the later years that at least partly compensates for the steadily rising temperatures.

Based on the average of all 12 climate models, and recognising the substantial uncertainties, most of the eastern and central-eastern regions of the State are projected to undergo a decrease in SOC stocks, with the most notable decline occurring over the alpine Snowy Mountain region of the far south east, while the central-western and
western regions undergo only a slight increase in stocks, as shown for the upper 0-30 cm interval by Figure 6.3. There are several exceptions to these trends, such as moderate increases in stocks in the central coastal regions around Sydney.

![Image of NSW with SOC change map](image)

Figure 6.3. Average change in SOC stock across NSW for 2nd change period (0-30 cm, Mg ha\(^{-1}\), mean of 12 NARClIM models)

The results can be used to gain an estimate of total change of soil organic carbon and equivalent carbon dioxide (CO\(_2\)) over NSW due to climate change, as presented in Table 6.3. Given the area of NSW is 80.36 million ha, the total SOC loss from the top 100 cm, based on the average rates of SOC loss from the 12 models, is estimated at 193 million Mg (tonnes) SOC or 737 million Mg CO\(_2\) equivalent for the far change period. The 95% confidence interval for CO\(_2\) equivalent change ranges from a gain of 321 million Mg to a loss of 1735 million Mg. For comparison purposes, the mean estimate of change (loss to atmosphere) is equivalent to approximately five times the total greenhouse gas emissions over NSW (estimated at 142 million Mg in 2013, including other greenhouse gases and agricultural emissions (OEH 2015), but with the 95% confidence interval ranging from over 2 years equivalent addition to the soil to over 12 years equivalent loss to the atmosphere.
In addition to uncertainty arising from the different climate models, there is uncertainty from the digital modelling and mapping process used to derive the change estimates. The RMSE of the initial models for the original five depth intervals is relatively high, ranging up to 0.66 (% log scale) in the deepest layer (Table 6.2). Additional uncertainty parameters associated with final map generation were not quantified, but are also likely to be significant. The predicted state-wide changes in SOC all appear to be within the envelope of uncertainty. Further treatment of sources of uncertainty is given later in the Discussion.

Table 6.3. Change in SOC stocks and CO₂ equivalent over NSW over 2nd change period (to approximately 2070)

<table>
<thead>
<tr>
<th>Depth (cm)</th>
<th>Change in SOC stock per ha (Mg ha⁻¹)</th>
<th>Total change (million Mg)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SOC</td>
<td>CO₂ equivalent</td>
</tr>
<tr>
<td>Upper 95%</td>
<td>0-30</td>
<td>0.116</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>0.974</td>
</tr>
<tr>
<td></td>
<td>0-100</td>
<td></td>
</tr>
<tr>
<td>Lower 95%</td>
<td>0-30</td>
<td>-4.186</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>-1.707</td>
</tr>
<tr>
<td></td>
<td>0-100</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0-30</td>
<td>-2.035</td>
</tr>
<tr>
<td></td>
<td>30-100</td>
<td>-0.367</td>
</tr>
<tr>
<td></td>
<td>0-100</td>
<td></td>
</tr>
</tbody>
</table>

Area of NSW: 80.36 million ha; Mg = tonne

6.3.2 Change by environmental regime

The projected SOC changes are primarily dependent on the balance between the changing temperatures and rainfall over any region, generally decreasing with rising temperatures and declining rainfall (Jenny, 1980; Badgery et al. 2013; Viscarra Rossel et al. 2014; Gray et al. 2015b). However the extent of the SOC change also varies depending on the environmental and land use regime, which adds complexity to the above trends. A breakdown in the SOC change for the upper (0-30 cm) depth intervals over the 2nd change period by current climate–parent material–land use sub-class is presented in Figure 6.4. Each column presents the mean plus the upper and lower 95% confidence interval derived from the 12 climate models for each sub-class. Only where the column does not intersect the zero change line can we be confident (at 95% level) of
a change based on the 12 climate models (but not considering the additional sources of uncertainty).

The plots reveal that the extent of SOC stock change varies with different current climate–parent material–land use regimes. For example, a mean loss of less than 1.0 Mg ha\(^{-1}\) is projected for the dry–highly siliceous–cropland regime, whilst a loss of 15.3 Mg ha\(^{-1}\) is projected for the moist–mafic–native vegetation regime. Although there are several anomalies, a trend of increasing loss of SOC stock with increasingly moist current conditions, less siliceous (more mafic) and less intensive land use is apparent.

![Figure 6.4. 95% confidence interval and mean change in SOC stocks by physical zone from the 12 NARClM models (Mg ha\(^{-1}\), 0-30 cm, 2\(^{nd}\) change period)](image)

The vertical spread of bars represents 95% confidence interval, dark line is mean. Note the greatest loss in wet-mafic-native vegetation regimes

### 6.4 Discussion

#### 6.4.1 Variation of predictions between models

The final predictions and spatial arrangement of SOC changes across NSW vary greatly with the different climate models and their differing projections, which hinder our ability to draw clear conclusions. There are some climate models that predict an increase in SOC magnitude but others that predict a decrease. For example, at the State-
Chapter 6: Change in soil carbon stocks under twelve climate-change projections, NSW, Australia

wide level, 2nd change period, 0-30 cm depth, the predicted change in SOC stocks varied from a 2.9 Mg ha\(^{-1}\) increase to an 8.7 Mg ha\(^{-1}\) decrease, depending on the model applied. Such variation applies at the State-wide level down to smaller regions and individual environmental regimes.

A wide range in future SOC stock projections derived by using different global climate models and emission scenarios has been widely demonstrated (Ito 2005; Lucht et al. 2006; Yurova et al. 2010; Gottschalk et al. 2012). The uncertainty in SOC introduced by different climate models is reported to be greater than the uncertainty introduced by CO\(_2\) emission pathway scenarios (Smith et al. 2006; Falloon et al. 2007; Gottschalk et al. 2012).

The variation in SOC projections between different climate models over NSW in this study appears to be mainly attributable to the wide variations in projected regional rainfall change, as against the more uniformly increasing regional temperatures, a situation also reported by Falloon et al. (2007) in their global study.

Additionally, the application of other IPCC emission scenarios apart from the intermediate A2 scenario selected in this study, would introduce further variations in the climate projections and associated soil property predictions. As the scientific community refine global climate-change projections we may acquire greater certainty in them, and thus obtain more reliable projections of soil property change into the future.

6.4.2 Broad trends in predicted soil property change

Despite the differences demonstrated by each of the twelve climate models, and the considerable modelling uncertainties, some broad trends are apparent. From the average of the models over NSW as a whole, an overall slight to moderate decline of SOC stocks was revealed. Declines are most pronounced in the east of the State, particularly in the highlands of the far south east. By approximately 2070 a mean estimate of 737 million Mg (tonnes) of carbon dioxide equivalent are projected to be lost from NSW soils to the atmosphere, however the 95% confidence interval of estimates are broad. When compared with the approximate total greenhouse gas emissions from NSW (142 million Mg in 2013), the change does not appear dramatic.

The results from this study reflect well established soil property–climate trends. Drier conditions result in lower SOC levels (Jenny 1980; Lal 2004; Badgery et al. 2013;
Viscarra Rossel et al. (2014; Gray et al. 2015b; Hobley et al. 2015). The direction and magnitude of change at any location will be defined by the integrated effects across all processes involved in emission and consumption or storage of carbon (Baldock et al. 2012; Gottschalk et al. 2012).

Additionally, this study has demonstrated that the extent of change is also influenced by the precise environmental regime, as represented by the current climate–parent material (soil type)–land use combination. A broad trend of increasing loss of SOC stock with increasingly moist current conditions, less siliceous (more mafic) and less intensive land use is apparent, despite some anomalies, a trend we also observed following large scale vegetation clearance over NSW (Gray et al. 2016). These trends suggest that the extent of absolute change in SOC with the projected climate change is broadly dependent on their initial levels in the soil. Soils with inherently high SOC levels will lose more SOC than those with inherently low SOC levels, at least in absolute terms. We recently demonstrated the inherently higher SOC storage levels of soils under moist, mafic parent material (with associated higher fertility soils) and high vegetation cover regimes across eastern Australia (Gray et al. 2015b).

The overall slight to moderate decline of SOC stocks in NSW, particularly in the east of the State, suggested by our study accords with the declining SOC levels over the small area of southern NSW reported by Minasny et al. (2013) who appeared to use climate projections from the Australian Government based on up to 40 global climate models (Climate Change in Australia 2015). The generally hotter and drier conditions in the future for Australia as reported by Baldock et al. (2012) using the above climate projections would suggest lower SOC levels, but those authors themselves state the direction and magnitude of SOC change is still under debate, due to the modelling uncertainties.

Our results are broadly consistent with those presented in a global map (0.5 degree resolution) by Gottschalk et al. (2012) using 10 climate model-emission scenario combinations to 2100. There is however, little consistency with the global maps presented at the same coarse scale and time frame by Ito (2005), who shows broad increases based on seven climate models and the A2 emission scenario, and Lucht et al. (2006) who show varying trends with two global climate models.
The trends of change in SOC vary substantially in different parts of the State, depending on the precise combination of key environmental factors, particularly current climate, parent material (reflecting soil type) and land use. The predictions vary from a notable rise in some regions of the State to a notable decline in other regions over the two change periods. The extent and direction of change tends to follow observable trends with respect to the above environmental factors, with greater declines being associated with wetter, more mafic and less disturbed land use regimes.

6.4.3 Application of the SOC change maps

Complete digital maps at 100-m pixel resolution and further detail on our results for the SOC change are available for public download through the Adapt NSW website (OEH, 2014). The predicted changes have implications for the future condition of NSW soils and climate change mitigation strategies. Soil condition and agricultural productivity generally improve with increased organic carbon content due to the enhanced physical, chemical and biological properties (Jenny 1980; Charman and Roper 2007; Lal et al. 2007; Baldock et al. 2012; Murphy 2015). The opposite effect applies with declines in SOC content. Farmers may need to consider measures to counter potential declines, such as implementing management practices that better retain soil moisture (Stokes and Howden 2010). Changes in soil condition due to climate change may also impact on native ecosystems (Steffen et al. 2009).

The maps deserve consideration during any climate change mitigation programs that are based upon increased soil carbon sequestration (Lal et al. 2007; Wilson et al. 2011; Baldock et al. 2012; Smith 2012). Those regions of the State where the soil carbon storage has been shown to be on a declining trend, such as in the east and south, will require even greater carbon enhancing actions to produce the desired soil carbon increases. On the other hand, those regions of the State with a projected slight increase in SOC, such as in the west and central west, will gain assistance towards their carbon sequestration programs.

