# Investigating soil change in Edgeroi, New South Wales using pedogenon mapping framework

By

**Ho Jun Jang**

A thesis submitted to fulfil requirements for the degree of

Doctor of Philosophy

2022

School of Life and Environmental Sciences

Faculty of Science

The University of Sydney

New South Wales, Australia

## Statement of originality

I hereby declare that the content presented in this thesis is solely the result of my own intellectual efforts and that I have not submitted it for any degree or other purposes before. All the sources and assistance received in the preparation of this thesis have been duly acknowledged.

Ho Jun Jang

Date: 25 December 2022

## Authorship attribution statement

**Chapter 1** of this thesis is published as:

JANG, H. J., DOBARCO, M. R., MINASNY, B. & MCBRATNEY, A. 2021. Creating a soil parent material map digitally using a combination of interpretation and statistical techniques. Soil Research.

I designed the study with my co-authors, performed the analysis, analysed the data, and wrote the manuscript draft.

**Chapter 2** of this thesis is published as:

JANG, H. J., DOBARCO, M. R., MINASNY, B., MCBRATNEY, A. & JONES, E. 2022. Developing and testing pedogenons in the lower Namoi valley, NSW, Australia. Geoderma, 116182.

I designed the study with my co-authors, performed the soil survey, executed the soil sample analysis, and analysed the data. I also wrote the manuscript drafts.

Student name: Ho Jun Jang

Date: 31 December 2022

In addition to the statement above, as a supervisor for this candidature, I can confirm that the authorship attribution statements above are correct.

Supervisor name: Budiman Minasny, Professor of Soil Science, The University of Sydney

Date: 31st December 2022

## Abstract

The impact of anthropogenic activities on soil has been significant in the last few hundred years, surpassing that of natural processes that occurred over thousands of years. Soil organic carbon (SOC) is especially vulnerable to anthropogenic forces and is critical to support soil functions such as nutrient and water cycling, crop production, and habitat for above- and below-ground biodiversity. It is therefore crucial to understand the current SOC status and predict how it has changed due to human influences. However, assessing changes in soil is challenging due to the various factors involved in soil formation. This thesis investigates the change in SOC in the Edgeroi area of New South Wales since European settlement using the pedogenon mapping concept. Pedogenon divides a landscape into unique soil entities based on homogeneous soil-forming factors, distinguishing soils that have been less affected by human activities (genosoil) and those that have been intensely affected (phenosoils). The primary hypothesis of this thesis is that pedogenon mapping can effectively stratify the landscape and be used to estimate soil change. To produce pedogenon classes, a parent material map of the study area was generated using interpretation and machine-learning techniques. Unsupervised classification of spatial layers representing soil-forming factors was then used to create the pedogenon map. Within each pedogenon, genosoil and phenosoil areas were identified using land use data. Multivariate data analysis confirmed that each pedogenon has unique soil properties from the surface down to 1 m. SOC data from genosoil and phenosoil areas were then compared and mapped to investigate soil change. The results showed that the variation of soil properties under phenosoil is half that of genosoil due to agricultural practices. Additionally, this approach enables the mapping of SOC sequestration potential using mineral-associated OC contents. This thesis improves the current digital soil mapping approach for assessing soil change in Australia.

## Acknowledgements

The past few years have been challenging for me to work in the office for long hours, but I am grateful to have had the support of so many wonderful people throughout my PhD journey. First and foremost, I want to express my sincere gratitude to my supervisor Budiman Minasny. His continuous advice, encouragement, and patience were instrumental in helping me overcome the challenges. His trust in me gave me the confidence I needed to succeed, and he has been a great emotional supporter throughout my studies.

I also want to extend my thanks to my co-supervisors, Alex McBratney and Mercedes Roman Dobarco. Their kind suggestions, patience, and support were invaluable, and Mercedes did an excellent job as a first-time PhD supervisor.

I am grateful to my co-authors, Edward Jones and Jose Padarian Campusano, for supporting my work, and I also want to thank Sumarni Sukiman for keeping me healthy with nutrient-rich foods.

When I started using R, I received invaluable assistance from Mario Fajardo Pedraza, Wartini Ng, and Patrick Filippi. I am also grateful to my friends in ATP, Katherine, Yuxin, Zongtang, Bright, Yijia, Tom, Sandra, Oscar, Iona, Liana, Niranjan, Hero, Vanessa, Marliana, Alexandre, Dhahi, and Andree for their amazing support.

Finally, I want to express my deepest appreciation to my parents and brother for their love, support, and understanding. I could not have made it without them. I also want to thank my best friends in the Korean temple (Jongbopsa), Gina, Hyera, Junghwan, Yeaji, Hyeongchan, Min, Julee, Jungwon, Yeojin, and Minji, for being a constant source of inspiration and encouragement.

# Table of contents

# Table of figures

# Table of tables

# Table of appendices

# Chapter 1. Introduction

## 1.1. Background

Soil plays a critical role in the earth's ecosystem, acting as a link between and providing feedback responses to other components such as water, atmosphere, and vegetation. It provides essential ecosystem services, including food and fibre production, carbon and nutrient cycling, climate regulation, and water delivery and purification (Brady et al., 2008, Robinson et al., 2012). Soil properties change over time due to various physical, chemical, and biological processes shaped by natural pedogenesis (Kuzyakov and Zamanian, 2019).

However, anthropogenic activities have increasingly compromised the sustainable use of soil resources, with the demand for food and clean water projected to double and increase by 50%, respectively (Banwart et al., 2013). The intensity of human activities over the last few hundred years has been significant, resulting in the proposal of a new geological epoch called the Anthropocene (Crutzen, 2016). The impact of anthropocene on soil conditions has gained much attention, and there is a discussion on its effect on the direction of soil change, which can significantly impact ecosystem services.

To ensure the sustainability of these soil functions, it is crucial to characterise the spatial and temporal variability in the chemical, biological, and physical processes that define the present state of the soil and evaluate how human activities impact the provision of these services. Quantifying soil changes is pivotal to calculating the extent to which anthropogenic factors affect the soil condition, predicting the time it takes to obtain full benefit from the soil and improving how we manage soil.

## 1.2. Soil change due to natural pedogenesis and anthropogenic activities

Pedology mainly studies soil change under natural pedogenesis. However, as human impact has intensified, the change of soil by anthropogenic activities, especially agricultural practices, has become a main topic in soil science research. Soil impacted by human activities is totally different from the soil formed by natural processes, as the processes occur much faster.

Richter and Yaalon (2012) revisited the essay "The changing model of soil" written by Cline (1961) and re-evaluated the concept from the perspective of the early 21$^{st}$ century. There is an ongoing change in the pedogenesis model, which tremendously impacts the future of soil science. In particular, Richter and Yaalon (2012) argue that soil has been transformed from a natural to a human-natural body. They proposed that studies on soil formation and change have been mostly on natural bodies up to the 19$^{th}$ century. However, since the 21$^{st}$ century, natural soil bodies have gradually disappeared due to the acceleration of anthropogenic activities. Today, human is the largest impacting factor on the change of soil across the diversity of earth's soils, and soil has been converted to the human-natural body. Previously, soil science focused on the soil formed under the natural system, but current soil science is changing due to the combination of soil and human activities (anthropedogenesis). Richter and Yaalon (2012) showed the example of soil organic carbon (SOC) change between natural pedogenesis and anthropedogenesis. There is a small variation in SOC carbon change under natural pedogenesis as vegetation and SOC pools have reached a steady-state equilibrium. However, under anthropedogenesis, SOC change varied up to 300 folds due to land use conversion. Therefore, there is a need to understand the interaction between soil and human activities to support sustaining soil and humanity.

Kuzyakov and Zamanian (2019) further looked at a special case of anthropedogenesis, i.e., agropedogenesis, soil formation under agricultural activities, especially on croplands. Kuzyakov and Zamanian (2019) mentioned that agropedogenesis and natural pedogenesis should be clearly separated because

1) Human activities have strong dominance of the factors among the other factors of soil formation.
2) Natural pedogenesis does not include new processes induced by humans.
3) The direction of soil development from agropedogenesis is different from natural pedogenesis.
4) The soil change from agropedogenesis usually shows an opposite trend of natural pedogenesis.
5) The changing intensity is very different between agropedogenesis and natural pedogenesis, and agropedogenesis shows a much higher intensity of the change in a short period.

Agropedogenesis results from the need to produce food for the world population. This process usually comes with a consequence of declining soil health and quality, reducing or even eliminating the effect of natural soil formation. The intensity and period of processes between the two pedogenesis are incredibly different. The process of soil formation from natural pedogenesis is over many centuries or millennia. According to Stockmann et al. (2014), the potential global average weathering rate is 0.1 mm per year and the soil residence time is about 47,000 years under steady-state conditions of soil formation. This is in contrast with the rate of soil erosion under cultivated cropland which is about 1 mm per year (Montgomery, 2007).

Agropedogenesis can happen almost yearly, depending on different agricultural practices (Richter Jr, 2007, Diamond, 2002). Natural pedogenesis leads to the divergence of soil

properties, while agropedogenesis leads to convergence or homogenisastion (Kuzyakov and Zamanian, 2019). To boost crop production, soil conditions need to be optimised (Richter et al., 2015, Richter Jr, 2007). Soil hydrological and physical properties are altered, including irrigation, drainage, terracing, removing stone and loosening soil by tillage. Chemical and biological conditions in the soil are also modified with fertilisation, desalinisation and liming, applying biocides and sowing domesticated plant species. Under natural conditions, diverse soils formed in different environmental conditions, but human-modified soil properties to maximise crop yield, resulting in homogenised soil. Homogenised soil is easier to manage. If the homogenised soil benefits humanity and the soil itself, then it is perfect for humanity and the earth. However, homogenised soil increases the risk of soil degradation. It is well known that soil degradation can lead to the collapse of a civilisation which can be witnessed in Mesopotamia (Diamond, 2002, Weiss et al., 1993).

Therefore, for producing food and getting other benefits from soil sustainably, it is critical to study how much human activities have been affecting the soil so far. With data on soil change due to human activities, it is possible to predict how long the soil can produce food sustainably without negative effect to the environment.

## 1.3. Mapping soil change using digital soil mapping techniques

In the past decade, digital soil mapping (DSM) has achieved outstanding results in predicting soil properties or classes at local and continental scales. McBratney et al. (2003) established the DSM SCORPAN model, which is written as S (soil classes or properties) = f (s, c, o, r, p, a, n) where Soil (s), Climate (c), Organisms (o), Topography (r), Parent material (p), Age (a) and geographical position (n) factors. It extends Jenny's state equation S = f (cl, o, r, p, t), where cl = climate and t = time (Jenny, 1994). Many papers use the SCORPAN framework to map different soil attributes (Dharumarajan et al., 2021, Sreenivas et al., 2016, Mitran et al., 2018) and soil classes (Shukla et al., 2018, Gomez et al., 2019). It has successfully quantified the spatial variation of soil properties at regional, national, and global extents.

However, most DSM studies only predict static soil properties or assume the properties are static in time. However, what makes soil work is that it is dynamic. While some soil properties, such as texture, are relatively stable with time, many properties, such as carbon content and pH, are dynamics. Current DSM techniques assume that these variables are constant with time. Although we know when soil observations were collected, maps produced by DSM rarely show the period the soil properties represent. Furthermore, maps are commonly produced using legacy soil data derived from different decadal observation periods. These are some unresolved issues regarding DSM for mapping dynamic soil properties.

While simulation modelling can be used to predict change, direct observations of the biophysical effects of the environment on soil are required to reduce the uncertainty of the model and provide robust evidence.

Conventionally, soil observations spanning a range of temporal scales are required to detect and calculate soil change. Most studies on soil change require resampling, either on the same

or new sites using design-based or model-based approaches. Several sampling methods exist (Filippi et al., 2016). However, this approach has disadvantages due to the high costs and effort.

Filippi et al. (2016) concluded that detecting and mapping soil change depends on available funding, existing soil data, soil properties of interest, long-term intentions, and extent. In the absence of soil monitoring networks or resampling projects, soil change can be detected from legacy soil data, space-for-time substitution, chronosequences or paired sites.

Studies have demonstrated using legacy soil data from soil surveys to study temporal trends at regional or national extents (Lindert et al., 1996). Minasny et al. (2011) utilised soil survey data collected at various times from Java, Indonesia, to show the dynamics of soil carbon over the island over 80 years. The legacy data is characterised by a relatively sparse space and time coverage. However, those studies lack a theoretical approach and do not provide any confidence in the prediction. In addition, numerous sources of bias and lack of location information are negative issues (Saby et al., 2008, Marchant et al., 2012).

Other studies utilised the change in the environmental factors used in DSM calibration (Gray et al., 2016, Yigini and Panagos, 2016) to extrapolate soil condition changes. Yigini and Panagos (2016) predicted current soil organic carbon content by using a regression kriging model using a set of environmental predictors (including climate and land use) and then projecting soil carbon to the near future by changing the climate and land use variables. However, such methods have to assume that the stationarity of the current model and other environmental factors are constant. This method is sometimes called space-for-time extrapolation. Barraclough et al. (2015) concluded that the space-for-time substitution method using regression method could be used to predict soil change influenced by different climatic conditions with constant soil farming factors and different climatic gradients. However the space-for-time extrapolation method relies on the representativeness of the soil samples. For

example, systematic preferential sampling of agricultural lands would make the model invalid to predict soil condition under land use change.

Space-for-time substitution is another method used to infer soil change by comparing natural systems and current land use, assuming that the drivers of the spatial patterns also drive temporal changes (Blois et al., 2013). Filippi et al. (2016) mentioned that using space-for-time substitution to observe soil change can be undoubtedly valuable, mainly when the changing time of land use is known. Many studies use this approach to assess how human activities have modified soil over time (Cattle et al., 1994, Tye et al., 2013). In the space for time substitution technique, soil conditions under the natural vegetation or less disturbed conditions are compared with soils under cropping or agricultural activities (Filippi et al., 2016). The space-for-time substitution model assumes that contemporary spatial phenomena can be used to model temporal processes. The main assumption is that drivers of spatial gradients of soil properties also drive temporal soil change (Blois et al., 2013). Problems can arise when other environmental factors (such as topography and parent materials) that vary spatially have a stronger influence on soil properties than time.

To reduce the effect of soil forming factors on the space for time substitution approach, it is important to understand the soil class distribution in the study area. If the soil type is different, the result of soil variation can be not only from anthropogenic activities but different soil types.

Identifying original soil classes can support understanding soil genesis and distribution and monitoring and assessing soil change because of the different types of human activities. Many studies produce maps of soil classes using machine learning and DSM techniques. Since the 1990s, two methods have been combined for mapping soil classes (Wadoux et al., 2020). Lagacherie and Holmes (1997) show an early example of using DSM to predict soil classes in a region. Cialella et al. (1997) predicted soil drainage classes using elevation and remote

sensing covariates. Sarmento et al. (2010) also compared four machine learning algorithms for DSM and predicted soil orders. Heung et al. (2016) mapped soil great groups and soil order using several methods of developing training datasets from the soil survey data.

However, there are a few drawbacks to predicting soil orders using DSM. Current soil orders according to national or international soil classification systems such as the Australian system, WRB, or USDA soil taxonomy mostly classify soil based on its currently observed morphological attributes, not soil forming factors. Still, the order can also be changed by human activities, including the loss of upper horizons due to accelerated erosion rates in farming lands and mixing with subsoil horizons using the operation of tillage. These unnatural activities can change the original soil order to different orders (Smeck and Balduff, 2002). Kuzyakov and Zamanian (2019) noted that soil properties undergo significant changes due to agropedogenesis. The human activities can transform a broad range of soil orders into Anthrosols, which are soils that have been significantly modified by human activities such as tillage, fertilization, and irrigation.

### 1.3.1. Pedogenon mapping

Investigating changes in soil properties due to human activities requires accurate soil classification that is not affected by those activities. However, current soil classification systems rely on soil properties data, which can be significantly modified by agricultural practices and may not accurately represent the original soil. Therefore, studying soil classification should rely on soil-forming factors that can provide a more representative assessment of the soil. By using soil-forming factors, researchers can more accurately classify soils and evaluate their properties in a way that is less influenced by human activities.

Román Dobarco et al. (2021b) generated soil classes applying unsupervised classification (k-means) to quantitative variables representing soil-forming or SCORPAN factors (soil, climate, organisms, relief, parent material and time) at a reference time. The assumption is that the dominant soil-forming processes occurring over pedogenetic time period in a class with homogeneous soil-forming factors would be similar, and hence the soil properties within the soil class will be similar. Pedogenon is defined as a conceptual soil taxon from a set of quantitative state variables to represent the soil-forming factors for a given reference time. The idea of the pedogenon is derived from the concept of the genon (Boulaine, 1969). The definition of genon is 'a soil volume comprising all the pedons that have the same structure, the same characteristics and result from the same pedogenesis'. The genons can be considered as soil mapping units that differ in composition and spatial distribution but are unrelated to any soil classification system.

The genon concept is akin to the US polypedon concept defined by Johnson (1963) as "a soil individual (polypedon) is also a real soil body; it is a parcel of contiguous pedons all of which have characteristics lying within the defined limits of a single soil series". Another soil mapping unit related to the concepts of polypedon and genon is the pedotop (Haase, 1968). It is a cartographic unit with homogenous pedological properties that results from uniform combinations of soil forming factors, even though it can be present in transition areas into neighbouring units and within unit variation (Campbell and Edmonds, 1984).

Most of the current soil classification systems define soil class from its current soil properties, which can be easily changed by human activities and thus cannot represent the original soil. The map pedogenon using soil-forming factor uses the information of soil-forming factors which is less impacted by human activities.

Pedogenon mapping addresses the current challenge in mapping soil change as soil varies according to the factors of soil formation. Pedogenons stratify a landscape into unique soil entities of homogeneous soil-forming factors. Within a pedogenon we differentiate soils that have been less affected by human activities (called genosoil) and soils that have been intensively affected by humans (called phenosoils). Application of the pedogenon mapping in NSW resulted in pedogenon classes that can explain 40% of the variation in stable soil properties (e.g. texture) (Román Dobarco et al., 2021a). In addition, Román Dobarco et al. (2021a) derived over 5000 pedophenons units in NSW that are under grazing and cropping. Comparing these pedophenons under agricultural influences and pedogenons that are still under native vegetation, they detected trends of soil acidification and organic carbon (OC) losses in topsoils. However, the concept of pedogenon has not been investigated locally. The study of the pedogenon concept on the local scale can be validated with actual soil data.

## 1.4. Soil security

The concept of 'Soil Security' was recently proposed for maintaining and strengthening global soil resources sustainably for humankind. Furthermore, the crucial point of soil security is considering biophysical stocks, functioning and ecosystem services, and economic, social and policy aspects. Five dimensions of soil security were defined (Capability, Condition, Capital, Connectivity, and Codification).

The terms genoform and phenoform were initially proposed by (Droogers and Bouma, 1997). Rossiter and Bouma (2018) further defined genoform is “soil class as recognised by the soil classification system used as the basis for detailed soil mapping in a given area.” and the phenoform is “continuous, non-cyclical variations of a soil genoform with sufficient chemical or physical differences affect soil function significantly” (Rossiter and Bouma, 2018). Linking back to the soil security dimensions, the capability is what soil can do for humans? or how much we can produce (food) from soil, which means the expected performance of soil, which can be linked to genoform (McBratney et al., 2017). The condition refers to the phenoform (current state of the soil), which agricultural practices change.

This study replaces the conceptual genoform and phenoform with operational genosoil and phenosoil (Huang et al., 2018). This study recognises areas of genosoil and phenosoil in every pedogenon to infer soil change. Hence, from the soil security point of view, we need to understand the capability and the condition of how soil changes to human activities. The ability to detect soil change promptly will minimise the risk of soil degradation. In addition, long-term data collected from soil change can support a better understanding of sustainable limits for soil use productivity.

## 1.5. Aims

This thesis aims to understand the current status of soil organic carbon (SOC) and predict how human activities have changed SOC in the Edgeroi region of NSW, Australia. Measuring and modelling anthropogenic signatures are essential to improve our understanding of the human impact on soil, quantify the change rate, and manage soil better. Measuring the soil properties change over a large area can support understanding current soil stocks and fluxes to provide ecosystem function. By understanding how vulnerable and resilient soil is to human-induced change, we can take proactive steps to protect and ensure the sustainability of soil functions. This thesis will test and improve a novel digital soil mapping using pedogenon approach for assessing soil change in Edgeroi, New South Wales. The main hypothesis is that pedogenon is an efficient method of stratifying the landscape and can be used to estimate soil organic carbon change.

Following the framework by Román Dobarco et al. (2021b), a pedogenon map is created at the local scale to investigate soil change. This new model synthesises national and local soil datasets utilising rich spatial datasets. Pedogenons will be derived in the Edgeroi area based on a set of quantitative state variables representing the soil forming factors (climate, organisms, relief, parent material, time) for a particular reference time. The underlying assumption is that these classes result from multimillennial natural soil formation processes and historic anthropedogenesis and hence would have developed similar soil properties. Subsequently, information on human forcings since the time chosen as a benchmark (drivers of contemporary soil change) is incorporated to stratify these pedogenon classes into subclasses (i.e., genosoils and phenosoils) with varying degrees degree on human pressure.

To achieve this aim, specifically this thesis:

(1) Optimising spatial layers for creating pedogenon map by creating a realistic parent material map as a SCORPAN factor (Chapter 2).
(2) Creating a pedogenon map and stratified pedophenon classes (genosoils and phenosoils) using the data of soil forming factors and land use (Chapter 3).
(3) Designing a sampling strategy to capture pedogenons and their pedophenon classes. And furthermore collecting and analysing field soil samples (0-1 m) to describe the pedogenon classes and validate the pedogenon map (Chapter 3).
(4) Mapping the change of soil organic carbon due to human cultivation using pedogenon and pedophenon classes (Chapter 4).
(5) Investigating SOC sequestration potential in the area by producing maps of MAOC (mineral-associated OC) potential and current conditions (Chapter 5).
(6) Based on the calculated soil change, setting target values of soil organic carbon sequestration for supporting further study or making the decision of better agriculture management approaches (Chapter 6).

This thesis presents a novel pedogenon mapping utilising rich remote-sensing data in combination with field sampling and machine learning models. The model will account for the impacts of land use and land management change. More importantly, this thesis aims to derive evidence-based information on soil change supporting soil security.

## 1.6. References


BANWART, S. A., CHOROVER, J., GAILLARDET, J., SPARKS, D., WHITE, T., ANDERSON, S., AUFDENKAMPE, A., BERNASCONI, S., BRANTLEY, S. & CHADWICK, O. 2013. Sustaining Earth's critical zone basic science and interdisciplinary solutions for global challenges. Univ. of Sheffield, Sheffield, UK.

BARRACLOUGH, D., SMITH, P., WORRALL, F., BLACK, H. & BHOGAL, A. 2015. Is there an impact of climate change on soil carbon contents in E ngland and W ales? European Journal of Soil Science, 66, 451-462.

BLOIS, J. L., WILLIAMS, J. W., FITZPATRICK, M. C., JACKSON, S. T. & FERRIER, S. 2013. Space can substitute for time in predicting climate-change effects on biodiversity. Proceedings of the National Academy of Sciences, 110, 9374-9379.

BOULAINE, J. 1969. Sol, pédon et genon. Concepts et définitions. Bull. AFES, 5, 7-10.

BRADY, N. C., WEIL, R. R. & WEIL, R. R. 2008. The nature and properties of soils, Prentice Hall Upper Saddle River, NJ.

CAMPBELL, J. B. & EDMONDS, W. J. 1984. The missing geographic dimension to soil taxonomy. Annals of the Association of American Geographers, 74, 83-97.