6.4.4 Modelling uncertainties

There are numerous sources of uncertainty associated with the application of the digital soil mapping – space-for-time substitution (DSM-SFTS) technique in this study in addition to the uncertainty arising from the different climate models, which need to
be recognised when assessing results and drawing final conclusions. Uncertainties arise from the initial establishment of the Cubist statistical models based on the training soil point dataset, as reflected in the relatively high RMSE values (Table 6.2). Further uncertainties are associated with creation of the actual digital map layers based on these models with available covariate grids. Final map uncertainties were not quantified in this study. These uncertainties reflect inherent weaknesses in our modelling process that are common to many DSM projects, such as: inherently poor relationships between the soil properties and selected environmental covariates; lack of representativeness of particular environments in the training dataset; errors and inconsistencies in laboratory data; weaknesses in covariate grid layers including downscaling of coarse climate and other grids to 100 m; and reliance on other modelled data such as the bulk density layer (McBratney et al. 2003, Nelson et al. 2011; Bishop et al. 2015; Robinson et al. 2015).

Through the use of the SFTS process, there is an assumption that the statistical relationships with climate developed under current conditions will hold true into the future with changing climate, but there is uncertainty in this. Some minor potential anomalies were evident in components of the Cubist decision tree models. Occasional rules of the decision trees, representing particular covariate combinations at particular depths, had climate trends in the terminal regression models that did not accord with well-established soil property–climate relationships, and thus gave questionable results when applied under future climate regimes. For example, the second of nine rules for SOC at the 5-15 cm interval, indicated a very slight increase in SOC with increasing temperatures rather than the expected opposite trend, a trend that carries through into the future maps for that covariate space. This may represent a weakness in the application of the SFTS process with the Cubist DSM approach. Some covariates included in the study, such as ground cover, may also have a certain climatic signal but these were kept constant throughout the future modelling periods, and thus may distort final results. The extent and consequences of such interactions should be explored prior to final covariate selection in future studies.

No account was taken of other related impacts associated with climate change such as changes to land use and management, vegetation patterns, erosion hazard, fire regimes and seasonal climatic patterns that may all impact on soil conditions. Smith et al. (2006) claim changes in land use and vegetation structure will outweigh the effects of climate change in influencing future SOC levels in European forests. The modelling
could be improved with better assumptions and detail on these conditions in the future time periods.

Feedback processes related to soil–atmosphere carbon dynamics were not included in the applied climate models. These include the exchange of carbon between the soil and atmosphere and possible increased vegetation growth and organic inputs arising from increased atmospheric CO₂ levels (Ito 2005; Jones et al. 2005; Friedlingstein et al. 2006; Smith et al. 2008; Ostle et al. 2009; Gottschalk et al. 2012). Below-ground processes in the carbon cycle–climate system are reported to be much less understood than the above-ground processes (Heimann and Reichstein 2008, Zhong and Xu 2014).

The modelling did not consider the period of time required by soils to re-equilibrate with the changing climatic conditions, an issue recognised by Baldock et al. (2012). A period of 20 years for soil carbon to approach equilibrium has been used by the IPCC, but equilibrium is reported to be reached in 100 years in temperate regions and even longer in boreal regions (IPCC 1997; Smith 2008). Periods of 20 to 50 years for SOC re-equilibration to altered land use were demonstrated by Guo and Gifford (2002) and Skjemstad et al. (2004). Nevertheless, the clear response of SOC and other key soil properties to the current prevailing climate has been remarked upon by several workers (Jenny 1980; Bui et al. 2006).

Future research in this field will need to quantify levels of uncertainty from all sources of potential error, particularly if predicted values are to be applied in carbon trading schemes (Baldock et al. 2012). The more dynamic simulation approaches, as adopted in most previous soil – climate change studies, may be more successful in addressing many of these uncertainties than the DSM-SFTS approach applied here. Nevertheless, the DSM-SFTS approach has been shown to present useful first approximations of the likely changes in SOC under different climate change scenarios.

6.5 Conclusion

This study has demonstrated that application of digital soil mapping - space-for-time substitution (DSM-SFTS) approaches can be useful in the fine scale spatial prediction of change in SOC content under the influence of climate change, although
several sources of uncertainty are identified. The approach is a feasible alternative to the more widely adopted dynamic simulation approaches.

It was demonstrated that the choice of climate-change projection has a great bearing on the final predictions, with the direction and magnitude of predicted SOC changes across NSW varying markedly between the twelve global and regional climate models applied. It is evident that greater consistency between climate-change projections is required before we can confidently predict SOC and soil behaviour generally under climate change.

Despite the uncertainties, the averaged results from the different climate models provide a useful first approximation of likely changes in SOC as a result of climate change across NSW over the coming decades. The results add to our knowledge of SOC behaviour under climate change. It was demonstrated that although the SOC changes are primarily controlled by the balance between changing temperatures and rainfall, the extent of change is also dependent on the precise environmental regime, as represented by the current climate–parent material (soil type)–land use combination.

Results such as presented here for NSW may assist land managers to more fully prepare for potential changes in soil condition due to ongoing climate change. They may provide valuable inputs into other environmental modelling programs and have implications for climate mitigation strategies based on soil carbon sequestration. Further refinement of global and regional climate-change projections will help to improve the quality and reliability of soil property change maps.

Acknowledgements

Yvonne Scorgie (NSW Office of Environment and Heritage, OEH) assisted in the coordination of this study as part of the NARClIM program. Polly Mitchell, Hamish Clarke and Terry Koen (OEH) and Brian Wilson (University of New England) provided constructive advice and assistance.
Chapter 6: Change in soil carbon stocks under twelve climate-change projections, NSW, Australia

6.6 References


Chapter 6: Change in soil carbon stocks under twelve climate-change projections, NSW, Australia


Chapter 7: Summary and discussion

This research project has used DSMM techniques to attempt to shed light on factors influencing soil distribution and behaviour under changing land use and climate conditions. The broad and specific research issues and questions as outlined in Chapter 1 have been addressed in Chapters 2 to 6, the contents of which have all been submitted as journal papers (published or under review).

This concluding chapter provides a summary and brief discussion of the main results and findings of the project and how these address the previously raised research issues. It begins with treatment of the 17 specific research issues before synthesising these into answers for the five broad research issues. This is followed by a listing of possible future research issues and then finally a broad conclusion.

7.1 Specific research issues

7.1.1 Influence of lithology in soil formation and its use in digital soil mapping

This research issue was primarily addressed in Chapter 2, but also in Chapter 3 and elsewhere. The specific research questions raised are answered below.

Can we further elucidate the relationship of lithology to key soil properties?

The proposed classification of parent material into twelve lithology classes, based on mineralogical and chemical composition facilitated the development of semi-quantitative relationships, including multiple linear regression models for multiple key soil properties over NSW and eastern Australia. The classification is believed to have more utility for DSM than the many existing schemes that emphasise origin, such FAO (2006) and Juilleret et al. (2016). The properties examined included soil organic carbon (SOC), pH, cation exchange capacity (CEC), sum-of-bases, total P and clay%. These models allowed for quantitative estimates of changes in these soil properties based on changes in lithology class (and silica index), such as the change in SOC in soils derived from basalt to granite.

The strong influence of lithology in controlling the distribution of these properties was demonstrated through the use of standardised regression coefficients and Random
Forest variable importance plots. These showed lithology to have the highest influence of all soil-forming factors for all six soil properties examined over NSW, but over eastern Australia, climate variables were slightly stronger, or of equivalent influence, for SOC and pH.

This research has helped to elucidate relationships between lithology and soil, an area which has received little effective treatment in the past. It goes further than the previous essentially qualitative descriptions, basic litho-functions or semi-quantitative relationships reported over specific regions, such as by Cline (1953), Cathcart et al. (2008), van de Wauw et al. (2008) and Gruba and Socha (2016). Some regional refinement of the 12 class classification scheme may be necessary to address local regional geological conditions.

How effective is properly organised lithology data relative to other geophysical covariates in digital soil model and map products developed for NSW?

The lithology covariate was demonstrated to exert considerably greater influence on all of the six soil properties examined than any of the geophysical parent material covariates applied, such as gamma radiometrics or hyperspectral imagery derived clay composition. The Random Forest variable importance plots showed lithology to have two to five times more influence than the next highest ranked geophysical covariate.

A strong improvement in both model and map quality was demonstrated when the lithology covariate was included in conjunction with the geophysical covariates. This was exemplified by the Lin’s concordance validation statistics for the Cubist sum-of-bases model, which rose from 0.46 with no parent material covariates to 0.58 with the continuous geophysical covariates to a high of 0.77 when lithology was also applied. The improvement was typically slightly less marked in the final digital maps than for the calibration models, probably due to the lower reliability of the lithology grid, derived from broad scale polygonal geological and soil data.

The results highlight the strong effectiveness of lithology as a covariate in DSM programs, as recently also suggested by Hengl et al. (2014) and Miller et al. (2015), particularly when properly organised data are applied. Although geophysical data are widely regarded as the primary source of parent material covariates for DSM (McBratney et al. 2003; Mulder et al. 2011), the results here suggest conventional geology/lithology data can also be valuable. In the call for new, more diverse covariates
App
endix 1: Background to digital soil mapping

(Hengl 2014; Brevik et al. 2016a,b; Minasny and McBratney 2016), the value of polygonal lithology should not be overlooked. The translation of conventionally described geological data into this simple pedological classification will allow the full value of lithological data to be realised in DSM program. Further research establishing the relationship between lithology and various geophysical data sources would be useful, for example establishing hyper spectral imagery signatures for different lithology classes.

Can we demonstrate an appropriate methodology for the use of lithology class data in DSM programs?

A process was suggested for the application of lithology data, typically derived from geological and soil maps, into DSM programs. Issues to consider were raised, such as the suitability of available data relative to the scale and purpose of final maps. Despite the potential drawbacks of using polygonal data, properly organised and suitably scaled categorical lithology data can strongly complement other continuous geophysical data sources in DSM programs, as recently also demonstrated by Miller et al. (2015).

7.1.2 Relationship of soil-forming factors to key soil properties and their use in digital soil mapping

This research issue was primarily addressed in Chapter 3, but also in Chapters 2, 4 and elsewhere. The specific research questions raised are answered below.

Can we develop effective pragmatic relationships, such as multiple linear regression (MLR) models, for key soil properties over eastern Australia, based on readily available pragmatic variables?

The study succeeded in developing relatively straightforward and easily applied MLR models between eight key soil properties and readily available covariates representing the major soil-forming factors, over three depth intervals to 100 cm. The covariates can all be obtained from field data and readily available climatic data. Validation results indicated the models are of variable but generally moderate statistical strength, with concordance correlation coefficients in the order of 0.7 for SOC upper
Appendix 1: Background to digital soil mapping

depth, pH, sum-of-bases, CEC and sand, but somewhat lower (0.4-0.6) for SOC lower depths, total phosphorous, clay and silt.

These models provide a sub-continental scale example of quantitative relationships between soil-forming factors and key soil properties, as called for by Heuvelink (2005). They would appear to represent at least a partial solving of the fundamental soil equation of Jenny and others in a readily interpreted manner, and to have gone further in this respect than previously reported. Further research should be undertaken to determine how closely these models relate to those prepared for other continents and at a universal (global) level. Grunwald (2005) asserts that it is not possible to develop a single universal model, only models for particular domains, but this assertion needs testing.

*Do they inform us on the quantitative influence of each soil forming factor?*

The models provide quantitative information on the influence of each covariate, representing the different soil-forming factors, on each of the eight soil properties. The standardised regression coefficients for each covariate inform on the relative influence of each covariate in the models. This showed that parent material (lithology) and climate were the key drivers for most soil properties over eastern Australia. The partial regression coefficients allow for quantitative estimates of the changes in each soil property with a given unit of change in the covariate, for example, a decrease in pH of 0.11 units with each 100 mm rise in rainfall (with other factors remaining constant) was demonstrated. Such estimates do, however, assume linear relationships, so they should only be considered first approximations.

The models provide valuable insights into the relative roles of the various soil-forming factors in controlling soil formation and distribution over eastern Australia. It needs to be recognised however, that the magnitude of influence and the relative order of influence varies according to the scale of the subject area, as recently stressed by Miller and Schaetzl (2016). Thus, for example, at the NSW state scale lithology had the highest influence on SOC and pH, however at the eastern Australian scale climate had the highest influence on these properties.

*What is the feasibility of preparing digital soil maps using this pragmatic approach over NSW?*
The relatively straightforward models were easily applied with readily available covariate grids to prepare DSMs with 100-m resolution for the eight soil properties over NSW. The predictive ability demonstrated by the maps was again broadly moderate, with Lin’s concordance generally between 0.4 and 0.7, but typically slightly lower than for the original models.

There is considerable potential for improvement in this pragmatic DSM approach and in final predictive performances, through for example, the application of finer scale, more reliable covariate data such as lithology and land use data, and the incorporation of an age related covariate.

How do the pragmatic models and maps compare with other more advanced DSM techniques?

The pragmatic models and derived maps appear to compare favourably with those derived from more sophisticated approaches such as the Cubist piecewise linear decision tree technique using advanced remotely sensed covariates, based on results from a comparison undertaken over the Hunter Valley Region. Overall, only a slight improvement in map validation statistics with the more sophisticated techniques was demonstrated.

More specifically, it was observed that: (i) the MLR technique is slightly less effective than the Cubist technique; (ii) for the development of models, the pragmatic covariates from site specific data used in this study appear to be equal to or outperform the remotely sensed covariates used; and (iii) for the preparation of digital soil maps, remotely sensed covariates appear slightly more robust than broader polygonal based pragmatic covariates, as they provide more accurate data at individual pixel level, however Miller et al. (2015) demonstrated a combination of both these provides the best outcomes for DSMs.

The maps prepared using this pragmatic approach show validation performance equal to most other DSMs prepared in other studies in Australia and internationally, Nevertheless, they do not quite match the performance of the recently prepared Australian DSMs produced using Cubist and multiple sophisticated covariates by Viscarra Rossel et al. (2014; 2015) which appear to have set a new benchmark of performance.
What are the main weaknesses and potential benefits of the pragmatic approach?

Weaknesses of the pragmatic DSMM approach include (i) reliance on polygonal data, such as mapped geological units; (ii) maps of slightly lower accuracy than possible with the more sophisticated approach with advanced continuous geophysical covariates; and (iii) less potential to precisely quantify the levels of uncertainty than possible with more sophisticated techniques.

Potential benefits of the pragmatic approach include: (i) ability to apply models to derive estimates over individual sites with field collected and climate data alone as might be utilised by a range of environmental and natural resource field scientists; (ii) the process logic and covariate data underlying them can be readily understood, accessed and interpreted, particularly by non-DSMM experts, thus facilitating their application by many end users (Bouma 2014; Brevik et al. 2016a,b; Minasny and McBratney 2016); (iii) potential to serve as a useful introduction to the more sophisticated DSMM approaches for many soil scientists and thereby encourage greater acceptance of DSMM products and strategies generally (Scull et al. 2003; Hartemink et al. 2008; Hempel et al. 2008; Moore et al. 2010); and (iv) the models provide useful quantitative data on the influence of each soil forming factor and more broad pedologic insights into the factors governing soil property distribution and change.

7.1.3 Factors controlling the distribution of soil organic carbon stocks, spatially and with depth

This research issue was primarily addressed in Chapter 4, but also in Chapters 2 and 3 and elsewhere. Useful quantitative information was derived on the factors controlling SOC storage and how these vary with depth using digital soil models and maps prepared over eastern Australia. The results contribute to our knowledge and understanding of the three dimensional spatial distribution of SOC across the landscape, a particular research challenge raised by McBratney et al. (2014). The specific research questions raised for the current project are answered below.

What are the key drivers of soil organic carbon stocks in the soils of eastern Australia and how do they vary with depth?

Climate (rainfall and maximum temperatures) was demonstrated to have the greatest influence in controlling SOC stocks over the entire eastern Australia province,
with parent material (lithology) and vegetation cover also being key drivers, while topography and aspect are of lesser influence, at least at this sub-continental scale. At the scale of NSW alone, parent material (lithology) was demonstrated to exert the greatest influence. The relative influence of temperature and land use/vegetation cover decreases with depth, while that of parent material increases, results that parallel those reported by Jobbágy and Jackson (2000), Albaladejo et al. (2013) and Wilson and Lonergan (2013). The contention that SOC in deeper soil layers might reflect historic rather than current land use (Schulp and Veldkamp 2008; Wilson and Lonergan 2013) suggests that this deep SOC may not be in equilibrium with current environmental conditions.

It was demonstrated that it is the combined influence of the key soil-forming factors that drives the soil carbon levels. Each factor has a different broad level of influence, being significant at differing scales (Miller and Schaetzl 2016), and they combine together to control final SOC stocks. Minor topographic influences are superimposed on the moderate land use/ground cover influences, which in turn are superimposed on the large parent material (soil type) influences which are ultimately superimposed on the very large climatic influences.

Are there systematic patterns of SOC stock levels according to climate – parent material/soil type – ground cover/land use?

Systematic patterns of SOC stock levels were graphically demonstrated over 45 different climate-parent material-vegetation cover regimes for upper soils (0-30 cm) and lower soils (30-100 cm) of eastern Australia. There are generally uniform trends of increasing SOC stocks with increasingly moist climate, increasing mafic character of parent material and increasing vegetation cover. Average SOC stocks in the 0-30 cm depth interval range from 16.3 Mg ha\(^{-1}\) (t/ha) in dry, highly siliceous parent material environments with low vegetation cover, up to over 145.0 Mg ha\(^{-1}\) in wet, mafic parent material environments with high vegetation cover.

It was shown that the use of only two of these three key factors would clearly result in unreliable SOC stock estimates. For example, SOC density over the 0-30 cm interval in a wet, high vegetation cover regime varies from 56.0 Mg ha\(^{-1}\) in soils from extremely siliceous parent material up to 145.0 Mg ha\(^{-1}\) for soils from mafic parent material, a 2.6 fold increase.
These results provide further insights into the issue of soil SOC saturation, the limiting capacity of a soil to accumulate carbon, an “unknown” of Stockmann et al. (2013) and a key research challenge raised by McBratney et al. (2014). Although those workers suggest this capacity is largely controlled by soil texture (quantity of clay) the results here reveal it is more strictly dependent on lithological composition, which controls clay type as well as clay content, in addition to climate and vegetation cover controls. The need to combine multiple soil-forming factors in order to derive meaningful estimates of potential SOC stocks forms the basis of the “carbon matrix zones” of Murphy et al. (2010) and the “potential capability index” for additional SOC storage of Baldock et al. (2009).

*Are there systematic trends in topsoil / subsoil SOC storage ratios?*

The results from this project support the widely held contention that subsoils contribute a substantial proportion of total SOC stocks despite their lower SOC concentration (Batjes 1996; Jobbágy and Jackson 2000; Rumpel and Kögel-Knabner 2011; Cotching 2012). The 30-100 cm interval was shown to contribute approximately half of the SOC stocks down to 1 metre, but the high uncertainty in the modelled results, particularly in the lower interval, means they have to be treated cautiously.

The results also revealed that the precise contributions of upper and subsoil levels vary depending on climatic influences. It was demonstrated that in the dry climate zones of eastern Australia the majority of carbon stock in the top metre is stored in the subsoil (30-100 cm), with an average 54%, whereas in wet climates a lower proportion is in the subsoil (average 43%). Climate appears to be the dominant driving influence of this ratio, although parent material and vegetation cover also have partial influence. These results parallel results reported by Jobbágy and Jackson (2000) and Chaplot et al. (2010). As the SOC present in deeper levels is believed to be older and more stable (Wilson and Lonergan 2013) this may be important knowledge in planning for increased long term carbon sequestration.

### 7.1.4 Change in soil organic carbon stocks following clearing of native vegetation

This research issue was primarily addressed in Chapter 5. The specific research questions raised are answered below.
Can we improve on currently available data on pre-clearing SOC stocks by applying DSMM techniques?

This research project demonstrated the viability of DSMM techniques, primarily Cubist piecewise linear decision trees, to develop maps of pre-clearing (pre-European) SOC stocks over NSW. These new maps display a greater level of detail than the existing conventional polygonal map of Banks and McKane (2002), with continuous values down to each 100-m pixel. In terms of statistical predictive ability, the Cubist derived map from this study significantly outperformed the earlier map, with Lin’s concordance being 13% higher and RMSE and median absolute error being approximately 33 and 50% lower respectively.

Pre-clearing SOC maps prepared using this or similar DSMM approaches are rare (for example, Mishra et al. 2012) but they have high potential to replace the existing conventionally produced polygonal maps. They could provide improved baseline data on SOC stocks for many carbon accounting schemes such as those prescribed by the Intergovernmental Panel on Climate Change internationally and in Australia (Richards 2001; IPCC 2006); carbon turnover models such as RothC (Coleman and Jenkinson 1999); and for soil condition monitoring programs (Chapman et al. 2011).