CATTLE, S., KOPPI, A. & MCBRATNEY, A. 1994. The effect of cultivation on the properties of a Rhodoxeralf from the wheat/sheep belt of New South Wales. Geoderma, 63, 215-225.

CIALELLA, A., DUBAYAH, R., LAWRENCE, W. & LEVINE, E. 1997. Predicting soil drainage class using remotely sensed and digital elevation data. Photogrammetric Engineering and Remote Sensing, 63, 171-177.

CLINE, M. G. 1961. The changing model of soil. Soil Science Society of America Journal, 25, 442-446.

CRUTZEN, P. J. 2016. Geology of mankind. Paul J. Crutzen: A pioneer on atmospheric chemistry and climate change in the Anthropocene. Springer.

DHARUMARAJAN, S., KALAISELVI, B., LALITHA, M., VASUNDHARA, R. & HEGDE, R. 2021. Defining fertility management units and land suitability analysis using digital soil mapping approach. Geocarto International, 1-21.

DIAMOND, J. 2002. Evolution, consequences and future of plant and animal domestication. Nature, 418, 700-707.

DROOGERS, P. & BOUMA, J. 1997. Soil survey input in exploratory modeling of sustainable soil management practices. Soil Science Society of America Journal, 61, 1704-1710.

FILIPPI, P., MINASNY, B., CATTLE, S. R. & BISHOP, T. F. A. 2016. Chapter Four - Monitoring and Modeling Soil Change: The Influence of Human Activity and Climatic Shifts on Aspects of Soil Spatiotemporally. In: SPARKS, D. L. (ed.) Advances in Agronomy. Academic Press.

GOMEZ, C., DHARUMARAJAN, S., FÉRET, J.-B., LAGACHERIE, P., RUIZ, L. & SEKHAR, M. 2019. Use of sentinel-2 time-series images for classification and uncertainty analysis of inherent biophysical property: Case of soil texture mapping. Remote Sensing, 11, 565.

GRAY, J. M., BISHOP, T. F. A. & WILFORD, J. R. 2016. Lithology and soil relationships for soil modelling and mapping. CATENA, 147, 429-440.

HAASE, G. 1968. Pedon und Pedotop-Bemerkungen zu Grundfragen der regionalen Bodengeographie. Landschaftsforschung-Beiträge zur Theorie und Anwendung. BARTHEL, H., Gotha/Leipzig, VEB Hermann Haack Geographisch-Kartographische Anstalt.

HEUNG, B., HO, H. C., ZHANG, J., KNUDBY, A., BULMER, C. E. & SCHMIDT, M. G. 2016. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. Geoderma, 265, 62-77.

HUANG, J., MCBRATNEY, A. B., MALONE, B. P. & FIELD, D. J. 2018. Mapping the transition from pre-European settlement to contemporary soil conditions in the Lower Hunter Valley, Australia. Geoderma, 329, 27-42.

JENNY, H. 1994. Factors of soil formation: a system of quantitative pedology, Courier Corporation.

JOHNSON, W. M. 1963. The pedon and the polypedon. Wiley Online Library.

KUZYAKOV, Y. & ZAMANIAN, K. 2019. Reviews and syntheses: Agropedogenesis–humankind as the sixth soil-forming factor and attractors of agricultural soil degradation. Biogeosciences, 16, 4783-4803.
LAGACHERIE, P. & HOLMES, S. 1997. Addressing geographical data errors in a classification tree for soil unit prediction. International Journal of Geographical Information Science, 11, 183-198.
LINDERT, P. H., LU, J. & WANLI, W. 1996. Trends in the soil chemistry of South China since the 1930s. Soil Science, 161, 329-342.
MARCHANT, B., CRAWFORD, D. & ROBINSON, N. Spatial and temporal prediction of soil properties from legacy data. Digital Soil Assessments and Beyond: Proceedings of the 5th Global Workshop on Digital Soil Mapping, 2012. 239-44.
MCBRATNEY, A. B., FIELD, D. J., MORGAN, C. L. & JARRETT, L. E. 2017. Soil security: A rationale. Global Soil Security. Springer.
MCBRATNEY, A. B., MENDONÇA SANTOS, M. L. & MINASNY, B. 2003. On digital soil mapping. Geoderma, 117, 3-52.
MINASNY, B., SULAEMAN, Y. & MCBRATNEY, A. B. 2011. Is soil carbon disappearing? The dynamics of soil organic carbon in Java. Global Change Biology, 17, 1917-1924.
MITRAN, T., MISHRA, U., LAL, R., RAVISANKAR, T. & SREENIVAS, K. 2018. Spatial distribution of soil carbon stocks in a semi-arid region of India. Geoderma Regional, 15, e00192.
MONTGOMERY, D. R. 2007. Soil erosion and agricultural sustainability. Proceedings of the National Academy of Sciences, 104, 13268-13272.
RICHTER, D. D., BACON, A. R., BRECHEISEN, Z. & MOBLEY, M. L. Soil in the Anthropocene. IOP Conference Series: Earth and Environmental Science, 2015. IOP Publishing, 012010.
RICHTER, D. D. & YAALON, D. H. 2012. "The changing model of soil" revisited. Soil Science Society of America Journal, 76, 766-778.
RICHTER JR, D. D. 2007. Humanity's transformation of Earth's soil: Pedology's new frontier. Soil science, 172, 957-967.
ROBINSON, D., HOCKLEY, N., DOMINATI, E., LEBRON, I., SCOW, K., REYNOLDS, B., EMMETT, B., KEITH, A., DE JONGE, L. W. & SCHJØNNING, P. 2012. Natural capital, ecosystem services, and soil change: Why soil science must embrace an ecosystems approach. Vadose Zone Journal, 11.
ROMÁN DOBARCO, M., MCBRATNEY, A., MINASNY, B. & MALONE, B. 2021a. A framework to assess changes in soil condition and capability over large areas. Soil Security, 4, 100011.
ROMÁN DOBARCO, M., MCBRATNEY, A., MINASNY, B. & MALONE, B. 2021b. A modelling framework for pedogenon mapping. Geoderma, 393, 115012.
ROSSITER, D. G. & BOUMA, J. 2018. A new look at soil phenoforms–Definition, identification, mapping. Geoderma, 314, 113-121.
SABY, N., ARROUAYS, D., ANTONI, V., LEMERCIER, B., FOLLAIN, S., WALTER, C., SCHVARTZ, C. J. S. U. & MANAGEMENT 2008. Changes in soil organic carbon in a mountainous French region, 1990–2004. 24, 254-262.
SARMENTO, E., GIASSON, E., WEBER, E., FLORES, C. & HASENACK, H. Comparison of four machine learning algorithms for digital soil mapping in the Vale dos Vinhedos, RS, Brasil. International workshop on digital soil mapping, 2010.
SHUKLA, G., GARG, R. D., SRIVASTAVA, H. S. & GARG, P. K. 2018. An effective implementation and assessment of a random forest classifier as a soil spatial predictive model. International Journal of Remote Sensing, 39, 2637-2669.
SMECK, N. E. & BALDUFF, D. Contrasting approaches for the classification of eroded soils in the USA. Proceedings of the Paper no. 616 in Transactions of the 17th World Congress of Soil Science: Confronting New Realities in the 21st Century, 2002. Citeseer.
SREENIVAS, K., DADHWAL, V., KUMAR, S., HARSHA, G. S., MITRAN, T., SUJATHA, G., SURESH, G. J. R., FYZEE, M. & RAVISANKAR, T. 2016. Digital mapping of soil organic and inorganic carbon status in India. Geoderma, 269, 160-173.
STOCKMANN, U., MINASNY, B. & MCBRATNEY, A. B. 2014. How fast does soil grow? Geoderma, 216, 48-61.
TYE, A., ROBINSON, D. & LARK, R. 2013. Gradual and anthropogenic soil change for fertility and carbon on marginal sandy soils. Geoderma, 207, 35-48.

WADOUX, A. M. J. C., MINASNY, B. & MCBRATNEY, A. B. 2020. Machine learning for digital soil mapping: Applications, challenges and suggested solutions. Earth-Science Reviews, 210, 103359.
WEISS, H., COURTY, M.-A., WETTERSTROM, W., GUICHARD, F., SENIOR, L., MEADOW, R. & CURNOW, A. 1993. The genesis and collapse of third millennium north Mesopotamian civilization. Science, 261, 995-1004.
YIGINI, Y. & PANAGOS, P. 2016. Assessment of soil organic carbon stocks under future climate and land cover changes in Europe. Science of The Total Environment, 557-558, 838-850.

# Chapter 2. Creating a soil parent material map digitally using a combination of manual interpretation and statistical techniques

This chapter is published as: JANG, H. J., DOBARCO, M. R., MINASNY, B. & MCBRATNEY, A. 2021. Creating a soil parent material map digitally using a combination of interpretation and statistical techniques. Soil Research.

## 2.1. Abstract

In this study, a map of soil parent material is created to support the delineation of soil properties and classes of the Narrabri Shire, NSW. Currently, available information in this study area are geological and lithological maps at a scale of 1:250 000 to 1: 1 million. These maps are not detailed, and the description in some areas is not accurate. Thus, this study created a new parent material map using geological and lithology information, barest earth satellite imagery, gamma radiometric, topography, prior soil map, and digital soil texture maps (clay and sand content). Based on manual interpretation and parent material observations, 18 parent material classes were delineated in the area. The 18 classes were then modelled using Linear Discriminant Analysis using Digital Elevation Model (DEM), Slope, Topographic Wetness Index (TWI), Gamma potassium (K) and Thorium (Th), and Ratio K to Th and soil VIS and NIR (created using RGB and Near Infrared (NIR) bands) as covariates. This modelling process was iterated 50 times, and the most frequently predicted class was assigned to each of the 90 m × 90 m pixels throughout the study area. A map of the frequency of the predicted classes was also created to assess modelling uncertainty. The new parent material map consists of sedimentary residuals (sandstone), volcanic materials (basalt), alluvium, and colluvium. The alluvium can be distinguished into six classes according to slope, soil information from satellite images and soil texture. The colluvium consists of three classes with a characteristic of high clay content (smectitic) and brown in colour (kaolinitic). Using similar approaches, such soil parent material or substrate maps could be developed for different regions in Australia. This method generated unique soil parent material classes combining stratigraphy, lithology and geomorphology.

## 2.2. Introduction

Soil-forming factors are critical to understanding soil formation and distribution and, consequently, soil mapping. In digital soil mapping, we recognised seven factors that can be used as covariates (s: soil, c: climate, o: organisms, r: topography, p: parent material, a: age and n: space (McBratney et al., 2003). Among the seven factors, the parent material is a significant factor for understanding the history of soil formation and origin of the soil. Moreover, there is a close relationship between parent material and soil properties (Ma et al., 2019). For example, Mora-Vallejo et al. (2008) concluded that using parent material information produced a highly accurate prediction of soil clay content.

The information on soil parent materials can be obtained from digitised geological or lithological map or other related information (McBratney et al., 2003). Some datasets of soil parent materials are available. For example, in the UK, a patent material map is available at 1 km × 1 km ground resolution (Lawley, 2014). In New South Wales (NSW), a lithology map was made based on a 1:250 000 geological map (Gray et al., 2016).

In one of the first digital soil mapping studies, Legros and Bonneric (1979) used geological and lithological maps as parent material information and consequently used it to derive soil classes maps. Legros and Bonneric (1979) also used a geological map to create a map of soil classes.

Several studies used remote sensing images to indicate parent materials. Dobos et al. (2013) developed and tested a digital mapping procedure to support SOTER (SOil and TERrain digital database (Oldeman and van Engelen, 1993)), requiring parent material and terrain information. They created a parent material map using remote sensing images (MODIS and SRTM). Boettinger et al. (2008) suggested that shortwave infrared (SWIR) range band can be used for the environmental covariates of the parent material. Their study also mentioned that, Landsat

spectral band ratios (3/7, and 5/7) (Landsat 7) can detect ferrous iron and hydroxyl radicals from the exposed soil and parent material. Moreover, Normalised difference ratios of the spectral bands can be indicative of certain parent materials such as gypsic soil (5–7)/(5+7) (Landsat 7) and calcareous rocks (5–2)/(5+2) (Landsat 7).

Other sources of information have also been used as parent material covariate for digital soil mapping. Maps of regolith and topography were commonly used to infer parent material (Dymond and Luckman, 1994). Many studies in Australia used airborne gamma radiometric data as parent material information for digital soil mapping (McKenzie and Ryan, 1999, Ryan et al., 2000). Wilford (2012) created a map of weathering intensity of Australian soils using gamma radiometric and elevation data. The map indicates the degree of weathering of soils.

Although parent material data contain very useful information for digital soil mapping, there is not much research on creating a parent material map itself. One of few examples is Heung et al. (2014), who used a digital elevation model (DEM) to calculate 27 topographical indices. They used the indices combined with soil survey data to create a parent material map. The results of their study showed a highly accurate prediction. Nevertheless, that study only used one type of data (topography). Lacoste et al. (2011) used Multiple Additive and Regression Trees to predict soil parent material at 50 m resolution using existing soil observations (20 classes of soil parent material, 11 were geological bedrock formation, and 9 were superficial deposits). The parent materials were predicted from 17 covariates represented by topography, gamma radiometric, geological information and remote sensing images. The geological information was the most important covariate, and the prediction showed a high accuracy. Moreover, a map of parent material in a tropical area was produced by Bonfatti et al. (2020). The study used remote sensing images and machine learning methods (decision tree, random forest, support vector machine, multinomial logistic regression, K-means, and object-based image analysis with maximum) as covariates. A map of parent material from 280 geological

observation points was predicted using a digital elevation model and multitemporal Landsat images.It appears that there are only a few studies that attempted to create digital parent materials map. Parent material maps should be produced at a high resolution using different types of covariates (information on topography, gamma radiometric, geological information and remote sensing). Thus, this study aims to create a high-resolution parent material map using different sources of information in the Edgeroi district of New South Wales (NSW). The approach used a combination of manual interpretation and statistical techniques.

## 2.3. Methods

### 2.3.1. Study area

The study area is the Edgeroi district of NSW, Australia (30.06°S, 149.48E). It covers an area of ~1700 $km^2$ (Figure 2-1). The climate is characterised by hot, dry summers and cold winters. The mean maximum temperature reaches 30°C in November, and the next 4 months, the maximum temperature keeps increasing (up to 40°C or higher) (Ward, 1999). The average minimum temperatures are 19°C in January and February and 3°C in July. The average annual rainfall is 712 mm, which is highly variable from month to month. Topographically, the west side of this study area is very flat, and the east side is a mountainous area. In terms of land use, the irrigated cotton industry is situated in the study area, and the main crops (wheat, sorghum and sunflower) are produced (McGarry et al., 1989). This study area has been frequently used as a basis for developing soil mapping methodologies (Malone et al., 2009, Triantafilis et al., 2016). Figure 2-2 shows the flowchart of this study in creating a new digital parent material map.

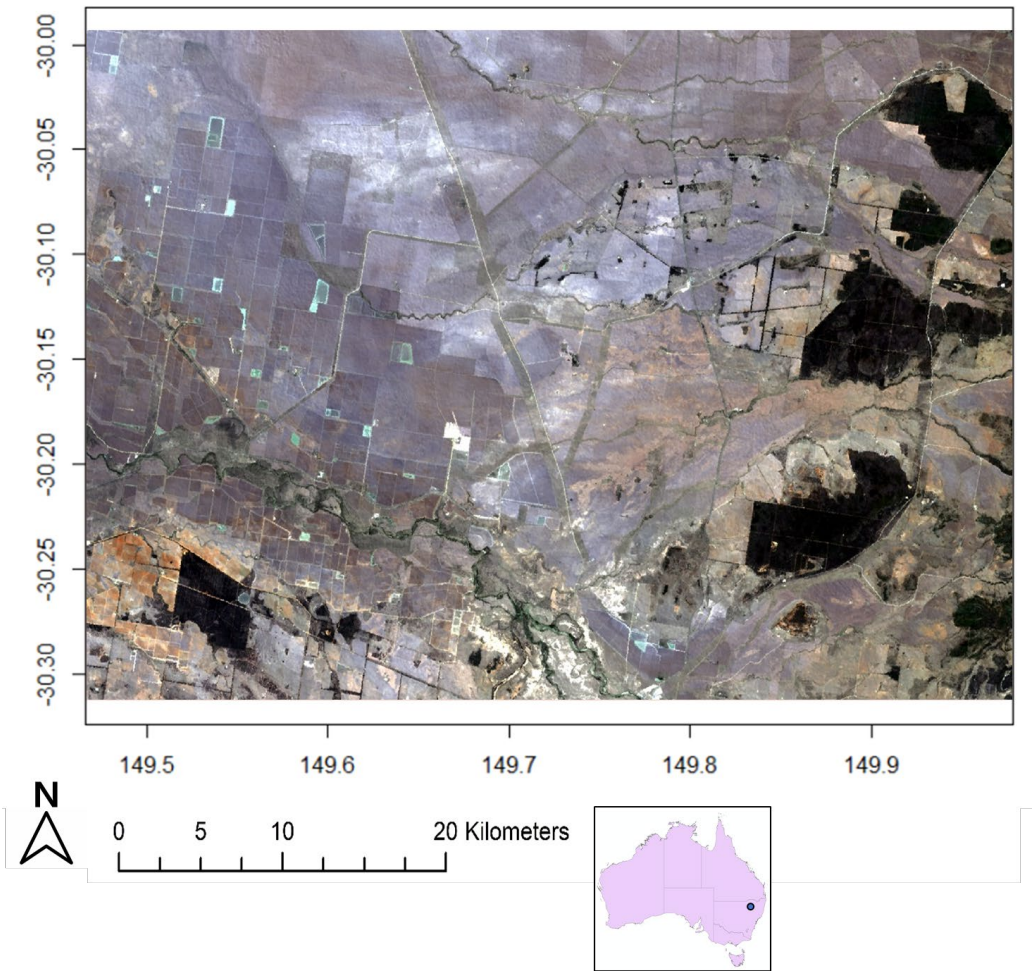



**Figure 2-1. Landsat image of the study area (Edgeroi, New South Wales, Australia).**

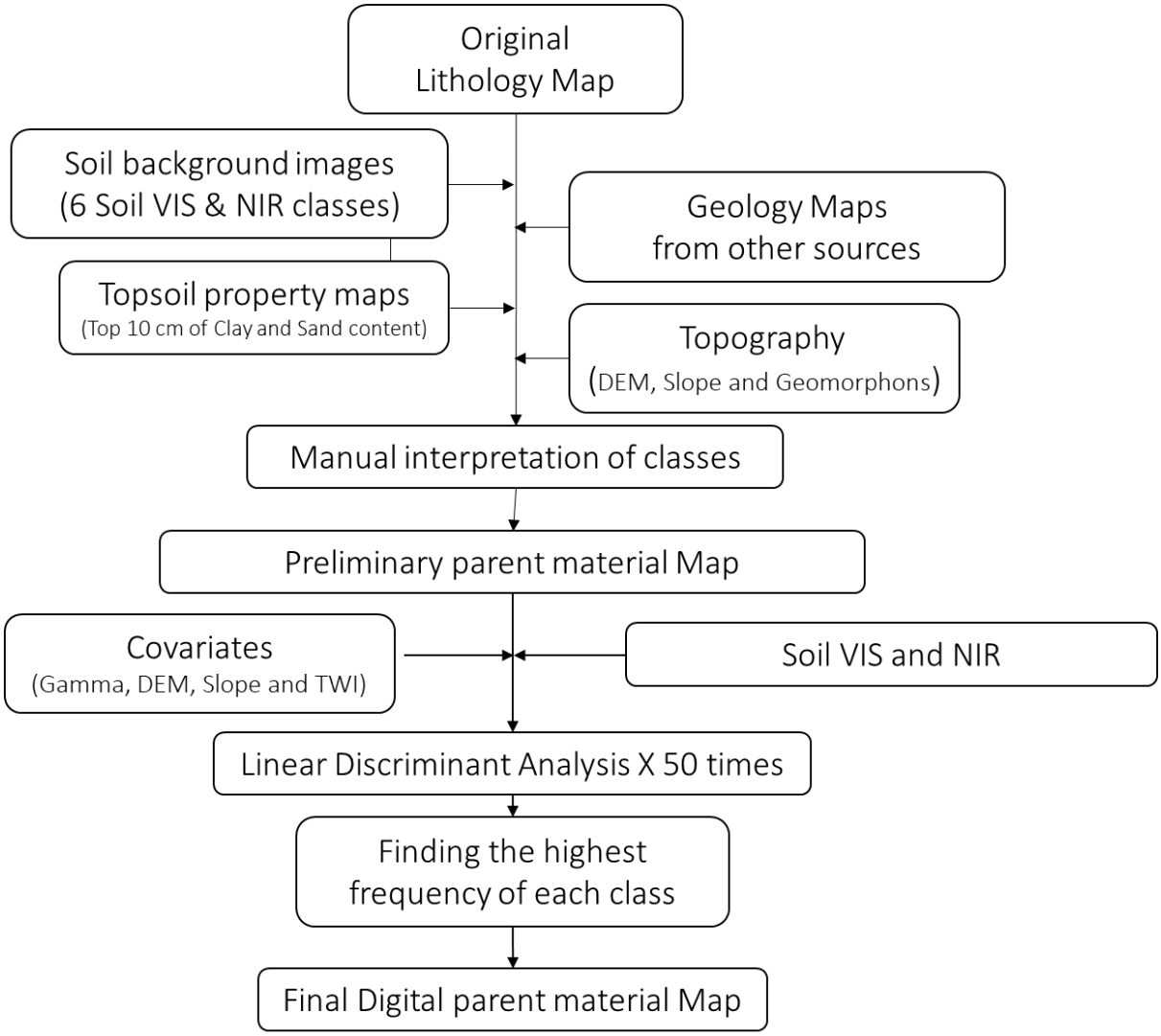



**Figure 2-2. Flowchart of the method to create a new digital parent material map.**

### 2.3.2. Creating a digital parent material map

Four types of information were used to create the soil parent material map: soil geological and lithological information, an image of the barest state of topsoil, topsoil properties and topography.

#### 2.3.2.1. Geological and lithological information

Four sources of soil parent material are available in this study area (Figure 2-3).