Can we use DSMM techniques to determine the loss in SOC over NSW since clearing?

Comparison of the pre-clearing SOC stock digital map against another prepared digital map of current SOC stock allowed a third map to be prepared of SOC change that has occurred across NSW since clearing took place. A decrease was demonstrated over most of the state where clearing has occurred except for localised areas mainly in the far west of the state, which were within the uncertainty ranges of the models, for example, within reported model RMSE values. Total pre-clearing SOC stocks amounted to 4.21 Gt in the top 30 cm which compared to the estimate of current stocks of 3.68 Gt, suggesting a total SOC loss of approximately 0.53 Gt (530 million Mg or tonnes), or 12.6% over the entire State since clearing. The decline in SOC density over the cleared areas only was 7.43 Mg ha⁻¹, which equates to an overall average decline of 16.3% over these lands. These results provide useful first approximations, but need to be treated with caution due to the associated high uncertainty levels, resulting from issues such as the inability to incorporate the period of time since clearance into the modelling.
The results confirm a significant decline in SOC following conversion of native vegetation to agricultural land, as widely reported in Australia and internationally. Globally a loss of 3-5% SOC is apparent based on estimates reported by Lal and Follet (2009). The SOC change maps from this project tell us, importantly, that the level of loss is not uniform across the state, but dependent on environmental conditions and land use as discussed below.

**Are there systematic patterns of change relative to climate, parent material and final land use?**

The project demonstrated a systematically greater decline in SOC, in both absolute and relative terms, following native vegetation clearing in cooler (moister) regimes, with more mafic (less siliceous) parent material, and to more intensive land uses. The greatest decline in SOC over the top 30 cm from 56 different climate–parent material–land use regimes was demonstrated in cooler (moist) conditions over mafic parent materials under intensive cropping land use, with a 44.3 Mg ha\(^{-1}\) (t/ha) or 50.0% loss. By comparison, the losses in warmer climates over highly siliceous parent materials under grazing land uses was less than 1 Mg ha\(^{-1}\) or 4%. It is evident that the higher the initial SOC storage level, the greater is the loss of SOC upon vegetation clearance for equivalent land use changes, in both absolute and relative terms.

It is clear that changes in SOC stocks following clearing are controlled by a combination of environmental and land management factors, and that individual factors such as the final land use cannot be considered in isolation. It is reasonable to postulate that similar but opposite trends may apply following conversion of agricultural land to native vegetation, given sufficient time for re-equilibration with the new environmental conditions. These results contribute to our understanding of SOC changes under different land use and management scenarios, which is so important if soil carbon sequestration is to be an effective strategy to mitigate climate change (Lal and Follet 2009; Stockmann et al. 2013).

### 7.1.5 Change in soil organic carbon stocks with projected climate change

This research issue was primarily addressed in Chapter 6. The specific research questions raised are answered below.
Appendix 1: Background to digital soil mapping

What is the feasibility of using DSM–SFTS techniques to spatially quantify changes in SOC due to the influence of future climate change over NSW?

This project demonstrated that digital soil mapping techniques involving Cubist piecewise linear decision trees in combination with a space-for-time substitution (SFTS) process can be effective in mapping the potential change in SOC due to projected climate change over NSW until approximately 2070. Digital maps with 100-m resolution of SOC change over two depth intervals: 0-30 and 30-100 cm, were produced for each of the 12 climate models and their mean. To date, almost all studies examining the influence of projected global climate change on soils have used simulation techniques such as RothC (Yurova et al. 2010; Gottschalk et al. 2012). DSM approaches appear to have not been previously adopted for this purpose, apart from the trial study of Minasny et al. (2013) who first suggested the approach.

Numerous sources of uncertainty in the process and final products were identified. Whilst some of these weaknesses are inherent to the DSM process such as inadequate modelling relationships, problems with covariate grid data and non-consideration of future changes in land use/management (Nelson et al. 2011; Bishop et al. 2015; Robinson et al. 2015); other sources of uncertainty are also common to the more widely used simulation techniques, such as differences between the various climate projections applied (discussed below); uncertain feedback processes in soil–atmosphere carbon dynamics that may also impact on climate change (Friedlingstein et al. 2006; Smith et al. 2008; Gottschalk et al. 2012), and the length of time required for soils to re-equilibrate with altered conditions (Smith 2008; Baldock et al. 2012).

Despite these uncertainties, the predictive maps provide a useful first approximation of likely changes in SOC as a result of climate change across NSW over the coming decades. They demonstrate the important role that DSM-SFTS techniques can play in helping to predict, understand and prepare for the impacts of global climate change. Further refinement of the approach, including greater quantification of levels of uncertainty, is recommended.

How consistent are the predictions of SOC change between the different global and regional climate models?

Considerable variation in both direction and magnitude of SOC change across NSW was demonstrated with application of the twelve different climate models with
their differing climate projections. There were some climate models that predicted an increase in SOC but others that predicted a decrease, with mean state-wide predictions for the 0-30 cm interval ranging from a 2.9 Mg ha\textsuperscript{-1} gain to an 8.7 Mg ha\textsuperscript{-1} loss. The 95% confidence intervals of change based on the 12 models intersected the zero change line, highlighting the uncertainty in the predictions. The application of other IPCC emission scenarios apart from the intermediate A2 scenario selected in this study, would introduce further variations in the predictions.

A similar wide range in future SOC stock projections following the use of different global climate models and emission scenarios was demonstrated in the simulation studies of Ito (2005), Lucht \textit{et al.} (2006), Yurova \textit{et al.} (2010) and Gottschalk \textit{et al.} (2012). It is evident that increased consistency between global climate-change projection trends is required to achieve more reliable predictions of soil property change into the future.

\textit{What factors drive the predicted changes in SOC with climate change? Do the changes vary systematically according to environmental conditions, based on current climate–parent material (soil type)–land use regimes?}

The overall responses of SOC demonstrated by this project reflect well established soil property–climate trends. Drier and warmer conditions result in lower SOC levels (Jenny 1980; Lal 2004; Badgery \textit{et al.} 2013; Viscarra Rossel \textit{et al.} 2014). Thus, the predicted changes are primarily controlled by the balance between changing temperatures and rainfall. However, the precise extent of change was also shown to be dependent on the particular environmental regime, with differing changes demonstrated over 36 current climate-parent material-land use combinations. Greater declines were demonstrated over wetter, less siliceous (more mafic) and less disturbed land use regimes: for example, the projected mean decline of SOC is less than 1 Mg ha\textsuperscript{-1} for dry-highly siliceous-cropping regimes but over 15 Mg ha\textsuperscript{-1} for wet-mafic-native vegetation regimes.

These results contribute to a better understanding of the potential for change in SOC with climate change and may enable more effective preparation for and adaption to the changes by land managers. The ability to spatially identify where the greatest changes in SOC may occur may particularly assist in monitoring changes in soil condition and for planning carbon sequestration programs. The results help to address
the challenge raised by McBratney et al. (2014) for improved modelling of SOC dynamics in space and time.

7.2 Broad research issues

Consideration of the above answers to the specific research questions, and to the overall results of the project, contribute to answers for the broad research questions as raised in the Introduction.

7.2.1 Can we better elucidate the influence of different factors in controlling the distribution of key soil properties?

This project used pragmatic soil – environment models to facilitate interpretation of the relative influence of different soil-forming factors. These models used readily interpreted variables to represent each factor, which better enabled the particular influence of each to be identified. Analysis was carried out over NSW and the entire eastern Australia, with reference to the key soil properties of SOC, pH, CEC, sum-of-bases, total P and clay, silt and sand, as covered in Chapters 2, 3 and 4.

The strong influence of the parent material factor became more evident following its classification into 12 lithology classes, based on mineral and chemical composition. When parent material was represented by other geophysical covariates such as gamma radiometrics or spectral derived clay composition, its relative influence in the models appeared much less significant (Chapter 2).

At this state and sub-continental scale, climate and parent material (lithology) were demonstrated to have the greatest influence for all these soil properties; ground cover and land use had intermediate influence and topography had the least influence. For SOC and pH, at the scale of eastern Australia, climate had the dominant or co-dominant influence, but at the NSW scale parent material had the dominant influence (Chapters 2 and 3). These levels of influence demonstrated at the sub-continental scale are similar to those reported by Hengl et al. (2014) in their global DSM project.

The results demonstrated that the relative level of influence is dependent on the scale of study, as particularly indicated by the results for SOC and pH. The influence of scale has been discussed by Miller and Schaeetzl (2016) who assert that debates over the most important soil forming factor are often moot, because the optimal predictor of soil
Appendix 1: Background to digital soil mapping

Spatial variability is usually a function of analysis scale: small cartographic scale maps emphasise bioclimatic relationships; medium cartographic scale maps emphasise parent material relationships, and large cartographic scale maps emphasise topographic and hydrologic relationships. Grunwald (2005) similarly noted that scaling behaviour of soil and environmental factors confounds quantitative relationships between factors. It has been suggested that digital soil modellers should use more multi-scale variables (mixture of fine and broad scale) to integrate scale phenomena (Miller et al. 2015; Miller and Schaetzl 2016).

The relative influence of each soil forming factor changes with depth. For SOC it was demonstrated that the relative influence of temperature and land use/vegetation cover decreases with depth, while that of parent material increases, findings also supported by Jobbágy and Jackson (2000) and Albaladejo et al. (2013) (Chapter 4). The need to better understand the spatial variation of SOC with depth was a research challenge raised by McBratney et al. (2014).

The use of partial regression coefficients from the MLR models allowed quantitative estimates to be derived for the degree of change per unit variation in the variable (Chapters 2, 3 and 4). For example, at the eastern state scale, it was revealed that over the 0-10 cm depth interval, for each 100 mm increase in annual precipitation there is a 0.062 loge % (6.4% proportional) increase in SOC% and a 0.11 unit decrease in pHca, assuming other factors remain constant. Thus, quantification of the influence of each variable was shown to be feasible by using this pragmatic approach of digital soil modelling. These estimates in degree of change were shown to be similar for SOC at both the NSW and eastern Australian scales. Further research should investigate the similarity of these estimates at different scales within Australia for other key soil properties, and also their similarity to estimates from other continents and at the global (universal) level.

There is a frequent call in the DSM literature for the discovery of new covariates to better represent soil-forming factors. (Hengl et al. 2014; Brevik et al. 2016b; Minasny and McBratney 2016). Although there is a preference for fine resolution continuous data from new geophysical sources, this project has demonstrated the value of using existing data sources in new, more effective ways, such as properly organised lithology class data. Other new pragmatic covariates also trialled in this project that were found to be effective included the topo-slope index (TSI, combining slope position
and gradient), an aspect index (Asp, combining compass orientation with slope gradient) and a land disturbance index (LDI, based on land use). Further investigation into the merit of these covariates for DSM programs is desirable.

7.2.2 How do these factors combine to control the distribution of soil properties?

The digital soil models developed in this project using MLR, Cubist and Random Forest approaches all demonstrated that the predicted magnitude for all soil properties is a result of the combined influence of all soil-forming factors. Variations in any of these variables would result in a different final quantitative estimate. This has been a fundamental underpinning of soil formation and soil geography thinking since the early days of pedology (Dokuchaev 1899; Hilgard 1906). However, results from this project have applied quantitative values to these combined factors in a pragmatic manner open to ready interpretation not previously widely undertaken (Chapter 3). They would appear to go a long way toward solving the fundamental soil equation, \( s = f(\text{clorpt}) \) (Jenny 1941) for these properties over eastern Australia. It was only recently that Heuvelink (2005) lamented that this fundamental equation had still not been solved, at least in a universal sense. The extent to which these relationships are applicable to other regions of the world with similar conditions to eastern Australia, or even universally, is an important question deserving further research.

The necessity of considering a combination of factors was also well illustrated by the results relating to the distribution of SOC stocks across eastern Australia (Chapter 4). Clear patterns of variation in stocks were demonstrated according to the precise combination of climate, parent material (or soil type) and ground cover, with stocks systematically increasing with increasingly moist climate, increasing mafic character of parent material and increasing ground cover. It was demonstrated that consideration of only two of these three factors would lead to a substantially high variation and uncertainty in the stock estimates. These results contribute to addressing the research challenges raised by Stockmann et al. (2013) and McBratney et al. (2014), particularly with respect to resolving soil SOC saturation issues – the limiting capacity of a soil to accumulate carbon. Future research should also investigate patterns of SOC fractions (particulate organic carbon, humus and resistant organic carbon) with these environmental combinations. Preliminary research undertaken during this project
suggests there are indeed strong relationships between these different SOC fractions and
the key soil-forming factors.

The results support the need for a holistic approach to understanding
environmental - soil landscape relationships as called for by Grunwald (2005).
However, that worker’s claim that there is no universal equation exists that fits all soil
landsapes, and that all models must be customised for specific domains, remains an
unresolved issue. Results from this study suggest there are indeed useful models that
apply over a sub-continental scales, and they may similarly be prepared at the global
scale, as demonstrated by Gray et al. (2009). Nevertheless, modelling approaches that
divide the landscape into domains, as carried out using the Cubist decision tree system
(Quinlan 1992) for much of this research project, do generally yield the most reliable
predictions. The results here offer more optimism on the potential for meaningful
environmental-soil landscape modelling than suggested by Phillips (2001) who argued
that intrinsic variability within homogeneous landscape units is more important in
determining pedo-diversity than is the extrinsic variability associated with measured
differences in topography, parent material and vegetation/land use. He claims we
currently cannot predict individual soil attributes accurately – only broad scale general
behaviour, but the success of the digital soil modelling in this project is reason to be
more positive about our prospects in this regard.

7.2.3 How does SOC respond to changes in the environment such as altered land
use and global climate change? Can readily interpreted relationships and patterns
in change be identified?

The pragmatic models controlling soil distribution presented in Chapter 3 provide
a useful and readily applied means to estimate changes in SOC over eastern Australia
due to variations in any of the applied variables. They could thus be applied to gain first
approximations of changes due to land use (varying the land disturbance index) or
climate change (varying the rainfall and maximum temperature values). However, this
project also applied more sophisticated DSM techniques such as Cubist piecewise linear
decision trees in innovative ways to address these research issues.

The change in SOC stocks following native vegetation clearing over NSW
(Chapter 5) was found to be highly dependent on the precise environmental regime,
being the combination of climate, parent material and final land use. The patterns
reflect the mean stock levels of each regime (demonstrated in Chapter 4) thus, the higher the initial SOC storage level, the greater is the loss in SOC upon vegetation clearance for equivalent land use changes. It is noteworthy that similar systematic patterns of SOC loss are evident in both absolute and relative terms, as also suggested by Baldock et al. (2009) and Powers et al. (2011). Perhaps the more fertile and productive soils are used relatively more intensively for the same land use, for example grazing with higher stocking rates or cropping at higher frequencies. These soils have the inherent potential to lose a greater proportion of their SOC and still be productive, thus are often worked harder until a higher proportion of their initial SOC is lost.

The potential change in SOC over NSW due to projected climate change was examined in Chapter 6. It was demonstrated that changes in this soil property are primarily controlled by the interactions of changing temperatures and rainfall, but the precise extent of change is also dependent on the particular environmental regime, with systematically differing quantitative changes demonstrated over 36 current climate-parent material-land use combinations. It is apparent that the impacts of a specified climate change on SOC stocks over a particular region are not uniform throughout that region.

An important observation from this study was the considerable variation in both direction and magnitude of change SOC across NSW demonstrated with application of the 12 different climate change models with their differing climate trajectories. For the mean state-wide change there were some climate models that predicted an increase in SOC magnitude but others that predicted a decrease. It is clear that reliable predictions of soil property change into the future will require increased consistency between global climate-change projections.

These findings represent important contribution to our understanding of SOC dynamics through space and time, a key research challenge raised by Stockmann et al. (2013) and McBratney et al. (2014). They allow us to identify locations and broader zones, being combinations of climate-parent material (soil type) and land use, with greater or lesser susceptibility to change under altering land use or climate conditions. They inform on the potential of different areas for enhanced carbon sequestration, a potentially important strategy for combating climate change (Lal and Follett 2009; Baldock et al. 2012; Smith 2012; IPCC 2014).
Carbon sequestration requires a stipulated duration timeframe (usually 100 years) in order to be considered a ‘permanent’ increase under managed agricultural systems (Smith et al. 2007, Stockmann et al. 2013). Useful research into the future may investigate linkages between the different environmental combinations utilised in this project, together with different depth intervals, and typical carbon residency periods. The significantly higher SOC storage potential of the moist-mafic-high ground cover regimes may possibly parallel longer residency times in the soil, but this hypothesis would need testing.

### 7.2.4 How effective are pragmatic soil relationships and data products, as derived in the study, in disseminating soil knowledge?

This project has developed and presented relationships governing soil distribution and soil change in forms that can be readily interpreted and applied by a wide range of end users, including policy makers, land use planners and land managers as well as other scientists. Many of the relationships presented are based on readily interpreted landscape and climatic features, rather than more complex and abstract variables such as geophysical data. For example, relationships and patterns of distribution and change were developed for SOC stocks in terms of climate – parent material (lithology/soil type) and land use/ground cover (Chapters 4, 5 and 6). These have the potential to provide clear spatial guidance on zones of differing potential SOC storage and change, which may be particularly important for possible future carbon sequestration programs, following the carbon matrix concept of Murphy et al. (2010).

The pragmatic multiple linear regression relationships derived in this project facilitate useful first approximations of key soil properties at individual sites with field collected and climate data alone, using as little as a hand held calculator. Access to sophisticated geophysical and remote sensed covariates, and computer technologies is not required, as for most other reported relationships. This option may be utilised by a range of soil and other environmental field scientists seeking to obtain estimates of soil properties, particularly where no suitable existing soil maps are available.

Components of this project (Chapters 4, 5 and 6) comprised the preparation of digital soil maps using sophisticated modelling tools and covariates, including complex geophysical data, but then synthesis of final results and key patterns in terms of readily understood variables such as described above. This may be a useful general strategy for
the more effective dissemination of results from digital soil modelling and mapping programs.

This project has attempted to meet ongoing demands for more accessible presentation of soil relationships through space and time. It should facilitate the call of Minasny and McBratney (2016) for dissemination of research outputs for practical use, not just for journal papers; and of Brevik et al. (2016) for optimising presentation of information for policy makers, land use planners and land owners, not just other scientists. It has addressed the concern of Bouma (2014) that soil information has to be more easily comprehended to raise awareness about the importance of soil for sustainable land use and other environmental problems.

7.3 Future research directions

A number of areas of further research continuing on from this project are identified. These broadly relate to the further exploration and elucidation of relationships between key soil properties and soil-forming factors. This includes further refinement of the pragmatic relationships and quantification of the precise influence of each factor.

Improved techniques of validation and estimation of uncertainty should be adopted. Throughout this project, validation sets were typically acquired by random data splitting of the original available legacy data, typically 15-20%, although stratification by State was performed for the eastern State datasets. This approach could be improved by stratification down to finer geographic levels to ensure all major elements of covariate space had been covered, for example using a Latin Hypercube approach (Minasny and McBratney 2006). The application of more advanced cross validation techniques such as those employed by Malone et al. (2014) and Kidd et al. (2015) or (if funding permitted) new design based probability sampling of validation points (de Gruitjer et al. 2006) could also strengthen validation results. The inclusion of additional validation criteria such as those presented by Malone et al. (2011) that measure the accuracy of both the predictions and their uncertainties would also be valuable.

More specific questions that could be addressed in future research are:
Appendix 1: Background to digital soil mapping

- How universal are the pragmatic relationships developed? How effective are they in other regions of the world? Are they only representative of other ancient continents such as Africa and South America?
- Are the 12 lithology classes identified here sufficient to adequately describe all parent material types? Should the sesquioxide class be divided into several separate classes to improve modelling in highly weathered ancient landscapes such as central and Western Australia?
- What is the relationship of different lithology classes to other geophysical data sources such as gamma radiometrics and hyper spectral imagery? Could we apply such geophysical data – lithology class relationships to improve DSMM products?
- Can we further elucidate relationships controlling the distribution of soil organic carbon fractions, including particulate organic carbon (POC), humus (HOC) and resistant organic carbon (ROC) using readily identifiable variables (for example, climate, lithology class and land use)? How do the different carbon fractions respond to changes such as change in land use or global climate?
- Can we establish linkages between typical carbon residency periods and combinations of readily identifiable variables (as above), together with different depth intervals?
- Can we effectively model the distribution of inorganic (mineral) soil carbon, such as carbonates, through space and time?
- Can we analyse historical aerial photos coverages to gain information on dates of vegetation clearance across different regions, and thereby derive estimates of annual rates of change in SOC following vegetation clearance?
- Can we expand our knowledge on key relationships and quantitative estimates of driving factors with other important soil properties? These could include: electrical conductivity (EC, for salinity issues), fine/coarse sand (for erosion hazard issues), sodicity (for agricultural productivity issues), and available water holding capacity (for agricultural and hydrological modelling issues).
- Are the pragmatic covariates introduced here of topo-slope index (TSI), aspect (Asp) and land disturbance index (LDI) worthy of further consideration as topographic and land use indices for wider use in DSMM programs? Could they be further improved to make them entirely continuous rather than ordinal categorical (1 to 6) as applied here? TSI combines landform position with slope...
gradient and proved moderately effective, typically displaying a higher relative influence in the models than other widely used topographic indicators such as topographic wetness index (TWI) or slope gradient. Similarly, the Asp index combined compass orientation with slope gradient and also proved moderately effective.