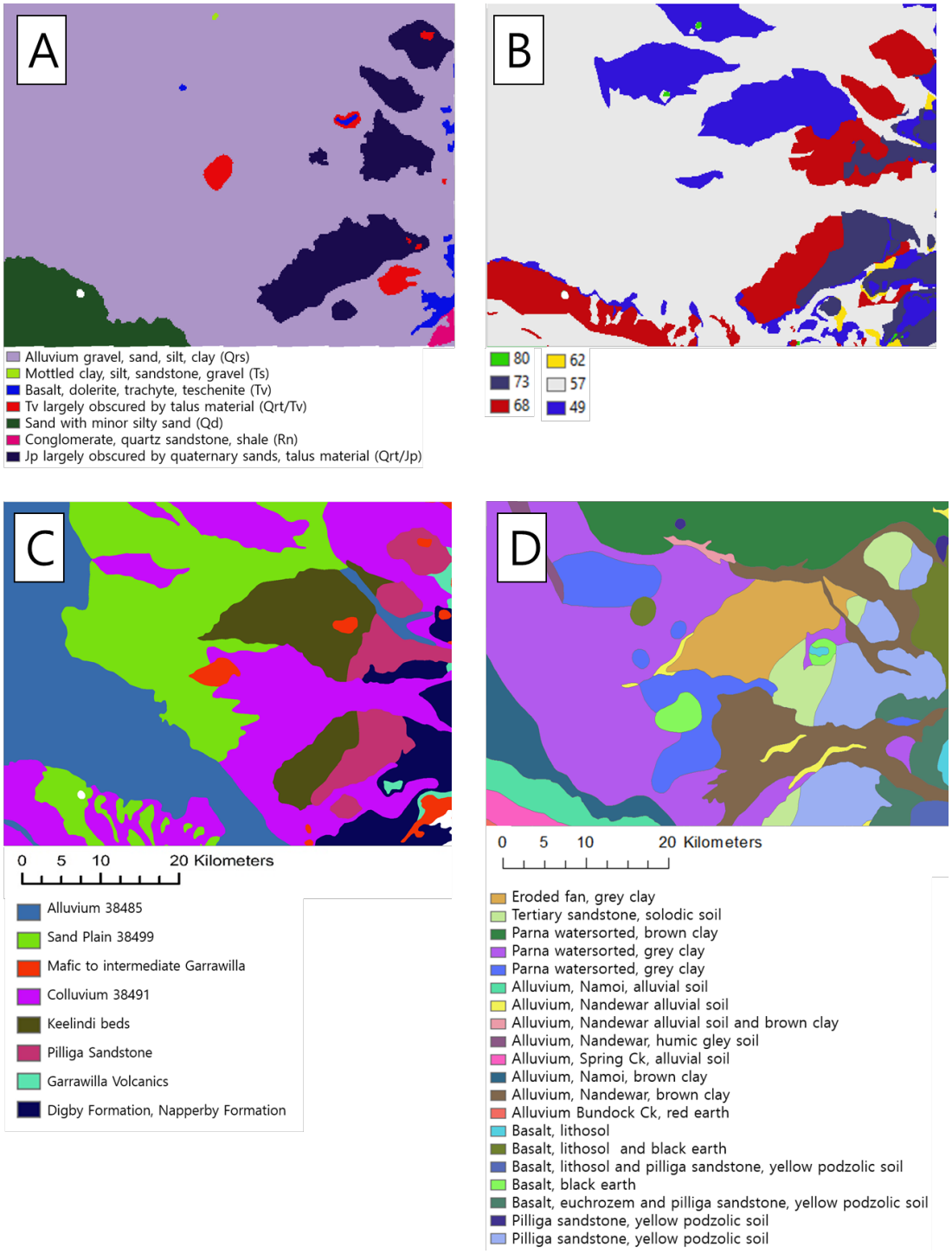



**Figure 2-3. the map (A) of 1:250 000 Geological Sheet SH-55-12 (Wallis, 2019). There are seven classes of the parent material. The image (B) of lithology based on silica index. the image of lithology information (C) (8 classes) (From Geoscience Australia). The white parts are removed classes due to small areas. The image (D) of soil, parent rocks and sediments information. It was digitalised from the paper map (Ward, 1999). The size of the image(D) is small than the study area. The A, B and C is the same size as the study area, but the D is smaller than the study area.**

The Narrabri 1:250 000 Geological Sheet SH-55–12 (Figure 2-3A) (Wallis, 2019) was compiled in 1966 using aerial photography. The image showed seven classes of residual and transported materials in the study area. The most significant class is alluvium, and there are two main areas where sand content is high. The limitation of this image is that it does not consider information about topography. In the eastern mountainous area, the image showed no relationship between alluvium and elevation.

A derivation of this image, called the lithology map of NSW was created by Gray et al. (2016). There are 12 lithology classes; eight classes were grouped based on the different level of silica (silica index). High silica index represents high sand content in the soil, and as the index decreases, the content of sand decreases. The other four classes are categorised based on the chemical characters: calcareous, evaporite, organic and sesqui-oxide materials. In this study area, the lithology information showed six classes of silica index (Figure 2-3B and Table 2-1). This image has further information than the geology map (Figure 2-3A), but it is too broad, not useful, and does not consider topography.

**Table 2-1. Lithology class, silica content (%), median value of silica (silica index) and examples. Source: Gray et al. (2016)**

| Lithology class | Silica, $SiO_2$ (%) | Median (silica index) | Examples |
|---|---|---|---|
| Siliceous upper | 77–85 | 80 | Quartz sandstone, quartz siltstone, unqualified quartzite and alluvial sands |
| Siliceous mid | 70–77 | 73 | Granite, rhyolite and siliceous tuff, arkose sandstone, most unqualified sandstone |
| Siliceous lower | 65–70 | 68 | Adamellite, granodiorite, dacite, monzogranite, siliceous/ intermediate tuff, most greywacke & lithic sandstone, unqualified siltstone |
| Intermediate upper | 60–65 | 62 | Syenite, trachyte, most argillaceous rocks (mudstone, claystone, shale, slate, phyllite and schist), alluvial loams and non-cracking clays |
| Intermediate lower | 52–60 | 57 | Monzonite, trachy-andesite, diorite, andesite, intermediate tuff, alluvial cracking clays (not black) |
| Mafic | 45–52 | 49 | Gabbro, dolerite, basalt, mafic tuff, amphibolite, alluvial black cracking clays |

A lithology map in the study area was retrieved from Geoscience Australia (Raymond et al., 2014). Th ough there are 10 classes, two classes ('Deriah Formation' and 'Millie Group') were unused due to their small size and location at the edge of the study area. Thus, eight classes were relevant to this study. These classes can be grouped as sandstones, colluvium, alluvium, sandplain and volcanic. The map is at 1:1 million scale (Figure 2-3C), which is coarser than the lithology map mentioned above (1:250 000) (Figure 2-3A), but there is a higher number of classes, and this map has more information than other lithology maps which is reflected in the topography and gamma radiometrics (Figure 2-3B and C).

Ward (1999) used soil, parent rocks and sediments information to create a map of parent rock and sediments of the Edgeroi area (Figure 2-3D). This map was created from a field survey, the type of information (20 classes) suitable for a base map for geological and lithological information. Unfortunately, the size of this map is smaller than the study area of this study. Therefore, this map is not suitable for use in this study. The 1:1 000 000 lithology map has a closer pattern to Ward's map. Hence, it was decided to use the lithology map (Figure 2-3C) as a base map.

#### 2.3.2.2. Maps of the barest ground soil visNIR

Satellite images (RGB and NIR) can be used to distinguish different soil types. The image of the barest state of topsoil of the area was obtained from Geoscience Australia (Roberts et al., 2019). The map was created from Landsat images which were taken over the last 30 years. All the images were combined to remove noises and clouds. The combined images predict the barest state of topsoil for the whole of Australia at 25 m × 25 m resolution. The visible and NIR bands of the image were used in this study (R, G, B and NIR).

Even though the map shows the barest state of topsoil, there were still areas with perennial vegetation. Therefore, areas with vegetation were removed based on NDVI (Normalised difference vegetation index, NIR – Red / NIR + Red). The range of NDVI in this study area is from –0.3 to 0.7. Different values of NDVI were first trialled to determine a threshold to exclude the vegetation (i.e. 0.45, 0.35, 0.30, 0.25). The trial showed that a value of 0.3 is effective as a threshold to exclude vegetated areas. NDVI values greater than 0.3 did not eliminate areas of woodlands and permanent vegetation, while a value less than 0.3 excluded soils of cultivated areas.

The spectral topsoil information (R, G, B and NIR) were then grouped into classes to allow for better interpretation of soil visNIR information. The bands were grouped into classes using the K-means method in the Google Earth Engine platform. The area was sampled for 50 000 pixels, and different numbers of clusters (2 to 8) and different input bands (only RGB, and RGB and NIR) were trialled.

Since soil information (Vis and NIR) on the vegetated area were missing, they were predicted using geostatistics. The RGB and NIR bands were interpolated using ordinary kriging with the 'gstat' package (Pebesma, 2004, Pebesma and Heuvelink, 2016). There are several studies using ordinary kriging for predicting soil information (Eldeiry and Garcia, 2010, Gia Pham et al., 2019). All sample points of non-vegetated areas were used for prediction. Different resolutions were trialled to observe the accuracy of the kriging interpolation results. Spatial resolution that produced less noise and artefacts, yet still differentiated soil types was used. The soil visNIR map of the area is shown in Figure 2-6 as a true colour map.

#### 2.3.2.3. Maps of topsoil properties (sand and clay)

Topsoil properties maps were created using information from 210 soil samples. The samples came from the study of McGarry et al. (1989), which collected the samples based on a systematic equilateral triangular grid with ~2.8 km spacing between sites. At each sampling point, a 3 m length and 0.1 m diameter core was collected and sampled at 0–10, 10–20, 30–40, 70–80, 120–130 and 250–260 cm for laboratory analysis. In this study, the top 10 cm of sand and clay data was used to create digital maps of surface soil texture as they coincide with the soil information from the RGB and NIR bands. These surface soil information were also captured by airborne gamma radiometric, which typically sensed the soil in the top 20–30 cm (Minty et al., 2009). The covariates used for modelling are DEM, Slope, topographic wetness index(TWI), Gamma K and Th, Ratio K-Th and Gamma dose (Figure 2-4A-C and Figure 2-5A, B, D and G). The modelling was done with the cubist tree model (Padarian et al., 2020) and more information about this model can be found in Kuhn et al. (2012) (package in R: Cubist; (Kuhn et al., 2020)). 70% of data were used for producing a model, and 30% of data were used for validation. The spatial predictions of sand and clay for the study area are at 30 m resolution.

#### 2.3.2.4. Topographic data

The data on topography (DEM, slope and geomorphon, Figure 2-4) is essential information for predicting parent materials. The 30 m DEM was obtained from SRTM, and the slope was calculated. The local geomorphological unit was calculated using the geomorphon approach on the DEM. Jasiewicz and Stepinski (2013) defined geomorphons (geomorphologic phonotype) as ‘a simple ternary pattern that serves as an archetype of a particular terrain morphology’. The geomorphon is calculated at each pixel of DEM raster by comparing the elevation with the neighbouring pixels. According to this information, different types of

geomorphon can be defined: flat, peak, ridge, shoulder, spur, slope, hollow, footslope, valley, and pit. The geomorphology of the area can be interpreted with the geomorphon classes.

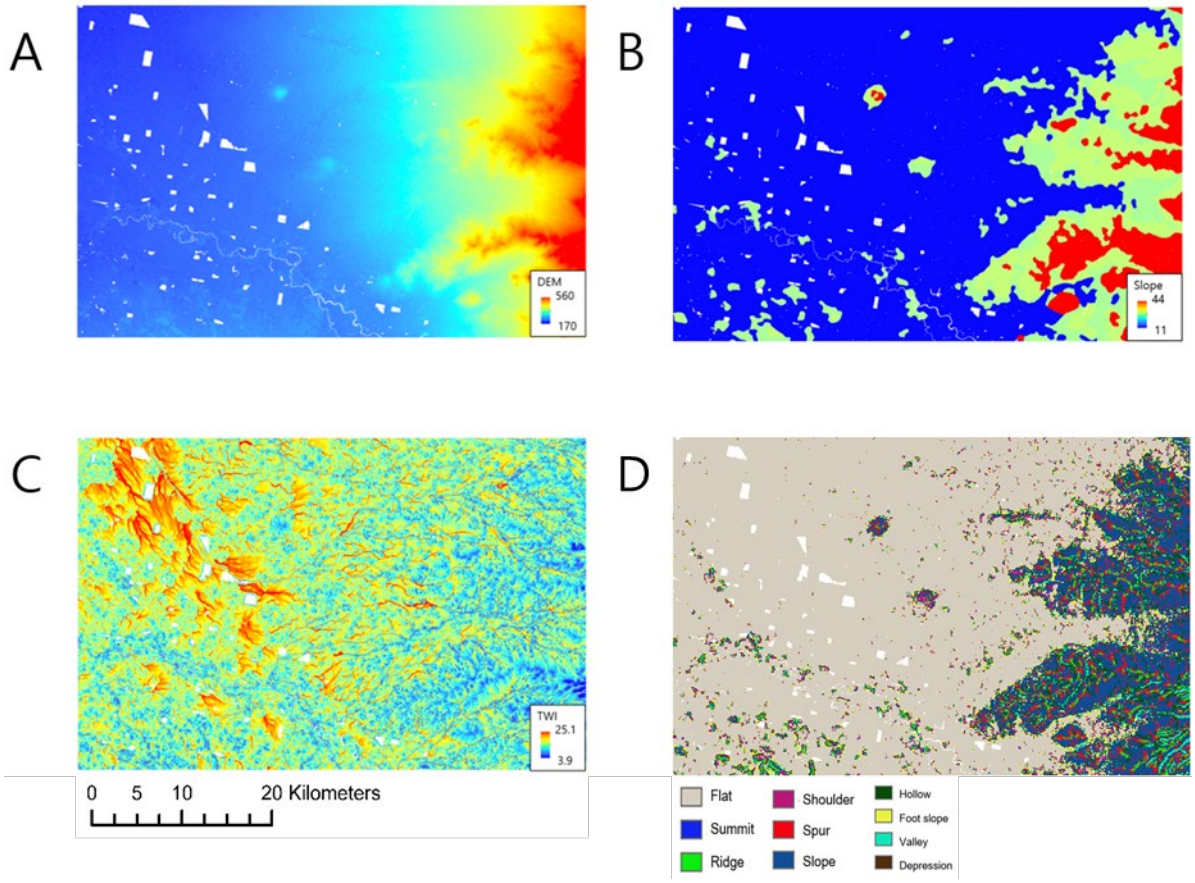



**Figure 2-4. Environmental covariates: Terrain information DEM (A), Slope (unit: percentage) (B) and TWI (C). The resolution of these maps is 30m. The map (D) of geomorphon at 90m resolution. There are ten different types of terrestrial landforms.**

### 2.3.3. Creating a provisional parent material map

Four different types of information (described above) were used to create new parent material classes. Based on the geology map, parent material classes were differentiated based on topsoil visNIR information, soil texture, gamma radiometric and topographic information.

### 2.3.4. Creating a new digital parent material map

A digital map was created based on the interpreted parent material classes (18 classes). For creating the new parent material map, the Linear Discriminant Analysis method (Padarian et al., 2020) was used with covariates (DEM, Slope, TWI, Gamma K and Th, and Ratio K and Th) (Figure 2-4A-C and Figure 2-5A,B and D) and soil visNIR (RGB map: Figure 2-6). Table 2-2 shows the information on the covariates. 10,000 sample points were selected randomly from the initial interpreted parent material map. The result of Linear Discriminant Analysis using more than 10 000 sample points shows very similar patterns as using only 10 000 sample points. Hence, the 10 000 points were used.

To avoid biased or random chance in selecting training data, the sampling and modelling processes were iterated 50 times, and the most frequently predicted class was assigned to each pixel, at a resolution of 90 m × 90 m. In addition, the frequency of that dominant class is also recorded to assess the uncertainty of the initial map delineation and modelling process. All processes were done using the R software (R Core Team, 2013). After the modelling, artefacts were minimised using the majority filter in ArcGIS.

**Table 2-2. Spatial covariates used in the modelling and mapping analysis.**

| Data Type | Data Description | Access Source | Resolution |
|---|---|---|---|
| Terrain | DEM | CSIRO Data Access Portal | 30m |
| | Slope | CSIRO Data Access Portal | 30m |
| | TWI | CSIRO Data Access Portal | 30m |
| Radiometric | Potassium | GADDS | 90m |
| | Thorium | GADDS | 90m |
| | Uranium | GADDS | 90m |
| | Dose | GADDS | 90m |

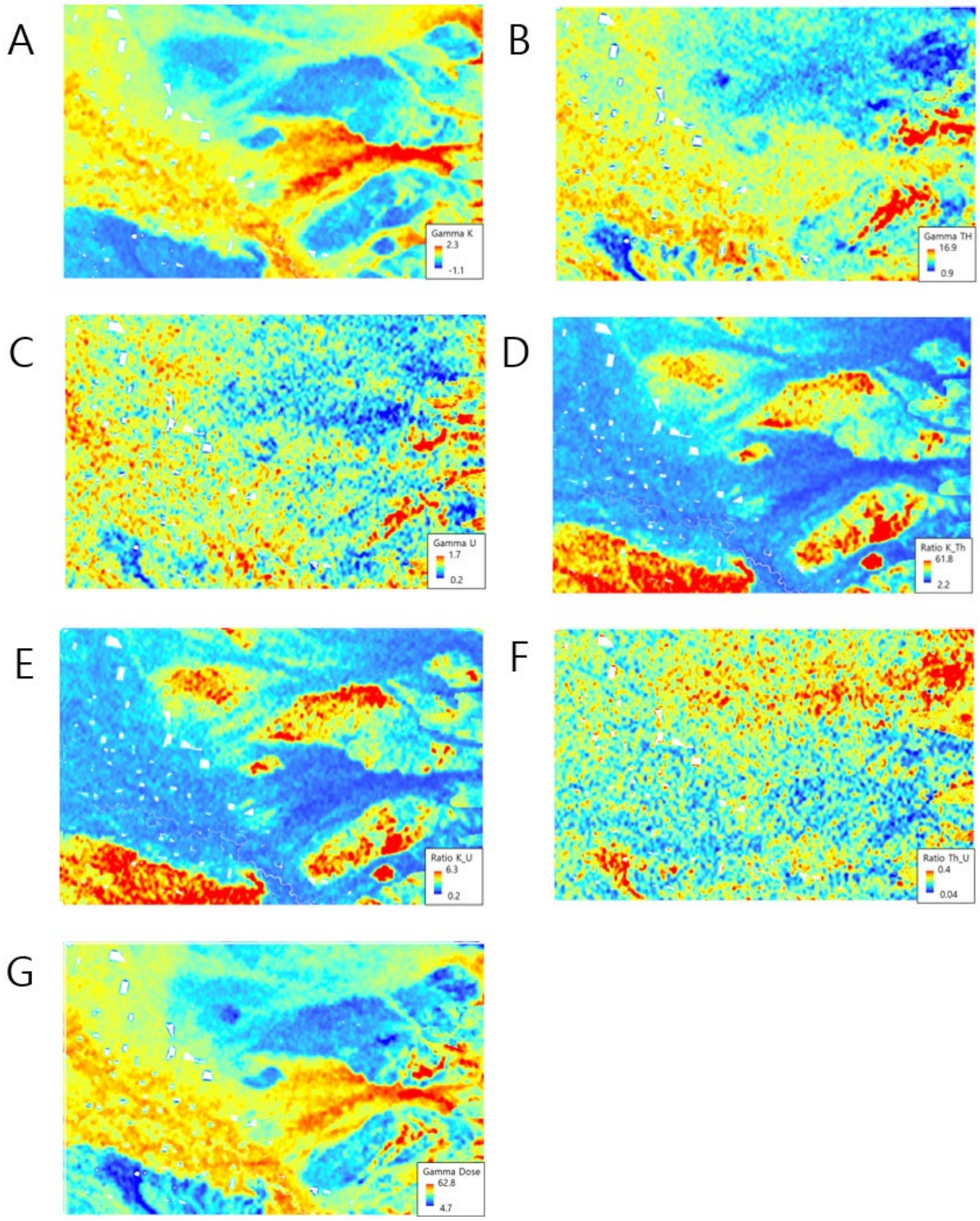



**Figure 2-5. Environmental covariates: Radiometric information Gamma K (A), Gamma Th (B), Gamma U (C), Ratio K/Th (D) Ratio K/U (E), Ratio Th/U (F) and Gamma Dose (G). The resolution of these maps is 90m.**

## 2.4. Result and discussion

### 2.4.1. Different information to create a new digital parent material map

The following sections present the digital layers that were created to support the development of the digital parent material map: existing lithology information, soil colour and soil texture maps.

#### 2.4.1.1. Existing lithology information

There are a few limitations of existing lithology maps. As described in the Methods section, we used the 1:1,000,000 geology map as the basis of the parent material map (Figure 2-3C). The geology classes described in Table 2-3 are reasonably distinct, except for the colluvium and sand plain. Both are the dominant parent material type within their class but can include elements of alluvium or interchangeable. The explanation on colluvium is unclear, e.g. 'may include minor **alluvial or sand plain deposits**, local calcrete and reworked laterite'. The sand plain is described as 'Sand or gravel plains; may include some residual **alluvium**; quartz sand sheets commonly with ferruginous pisoliths or pebbles; local clay, calcrete, laterite, silcrete, silt, **colluvium**'. Furthermore, some classes do not have any information about their origin (rock type). For example, Pilliga Sandstone clearly shows a sandstone origin, whereas other classes, especially colluvium and alluvium, do not provide information about their lithological or stratigraphic origin.

The other limitation is that the soils within the alluvium and colluvium area differed in colour, as seen in the satellite image (Figure 2-1). It means that this lithology map is very broad.

**Table 2-3. The name and description of classes from the Geosciences Australia lithology map.**

| Name of classes | Description of classes |
|---|---|
| Pilliga Sandstone | Medium to very coarse-grained, well-sorted, angular to subangular quartzose sandstone and conglomerate. Minor interbeds of mudstone, siltstone and fine-grained sandstone and coal. Common carbonaceous fragments and iron staining. Rare lithic fragments. |
| Keelindi Beds | Off-white, fine to coarse-grained, poorly to well-sorted, quartzose sandstone, pebbly sandstone and conglomerate interbedded with minor shale, siltstone and coal. Cross-bedded, kaolinitic and iron-stained. |
| Colluvium 38491 | Colluvium and/or residual deposits, sheetwash, talus, scree; boulder, gravel, sand; may include minor alluvial or sand plain deposits, local calcrete and reworked laterite |
| Garrawilla Volcanics | Dolerite, basalt, trachyte, tuff, breccia. |
| Digby Formation, Napper by Formation | Conglomerate at base, overlain by quartz-lithic sandstone which gradually changes into quartzose sandstone unit of cross-bedded sandstone with well-rounded quartz pebbles; a siltstone/sandstone and grey/purple claystone palaeosol is present on top. |
| Alluvium 38485 | Channel and flood plain alluvium; gravel, sand, silt, clay; may be locally calcreted |
| Mafic to Intermediate Volcanics 68000 | Hawaiite, trachyandesite, tristanite, trachyte, minor peralkaline trachyte, tuff |
| Sand plain 38499 | Sand or gravel plains; may include some residual alluvium; quartz sand sheets commonly with ferruginous pisoliths or pebbles; local clay, calcrete, laterite, silcrete, silt, colluvium |

#### 2.4.1.2. Maps of the barest ground and soil visNIR

The map of soil visNIR was created from the near-bareground image. As the image still contains vegetation, the vegetated area was removed using NDVI = 0.3 as a threshold (see Methods). Loiseau et al. (2019) used a threshold of 0.27 because in cultivated soils in France it gave the best trade-offs between the maximum area covered and minimum vegetation effect. The optimum threshold will depend on the type of natural vegetation (e.g. herbaceous vs shrubs and perennial woody vegetation), the percentage of ground covered by vegetation, the proportion of different land uses, and precipitation (Drori et al., 2020). This study area has a high proportion of perennial woody vegetation, open woodland and sparse shrubland with NDVI values ranging between 0.2 and 0.3. Whereas NDVI $\leq$ 0.20 is often used to discriminate bare soil (Gomez et al., 2016), a more conservative threshold would have excluded some cultivated soils and areas with sparse vegetation. Following this exclusion, visible and NIR bands (R, G, B, and NIR) of the vegetated area were interpolated using kriging (Figure 2-6). As there was a large area with vegetation, kriging can still show some artefacts.