- Can the climate modelling community provide more consistency between climate change models? Can we reach greater certainty regarding projected climate change over NSW and Australia?
- Can we enhance the digital soil mapping – space-for-time substitution (DSM-SFTS) approach for predicting soil behaviour under climate change as trialled in this study, so as to improve its performance and overcome some of its inherent weaknesses?

7.4 Conclusion

This research project has used digital soil modelling and mapping techniques over eastern Australia under past, present and future conditions to further elucidate important relationships between a number of key soil properties and the main soil-forming factors. It has developed pragmatic relationships and revealed patterns of change through space and time that can be readily applied and interpreted by a large range of end users. The relationships would seem to go further toward quantitatively solving the fundamental soil equation of Jenny and others in a universal and readily interpreted manner than has been previously reported.

The necessity of considering the combination of factors when quantitative estimates are being derived for soil property distribution and behaviour under changing conditions has been demonstrated. This was shown to be essential for the estimation of current and potential future SOC stocks, where clear quantitative patterns in SOC distribution and change according to climate–parent material (soil type)–land use combinations were revealed.

Novel applications of DSMM were demonstrated in areas where it has not been widely used to date, including the mapping of SOC under pre-clearing (pre-European) and projected future climate change conditions across NSW. Detailed examination of
the associated changes revealed several interesting and as yet unreported patterns of change.

Further research into elucidating soil-environment relationships is suggested, with priorities that include establishing the applicability of the derived pragmatic relationships beyond eastern Australia; the precise relationships between the 12 lithology classes and emerging geophysical techniques such as hyper-spectral imagery; relationships of soil carbon fractions and residency periods with key soil-forming factors, and refining the DSM-SFTS approach into predicting soil property change under climate change.

The presented and ongoing research should assist in improving our understanding and knowledge of the factors driving the distribution of key soil properties through space and time. Ultimately this knowledge may allow us to better manage and protect our vital soil resources and to better adapt to a wide range of environmental challenges that face humankind in the future.

### 7.5 References


Appendix 1: Background to digital soil mapping


Appendix 1: Background to digital soil mapping

Dokuchaev VV, 1899. On the Theory of Natural Zones. St. Petersburg, Russia
Gruba P, Socha J, 2016. Effect of parent material on soil acidity and carbon content in soils under silver fir (Abies alba Mill.) stands in Poland. Catena 140, 90–95


Appendix 1: Background to digital soil mapping


Appendix 1: Background to digital soil mapping


Appendix 1: Background to development of soil modelling and digital soil mapping

A1.1 The State Factor model

At the base of all soil mapping, be it conventional or digital soil mapping, are conceptual or quantitative models that describe the relationship of soil type and soil properties to various soil-forming factors.

The State Factor model of soil formation has been, and remains, the most widely used model to explain the spatial distribution of soils. The Russian agronomist Vasilli Dokuchaev is generally credited as the first person to put forward this “factorial” model when, in the late 1800s, he presented a relationship showing soil to be the product of the combined influence of climate, organisms, parent material (“subsoil”), age and relief (Dokuchaev 1899; Florinsky 2011). The model was further advanced by several other workers in the early 1900s workers including Hilgard (1906), who developed similar ideas independently, and then Glinka (1914), Kellogg (1934) and Joffe (1936) and others, as recounted by Tandarich and Sprecher (1994), Paton and Humphreys (2007a and b) and Bockheim et al. (2005). It was eventually popularised by Jenny (1941) in his influential “Factors of Soil Formation”. This work attempted to explicitly state and further quantify the model with the now familiar fundamental soil equation: \( s = f(cl, o, r, p, t,...) \).

The widespread adoption of the model was attributable to its ease of comprehension and because information on each factor was generally readily gained and measurable (Schaetzl and Anderson 2005). It has been described as “a blueprint from which we may begin to unravel the mysteries of our surroundings” (Amundsen 1998, as quoted in Heuvelink and Webster 2001). It fostered the development of quantitative soil - environment relationships, which until the later part of the century were generally relatively simple mono- or bi-variate pedo-functions (Jenny 1961), such as “climo-functions” (Jenny 1980; Webb et al. 1986) or “topo-functions (Furley 1971).

More recently, the model was further advanced and adapted for digital soil mapping applications with the “scorpan” model of McBratney et al. (2003). This has additional factors of \( s \) (a soil attribute predictor) and \( n \) (a geographic position predictor) as well as the original five factors of the “clorpt” model.
A1.2 Development of digital soil mapping

The acquisition of soil data around the globe has, to date, predominantly been by conventional methods. This has involved collection of data in the field, together with aerial photography and other map sources; the application of broad conceptual models of soil distribution; and the production of maps showing different soil units with distinct boundaries, i.e., choropleth maps. Conventional soil maps have been described as “involving scientific methods and an element of art” (Wilding 1985) and being “representations of soil surveyor knowledge” about soil distribution across the landscape (McBratney et al. 2003).

The limitations of the conventional approach are increasingly being recognised, these mostly relating to it being too qualitative, with the resulting products being inadequate to meet the needs of many potential users of soil data.

To overcome these limitations, calls for more quantitative approaches became stronger in the early 1990s (McBratney 1992). These approaches belonged to the emerging field of pedometrics – the mathematical and statistical study of pedology (Webster 1994). The mapping approach typically involves the development of quantitative models representing the relationships between environmental variables and soil properties or classes, which are then applied to readily available quantitative environmental data to produce predictive soil maps. The approach gained widespread recognition and acceptance following the publication of the landmark paper “On digital soil mapping” (McBratney et al. 2003).

The new digital soil maps (DSMs) overcome many of the inherent limitations of the conventional approach including:

- being more objective, with all rules and assumptions made explicit, rather than being implicit
- involving quantitative data and relationships with measured accuracy
- avoiding the concept of soil as a spatial entity with distinct unit boundaries and being better able to deal with the continuously varying nature of soils
- being easier to compare, avoiding problems of different legend structures and soil classification schemes
- allowing results to be more easily incorporated into other modelling processes
• having the potential to be produced in less time and at lower cost, as they typically have lower human resource requirements
• having the potential to be more easily updated as new data or more reliable relationships become available.

The rise in application of DSM techniques has been aided by enormous advances in information technology, including powerful computers, mathematical and statistical modelling techniques, geographic information system (GIS) technology and environmental databases. Sophisticated data sources are available from geophysical techniques (e.g., radiometrics), remote sensing (e.g., Landsat imagery) and digital elevation models (DEMs), as described by McBratney et al. (2003) and Mulder et al. (2011) and in the following section. Also driving the rise in DSM has been the concurrent increase in demand for digital soil information from a range of other agricultural, climate, ecological, hydrological and other modelling applications.

Digital soil mapping moved from a primarily research phase in the 1990s to actual map production phase by early 2000s. Digital maps have now been produced all around the world at a range of spatial coverages. For example, in Australia it used for field scale studies in precision agriculture (Triantafilis et al. 2009), sub-catchments (Malone et al. 2009), regional catchments (Payne and Pringle 2012; Holmes et al. 2014; Kidd et al. 2014; Liddicoat et al. 2014), sub-continental scale (Henderson et al. 2005, Bui et al. 2009) to full continental coverage in the 90 m grids of the Soil and Landscape Grid of Australia (Grundy et al. 2015; Viscarra Rossel et al. 2015). The maturity of DSM is demonstrated by progress towards global high resolution coverage in GlobalSoilMap.net (Sanchez et al. 2009), of which the above mentioned Australian grid is a component.

Digital soil map products continuing to increase in reliability and sophistication. The future of soil science lies in this direction, as emphasised by many prominent world soil scientists in the booklet “The Future of Soil Science” (Hartemink 2006). The generally exponential increase in journal papers and citations dealing with DSM issues over the last 25 years (Minasny and McBratney 2016) bear testament to the continuing rise of this science.
A1.3 Use of geophysical covariates to represent soil and parent material in DSM

The ever rising sophistication of geophysical techniques means these data sources are being increasingly applied as covariates to represent the parent material \((p)\) or soil \((s)\) factors in DSM programs. These include gamma radiometrics, multi- and hyperspectroscopy, electromagnetic induction and others, collected either as proximal (laboratory or field based) or remote sensed (airborne or satellite) forms. Reviews on the use of geophysical techniques in DSM programs are given in Mulder et al. (2011) and McBratney et al. (2003).

Gamma radiometrics provide an important indicator of lithology, soil composition and degree of weathering (Cook et al. 1996; Wilford and Minty 2007; Wilford 2012; Martelet et al. 2014). Remote sensed radiometric data is widely used in many DSM programs, being applied in 11% of a compilation of 267 DSM studies derived from the meta-studies of McBratney et al. (2003), Grunwald (2009), Minasny et al. (2012), Minasny et al. (2013) and Arrouays et al. (2014). Recent examples include programs by Rawlins et al. (2009), Karunaratne et al. (2014) and Lugumira et al. (2014). The use of proximally sensed radiometric data is also becoming more widespread. Its merits for mapping and understanding the distribution of soil materials in complex geological regions was recently demonstrated by Viscarra Rossel et al. (2014) and Stockmann et al. (2015).

Multi-spectral imagery from Landsat Thematic Mapper (TM) with current resolutions of 30m is widely used, being applied in 9% of the compiled 267 DSM studies. Recent examples include those by Ciampalini et al. (2012), Odgers et al. (2014) and Padarian et al. (2014). Particular colour intervals and band ratios such as 3/2, 3/7, 5/7 and \((5-2)/(5+2)\) (for Landsat 7 and earlier) are valuable in reflecting mineralogy rather than vegetation character (Boettinger et al. 2008). Other satellite sourced multi-spectral imagery such as ASTER with resolution to 15 m (Mulder et al. 2013; Aichi et al. 2014) and IKONOS with resolution to 1 m (Simbahan et al. 2006) are also being applied. The Sentinel missions of the European Space Agency are providing a new valuable source of multi-spectral imagery for soil mapping.