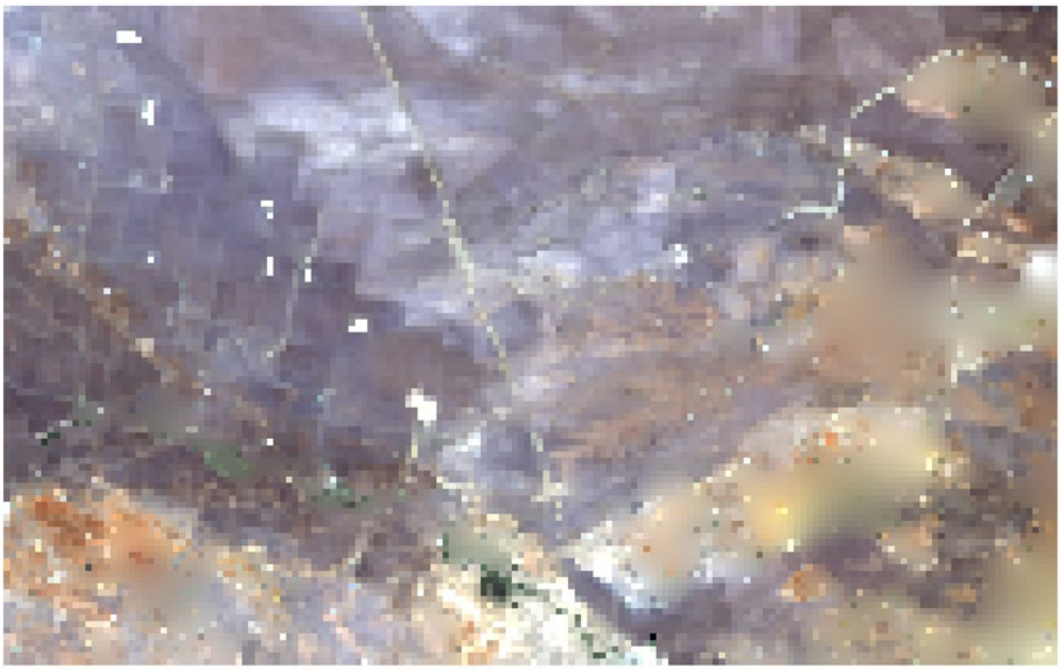


**Figure 2-6. Soil visNIR (the combination of the map of the barest state of topsoil obtained from Geoscience Australia and predicted the topsoil information using ordinary kriging) (300m resolution).**

The R, G, B, and NIR bands of the area were then grouped using K-means clustering to produce soil visNIR classes to allow for easier interpretation. Soil visNIR information from multispectral data can provide useful soil information. For example, Aldana-Jague et al. (2016) used multispectral images to predict soil carbon.

Six soil visNIR classes were found to be optimum in minimising within class variation. The six classes were named according to their dominant colour (Brown, Dark Brown Dark Grey, Greyish, Yellow and White) (Figure 2-7A-F respectively). The clustering method separated dark and light coloured soils. Moreover, with this information, it is possible to recognise the different types of parent materials.

The dark coloured soils (Brown and Dark brown) (Figure 2-7A and B) corresponded to the area of alluvium and colluvium (Figure 2-3C). According to the soil texture map (Figure 2-8), these areas had a high clay content. The light coloured soils (Dark Grey, Greyish, Yellow and White) (Figure 2-7C, D, E, F and B) corresponded to the area of sand plain or areas where the parent materials were sandstone. These areas had high sand content (Figure 2-8). The brighter soil colour implied a high sand content.

There are limitations of using the so-called bare image as some areas with field boundaries are visible. Nevertheless, the grouping of soil visNIR provides a pattern of soil types that can help in separating parent materials in the broad lithology class.

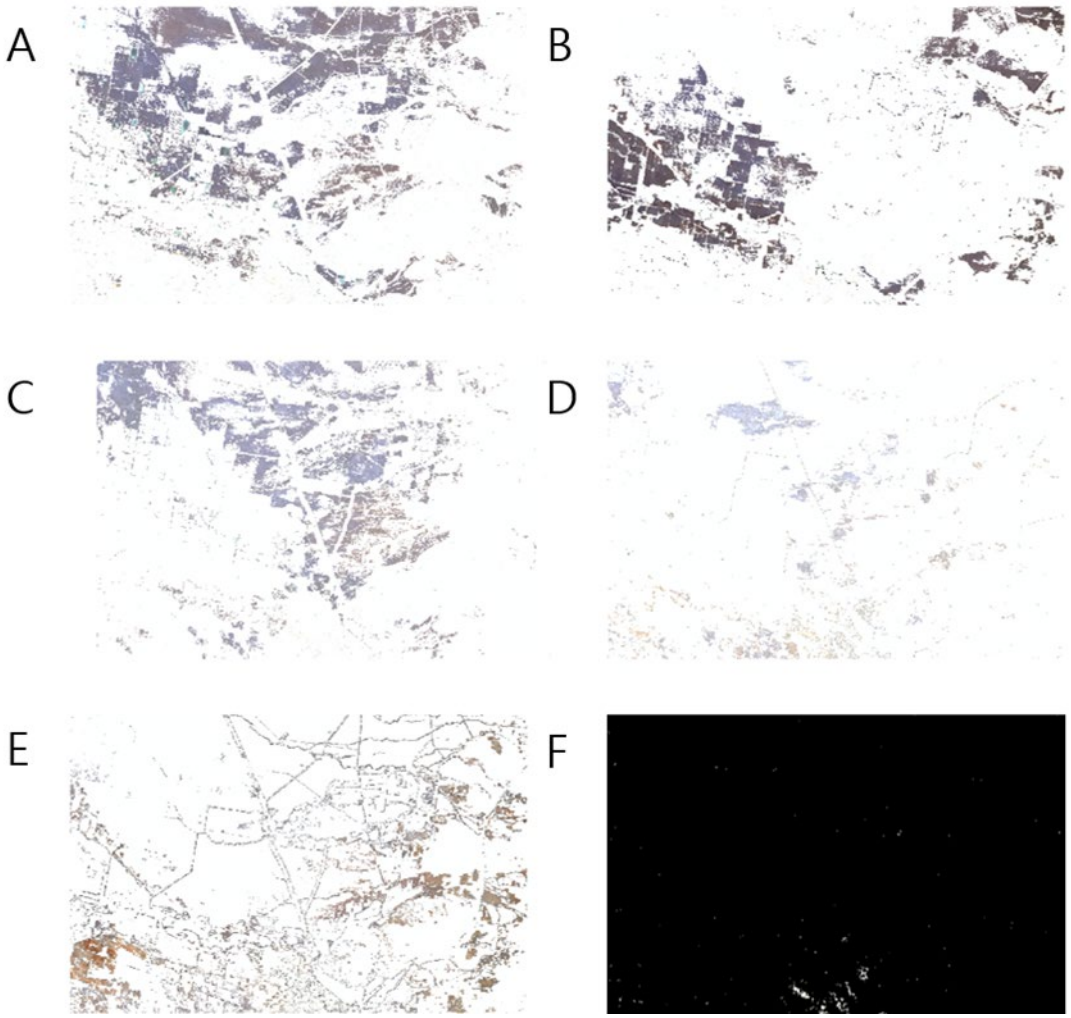



**Figure 2-7. Maps of the 6 VisNIR classes generated with the barest state of topsoil imagery (R, G, B and NIR) using K-means clustering method. The classes were designated with the dominant colour: Brown (A), Dark Brow (B), Dark Grey (C), Greyish (D), Yellow (E) and White (F) (for the map of white colour, black is the background colour to show white colour more clearly).**

#### 2.4.1.3. Map of topsoil texture (sand and clay)

Both clay and sand maps were predicted using the cubist regression tree model (Figure 2-8). The $R^2$ of the model calibration for sand content was 0.70 and decreased to 0.48 for validation data. For clay content, $R^2$ was 0.65 for model calibration and 0.43 for validation. Gamma radiometric data had more influence on the predicted maps than other covariates. It is because gamma radiometric data and soil texture data have a close relationship. Pracilio et al. (2006) demonstrate a relationship between gamma and soil texture from predicting clay content using high-resolution gamma ray spectrometry.

According to the lithology map, the main parent material was sandstone on the east side of the study area, suggesting there should be higher sand content. The predicted map confirmed the high sand content in the east part of the area. The clay map showed the opposite pattern to the sand map. On the western side, there are many areas with high clay content. Only some volcanic areas on the east side showed very high clay content.

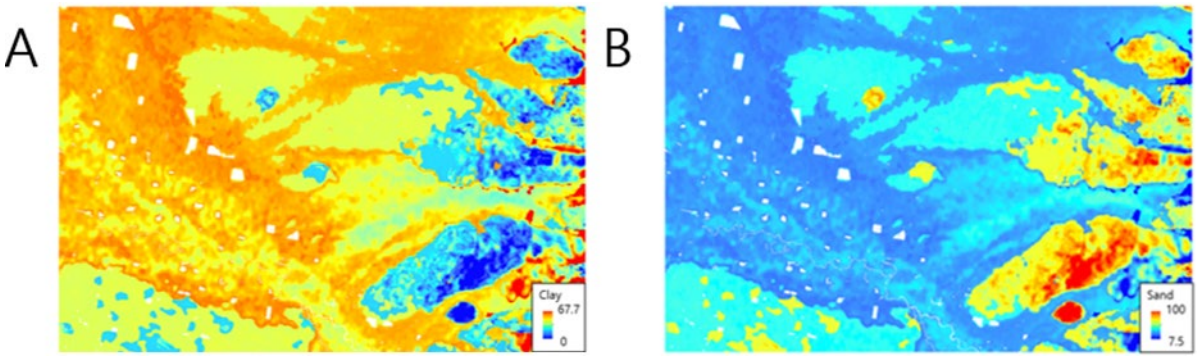



**Figure 2-8. Maps of topsoil clay (A) and sand (B) content created using the cubist model. The resolution of these maps is 30m.**

### 2.4.2. Creating a provisional parent material map

The four sources of information (soil visNIR classes, soil texture (sand and clay content), topography data and parent material information) described above were used to create a provisional parent material map using manual interpretation (Figure 2-9).

Existing classes from the geological map were divided based on the soil information. In addition, new classes were created because their characteristics are very different from their surrounding classes.

Most areas identified as residual materials were kept, but some areas were changed to transported material or become new classes. The classes of transported material were mostly separated or replaced with other classes.

**Sand plain 1** was designated as colluvium in the lithology map (Figure 2-3C) but is located in flat relief (Figure 2-4D), and greyish area (Figure 2-7C). Moreover, soil texture (Figure 2-8B) indicates a high sand content in the area. Therefore, the area was designated as Sand plain 1.

**Sand plain 2** corresponds to Keelindi beds in the lithology map (Figure 2-3C), but the colour is different from Keelindi beds (Figure 2-7A and C), and the area is flat (Figure 2-4). Hence, it was called the Sand plain 2.

The classes **Alluvium 1** and **2** are located near the river, and the soil colour is bright (Figure 2-7F). Soil texture map shows a high content of sand (Figure 2-8B), so they were classified as alluvium.

Both **Alluvium 3.1** and **3.2** areas were formed in one alluvium area (Figure 2-3C), but they are two different visNIR groups (Figure 2-7 A and B). Therefore, it was separated into two classes.

**Alluvium 4** was called sand plain in the lithology map (Figure 2-3C), but the area is located in Alluvium 1.1 (Figure 2-7A). Figure 2-8 A also shows the high clay content in that area, so it was called alluvium instead of sand plain.

**Alluvium 5** was designated as colluvium on the lithology map (Figure 2-3C). While this area is located at a higher elevation, the soil visNIR class is different from the one located in the colluvium area. In addition, the river channel can be observed from the satellite image, so it was allocated to Alluvium 5.

The sand content in classes **Sand plain 3** and **Sand plain 4** is high (Figure 2-8B), but they were separated into two classes due to difference in soil visNIR class (Figure 2-7D and E). Hence, they were named Sand plain 3 and 4.

The classes **Colluvium 1.1** and **1.2** belong to the same colluvium class in the lithology map (Figure 2-3C). However, they were separated into two classes because of the difference in soil visNIR class (Figure 2-7A and B) and clay content (Figure 2-8A).

**Colluvium 2** was kept as described in the lithology map (Figure 2-3C) because it is located at a higher elevation and has higher clay content than the alluvium.

The residual materials maintained the same extension as in the lithology map (Figure 2-3C) (Digby Formation, Napperby Formation, Garrawilla Volcanics, Pilliga Sandstone and Keelindi Beds) except for Mafic to Intermediate Volcanics. According to Ward (1999), the Pilliga Sandstone is also located in the top right of the forest area. Ward (1999) shows, in that forest area, there are two types of material. However, the clay content (Figure 2-8A) and soil visNIR class (Figure 2-6) show a similar pattern in that area, so it was decided to keep them as one class (Figure 2-3C).

Most areas designated as **Keelindi Beds** were kept. However, some areas were designated as **Sand plain 2** due to differences in soil visNIR class (brown and dark grey). The area

**Garrawilla Volcanics** did not change because it has a high clay content (Figure 2-3B) and the base lithology map (Figure 2-3C) already shows detailed information. **Digby Formation, Napperby Formation** is the oldest material and there is not enough information from other lithology maps (Figure 2-3A, B and D). Therefore, it was kept. However, the information from Figure 2-3D and Figure 2-4 shows additional **Mafic to Intermediate Volcanics**.

The gamma radiometric data (Figure 2-5) showed the spatial pattern is consistent with the new lithology map, supporting that the residual materials were kept unmodified.


**Figure 2-9. Map with the preliminary new parent material data (18 classes). The order of legend represents the age of formation, from oldest to youngest (bottom to top).**

### 2.4.3. A new digital parent material map

Based on the interpreted map, a new digital parent material map (Figure 2-11) was created. The provisional map was manually created by interpretation, so it had many unnatural boundaries. Whereas some of the provisional map's subjectivity could influence the LDA model, this was likely to be refined during the DSM process, under the assumption that the LDA model identified the most relevant relationships between the environmental covariates and the defined parent material classes. The provisional map was sampled at 10 000 points, and a digital map was created using covariates (DEM, Slope, TWI, Gamma K and Th, and Ratio K and Th) and soil visNIR. Fifty iterations of sampling and mapping were conducted, and the most frequently predicted class for each pixel was assigned as the likely class (Figure 2-11).

Figure 2-11 shows the final parent material map, and Table 2-4 presents the description of the 18 classes from the new parent material map. The number of classes in the final parent material map increased from eight (lithology map, Figure 2-3C). Increasing the number of classes means having more detailed information and complexity. The classes were also ordered based on the age of formation (Table 2-4), that support understanding the formation of the parent material. The age information was derived from several studies in the area (Young et al., 2002, Raymond et al., 2014, Triantafilis et al., 2013).

Figure 2-10 shows the most predicted class frequency from the 50 modelling iterations. The frequency map shows that the LDA method consistently predicts a class across the 50 iterations. The mean value of accuracy from the 50 modelling iterations is 0.62 (the range of accuracy is between 0.60 and 0.63). The mean frequency for the area is close to 1 (= 0.96), even though the location of sampling points was different in every modelling iteration. Areas with higher uncertainty (lower frequency of occurrence) were located near the boundaries of the class. The results imply that the modelling is robust and suitable to use.

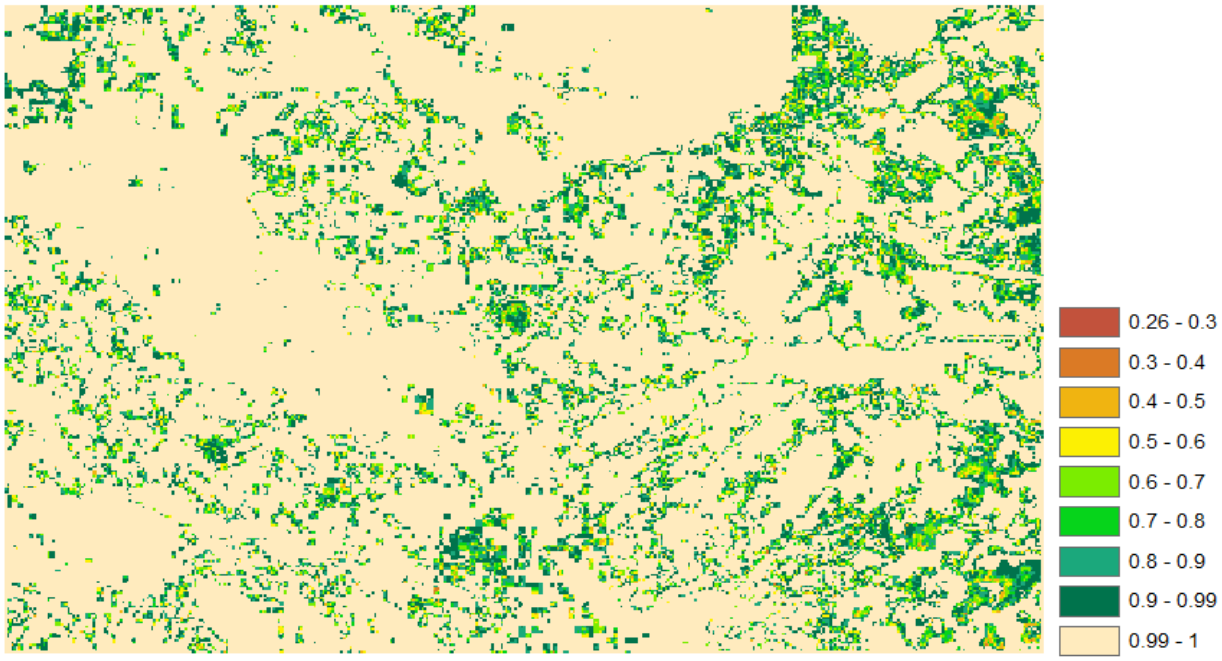



**Figure 2-10. Frequency of the most predicted classes for each pixel calculated from the 50 iteration modelling results.**

The final parent material map (Figure 2-11) shows a similar pattern to the provisional lithology map (Figure 2-3C). Most classes of the digital map are located at a similar position as the ones shown in the lithology map. The reason for the similarity is that the gamma radiometric data significantly influenced both maps. The digital map also showed similarity with Ward's parent material map, but there were some differences in the smaller classes (e.g. **Alluvium 5**). Compared to the manual interpreted map, classes such as **Alluvium 3.2** and **Sand plain 2** were much larger according to LDA. However, **Keelindi Beds** occupied a smaller area according to LDA.

Past papers commonly used only one type of information (Geological and lithological maps, gamma spectrometry or topographic area (Young et al., 2002, Raymond et al., 2014, Triantafilis et al., 2013). However, this paper integrated information on soil properties and soil colour to create a parent material map. It means that the map using different types of data encompasses the complexity of the landscape. On the other hand, the map has not been validated in the field, and thus a field survey is required.

There are several sources of uncertainty derived from the successive decisions integrated in our modelling approach. The choice of the input variables, number of classes in K-means clustering and machine learning algorithms influence the manual interpretation and the output map. Improvement of the parent material map could be produced when more diverse and accurate field data were included.

The method proposed in this study can be applied in a different part of Australia. Especially in areas where access is limited. It used geological and lithological information, soil visNIR information (which can be obtained from barest ground images), soil information (such as previous or digital soil maps) and topography (DEM and geomorphons). All this information is now available for public use. It combined manual interpretation and statistical techniques.

This parent material map can also be used as a layer among soil-forming factors to investigate soil distribution and digital soil mapping.

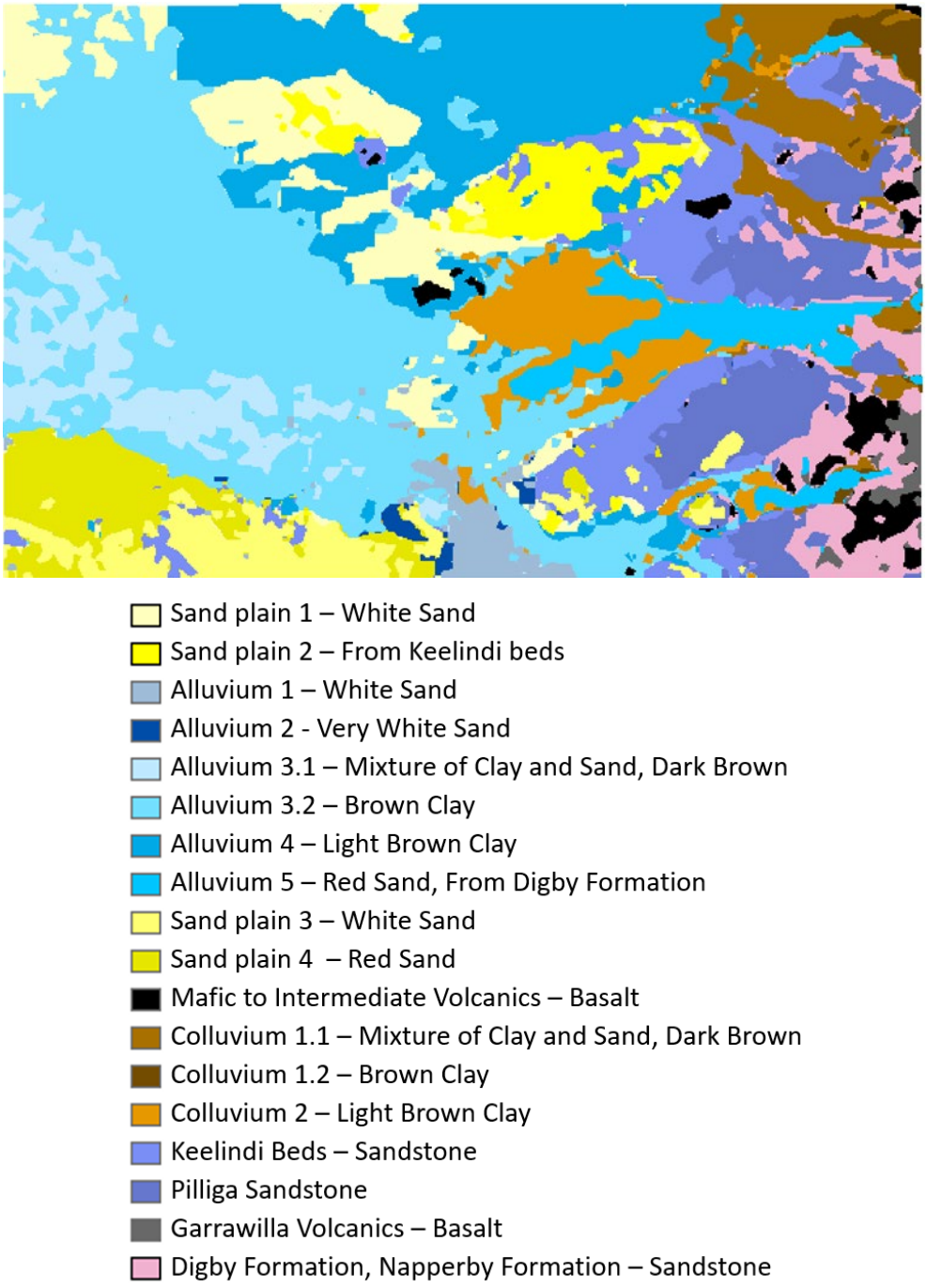



**Figure 2-11. The map of Linear Discriminant Analysis with the new parent material data (18 classes). The order of the legend represents the age of formation, from oldest to youngest (bottom to top).**

**Table 2-4. The description of 18 classes from the new parent material map.**

| Class name | Description | Period |
|---|---|---|
| **Sand plain 1** White Sand | A mixture of sand and clay soil, mildly alkaline, quartz and kaolinite and white colour. | Holocene |
| **Sand plain 2** From Keelindi beds | Sandy, mildly alkaline, quartz and white colour. | Holocene |
| **Alluvium 1** White Sand | Sandy, moderately acid, quartz and white colour. | Holocene |
| **Alluvium 2** Very White Sand | Sandy, moderately acid, quartz and very white colour. | Holocene |
| **Alluvium 3.1** Mixture of Clay and Sand, Dark Brown | Clay soil, moderately alkaline, smectite and dark brown colour. | Pleistocene |
| **Alluvium 3.2** Brown Clay | Clay, moderately alkaline, smectite and brown colour. | Pleistocene |
| **Alluvium 4** Light Brown Clay | Clay, moderately alkaline, smectite and light brown colour. | Pleistocene |
| **Alluvium 5** Red Sand, From Digby Formation | Sandy, neutral, quartz and red colour. | Pleistocene |
| **Sand plain 3** White Sand | A mixture of sand and clay soil, neutral, quartz and kaolinite and white colour. | Neogene |
| **Sand plain 4** Red sand | Sandy, neutral, quartz and red colour. | Neogene |
| **Mafic to Intermediate Volcanics** Basalt | Hawaiite, trachyandesite, tristanite, trachyte, minor peralkaline trachyte, tuff | Neogene |
| **Colluvium 1.1** Mixture of Clay and Sand, Dark Brown | Clay, moderately alkaline, smectite, dark brown colour. | Paleogene |
| **Colluvium 1.2** Brown Clay | Clay, moderately alkaline, smectite and brown colour. | Paleogene |
| **Colluvium 2** Light Brown Clay | Clay, neutral, illite and smectite and light brown colour. | Paleogene |
| **Keelindi Beds** Sandstone | Off-white, fine to coarse-grained, poorly to well-sorted, quartzose sandstone, pebbly sandstone and conglomerate interbedded with minor shale, siltstone and coal. Cross-bedded, kaolinitic and iron-stained. | Cretaceous Jurassic |
| **Pilliga Sandstone** | Medium to very coarse-grained, well-sorted, angular to subangular quartzose sandstone and conglomerate. Minor interbeds of mudstone, siltstone and fine-grained sandstone and coal. Common carbonaceous fragments and iron staining. Rare lithic fragments. | Jurassic |
| **Garrawilla Volcanics** Basalt | Dolerite, basalt, trachyte, tuff, breccia. | Jurassic Triassic |
| **Digby Formation, Napperby Formation** Sandstone | Conglomerate at base, overlain by quartz-lithic sandstone which gradually changes into quartzose sandstone unit of cross-bedded sandstone with well-rounded quartz pebbles; a siltstone/sandstone and grey/purple claystone palaeosol is present on top. | Permian |

## 2.5. Conclusions

A new parent material map was created based on the lithology map at 1:1 000 000 with four spatial layers (maps of the barest state of topsoil, topsoil properties, topography, and geology map from other sources), which increased parent material classes from 8 to 18. This paper used several sources of information to create this new parent material map at 90 m resolution. A combination of manual interpretation and digital mapping allows creating a robust map.