The use of hyper-spectral VNIR imagery, involving continuous rather than discrete spectral bands, for mapping soils has developed since the early 1980s and with
the expected increase in availability of remote sensed hyper-spectral imagery it is likely to increase in importance (Viscarra Rossel et al. 2011; Lagacherie and Gomez 2014). This data was applied in 7% of the 267 compiled DSM studies, but most are in the more recent years. The imagery has been used to map a range of soil properties including clay content, sand content, pH and CEC at the surface (Gomez et al. 2012; Lagacherie et al. 2012) and in the subsurface with DSM techniques (Lagacherie et al. 2013). VNIR spectral libraries provide valuable soil point datasets that have been modelled against other environmental covariates in DSM programs to successfully predict a wide range of soil properties (Viscarra Rossel 2011; Viscarra Rossel et al. 2011; Viscarra Rossel and Webster 2012).

Electromagnetic induction (EMI) and electrical conductivity techniques measure the soil’s apparent electrical conductivity and have been useful for estimating various soil properties, particularly salinity, clay content, soil texture class, soil moisture, available water holding capacity, cation exchange capacity and depth to bedrock or other layers (Zhu et al. 2010; Buchanan et al. 2012; Triantafilis et al. 2009, 2013). They are typically applied proximally at field scales but have been adapted for airborne regional remote sensing applications (Everett 2012). They were used in 6% of the combined DSM studies, but this rate would be higher if only field scale studies were considered.

Other geophysical data sources such as magnetometry (Ryan et al. 2000; Jordanova et al. 2008) and gravity anomalies (Viscarra Rossel et al. 2015) are also occasionally used.

A1.4 References


Appendix 1: Background to digital soil mapping


Appendix 1: Background to digital soil mapping


Appendix 1: Background to digital soil mapping


Appendix 1: Background to digital soil mapping


Appendix 1: Background to digital soil mapping


Appendix 2: Behaviour of pH and sum-of-bases under projected climate change over NSW

This appendix presents results on the predicted change in pH and sum-of-bases under projected climate change over NSW to approximately 2070. These results supplement those presented for soil organic carbon (SOC) in Chapter 6.

A2.1 The soil properties

The soil properties of pH and sum-of-bases (representing major macro-nutrients) are, like SOC, important indicators of a soil’s chemical and physical character and condition, and are vital for the soil’s agricultural productivity and the ecosystem health more broadly. The change in pH, together with SOC, is considered a priority for Australian soil monitoring programs (McKenzie and Dixon 2006). The laboratory test methods used and sample size in the final datasets for these two soil properties are given in Table A2.1.

Table A2.1 Soil properties: laboratory methods and sample numbers

<table>
<thead>
<tr>
<th>Soil property</th>
<th>Units</th>
<th>Laboratory method with test number from Rayment and Lyons (2011)</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH</td>
<td>pH units</td>
<td>pH of 1:5 soil/0.01M calcium chloride extract (4B1, 4B2). Includes conversions from pH 1:5 soil/water suspension (4A1)</td>
<td>7682</td>
</tr>
<tr>
<td>Sum-of-bases$^1$</td>
<td>cmol$_c$/kg</td>
<td>Various methods (15A – 15F)</td>
<td>4315</td>
</tr>
</tbody>
</table>

$^1$ exchangeable cations of calcium, magnesium, sodium and potassium

The variation in different laboratory methods for the same soil property, due to the different dates and jurisdictions of the analyses, results in a degree of inconsistency in the test results and potential error in the predictive models. To rationalise all pH test results into one consistent methodology, the original pH$_{water}$ values were converted into pH$_{CaCl_2}$ values using the correlation tables of Henderson and Bui (2002). The latter mode is preferred in Australia as it more closely represents the ionic soil solutions typically found in the field, and thus gives more reliable results. Changes are expressed in absolute terms for pH, but in relative terms for sum-of-bases, which is considered to be
more reflective of meaningful change in soil fertility, because of the large variation in original absolute values.

### A2.2 Results

Validation results for the initial Cubist models developed using the baseline (1961-1990) climate data are presented in Table A2.2. For both soil properties the strength of the models increases slightly with depth, with concordance values reaching highs of 0.79 and 0.74 at the lower depths, but the standardised RMSE values remain essentially constant.

<table>
<thead>
<tr>
<th>Soil property</th>
<th>Depth interval (cm)</th>
<th>N</th>
<th>CCC</th>
<th>$R^2$</th>
<th>RMSE</th>
<th>Std RMSE</th>
<th>ME</th>
</tr>
</thead>
<tbody>
<tr>
<td>pH</td>
<td>0-5</td>
<td>1582</td>
<td>0.72</td>
<td>0.57</td>
<td>0.83</td>
<td>0.15</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>5-15</td>
<td>1550</td>
<td>0.75</td>
<td>0.60</td>
<td>0.81</td>
<td>0.14</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>15-30</td>
<td>1405</td>
<td>0.76</td>
<td>0.61</td>
<td>0.85</td>
<td>0.15</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>30-60</td>
<td>1317</td>
<td>0.79</td>
<td>0.64</td>
<td>0.89</td>
<td>0.15</td>
<td>-0.0026</td>
</tr>
<tr>
<td></td>
<td>60-100</td>
<td>1067</td>
<td>0.78</td>
<td>0.63</td>
<td>0.96</td>
<td>0.15</td>
<td>-0.0034</td>
</tr>
<tr>
<td>Sum bases (log cmolc/kg)</td>
<td>0-5</td>
<td>1256</td>
<td>0.65</td>
<td>0.48</td>
<td>0.81</td>
<td>0.14</td>
<td>-0.028</td>
</tr>
<tr>
<td></td>
<td>5-15</td>
<td>1238</td>
<td>0.67</td>
<td>0.51</td>
<td>0.79</td>
<td>0.14</td>
<td>-0.022</td>
</tr>
<tr>
<td></td>
<td>15-30</td>
<td>1169</td>
<td>0.72</td>
<td>0.56</td>
<td>0.80</td>
<td>0.16</td>
<td>-0.00048</td>
</tr>
<tr>
<td></td>
<td>30-60</td>
<td>1094</td>
<td>0.72</td>
<td>0.55</td>
<td>0.86</td>
<td>0.15</td>
<td>-0.022</td>
</tr>
<tr>
<td></td>
<td>60-100</td>
<td>883</td>
<td>0.74</td>
<td>0.58</td>
<td>0.83</td>
<td>0.12</td>
<td>-0.027</td>
</tr>
</tbody>
</table>

CCC: Lin’s concordance correlation coefficient; $R^2$: Coefficient of determination; RMSE: Root mean square error; Std RMSE: Standardised RMSE (RMSE/mean of estimate); ME: Mean error (prediction – observed)

Results for the predicted change in the two soil properties, based on the above models, are presented over the two combined depth intervals: upper soil (0-30 cm), and lower soil (30-100 cm) in the following sections. Primary focus is given to the results for the second change period (1990-2009 to 2060-2079). More detailed results and full digital maps (100-m pixels) will be made available for public download through the Adapt NSW website (OEH 2014). For each soil property, change is reported on a state-wide basis and by physical zones including climate–parent material (soil type)–land use regime.
Appendix 2: pH and sum-of-bases under projected climate change in NSW

A2.2.1 pH

State-wide change

The absolute change in pH across NSW for both depth intervals and change periods as derived using each of the 12 climate models is presented in Figure A2.1. The results suggest that from the average of all 12 climate models for the State as a whole, there is no significant change in pH levels, except for a modest decline of 0.15 pH units in the lower depth interval, 2nd change period.

Predictions vary substantially, however, with the different climate models. The 95% confidence interval are almost centred over the zero change point in all except the fourth column of Figure A2.1. The MIROC and CCC models, the wetter models, almost all suggest a decrease over both depth intervals and change periods, with 0.34 pH units decline being predicted by the CCC2 model for the lower depth interval, 2nd change period. By contrast the CSIRO models, the driest models, suggest notable increases up to 0.17 pH units for the CSIRO1 model for the lower depth interval, 2nd change period. The ECHAM models generally reveal a slight increase in pH, apart from a slight decrease for the lower depth interval, 2nd change period.

Figure A2.1. Absolute change in pH from the 12 climate models for both change periods

Notes: yellow for change period 1 (1990-2009 to 2020-2039); orange for change period 2 (1990-2009 to 2060-2079); thin black lines: the mean change across NSW for each climate model; thick red lines: mean from all 12 climate models
In addition to uncertainty arising from the different climate models, there is uncertainty from the digital modelling and mapping process used to derive the change estimates. The RMSEs of the initial models for the original five depth intervals are relatively high, ranging up to 0.96 pH units in the deepest layer (Table A2.2). Additional uncertainty parameters associated with final map generation were not quantified, but are also likely to be significant. The predicted state-wide changes in pH all appear to be within the envelope of uncertainty.

Based on the average of all 12 climate models, and recognising the substantial uncertainties, most of the eastern and central-eastern regions of the State are projected to have a slight increase in pH (generally less than 0.2 pH units) for both depth intervals, as shown for the upper (0-30 cm) interval by the map of Figure A2.2. In the central-western regions there is typically a shift to slight decreases in pH (generally less than 0.2 pH units), particularly evident in the lower depth interval. For most of the western regions a notable decline in pH is projected for both depth intervals, ranging up to more than 0.8 pH units, however the extent of this decrease appears anomalously high, and warrants further investigation.

Figure A2.2. Change in pH across NSW for 2nd change period (0-30 cm)
**Appendix 2: pH and sum-of-bases under projected climate change in NSW**

*Change by environmental regime*

The projected pH changes are primarily controlled by the balance between the changing temperatures and rainfall, normally increasing with rising temperatures and declining rainfall (Gray *et al.* 2015; Kopittke *et al.* 2012; Rubinic *et al.* 2015). However the extent of the pH change also varies depending on the environmental and land use regime, which adds complexity to the above trends.

A breakdown in the pH change results for the 0-30 cm depth intervals over the 2nd change period by climate–parent material–land use sub-classes is presented in Figure A2.3. It demonstrates variation in the extent of change over different environmental regimes, for example, the change varies from a mean rise of 0.05 pH units in moist–mafic–native vegetation regimes to a mean decline of over 0.3 pH units for dry–highly siliceous–native vegetation regimes. Many regimes do not demonstrate a significant change at the 95% confidence level based on the 12 climate models; which was also observed for the 30-100 cm depth interval (plot not presented here). Nevertheless, some broad trends are apparent, including greater declines in pH with increasingly drier current conditions, more siliceous parent materials and more intensive land uses.

pH buffering capacity is a factor that influences the extent of pH change at any site, and may vary spatially across the state, generally increasing with higher clay (less siliceous) soils (Helyar *et al.* 1990; Nelson and Su 2010). Whilst this factor was not directly included as an input variable in the modelling process, it should be at least partially represented by the range of environmental covariates that were applied.