This method can be used in a different parts of Australia, and it is of special interest to areas with limited access. Moreover, with the results of this study, it is possible to understand soil formation and use this as a covariate in future digital soil mapping studies.

## 2.6. References


ALDANA-JAGUE, E., HECKRATH, G., MACDONALD, A., VAN WESEMAEL, B. & VAN OOST, K. 2016. UAS-based soil carbon mapping using VIS-NIR (480–1000nm) multi-spectral imaging: Potential and limitations. Geoderma, 275, 55-66.

BOETTINGER, J., RAMSEY, R., BODILY, J., COLE, N., KIENAST-BROWN, S., NIELD, S., SAUNDERS, A. & STUM, A. 2008. Landsat spectral data for digital soil mapping. Digital soil mapping with limited data. Springer.

BONFATTI, B. R., DEMATTÊ, J. A. M., MARQUES, K. P. P., POPPIEL, R. R., RIZZO, R., MENDES, W. D. S., SILVERO, N. E. Q. & SAFANELLI, J. L. 2020. Digital mapping of soil parent material in a heterogeneous tropical area. Geomorphology, 367, 107305.

DOBOS, E., SERES, A., VADNAI, P., MICHÉLI, E., FUCHS, M., LÁNG, V., BERTÓTI, R. D. & KOVÁCS, K. 2013. Soil parent material delineation using MODIS and SRTM data. Hungarian Geographical Bulletin, 62, 133-156.

DRORI, R., DAN, H., SPRINTSIN, M. & SHEFFER, E. 2020. Precipitation-Sensitive Dynamic Threshold: A New and Simple Method to Detect and Monitor Forest and Woody Vegetation Cover in Sub-Humid to Arid Areas. Remote Sensing, 12, 1231.

DYMOND, J. R. & LUCKMAN, P. G. 1994. Direct induction of compact rule-based classifiers for resource mapping. International Journal of Geographical Information Systems, 8, 357-367.

ELDEIRY, A. A. & GARCIA, L. A. 2010. Comparison of Ordinary Kriging, Regression Kriging, and Cokriging Techniques to Estimate Soil Salinity Using LANDSAT Images. Journal of Irrigation and Drainage Engineering, 136, 355-364.

GIA PHAM, T., KAPPAS, M., VAN HUYNH, C. & HOANG KHANH NGUYEN, L. 2019. Application of ordinary kriging and regression kriging method for soil properties mapping in hilly region of Central Vietnam. ISPRS International Journal of Geo-Information, 8, 147.

GOMEZ, C., GHOLIZADEH, A., BORŮVKA, L. & LAGACHERIE, P. 2016. Using legacy data for correction of soil surface clay content predicted from VNIR/SWIR hyperspectral airborne images. Geoderma, 276, 84-92.

GRAY, J. M., BISHOP, T. F. A. & WILFORD, J. R. 2016. Lithology and soil relationships for soil modelling and mapping. CATENA, 147, 429-440.

HEUNG, B., BULMER, C. E. & SCHMIDT, M. G. 2014. Predictive soil parent material mapping at a regional-scale: A Random Forest approach. Geoderma, 214-215, 141-154.

JASIEWICZ, J. & STEPINSKI, T. F. 2013. Geomorphons — a pattern recognition approach to classification and mapping of landforms. Geomorphology, 182, 147-156.

KUHN, M., WESTON, S., KEEFER, C. & COULTER, N. 2012. Cubist models for regression. R package Vignette R package version 0.0, 18.

KUHN, M., WESTON, S., KEEFER, C. & KUHN, M. M. 2020. Package 'Cubist'.

LACOSTE, M., LEMERCIER, B. & WALTER, C. 2011. Regional mapping of soil parent material by machine learning based on point data. Geomorphology, 133, 90-99.

LAGACHERIE, P. & HOLMES, S. 1997. Addressing geographical data errors in a classification tree for soil unit prediction. International Journal of Geographical Information Science, 11, 183-198.

LAWLEY, R. 2014. User guide: soil parent material 1 kilometre dataset.

LEGROS, J. & BONNERIC, P. 1979. Modelisation informatique de la repartition des sols dans le Parc Naturel Régional du Pilat. Annales de l'Université de Savoie, 4, 63-68.

LOISEAU, T., CHEN, S., MULDER, V. L., ROMÁN DOBARCO, M., RICHER-DE-FORGES, A. C., LEHMANN, S., BOURENNANE, H., SABY, N. P. A., MARTIN, M. P., VAUDOUR, E., GOMEZ, C., LAGACHERIE, P. & ARROUAYS, D. 2019. Satellite data integration for soil clay content modelling at a national scale. International Journal of Applied Earth Observation and Geoinformation, 82, 101905.

MA, Y., MINASNY, B., MALONE, B. P. & MCBRATNEY, A. B. 2019. Pedology and digital soil mapping (DSM). European Journal of Soil Science, 70, 216-235.

MALONE, B., MINASNY, B. & MCBRATNEY, A. 2009. Mapping continuous soil depth functions in the Edgeroi District, NSW, Australia, using terrain attributes and other environmental factors. Proceedings of geomorphometry, 90-7.

MCBRATNEY, A. B., MENDONÇA SANTOS, M. L. & MINASNY, B. 2003. On digital soil mapping. Geoderma, 117, 3-52.
MCGARRY, D., WARD, W. T., MCBRATNEY, A. B. & SOILS, C. D. O. 1989. Soil Studies in the Lower Namoi Valley: Methods and Data. 1, The Edgeroi Data Set, CSIRO Division of Soils.
MCKENZIE, N. J. & RYAN, P. J. 1999. Spatial prediction of soil properties using environmental correlation. Geoderma, 89, 67-94.
MINTY, B., FRANKLIN, R., MILLIGAN, P., RICHARDSON, M. & WILFORD, J. 2009. The radiometric map of Australia. Exploration Geophysics, 40, 325-333.
MORA-VALLEJO, A., CLAESSENS, L., STOORVOGEL, J. & HEUVELINK, G. B. M. 2008. Small scale digital soil mapping in Southeastern Kenya. CATENA, 76, 44-53.
OLDEMAN, L. R. & VAN ENGELEN, V. W. P. 1993. A world soils and terrain digital database (SOTER) — An improved assessment of land resources. Geoderma, 60, 309-325.
PADARIAN, J., MINASNY, B. & MCBRATNEY, A. B. 2020. Machine learning and soil sciences: A review aided by machine learning tools. Soil, 6, 35-52.
PEBESMA, E. & HEUVELINK, G. 2016. Spatio-temporal interpolation using gstat. RFID Journal, 8, 204-218.
PEBESMA, E. J. 2004. Multivariable geostatistics in S: the gstat package. Computers & geosciences, 30, 683-691.
PRACILIO, G., ADAMS, M. L., SMETTEM, K. R. J. & HARPER, R. J. 2006. Determination of Spatial Distribution Patterns of Clay and Plant Available Potassium Contents in Surface Soils at the Farm Scale using High Resolution Gamma Ray Spectrometry. Plant and Soil, 282, 67-82.
R CORE TEAM 2013. R: A language and environment for statistical computing. Vienna, Austria.
RAYMOND, O. L., LIU, S., GALLAGHER, R., ZHANG, W. & HIGHET, L. M. 2014. Surface Geology of Australia 1:1 million scale dataset 2012 edition [Online]. Geoscience Australia,. Available: https://ecat.ga.gov.au/geonetwork/srv/eng/catalog.search#/metadata/74619 [Accessed 28/04 2020].
ROBERTS, D., WILFORD, J. & GHATTAS, O. 2019. Exposed soil and mineral map of the Australian continent revealing the land at its barest. Nature Communications, 10, 5297.
RYAN, P., MCKENZIE, N., O'CONNELL, D., LOUGHHEAD, A., LEPPERT, P., 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.
TRIANTAFILIS, J., EARL, N. & GIBBS, I. 2016. Digital soil-class mapping across the Edgeroi district using numerical clustering and gamma-ray spectrometry data. Computing Ethics: A Multicultural Approach, 187.
TRIANTAFILIS, J., GIBBS, I. & EARL, N. 2013. Digital soil pattern recognition in the lower Namoi valley using numerical clustering of gamma-ray spectrometry data. Geoderma, 192, 407-421.
WALLIS, G. R. 2019. Narrabri 1:250 000 Geological Sheet SH-55-12, 1st edition, Geological Survey of New South Wales, Sydney [Online]. Available: https://search.geoscience.nsw.gov.au/product/185 [Accessed 01/05 2020].
WARD, W. 1999. Soils and landscapes near Narrabri and Edgeroi, NSW, with data analysis and using fuzzy k-means.
WILFORD, J. 2012. A weathering intensity index for the Australian continent using airborne gamma-ray spectrometry and digital terrain analysis. Geoderma, 183-184, 124-142.
YOUNG, R. W., YOUNG, A. R. M., PRICE, D. M. & WRAY, R. A. L. 2002. Geomorphology of the Namoi alluvial plain, northwestern New South Wales. Australian Journal of Earth Sciences, 49, 509-523.

# Chapter 3. Developing and testing of pedogenons in the lower Namoi valley, NSW, Australia

This chapter is published as: JANG, H. J., DOBARCO, M. R., MINASNY, B., MCBRATNEY, A. & JONES, E. 2022b. Developing and testing of pedogenons in the lower Namoi valley, NSW, Australia. Geoderma, 116182.

## 3.1. Abstract

Mapping soil classes can support the understanding of soil origin and development, subsequently, the soil classes can be used to support monitoring and assessing soil change due to human influence. Pedogenon was proposed as a conceptual soil taxon derived from a set of quantitative state variables representing the soil-forming factors for a given reference time. This study aims to test the pedogenon concept in the Edgeroi region in New South Wales. This paper developed local pedogenons, designed a sampling scheme to capture soil variation under natural conditions and under intensive human activities, and tested the hypothesis that pedogenon is an efficient method of stratifying the landscape to capture soil variation. This study derived the 14 pedogenons by employing layers of soil-forming factors (soil, climate, organism, topography, and parent material and age) using an unsupervised classification technique (k-means clustering). Within each pedogenon, genosoils were identified based on areas with native vegetation, while phenosoils were identified as areas with cropping practises. One meter soil cores were collected for each genosoil and phenosoil, and scanned using Vis-NIR spectrometer for predicting soil properties (clay, sand, cation exchange capacity, pH, and organic carbon). Results show that each pedogenon was characterised by a soil type formed under a dominant parent material occupying a unique position in the landscape. Redundancy discriminant analysis of the soil properties as a function of pedogenon and depth of observations show that pedogenon significantly explained the variation in soil properties. Variance partitioning analysis confirmed that pedogenon explained a large proportion of the variation (49 %) as opposed to landuse (5 %). Principal component analysis of the soil properties shows that genosoils had twice the variation of phenosoils. The results indicate that agricultural activities homogenised the variation of soil profiles. This study demonstrated that pedogenon clasess can effectively characterise soil variation and be used as a benchmark to compare how human activities have altered soil conditions.

## 3.2. Introduction

Accurate spatial soil information is required to address soil security, sustaining soil functions to provide planetary services and human wellbeing. Digital maps of soil classes can help with the understanding of soil genesis and distribution (Ma et al., 2019). Identifying diverse soil classes with soil maps can support monitoring and assessing soil change due to different ways humans have impacted via management practices. In recent years, digital soil mapping (DSM) techniques have been applied to create quantitative soil and environmental classes that reflect soil-forming factors over a sizeable geographical extent (Chen et al., 2022). Digital maps of soil class were commonly created based on current soil observations. These soil observations are products of pedogenesis processes, including anthropogenic effects (Richter Jr, 2007). Román Dobarco et al. (2021b) recently proposed a framework for mapping pedogenon, "a conceptual soil taxon defined from a set of quantitative state variables representing the soil-forming factors for a given reference time". As the framework has not been tested locally, this study aims to test the pedogenon concept in a region in New South Wales. This paper aims to derive local pedogenons and design a sampling scheme to capture soil variation under natural conditions and under intensive human activities. This information allows us to test if pedogenons can capture soil variation effectively and help us understand how human activities have altered the soil profiles.

McBratney et al. (2003) devised the SCORPAN factors as a way to map soil quantitatively, which include soil (s), climate (c), organisms (o), topography (R), parent material (p), age (a), and geographical position (n) factors, and represented as S (soil classes) = f(s, c, o, r, p, a, n). Soil class maps can be created using supervised, unsupervised and knowledge-based approaches (Ma et al., 2019). In supervised mapping, DSM is based on field observations allocated to an established classification system. For example, Heung et al. (2016) created machine learning models to relate soil great groups and orders from a conventional soil survey

observation with a suite of environmental covariates representing the topography, climate, and vegetation. In unsupervised classification, soil classes can be created using soil-forming factors represented as spatial layers using numerical classification techniques. Triantafilis et al. (2013) used gamma radiometric data as a proxy of soil-forming factors (parent material) and created soil classes using fuzzy k-means clustering of the gamma data. As a result, erosional and depositional areas could be better described. Whereas knowledge-based approaches use expert knowledge rules involving few covariates (Zhang et al., 2017).

The advantage of the unsupervised classification approach is that it can be carried out prior to the soil survey. Román Dobarco et al. (2021b) created a soil class map following the state-factor model (Dokuchaev, 1883, Jenny, 1994), in which the soil classes were named 'pedogenons'. They applied an unsupervised classification method to identify the different types of pedogenon in New South Wales (NSW), Australia, using environmental covariates (climate, parent material, topography, and organisms). They demonstrated that pedogenons can be identified using soil-forming spatial layers. It is beneficial not to use an established classification system because the established system is commonly based on soil attributes that may result from either natural processes or human activities (Fitzpatrick, 2013).

Within a pedogenon, Román Dobarco et al. (2021b) also differentiated genosoils and phenosoils (the two terms are from Huang et al. (2018)). These terms are derived from genoform and phenoform, as proposed by Rossiter and Bouma (2018). Genoforms are defined as “soil class as recognised by the soil classification system used as the basis for detailed soil mapping in a given area.” While phenoforms are “continuous, non-cyclical variations of a soil genoform with sufficient chemical or physical differences affect soil function significantly” (Rossiter and Bouma, 2018). Thus, within a pedogenon we can recognise areas that haven’t been severely influenced by human activities as genosoil while phenosoils refer to the different types and levels of landuse and agricultural practises. The understanding of soil variation

within pedogenon can provide insight to better managing agricultural lands, and it is possible to identify which management methods maximise yield and make less variation from genosoil.

The study of Román Dobarco et al. (2021b) was conducted over a large geographical extent and was not validated with field data. They also emphasised the need of determining the optimal number of clusters for classification and selecting meaningful environmental covariates in a local region. Thus this research aims to evaluate the generation of pedogenon maps on a local scale. In addition, we will design a sampling scheme to sample the pedogenons along with their genosoils and phenosoils. Based on this sampling scheme, we collected the samples in the field and characterised soils within each pedogenon. We then investigate the effectiveness of pedogenons in capturing soil profile variation and quantify the difference between genosoil and phenosoil using multivariate statistical analyses.

## 3.3. Methods

### 3.3.1. Description of the study area

The study area is the Edgeroi district in New South Wales (NSW), Australia. The site is approximately 1700 $km^2$ with a climate that is dry, hot in the summer, and cold in the winter. The mean annual precipitation is 712 mm, with significant monthly variations (Ward, 1999). The average maximum temperature in November is 30 °C, and it continues to rise to 40 °C until February. In July, the mean minimum temperature is 3°C. Agriculture has been practised for about 200 years since European settlers came in the early 1800s (Narrabri Shire Council, 2021). Because grazing was the predominant agricultural activity in the past, enormous regions were used as stock routes. Wheat, sorghum, and irrigated cotton farming were introduced in the 1960s (McGarry et al., 1989). Few national parks have remained relatively unaffected by human activities for an extended period. This research area has frequently been used as

exemplar in developing digital soil mapping methodologies (Minasny et al., 2006, Malone et al., 2009, Triantafilis et al., 2016).

### 3.3.2. Environmental covariates data for creating pedogenons

Based on the approach of Román Dobarco et al. (2021b), pedogenons were developed using covariates that characterise the soil-forming factors in the area. The covariates reflect relatively stable variables that would influence the formation of the soils.

#### 3.3.2.1. Soil

The gamma radiometric data were used to represent soil mineral information. The gamma data have implicit information on soil geochemistry, soil properties, and parent material and were used in numerous digital soil mapping studies (Lacoste et al., 2011, Dierke and Werban, 2013, Triantafilis et al., 2013). Here we used four gamma radiometric data (gamma K, gamma Th, Gamma K/Th and total gamma dose) based on aerial survey conducted by Geoscience Australia (Minty et al., 2009).

#### 3.3.2.2. Climate

Rainfall and temperature are the most significant factors to consider because they affect soil properties directly or indirectly. Precipitation impacts soil properties (level of pH and soil salinity, soil texture and structure). Depending on the amount of rainfall, these processes can be varied (Sharpley, 1985). Annual precipitation (mm) and average temperature (Annual mean) (°C) were used to represent climatic factors. In addition, the Prescott Index (PI) representing water balance was utilised. This index was calculated based on long-term average precipitation

(P) and potential evaporation (E) as PI = 0.445P / E0.75 (Gallant and Austin, 2015). Prescott Index was found to be one of the most effective climate factors representing soil formation in Australia (Prescott, 1950).

As seen in Figure 3-1, the mountain ranges on the eastern part of the study area gradually flatten towards the west. Therefore, the spatial climate pattern tends to follow the topography. The average annual temperature, precipitation, and PI are 23 °C, 646 mm and 0.62, respectively.

#### 3.3.2.3. Organisms

Pedogenon represents a conceptual soil taxon derived from a set of quantitative soil-forming factors for a given reference time. Here we define the pedogenon classes at the time of the European settlement in New South Wales (Australia) (1788) (Parliament of New South Wales, 2021). As a proxy for the organisms, we used a pre-European vegetation map in the Namoi Catchment (Eco Logical Australia, 2013).

The pre-European vegetation map only covered half of the study area, and thus we expanded it to the whole area using a machine learning model. Three types of covariates (nine climate, three topography, and one gamma-ray data) were used. The spatial layers were generated at a resolution of 90 m. The mapped area was used as training data and sampled randomly. The total number of sampling points was 29,249 (70 percent of data used for training and 30 percent for testing). The random forest model (Breiman, 2001)was used to predict vegetation classes based on the covariates. After prediction, the final vegetation map (thirteen classes) (Figure 3-1) was spatially filtered to remove the noise.

The vegetation class map was then converted to continuous variables for the clustering requirement. First, the vegetation classes were converted into binary variables (for each class, 0=absence and 1 = presence). The vegetation variables were then combined using the Principal

Component Analysis (PCA). A grid sample of 1931 sample points was selected for PCA. The prcomp function of the MASS package in R was used for PCA (Venables and Ripley, 2002). Areas of water (dams or irrigation channels) were removed before PCA. The water area data were from the State Vegetation Type Map (SVTM) program's NSW Native Vegetation Extent 5m Raster v1.2 (Department of Planning Industry and Environment, 2019).

#### 3.3.2.4. Topography

Three topographical data were derived from the SRTM digital elevation model (DEM). Topographical variables of slope, and topographic wetness index (TWI) were calculated. Studies on mapping soil properties have extensively used DEM and TWI data (Pei et al., 2010, Xue et al., 2018, Filippi et al., 2018b). Slope and soil formation are inextricably linked (Bunting, 1964). There are many studies using DEM and slope as covariates with Digital soil mapping techniques (Carré et al., 2007, Padarian et al., 2019, Vaysse and Lagacherie, 2015). Jones et al. (2021) have found DEM and slope to be powerful predictors for soil variables. The majority of the study area is flat, while the eastern part (one-fourth of the study area) is occupied by mountainous terrain. The DEM is shown in Figure 3-1.

#### 3.3.2.5. Parent material and Age

For the parent material map in this study area, we used the map created from Chapter 2 (Figure 3-1). The map recognised eighteen parent material classes arranged according to the age of formation, which aids in comprehending parent material formation history. The parent material map is categorical data, so the eighteen classes were converted into binary variables (each parent material, 0=absence and 1 = presence). PCA was performed on these binary variables using a regular grid sample of 1960 sample points.

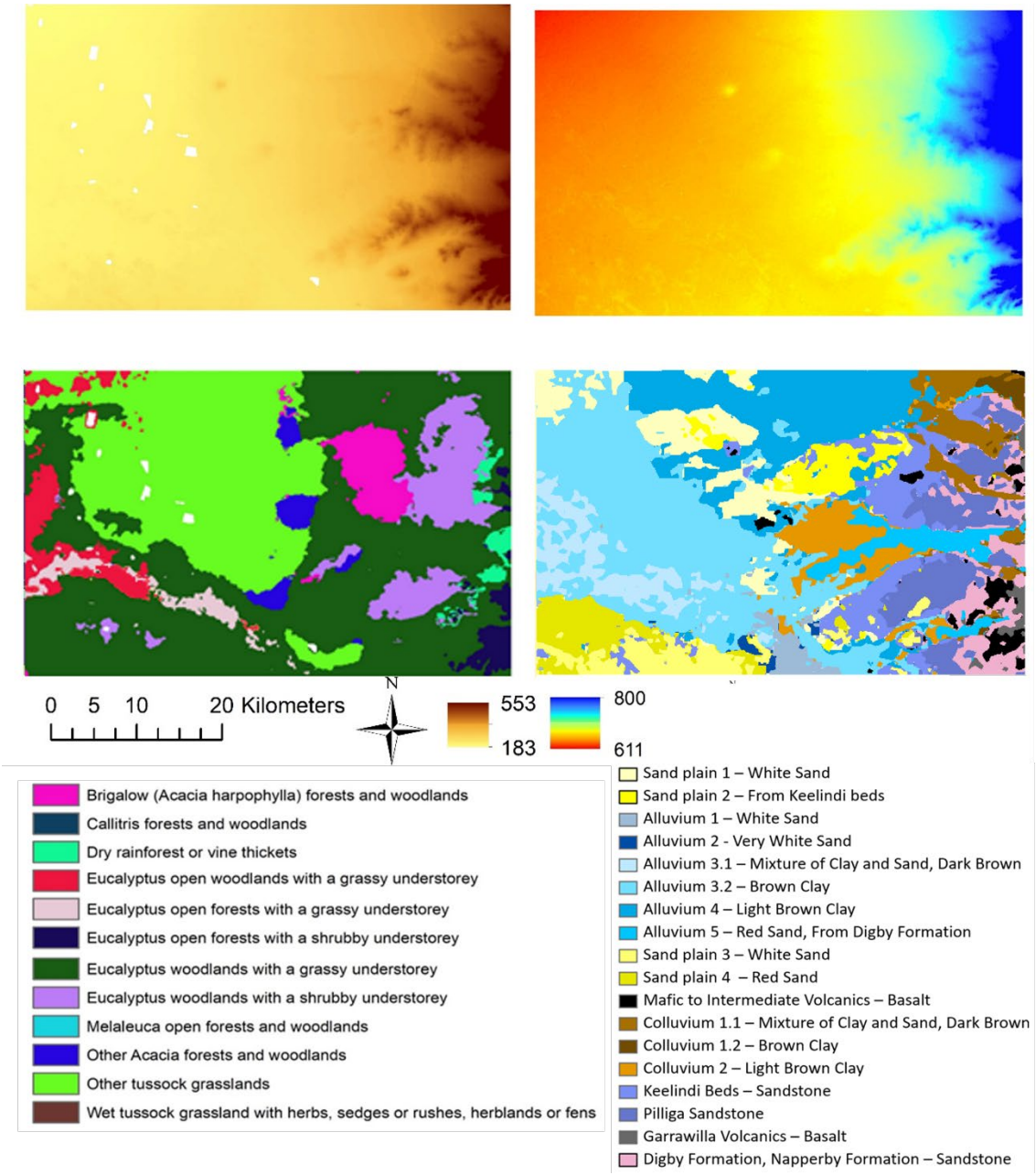



**Figure 3-1. four soil-forming factors – top left: DEM (range: 183–553m), top right: annual precipitation (range: 611–800 mm), bottom left: vegetation map (13 vegetations), and bottom right: parent material map (18 parent materials)**

### 3.3.3. Creation of pedogenons

Soil, climate, organism, topography, and parent material and age layers were used as soil-forming factors in creating pedogenons (Table 3-1). Covariates that were not normally distributed were log-transformed (annual precipitation, Prescott Index, DEM, and Gamma K/Th). All data were then scaled, and water regions were deleted (dams or irrigation channels). The scaled data were then decorrelated using the inverse of the Cholesky transformation (Wicklin, 2012).

The study area was sampled with a regular grid of 90 m, resulting in 237390 pixels. K-means clustering (Lloyd, 1982, MacQueen, 1967) was used to group the data into clusters. The K-means was calculated using Euclidean distance on scaled and decorrelated data. We used the "elbow method" to determine the optimal number of classes using adjusted R-squared and WCSSE (Within cluster sum of squares) measures.

**Table 3-1. Environmental covariates used for the pedogenon classification.**

| Soil-forming factors | Covariate name | Units | Reference |
|---|---|---|---|
| **Soil** | Gamma K | | Wilford and Kroll (2020) |
| | Gamma Th | | Wilford and Kroll (2020) |
| | Gamma K/Th | | Wilford and Kroll (2020) |
| | Gamma dose | | Wilford and Kroll (2020) |
| **Climate** | Mean annual temperature | °C | Williams et al. (2012) |
| | Mean total annual precipitation | mm | Williams et al. (2012) |
| | Prescott Index | | |
| **Organism** | Vegetation<br>(5 PCs, explained 91% variation) | | Eco Logical Australia (2013) |
| **Topography** | Digital elevation model (DEM) | m | Gallant et al. (2009) |
| | Slope | % | Gallant et al. (2009) |
| | Topographic wetness index (TWI) | | Quinn et al. (1991) |
| **Parent material & Age** | Parent material<br>(7 PCs, explained 71% variation) | | Form Chapter 2 |

### 3.3.4. Creation of the landuse map for distinguishing genosoils and phenosoils

With the map of pedogenons, genosoils and phenosoils could be distinguished using landuse information. To create a landuse map for the area, a supervised classification was conducted in the Google Earth Engine. Two hundred fifty polygons (from woodland, pasture and cropping areas) were collected by investigating Landsat images and used as training data. The landuse map was derived using random forest modelling using chlorophyll index (CI) and the soil adjusted vegetation index (SAVI) as predictors. Other detailed types of landuse (irrigated cropping, production forest, water and cropping/modified pasture lands) were gathered from the Catchment Scale Land Use of Australia map (Department of Agriculture Water and the Environment (2019) and added to the map.

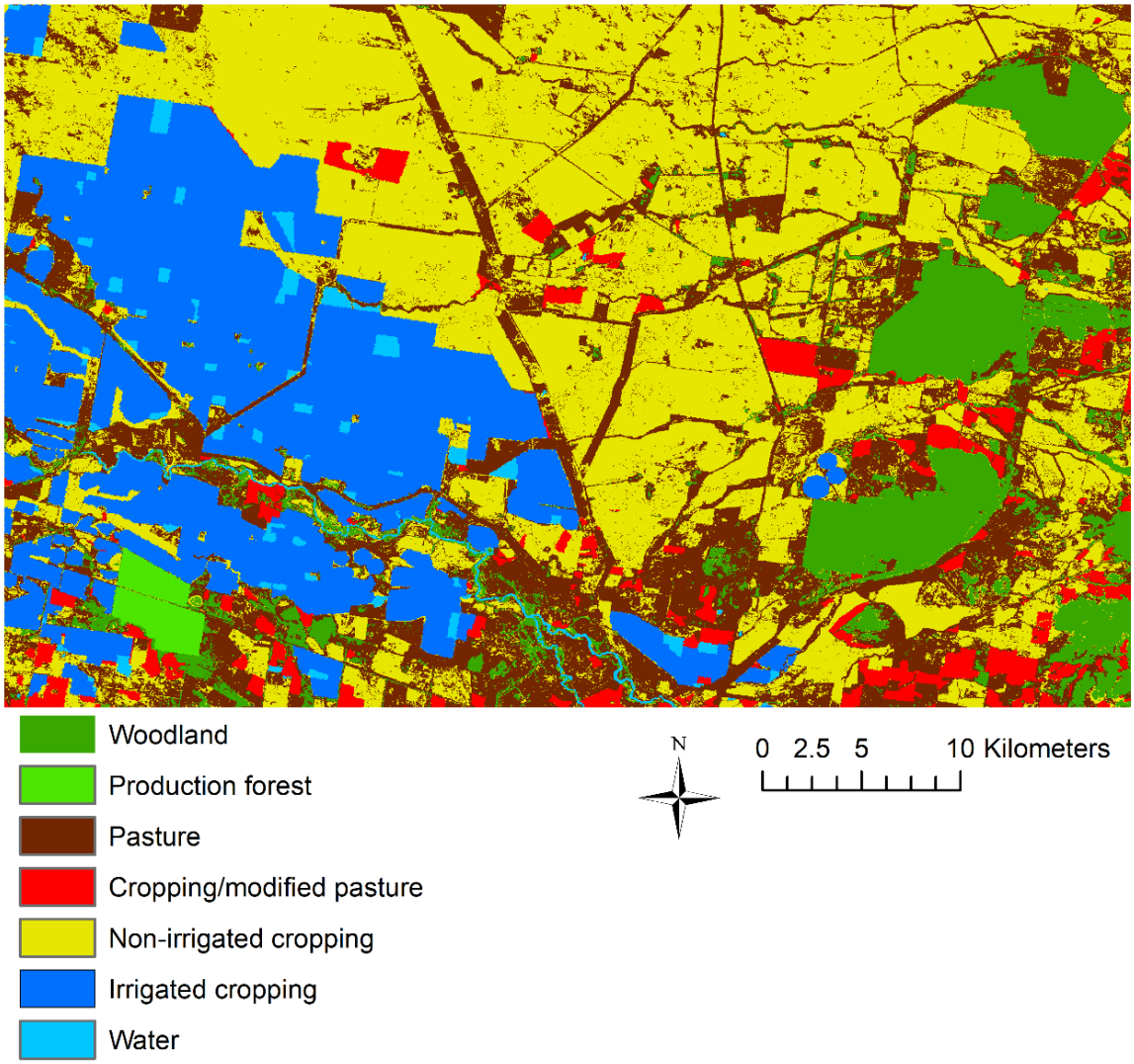



**Figure 3-2. Map of landuse or cover. There are seven broad landuse/cover classes in the area.**

### 3.3.5. Soil sampling design

To validate the pedogenons, we designed a sampling strategy to capture genosoils and phenosoils. Within each pedogenon, we first determined the genosoils (remnant genosoil) by examining large patches of woodland areas. If there was no woodland, broad grazing fields were considered the state between genosoil and phenosoil, which was named phenosoil_1. While some of these grazing fields may have native vegetation and overlap with areas that were considered as grasslands in the pre-European vegetation map, the intensity and frequency of disturbance by human activity differ between the genosoil and phenosoil_1. Compared with other landuses, the pasture area is much less disturbed than non-irrigated or irrigated cropping, but more disturbed than woodlands (genosoil). The main landuse in the pasture areas are stock routes, and thus impacted by the movement of cattle. Three sampling locations were placed on the woodlands or pastures randomly to represent remnant genosoil or phenosoil_1, respectively. After the remnant genosoil or phenosoil_1 sampling point was fixed, other phenosoils or soils under intensive agricultural use were selected, and it was named phenosoil_2. A 1-5 km buffer was placed on the remnant genosoil&phenosoil 1 and phenosoil 2 sampling point was determined randomly within that buffer according to the land use map.

### 3.3.6. Soil sampling and VisNIR spectra acquisition

Sampling locations were determined using a handheld GPS. A drill rig was used to extract two soil cores to 1 m depth at each sampling location. The soil cores were transported back to the laboratory, where they were split vertically and air-dried for 48 hours. Splitting the core produced a relatively flat surface that was free from contamination of falling debris that had settled on the external surface of many cores during the extraction process. Diffuse reflectance VisNIR spectra (350-2500 nm) were acquired on one of the split surfaces of each core using a contact probe attachment connected to an ASD AgriSpec (Malvern Panalytical, Boulder, CO,

USA). Spectra were obtained at 5 cm depth intervals with a wavelength range of 350 to 2,500 nm. Illumination was provided by an inbuilt halogen lamp housed in the contact probe attachment. A baseline reading was taken on a Spectralon (Labsphere Inc., North Sutton, NH, USA) white panel before scanning each core.

### 3.3.7. Estimating soil properties using VisNIR spectra

A regional soil spectral library was used to develop machine learning models to estimate the properties of the soil cores utilising the acquired VisNIR spectra. The regional spectral library is a composite library containing 3,878 samples, covering an area of ~100,000 $km^2$, predominantly from within agricultural regions of New South Wales, Australia. The component datasets used to build the library have previously been utilised to successfully estimate clay, CEC, pH, OC and sand content (Chen et al., 2021, Filippi et al., 2018a, Ng et al., 2022, Tang et al., 2019). Samples in the library were collected from various depths from the soil surface down to 1.2 m depth and represented both agricultural and native/natural land use systems. To develop the spectra predictive models, the VisNIR spectra were first pre-processed as follows: the step in the spectra observed at the 1,000 and 1,800 nm detector junctions was remove using the *splice* function from the "spectacles" R package (Roudier, 2020)); the diffuse reflectance spectra were then converted to absorbance, A= log(1/R); the spectra were smoothed using a Savitzky−Golay filter with a window size of 11 and fitting a second order polynomial; wavelengths with a low signal-to-noise ratio were removed to leave only the 500−2450 nm wavelength range; a standard normal variate baseline correction was performed; the spectra were resampled at 10 nm intervals to reduce the overall data size; and finally principal component analysis (PCA) was performed to further reduce the dimensionality of the data, with the minimum number of principal components that accounted for more than 95% of the

observed variability retained. The principal component scores were then used as input variables to develop Cubist models to estimate clay, CEC, pH, OC and sand content (Kuhn et al., 2012).

Pre-processing spectra from the soil cores proceeded in a similar fashion with the exception that the core spectra were projected into the principal component space developed from the spectral library. A bootstrapping methodology utilising 100 iterations was used to understand the prediction uncertainty, with final estimates representing the mean of the 100 iterations. Calibration and 10-fold cross-validation statistics demonstrated that the regional spectral library was able to estimate soil properties with a good level of accuracy (cross-validated $R^2$ from 0.6-0.8, Table 3-2). The developed Cubist models were applied to all spectra from the soil cores. After prediction, estimates were averaged into 10 cm depth intervals, and data from the two cores were averaged into one profile.

**Table 3-2. 10-fold cross validation statistics for the estimation of clay (%), CEC (mmol/kg), pH, OC (%) and sand (%) from Cubist models developed utilising VisNIR spectral data.**

| | | $R^2$ | concordance | RMSE | bias | RPIQ |
|---|---|---|---|---|---|---|
| **Clay** | calibrated | 0.85 | 0.92 | 5.66 | 0.06 | 1.50 |
| | 10-fold CV | 0.82 | 0.90 | 6.11 | 0.09 | 1.39 |
| **CEC** | calibrated | 0.77 | 0.85 | 64.21 | -5.58 | 1.23 |
| | 10-fold CV | 0.68 | 0.80 | 73.86 | -5.44 | 1.06 |
| **pH** | calibrated | 0.72 | 0.82 | 0.51 | 0.04 | 1.36 |
| | 10-fold CV | 0.64 | 0.77 | 0.58 | 0.04 | 1.20 |
| **OC** | calibrated | 0.69 | 0.80 | 0.29 | -0.03 | 0.74 |
| | 10-fold CV | 0.60 | 0.74 | 0.32 | | 0.66 |
| **Sand** | calibrated | 0.83 | 0.90 | 7.10 | -0.29 | 1.18 |
| | 10-fold CV | 0.77 | 0.87 | 8.31 | -0.33 | 1.00 |

### 3.3.8. Statistical analysis of genosoil and phenosoil

The five soil properties (clay, CEC, pH, OC and sand) predicted using NIR data were used for statistical analyses. To investigate the patterns of soil properties variation and the amount of variation explained by pedogenon classes and landuse, we utilised principal component analysis (PCA), and redundancy analysis (RDA).

PCA is a statistical technique for investigating the interrelationships among a set of variables and determining their underlying structure (Malone et al., 2021). PCA can be a valuable tool for soil researchers investigating to link the state of soil function, soil properties and agroecosystem management (Islam et al., 2003, Zeraatpisheh et al., 2020). The difference between PCA and RDA is that PCA requires only response variables, but both response and explanatory variables were needed for RDA (Zuur et al., 2007). This study used soil properties as response variables, whereas pedogenon class and depth were used as explanatory variables for RDA. The exploratory variables consisted pedogenon classes and 10 depth intervals as factor levels. In addition, we performed a variance partitioning of the RDA model using the adjusted $R^2$ to differentiate the pure effects of pedogenon class, landuse, and their shared effect on the variation of soil properties (Borcard et al., 1992, Legendre and Legendre, 2012).

## 3.4. Results

### 3.4.1. Deriving pedogenons for the Edgeroi area

#### 3.4.1.1. Obtaining an optimal number of pedogenon classes

Román Dobarco et al. (2021b) mentioned that finding an optimal number of classes is essential for creating the pedogenon map. This part attempted to find the optimal number and improve the classification result. The number of pedogenon classes from clustering depends on numerical and knowledge-based criteria. One issue that arose during the clustering process was

that one of the classes included both flat and mountainous areas. As a result, several strategies were trialled to optimise resulting pedogenons that separated mountain and flat areas.

The first strategy was increasing the number of pixels for clustering. We initially used 20,000 pixels for clustering and found that it gave unsatisfactory results. Thus, we trialled and increased the number to 230,000 pixels, where a clear separation of pedogenons can be found. To obtain an optimal number of clusters, the “elbow method” was used. The elbow method (Adjusted R-squared and WCSSE (Within cluster sum of squares)) suggested that the optimal number of clusters is between nine and seventeen (Figure 3-3). However, these two clusters show one pedogenon located in both flat and mountainous areas. Hence, we qualitatively examined pedogenon classes that can separate flat and mountainous areas from nine to seventeen. Thus, we determine fourteen classes as an optimum number.

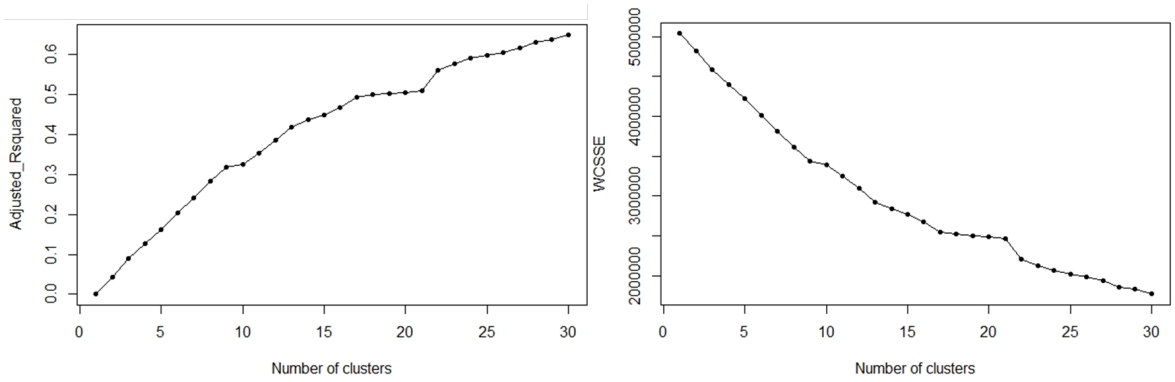



**Figure 3-3. k-means plot of the adjusted R-squared (left) and the within-cluster sum of squares (WCSSE) (right) created for each cluster number. These two graphs recommend that the optimal numbers are nine and seventeen.**

### 3.4.2. Pedogenon and pedophenon classes

The resulting pedogenon map is in Figure 3-4. Based on the pedogenon and landuse maps, we designed soil sampling that captured remnant genosoils and phenosoils (phenosoil_1 and phenosoil_2). Only 13 pedogenons were sampled because one class was located in the mountain and occupied a small percentage of the area (approximately 1%). Moreover, the map of pedophenon (Figure 3-5) shows no phenosoil 2 in pedogenon N. Román Dobarco et al. (2021a) proposed the definition of pedophenons, which are subclasses of pedogenon, created with the combination of pedogenon, vegetation and landuse information. In this study, the pedophenon map (Figure 3-5) was created using pedogenon and landuse maps together. This map clearly shows the distribution of different landuses (remnant genosoil, phenosoil 1, and 2) in each pedogenon. The sample locations are given in Figure 3-4 and Figure 3-5.

The general pattern of the pedogenon map is similar to the map of the parent material (Figure 3-4). Generally, soil properties were determined by the characteristics of the parent materials. Hence, it is reasonable that both pedogenon and parent material maps were similar. Ward (1999) also stated that soil variation on the local scale is mainly associated with the parent rock and sediment types. Most of the areas were covered by alluvium material on the western part. This area is flat, and the pre-European vegetation was Tussock grasslands and eucalyptus woodlands (Figure 3-1). At the same time, pedogenons in the eastern part of the study area are heavily influenced by topography and climate.