*Summary*

From the average of the twelve models, and recognising the sources of uncertainty, only a minor change in pH over most of the State is predicted. Most eastern and central regions are predicted to undergo a slight increase in pH (ie, become more alkaline), while the western regions are predicted to undergo a notable decrease in pH, ie, become more acidic, however, some of the larger decreases appear anomalously high, particularly in the lower depth interval. Greater declines or lower increases in pH
Appendix 2: *pH and sum-of-bases under projected climate change in NSW*

...are evident over currently drier climates and more siliceous parent materials, but no clear trends with land use are observed.

![Figure A2.3. 95% confidence interval and mean change in pHca by physical zone from the 12 NARClM models (pH units, 0-30 cm, 2nd change period)](image)

### A2.2.2 Sum-of-bases

**State-wide change**

The relative change in sum-of-bases, representing key macro-nutrients, across NSW for both depth intervals and change periods as derived using each of the 12 climate models is presented in Figure A2.4. The results suggest an overall increase in this soil property across the State from the average of the 12 models. The 95% confidence intervals are clearly above or only just coincide with the zero change mark. The increase is particularly evident over the 2nd change period where there is a mean increase of 6.5% in the upper depth interval and 8.1% in the lower depth interval. The changes are generally slightly more pronounced in the lower depth interval and the 2nd change period.
The predictions nevertheless still vary substantially between the different climate models, in terms of both the direction and extent of change. The drier CSIRO and ECHAM models reveal the greatest levels of increase in sum-of-bases with CSIRO1 displaying a 16.5% increase over the lower depth interval, 2nd change period, while the wetter MIROC and CCC models generally display the greatest declines, with MIROC3 displaying a 7.7% decrease over the upper depth interval, 2nd change period. There is an apparent anomaly however, with the MIROC2 model displaying a notable increase in sum-of-bases for the last mentioned depth interval, change period.

![Figure A2.4. Relative change in sum-of-bases from the 12 climate models for both change periods](image)

In addition to uncertainty arising from the different climate models, there is uncertainty from the digital modelling and mapping process used to derive the change estimates. The RMSE of the initial models for the original five depth intervals is relatively high, ranging up to 0.86 (cmol./kg, log scale) (Table A2.2). Additional uncertainty parameters associated with final map generation were not quantified, but are also likely to be significant. The predicted state-wide changes in sum-of-bases all appear to be within the envelope of uncertainty.
Appendix 2: pH and sum-of-bases under projected climate change in NSW

Based on the average of all 12 climate models, and recognising the substantial uncertainties, most of the State is projected to undergo a moderate increase (up to 30% or more) in sum-of-bases over both depth intervals over the 2nd change period, as shown for the upper (0-30 cm) interval by the map of Figure A2.5. Many northern regions and isolated central regions are however projected to undergo minor decline (generally less than 10%).

![Figure A2.5. Relative change in sum-of-bases across NSW for 2nd change period (0-30 cm)](image)

Change by environmental regime

The projected changes in sum-of-bases are primarily controlled by the balance between the changing temperatures and rainfall, normally increasing with rising temperatures and declining rainfall (Gray et al. 2015; Rubinic et al. 2015). However the extent of the change also varies depending on the environmental and land use regime. Figure A2.6 presents a breakdown in the relative sum-of-bases change results for the upper depth interval over the 2nd change period by climate–parent material–land use sub-class. The plot demonstrates variation in the extent of change over different environmental regimes, for example, a 22% mean increase over dry–lower siliceous–grazing regimes but a 3% mean decrease over wet–upper siliceous–native vegetation regimes.
Complex patterns of change with different environmental regimes are revealed. In the upper depth interval there is a trend towards larger increases (and broader confidence intervals) in progressively drier sub-classes, however for the lower depth interval (not presented here) the reverse trend was apparent. This may reflect the greater level of leaching of nutrients from the upper soils to lower soils in the wetter climate regimes such as the north and central coasts, a much weaker process in the drier regions such as in the west. A faster breakdown of parent material in the wetter climate, with release of bases into the subsoil, might also be a contributing factor. A curious trend is apparent with respect to parent material over both depth intervals. In the drier climate regimes there is a greater increase in sum-of-bases with more siliceous materials, but in the wet regimes there is generally less decrease with more siliceous materials.

Figure A2.6. 95% confidence interval and mean relative change in sum-of-bases by physical zone from the 12 NARClIIM models (% , 0-30 cm, 2nd change period)

Summary

From the average of the twelve models, and recognising the sources of uncertainty, a moderate and somewhat complex pattern of change is predicted in sum-of-bases over most of the State. Most regions are predicted to undergo a modest increase while many northern and some central areas are predicted to undergo a notable
Appendix 2: pH and sum-of-bases under projected climate change in NSW

decrease. Changes are most pronounced in the lower 30-100 cm depth interval. In the upper (0-30 cm) interval, the relative increase is highest in the currently drier climate, however, in the lower (30-100 cm) interval the increase is highest in the wetter climates, presumably reflecting increased levels of leaching.

### A2.3 Discussion

**Application of results**

Complete digital maps at 100-m pixel size and further detail on results for the two soil properties will be made available for public download through the *Adapt NSW* website (OEH 2014). The predicted changes in these soil properties have implications for the future health and character of NSW soils, and consequent effects on agriculture, natural ecosystems and climate change mitigation strategies.

Soil condition and agricultural productivity generally improve with major nutrient content (McKenzie *et al.* 2004) and, for much of NSW, with increasing pH (alkalinity), however this is dependent on the desired pH and nutrient ranges of different crop and pasture species (Russell and Russell 1988; Hazelton and Murphy 2007). Significant changes in pH and macro-nutrients may also indicate important changes in other minor and trace elements, with associated impacts on fertility and toxicity levels (Mulvey and Elliott 2007). Farmers may need to adjust application rates of fertilisers and conditioners, or possibly modify their selection of crop or pasture species to better match the altered soil conditions.

Thus, consideration of the changes in these soil properties provides potentially important guidance for the productive management of soils across NSW in future decades (Stokes and Howden 2010). Estimates of wheat and corn production under the effect of the climate changes in Mexico were shown to vary by 15-20% when changes to soil fertility were included (Nikolskii *et al.* 2010). Those authors lament the typical omission of this issue in most studies on the impacts of climate change on agricultural productivity.

Changes in soil properties, particularly pH, macro-nutrients and associated minor and trace elements, may impact on natural ecosystems, which often have narrow chemical tolerance ranges. Where significant increases or decreases are revealed there is likelihood for an introduction of environmental weeds and alterations in floral and
faunal compositions (Prober and Wiehl 2012; Steffen et al. 2009). Such changes may need to be considered and addressed by managers of these ecosystems, for example through increased weed management programs and by allowing for the gradual migration of particular ecological communities to areas with the required soil conditions.

Anomalies in results

The prediction of change in pH and sum-of-bases is subject to the same uncertainties as those discussed in Chapter 6 for SOC, including the high variability of predictions arising between the 12 climate models applied.

As noted in section 6.4.4 of Chapter 6, possible anomalies in the soil property predictions may be attributable to potential weaknesses in the initial Cubist piecewise linear decision tree models, which were developed using a 7000 point dataset over eastern Australia. For each of the five depths (later amalgamated into just two depths) there were typically 10 rules sets, covering different combinations of covariate space. In most cases there were one or two isolated rules with terminal linear regression equations that had coefficients for rainfall or temperature with signs opposite to that normally expected. For example, a rule may indicate an increase in pH with increasing rainfall, rather than the expected decrease. When this rule is applied into the future with an altered climate, the trends with respect to climate may be opposite to that expected for that combination of covariate space.

For this reason, the derivation of single multiple linear regression models over a study region (for each soil property and depth interval), where one has high confidence in the signs (positive or negative) of the climate coefficients, may in fact be more reliable at revealing the key trends in the change than other more sophisticated decision tree approaches such as Cubist. This message and potential warning may be of interest to others who are considering adoption of a DSM - space-for-time substitution (SFTS) approach, as applied here and in the initial trial study of Minasny et al. (2013).

A2.4 References

Appendix 2: pH and sum-of-bases under projected climate change in NSW


Appendix 2: pH and sum-of-bases under projected climate change in NSW


Appendix 3: Digital soil maps of bulk density for NSW

Digital soil map of bulk density over NSW for the 0-10, 10-30 and 0-30 cm intervals were prepared, as reported in Chapter 5. This allowed the determination of soil organic carbon mass and stocks (Mg/ha). At the time of analysis for Chapter 5, the bulk density layers from the Soil and Landscape Grid of Australia were not available, as applied in later analysis.

Data was derived from the NSW Monitoring Evaluation and Reporting dataset (refs). This contained bulk density data over 888 sites down to 30 cm. From this a validation subset of 168 points was randomly selected. Cubist decision tree models were prepared over the calibration data, which was then applied over the entire state to produce the digital soil maps at 100-m resolution (0-30 cm map presented in Figure A3.1. Validation of the map was carried out using the validation data, with results presented in Table A3.1 and Figure A3.2.

Figure A3.1. Bulk density layer for NSW (0-30 cm)
### Table A3.1. Validation of NSW bulk density maps

<table>
<thead>
<tr>
<th>Depth (cm)</th>
<th>N</th>
<th>Lin’s CCC</th>
<th>RMSE</th>
<th>Mean error</th>
<th>Median absolute error</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-10</td>
<td>149</td>
<td>0.41</td>
<td>0.20</td>
<td>0.012</td>
<td>0.12</td>
</tr>
<tr>
<td>10-30</td>
<td>143</td>
<td>0.45</td>
<td>0.20</td>
<td>-0.02</td>
<td>0.12</td>
</tr>
<tr>
<td>0-30</td>
<td>143</td>
<td>0.52</td>
<td>0.18</td>
<td>0.012</td>
<td>0.10</td>
</tr>
</tbody>
</table>

CCC: concordance correlation coefficient; RMSE: root mean square error; Mean error (predicted - observed value)

![Validation plot of NSW bulk density layer (0-30 cm)](image)

**Figure A3.2.** Validation plot of NSW bulk density layer (0-30 cm)