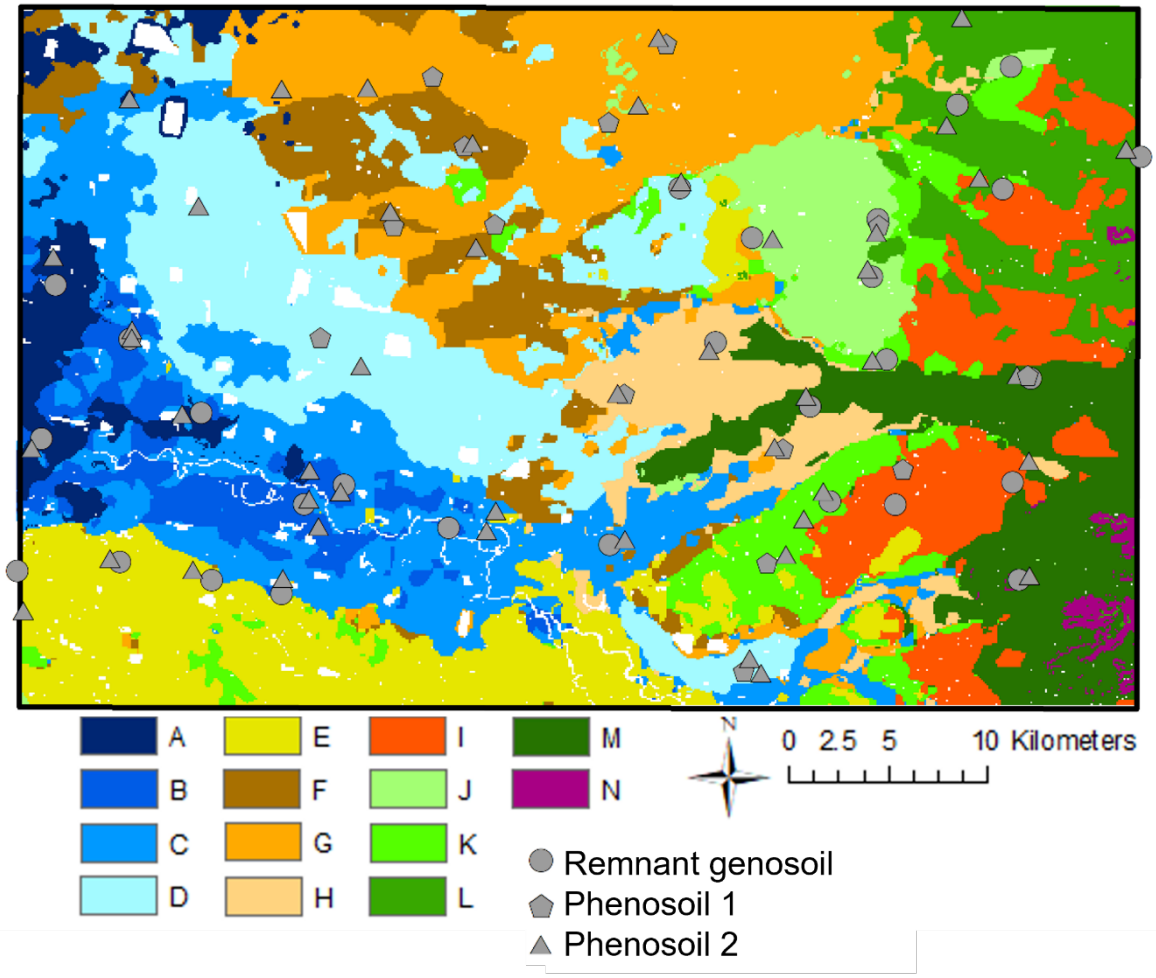



**Figure 3-4. Fourteen different pedogenons from clustering and sampling points remnant genosoil, phenosoil 1 and 2. Soil-forming factors were used for generating the pedogenon maps. There are a total of 89 sample points (28 remnant genosoil points, 14 phenosoil 1 and 47 phenosoil 2 points).**

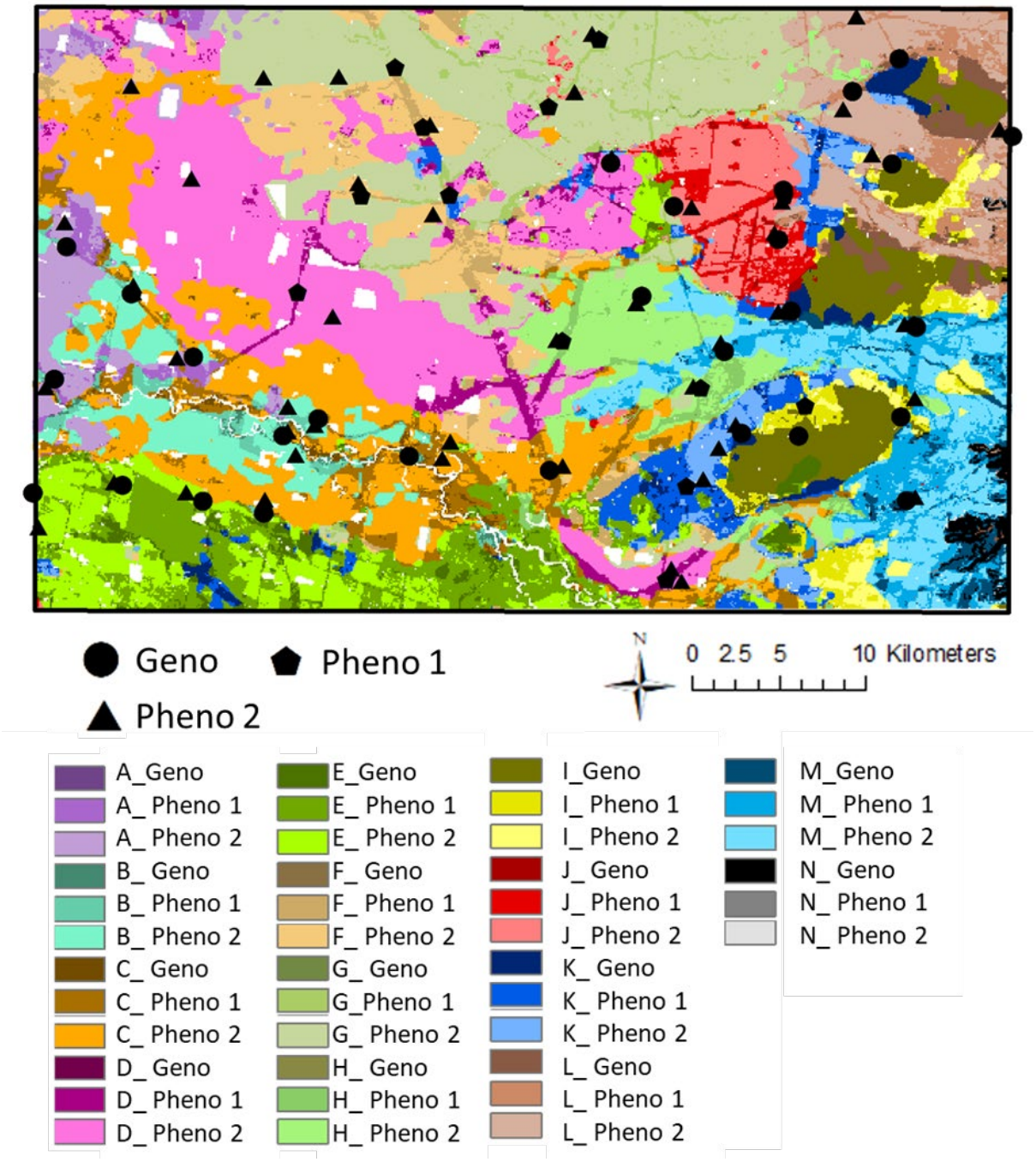



**Figure 3-5. Forty-two different pedophenons from the combination of pedogenon and landuse, and sampling points of genosoil (geno) and phenosoil (pheno 1 and 2). There are a total of 89 sample points (28 remnant genosoil points, 14 phenosoil and 47 phenosoil 2 points).**

#### 3.4.2.1. Description of the pedogenons

Based on the soil description and soil-forming factors, we describe the characteristics of each pedogenon as follows, including dominant types of parent material and vegetation, the topography, phenosoil landuse, genosoil landuse, soil texture, colour and corresponding soil class (Australian Soil Classification, ASC, Isbell (2016)). The detail of each pedogenon is given in Appendix 3-1 and the soil properties of each pedogenon is also shown in Appendix 3-2.

**Pedogenon A** – The dominant parent material is alluvium, and the current vegetation is native grasslands and forested wetlands. The terrain is flat. The phenosoil landuse is irrigated cropping, and the genosoil landuse is pasture. The soil texture is heavy clay, and the soil colour is black. The ASC soil class is a lack Vertosol.

**Pedogenon B** – The dominant parent material is alluvium, and the current vegetation is native grasslands. The terrain is flat. The phenosoil landuse is irrigated cropping, and the genosoil landuse is woodland. The soil texture is medium clay, and the soil colour is dark brown. The ASC soil class is a Dark Brown Vertosol.

**Pedogenon C** – The dominant parent material is alluvium, and the current vegetation is native grasslands. The terrain is flat. The phenosoil landuse is non-irrigated cropping and irrigated cropping, and the genosoil landuse is woodland and pasture. The soil texture is heavy clay, and the soil colour is dark brown. The ASC soil class is a Dark Brown Vertosol.

**Pedogenon D** – The dominant parent material is alluvium, and the current vegetation is native grasslands. The terrain is flat. The phenosoil landuse is irrigated cropping, and the genosoil landuse is pasture. The soil texture is heavy clay, and the soil colour is black. The ASC soil class is a Black Vertosol.

**Pedogenon E** – The dominant parent material is sand plain, and the current vegetation are native grasslands and dry sclerophyll forests (Shrub/grass sub-formation or shrub sub-formation). The terrain is flat. The phenosoil landuse is irrigated cropping, and the genosoil landuse is woodland. The soil texture is sandy clay loam, and the soil colour is red. The ASC soil class is a Red Chromosol.

**Pedogenon F** – The dominant parent material is sand plain, and the current vegetation is native grasslands. The terrain is flat. The phenosoil landuse is non-irrigated cropping, and the genosoil landuse is pasture. The soil texture is heavy clay, and the soil colour is black. The ASC soil class is a Black Vertosol.

**Pedogenon G** – The dominant parent material is alluvium, and the current vegetation is native grasslands. The terrain is flat. The phenosoil landuse is non-irrigated cropping, and the genosoil landuse is pasture. The soil texture is heavy clay, and the soil colour is black. The ASC soil class is a Black Vertosol.

**Pedogenon H** – The dominant parent material is colluvium, and the current vegetation is native grasslands. The terrain is flat. The phenosoil landuse is non-irrigated cropping, and the genosoil landuse is pasture. The soil texture is heavy clay, and the soil colour is dark brown. The ASC soil class is a Dark Brown Vertosol.

**Pedogenon I** – The dominant parent material is Pilliga Sandstone, and the current vegetation is dry sclerophyll forests (Shrub/grass sub-formation or shrub sub-formation). The topography is characterised by gentle slopes. The phenosoil landuse is non-irrigated cropping, and the genosoil landuse is woodland. The soil texture is sandy clay loam, and the soil colour is dark brown. The ASC soil class is a Dark Brown Tenosol.

**Pedogenon J** – The dominant parent material is Keelindi beds and sand plain, and the current vegetation are native grasslands and western vine thickets. The topography is characterised by

gentle slopes. The phenosoil landuse is non-irrigated cropping, and the genosoil landuse is pasture. The soil texture is medium-heavy clay, and the soil colour is black. The ASC soil class is a Black Vertosol.

**Pedogenon K** – The dominant parent material is Keelindi beds, and the current vegetation is consists of native grasslands and dry sclerophyll forests (Shrub/grass sub-formation). The topography is characterised by gentle slopes. The phenosoil landuse is non-irrigated cropping and irrigated cropping, and the genosoil landuse is woodland and pasture. The soil texture is medium clay, and the soil colour is brown. The ASC soil class is a Brown Vertosol.

**Pedogenon L** – The dominant parent material is colluvium, and the current vegetation is native grasslands and dry sclerophyll forests (Shrub/grass sub-formation). The topography is characterised by gentle slopes. The phenosoil landuse is non-irrigated cropping, and the genosoil landuse is woodland. The soil texture is heavy clay, and the soil colour is black. The ASC soil class is a Black Vertosol.

**Pedogenon M** – The dominant parent material is Digby formation, Napperby formation – sandstone and alluvium, and the current vegetation are native grasslands and western vine thickets. The topography is characterised by gentle slopes. The phenosoil landuse is non-irrigated cropping, and the genosoil landuse is woodland and pasture. The soil texture is light-medium clay, and the soil colour is dark brown. The ASC soil class is a Dark Brown Vertosol.

### 3.4.3. Comparing pedogenons with existing soil maps

The pedogenon map was compared with previous studies conducted in the area. Ward (1999) created a geology map of the Edgeroi area using soil survey and fuzzy K-means clustering. Their study area was slightly smaller than ours, but the pattern of their result is similar to our pedogenon map (especially pedogenon F and K). Triantafilis et al. (2013) used gamma-ray

spectrometry data and coupled it with the fuzzy K-means algorithm to identify geological and geomorphological units of the area. They identified 11 classes, and some of their results are similar to some of our pedogenons. Soils developed over alluvium show a similar pattern. Some of their alluvial soils classes corresponded to pedogenon D and G. While one of their class can be further differentiated in pedogenons A, B, C and H. Overall, the qualitative comparison between soil classes generated by Triantafilis et al. (2013) and pedogenons confirmed that both maps had similarities, even though the two maps were created differently.

### 3.4.4. Statistical analysis of soil properties of pedogenons.

Soil properties derived from NIR spectra were analysed using multivariate analysis, i.e. PCA and RDA to reveal the pattern of pedogenons, genosoil and phenosoil.

#### 3.4.4.1. Principal component analysis

PCA can be a valuable tool for soil researchers investigating to link the state of soil properties and soil classes. Here we combined 5 soil properties which were measured every 10 cm from the soil surface to 1 m depth. The first two PCs explained 89.1% of data variance, PC1 explained 67.8%, and PC2 explained 21.3% of data variance. Figure 3-6 shows that the PC1 was controlled mainly by clay, CEC, pH and sand, while PC2 was affected by OC. Clay, pH and CEC have significant positive loadings on component 1, while sand has large negative loadings on component 1. On the other hand, OC has negative loadings on component 2. Figure 3-6 A shows that most data points of topsoil (0-10 and 10-20 cm) were very close to OC axes, while subsoil data were away from the OC axes. This confirmed the observation that OC was concentrated in the top layers of soil, and its content decreased with depth.

The ellipses of the PCA represent the 95% confidence interval of the data. Figure 3-6 B shows the PCA scores of soil properties grouped by pedogenons where we can see two distinct groups. Most pedogenons were concentrated on the right side of the graph, which indicated high clay content, pH, and CEC values. While the other pedogenons (Pedogenon E, I and K) were located on the left side of the PC plot, indicated by high sand content.

The PC plot indicated that the ellipse representing 95% of remnant genosoil and phenosoil 1 data was larger than that of phenosoil 2 (Figure 3-6 C). The ellipse area of remnant genosoil and phenosoil 1 is twice the size of phenosoil 2. Remnant genosoil and phenosoil (phenosoil 1 and 2) of a pedogenon were initially assumed to be formed by the same pedogenetic processes. However, since human cultivation, phenosoil_1 and phenosoil_2 were affected by agropedogenesis (Kuzyakov and Zamanian, 2019), resulting in different variability. Figure 3-6 D shows the variation of soil properties as a function of land use. It also shows that the variation of soils in woodland, represented by the 95% confidence ellipse, was larger than pasture soils, followed by non-irrigated cropping soils, and finally, irrigated soils.

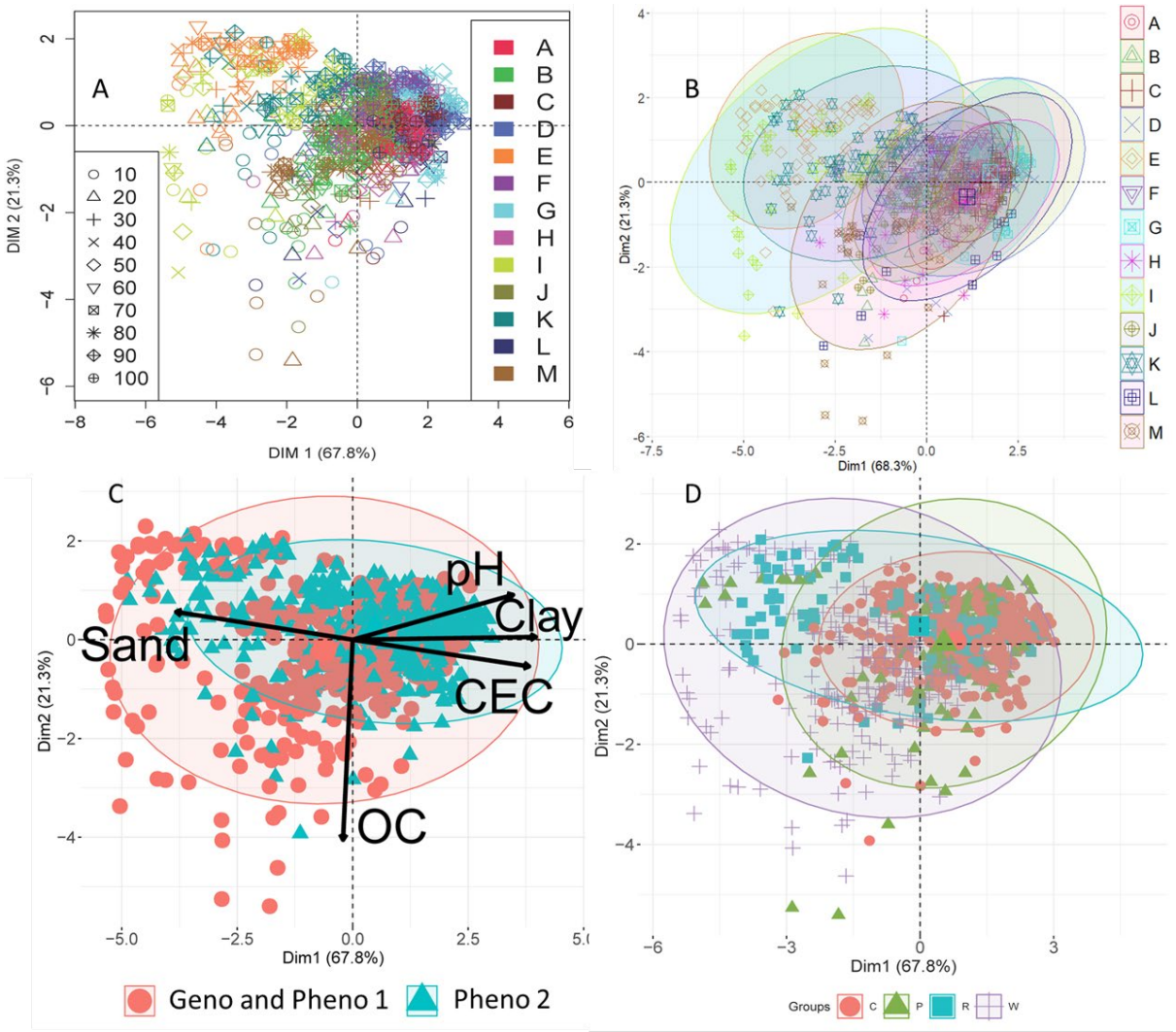



**Figure 3-6. a) Scores plot of the first two PC using 89 cores (every 10 cm) from all thirteen pedogenons. The different shapes represent the depth in every 10cm (10-100cm), and the colour of points represents each pedogenon (Pedogenon A-M). b) Scores plot of the first two PC using 89 cores (every 10 cm) from all thirteen pedogenons with the ellipse. c) Biplot of the first two PC using remnant genosoil (Geno) and phenosoil 1 (Pheno 1) (red circle) and phenosoil 2 (Pheno 2) (blue triangle) with 95 % prediction ellipse. The black arrows are the active variables (Clay, CEC, pH, OC and Sand) used to build the PCA. d) Scores plot of four types of landuse (C: Nonirrigating Cropping, P: Pasture, R: Irrigating cropping and W: Woodland) with the ellipse.**

#### 3.4.4.2. Redundancy discriminant analysis

To understand the characteristics of each pedogenon, RDA is applied to investigate the relationship between each pedogenon and soil properties and the amount of variation explained by pedogenons. The first two axes of RDA explained 71.8% of the total variance, leaving 28.2% unexplained. The model and axes of individual RDA were statistically significant (P-value: 0.001). The first and second axes of the RDA explained 57.4 % and 14.4 % of the variation, respectively. Based on the direction of the arrows in the RDA biplot, CEC, clay, pH, and sand explained a substantial amount of the variance in RDA 1, whereas OC explained the variance of RDA 2 (Figure 3-7 III). The length of the arrow indicated the strength of correlation between the soil properties and sampling points of pedogenons. The points of each pedogenon represented every 10 cm of observation, with the most surface points indicated by a small arrow and the name of pedogenon. Based on Figure 3-7, most pedogenons started (from 0-10 cm) from the top left of the graph to the bottom right. In addition to that, the OC axes explained this trend indicating that OC content decreased with depth across all pedogenons.

Figure 3-7 shows the difference in soil properties between genosoil&phenosoil 1 and phenosoil 2. The patterns of soil texture and other properties (pH and CEC) from both genosoil&phenosoil 1 and phenosoil 2 (Figure 3-7 I and II) were similar. However, the OC content shows a distinct pattern between remnant genosoil&phenosoil 1 and phenosoil 2. The topsoil OC of remnant genosoil&phenosoil 1 was larger and gradually decreased as the depth increased. However, OC contents in phenosoil 2 were similar throughout the profile. RDA of phenosoil 2 profiles tended to cluster together as opposed to genosoil&phenosoil 1.

The RDA and PCA plots show a similar pattern, where we can distinguish two main groups of soils. The first group occupied on the bottom right-hand side of the RDA plot (Figure 3-7), where most observations were correlated with clay, pH, and CEC. The second group was formed by pedogenon E, I and K, which were correlated with sand content. The soil-forming

factors of pedogenon I and K were similar and were located nearby in the pedogenon map (Figure 3-4). This means that their soil formation is influenced by similar climate, topography, and parent material. Both pedogenons were derived from sandstone, on a gentle slope. The range of soil properties from both pedogenons was also similar. In contrast, pedogenon E was located in the southwest part of the study area (Figure 3-4). The parent material of this pedogenon was a sand plain, and the area is very flat. The soil type is a Chromosol (soil with a texture contrast between topsoil and subsoil). This was reflected in the RDA, as depth increased, and the observations were moving towards the clay axes (Figure 3-7). Overall the sand content in pedogenon E is less than pedogenon I and K.

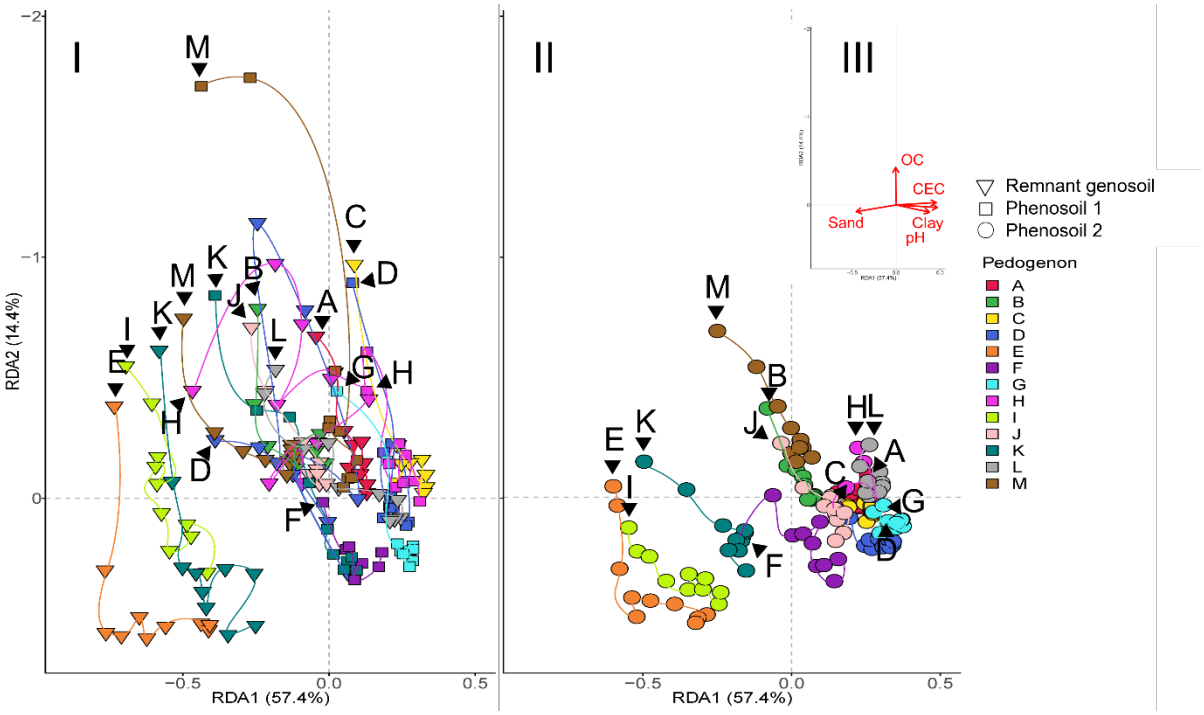



**Figure 3-7. Graph I (remnant genosoil and phenosoil 1) and II (phenosoil 2) are the RDA plot with pedogenon and soil depth as explanatory variables. The square or triangle points(Graph I) and circle(Graph II) with black arrows and the name of pedogenon are 0-10 cm depth, and after that point, every 10cm depth is increased until 100 cm. Graph III is the plot with five soil properties as response variables (red arrows: CEC, clay, pH, OC and sand).**

#### 3.4.4.3. Variance partitioning

The variance partitioning method determined the effect of pedogenon, landuse, and their shared effect on RDA. The pure pedogenon effect was 49 % of the variation, and the pure landuse effect was 5 %. The shared amount of variation was 20 %, and the residual was 25 % (Figure 3-8). The results again stressed the success of pedogenons in explaining the variation of soil properties.

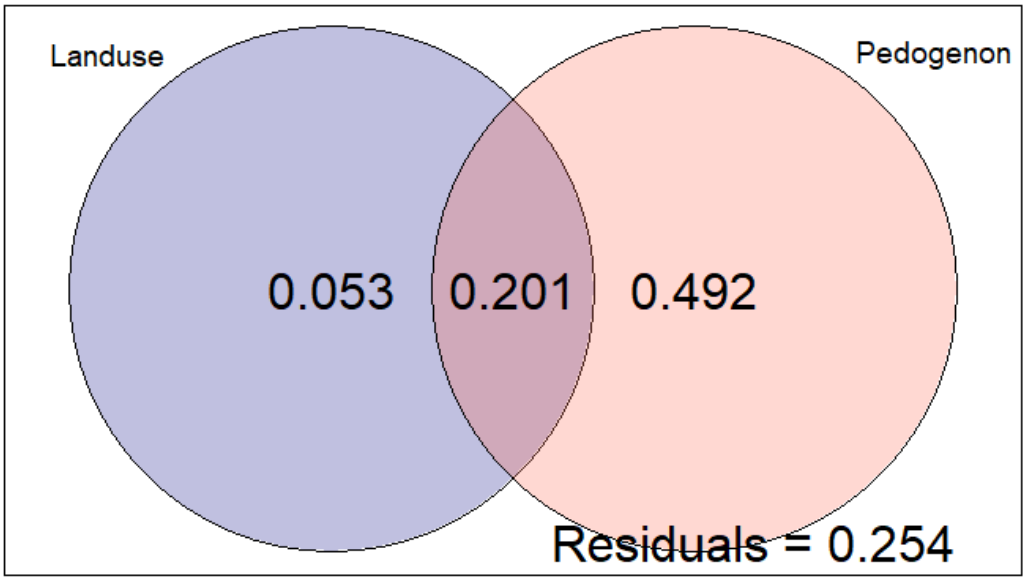



**Figure 3-8. Venn diagram of variance partitioning. The explanatory variables are pedogenon and landuse, and the numbers in the figure represent the explained variance. The percentage of variation can be calculated by dividing the explained variance by total variance (1).**

## 3.5. Discussion

Román Dobarco et al. (2021b) identified the different types of pedogenon in NSW with an unsupervised classification method. The benefit of this approach is that the pedogenon map can be generated without the reliance on legacy soil maps, such as done by Roudier et al. (2022). However, due to a large geographical extent, the validation with field data is limited. Therefore, this study created a local pedogenon map using the framework proposed by Román Dobarco et al. (2021b) and validated it. The resulting map was validated using PCA and RDA with five soil properties (clay, CEC, OC, pH and sand) from surface to 1 m.

The framework using unsupervised classification for creating pedogenons based on local data is straightforward. Unsupervised classification has been used for creating soil classes for different applications, such as Triantafilis et al. (2013), who focused just on using gamma radiometric data. While Roell et al. (2020) used soil, climate and landscape variables to model terron classes. In this study, we used layers that represent the *scorpan* factors. Applying the methodology at a local scale allows incorporating more detailed information on soil geology and landform.

Moreover, the produced pedogenons were more compact for the study area than those from the classification at the state scale (Román Dobarco et al., 2021b). However, determining the number of covariates and an optimal number of pedogenon classes cannot be fully automated yet. Manual interpretation of the resulting classes was still required to ensure the pedogenon classes were meaningful. The ambiguity or sensitivity of some classes may be caused by the less detailed covariates used in this study (at a resolution of 90 m). Future studies should use finer resolution covariates.

The pedogenons also serve as a local soil map equivalent to a soil series. The qualitative description of the pedogenons shows that each pedogenon was derived by unique parent

material and occupied certain parts of the landscapes and the resulting soil profiles are unique. This is further supported by quantitative soil information. The results from redundancy discriminant analysis (RDA) proved that the pedogenons effectively explained the variation of soil properties from surface to 1 m. Based on variation partitioning analysis, the effect of pedogenon is much more significant than landuse (49% compared to 5%). This result explains that soil types described by each pedogenon were unique, and landuse explained the variation within each pedogenon.

The variability of soil properties of the phenosoil_2 was reduced to half of the genosoil and phenosoil_1. The results indicated that natural pedogenesis created soils with high variability, and decades of agricultural practices have modified the soil properties to become less variable. Intensive human activities have changed the characteristics of the genosoils into phenosoils. The processes in phenosoils tend to make soil homogenised. This homogenisation can also be visualised in RDA, where the variation in OC tends to be less in phenosoils compared to genosoils. This result agrees with the hypothesis presented by (Kuzyakov and Zamanian, 2019).

Although the present results clearly support the uniqueness of pedogenons, it is appropriate to recognise several potential limitations. First, we assume that the topography and gamma radiometric data do not change significantly over the past 200 years. In addition, the pre-European vegetation maps were modelled. Nevertheless, the resulting pedogenons generally agree with past studies that were based on parent material classification (Ward, 1999). This study used detailed soil information based on Vis-NIR spectra that were collected every 10 cm of the soil cores. While the soil properties predicted by Vis-NIR are subjected to uncertainty, many studies have shown the effectiveness of soil specta in inferring soil properties. In addition, this method is faster and less laborious than conventional laboratory methods (Jones and McBratney, 2016). Still, the pattern of predicted soil properties using NIR spectra is interpretable for differentiating pedogenons.

In terms of future research, it would be useful to investigate in more detail the difference in soil properties between genosoil and phenosoil to understand how the soils have changed due to intensive cultivation. These data would have some potential intervention implications. For example, the genosoils would also provide a benchmark against the phenosoils on the achievable level of soil conditions by changing management practices.

## 3.6. Conclusion

Pedogenon map is produced based on soil-forming factors (soil, climate, organism, topography, and parent material) informing genosoils. In addition, the information of landuse within each pedogenon supports the distribution of phenosoils. Multivariate statistical analyses (PCA, RDA and Variance partitioning) of soil profile properties of pedogenons validated the pedogenon map. Soils of each pedogenon are unique and the pedogenons characterise the pedodiversity of the area. Soil charactersitics derived from genosoils and phenosoils indicate that agropedogenesis have modified soil profiles. The variation of soil profile properties of genosoil was twice larger than phenosoil. This supports the idea that agropedogenesis reduced the variation of natural pedogenesis. The application of pedogenon map as spatial soil information can support addressing soil security, sustaining soil functions to provide planetary service and human wellbeing.

## 3.7. References


BORCARD, D., LEGENDRE, P. & DRAPEAU, P. 1992. Partialling out the spatial component of ecological variation. *Ecology,* 73**,** 1045-1055.

BREIMAN, L. 2001. Random forests. *Machine learning,* 45**,** 5-32.

BUNTING, B. T. 1964. Slope Development and Soil Formation on Some British Sandstones. *The Geographical Journal,* 130**,** 73-79.

CARRÉ, F., MCBRATNEY, A. B. & MINASNY, B. 2007. Estimation and potential improvement of the quality of legacy soil samples for digital soil mapping. *Geoderma,* 141**,** 1-14.

CHEN, S., ARROUAYS, D., LEATITIA MULDER, V., POGGIO, L., MINASNY, B., ROUDIER, P., LIBOHOVA, Z., LAGACHERIE, P., SHI, Z., HANNAM, J., MEERSMANS, J., RICHER-DE-FORGES, A. C. & WALTER, C. 2022. Digital mapping of GlobalSoilMap soil properties at a broad scale: A review. *Geoderma,* 409**,** 115567.

CHEN, Y., GAO, S., JONES, E. J. & SINGH, B. 2021. Prediction of Soil Clay Content and Cation Exchange Capacity Using Visible Near-Infrared Spectroscopy, Portable X-ray Fluorescence, and X-ray Diffraction Techniques. *Environmental Science & Technology,* 55**,** 4629-4637.

DEPARTMENT OF AGRICULTURE WATER AND THE ENVIRONMENT 2019. Catchment scale land use of Australia - Update December 2018.

DEPARTMENT OF PLANNING INDUSTRY AND ENVIRONMENT 2019. NSW Native Vegetation Extent 5m Raster v1.2.

DIERKE, C. & WERBAN, U. 2013. Relationships between gamma-ray data and soil properties at an agricultural test site. *Geoderma,* 199**,** 90-98.

DOKUCHAEV, V. 1883. Russian Chernozem. Selected Works of VV Dokuchaev. Volume I (Translated in 1967). Israel Program for Scientific Translations. *US Department of Agriculture, Washington DC*.

ECO LOGICAL AUSTRALIA 2013. Refinement of vegetation mapping in the Namoi Catchment: Extant and pre-European. Armadale NSW.

FILIPPI, P., CATTLE, S. R., BISHOP, T. F. A., JONES, E. J. & MINASNY, B. 2018a. Combining ancillary soil data with VisNIR spectra to improve predictions of organic and inorganic carbon content of soils. *MethodsX,* 5**,** 551-560.

FILIPPI, P., CATTLE, S. R., BISHOP, T. F. A., ODEH, I. O. A. & PRINGLE, M. J. 2018b. Digital soil monitoring of top- and sub-soil pH with bivariate linear mixed models. *Geoderma,* 322**,** 149-162.

FITZPATRICK, R. 2013. Demands on soil classification and soil survey strategies: special-purpose soil classification systems for local practical use. *Developments in soil classification, land use planning and policy implications.* Springer.

GALLANT, J., WILSON, N., PK, T., DOWLING, T. & READ, A. 2009. 3 second SRTM Derived Digital Elevation Model (DEM) Version 1.0. Canberra: Geoscience Australia.

GALLANT, J. C. & AUSTIN, J. M. 2015. Derivation of terrain covariates for digital soil mapping in Australia. *Soil Research,* 53**,** 895-906.

HEUNG, B., HO, H. C., ZHANG, J., KNUDBY, A., BULMER, C. E. & SCHMIDT, M. G. 2016. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma,* 265**,** 62-77.

HUANG, J., MCBRATNEY, A. B., MALONE, B. P. & FIELD, D. J. 2018. Mapping the transition from pre-European settlement to contemporary soil conditions in the Lower Hunter Valley, Australia. *Geoderma,* 329**,** 27-42.

ISBELL, R. 2016. *The Australian soil classification*, CSIRO publishing.

ISLAM, K., SINGH, B. & MCBRATNEY, A. 2003. Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared reflectance spectroscopy. *Soil Research,* 41**,** 1101-1114.

JENNY, H. 1994. *Factors of soil formation: a system of quantitative pedology*, Courier Corporation.

JONES, E. J., FILIPPI, P., WITTIG, R., FAJARDO, M., PINO, V. & MCBRATNEY, A. B. 2021. Mapping soil slaking index and assessing the impact of management in a mixed agricultural landscape. *SOIL,* 7**,** 33-46.

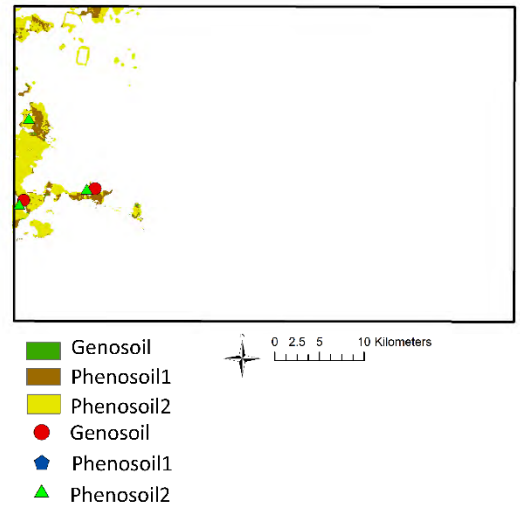
JONES, E. J. & MCBRATNEY, A. B. 2016. In situ analysis of soil mineral composition through conjoint use of visible, near-infrared and X-ray fluorescence spectroscopy. *Digital soil morphometrics.* Springer.
KUHN, M., WESTON, S., KEEFER, C. & COULTER, N. 2012. Cubist models for regression. *R package Vignette R package version 0.0,* 18.
KUZYAKOV, Y. & ZAMANIAN, K. 2019. Reviews and syntheses: Agropedogenesis–humankind as the sixth soil-forming factor and attractors of agricultural soil degradation. *Biogeosciences,* 16**,** 4783-4803.
LACOSTE, M., LEMERCIER, B. & WALTER, C. 2011. Regional mapping of soil parent material by machine learning based on point data. *Geomorphology,* 133**,** 90-99.
LEGENDRE, P. & LEGENDRE, L. 2012. *Numerical ecology*, Elsevier.
LLOYD, S. 1982. Least squares quantization in PCM. *IEEE transactions on information theory,* 28**,** 129-137.
MA, Y., MINASNY, B., MALONE, B. P. & MCBRATNEY, A. B. 2019. Pedology and digital soil mapping (DSM). *European Journal of Soil Science,* 70**,** 216-235.
MACQUEEN, J. Some methods for classification and analysis of multivariate observations. Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, 1967. Oakland, CA, USA, 281-297.
MALONE, B., MINASNY, B. & MCBRATNEY, A. 2009. Mapping continuous soil depth functions in the Edgeroi District, NSW, Australia, using terrain attributes and other environmental factors. *Proceedings of geomorphometry***,** 90-7.
MALONE, B., WADOUX, A. M.-C., MCBRATNEY, A. B., FAJARDO, M. & MINASNY, B. 2021. *Soil Spectral Inference with R: Analysing Digital Soil Spectra Using the R Programming Environment*, Springer Nature.
MCBRATNEY, A. B., MENDONÇA SANTOS, M. L. & MINASNY, B. 2003. On digital soil mapping. *Geoderma,* 117**,** 3-52.
MCGARRY, D., WARD, W. T., MCBRATNEY, A. B. & SOILS, C. D. O. 1989. *Soil Studies in the Lower Namoi Valley: Methods and Data. 1, The Edgeroi Data Set*, CSIRO Division of Soils.
MINASNY, B., MCBRATNEY, A. B., MENDONÇA-SANTOS, M., ODEH, I. & GUYON, B. 2006. Prediction and digital mapping of soil carbon storage in the Lower Namoi Valley. *Soil Research,* 44**,** 233-244.
MINTY, B., FRANKLIN, R., MILLIGAN, P., RICHARDSON, M. & WILFORD, J. 2009. The radiometric map of Australia. *Exploration Geophysics,* 40**,** 325-333.
NARRABRI SHIRE COUNCIL. 2021. *Historical Walks* [Online]. [Accessed].
NG, W., MINASNY, B., JONES, E. & MCBRATNEY, A. 2022. To spike or to localize? Strategies to improve the prediction of local soil properties using regional spectral library. *Geoderma,* 406**,** 115501.
PADARIAN, J., MINASNY, B. & MCBRATNEY, A. B. 2019. Using deep learning for digital soil mapping. *Soil,* 5**,** 79-89.
PARLIAMENT OF NEW SOUTH WALES. 2021. *1788 to 1810 - Early European Settlement* [Online]. Available: https://www.parliament.nsw.gov.au/about/Pages/1788-to-1810-Early-European-Settlement.aspx [Accessed].
PEI, T., QIN, C.-Z., ZHU, A. X., YANG, L., LUO, M., LI, B. & ZHOU, C. 2010. Mapping soil organic matter using the topographic wetness index: A comparative study based on different flow-direction algorithms and kriging methods. *Ecological Indicators,* 10**,** 610-619.
PRESCOTT, J. A. 1950. A climatic index for the leaching factor in soil formation. *Journal of Soil Science,* 1**,** 9-19.
QUINN, P., BEVEN, K., CHEVALLIER, P. & PLANCHON, O. 1991. The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. *Hydrological processes,* 5**,** 59-79.
RICHTER JR, D. D. 2007. Humanity's transformation of Earth's soil: Pedology's new frontier. *Soil science,* 172**,** 957-967.
ROELL, Y. E., PENG, Y., BEUCHER, A., GREVE, M. B. & GREVE, M. H. 2020. Development of hierarchical terron workflow based on gridded data – A case study in Denmark. *Computers & Geosciences,* 138**,** 104454.

ROMÁN DOBARCO, M., MCBRATNEY, A., MINASNY, B. & MALONE, B. 2021a. A framework to assess changes in soil condition and capability over large areas. *Soil Security,* 4**,** 100011.
ROMÁN DOBARCO, M., MCBRATNEY, A., MINASNY, B. & MALONE, B. 2021b. A modelling framework for pedogenon mapping. *Geoderma,* 393**,** 115012.
ROSSITER, D. G. & BOUMA, J. 2018. A new look at soil phenoforms–Definition, identification, mapping. *Geoderma,* 314**,** 113-121.
ROUDIER, P. 2020. spectacles: storing and manipulating spectroscopy data in R.
ROUDIER, P., ODGERS, N., CARRICK, S., EGER, A., HAINSWORTH, S. & BEAUDETTE, D. 2022. Soilscapes of New Zealand: Pedologic diversity as organised along environmental gradients. *Geoderma,* 409**,** 115637.
SHARPLEY, A. 1985. Depth of surface soil-runoff interaction as affected by rainfall, soil slope, and management. *Soil Science Society of America Journal,* 49**,** 1010-1015.
TANG, Y., JONES, E. & MINASNY, B. 2019. Evaluating low-cost portable near infrared sensors for rapid analysis of soils from South Eastern Australia. *Geoderma Regional,* 20**,** e00240.
TRIANTAFILIS, J., EARL, N. & GIBBS, I. 2016. Digital soil-class mapping across the Edgeroi district using numerical clustering and gamma-ray spectrometry data. *Computing Ethics: A Multicultural Approach***,** 187.
TRIANTAFILIS, J., GIBBS, I. & EARL, N. 2013. Digital soil pattern recognition in the lower Namoi valley using numerical clustering of gamma-ray spectrometry data. *Geoderma,* 192**,** 407-421.
VAYSSE, K. & LAGACHERIE, P. 2015. Evaluating digital soil mapping approaches for mapping GlobalSoilMap soil properties from legacy data in Languedoc-Roussillon (France). *Geoderma Regional,* 4**,** 20-30.
VENABLES, W. & RIPLEY, B. 2002. Modern Applied Statistics With S., 4th edn.(Springer-Verlag: New York.).
WARD, W. 1999. Soils and landscapes near Narrabri and Edgeroi, NSW, with data analysis and using fuzzy k-means.
WICKLIN, R. 2012. *What is Mahalanobis distance?* [Online]. Available: https://blogs.sas.com/content/iml/2012/02/15/what-is-mahalanobis-distance.html [Accessed].
WILFORD, J. & KROLL, A. 2020. Complete Radiometric Grid of Australia (Radmap) v4 2019 with modelled infill. Canberra: Geoscience Australia.
WILLIAMS, K. J., BELBIN, L., AUSTIN, M. P., STEIN, J. L. & FERRIER, S. 2012. Which environmental variables should I use in my biodiversity model? *International Journal of Geographical Information Science,* 26**,** 2009-2047.
XUE, P.-P., CARRILLO, Y., PINO, V., MINASNY, B. & MCBRATNEY, A. B. 2018. Soil Properties Drive Microbial Community Structure in a Large Scale Transect in South Eastern Australia. *Scientific Reports,* 8**,** 11725.
ZERAATPISHEH, M., BAKHSHANDEH, E., EMADI, M., LI, T. & XU, M. 2020. Integration of PCA and Fuzzy Clustering for Delineation of Soil Management Zones and Cost-Efficiency Analysis in a Citrus Plantation. *Sustainability,* 12**,** 5809.
ZHANG, G.-L., FENG, L. & SONG, X.-D. 2017. Recent progress and future prospect of digital soil mapping: A review. *Journal of Integrative Agriculture,* 16**,** 2871-2885.
ZUUR, A., IENO, E. N. & SMITH, G. M. 2007. *Analyzing ecological data*, Springer.

## 3.8. Appendices

**Appendix 3-1. Soil information for each pedogenon.**

### Pedogenon A




Geno Pheno2

**Environment and landuse**

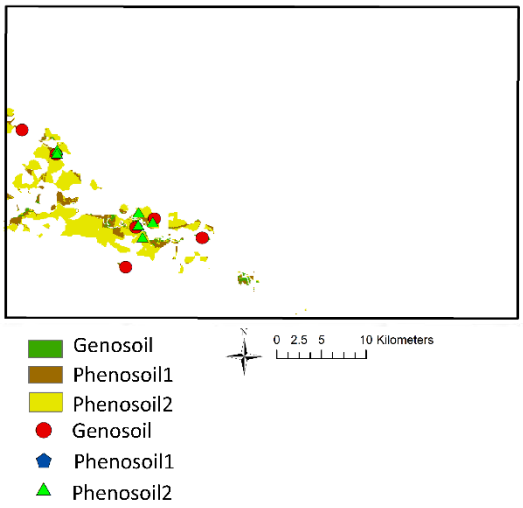
- **Parent material:** Alluvium
- **Topography:** Flat
- **Dominant Vegetation:** Native Grasslands and Forested Wetlands
- **Genosoil (Geno) landuse:** Woodland
- **Phenosoil2 (Pheno2) landuse:** Irrigated Cropping

**Soil**

- **Soil texture:** Heavy clay
- **Soil colour:** 10YR 2.5/1
- **Soil class:** Black Vertosol

| A | 0-30 (cm) | | 30-60 (cm) | | 60-100 (cm) | |
|---|---|---|---|---|---|---|
| | Geno | Pheno2 | Geno | Pheno2 | Geno | Pheno2 |
| Clay (%) | 50 | 59 | 53 | 63 | 53 | 64 |
| pH | 7.6 | 8.1 | 8.1 | 8.3 | 8.2 | 8.3 |
| CEC (mmol/kg) | 346.4 | 357.5 | 358.2 | 392.6 | 367.9 | 383.6 |
| OC (%) | 1.5 | 1.0 | 1.2 | 0.9 | 1.1 | 1.0 |

### Pedogenon B




Geno Pheno2

**Environment and landuse**

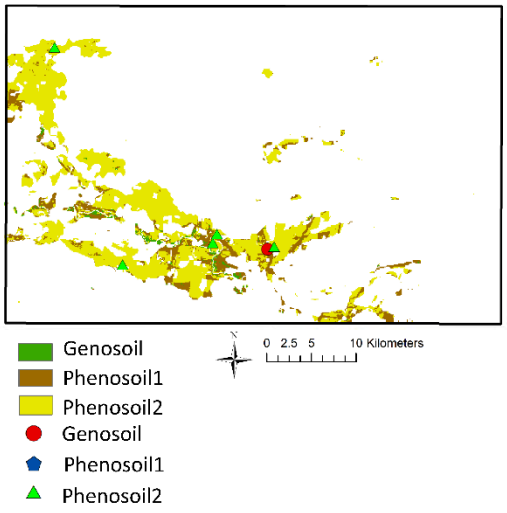
- **Parent material:** Alluvium
- **Topography:** Flat
- **Dominant Vegetation:** Native Grasslands
- **Genosoil (Geno) landuse:** Woodland
- **Phenosoil2 (Pheno2) landuse:** Irrigated Cropping

**Soil**

- **Soil texture:** Medium clay
- **Soil colour:** 7.5YR 3/3
- **Soil class:** Dark Brown Vertosol

| B | 0-30 (cm) | | 30-60 (cm) | | 60-100 (cm) | |
|---|---|---|---|---|---|---|
| | Geno | Pheno2 | Geno | Pheno2 | Geno | Pheno2 |
| Clay (%) | 37 | 46 | 42 | 50 | 47 | 51 |
| pH | 7.2 | 7.9 | 7.7 | 8.1 | 7.8 | 8.2 |
| CEC (mmol/kg) | 241.2 | 302.8 | 275.8 | 319.1 | 308.7 | 322.2 |
| OC (%) | 1.5 | 1.2 | 1.2 | 1.0 | 1.2 | 1.0 |

# Pedogenon C




Geno Pheno2

**Environment and landuse**

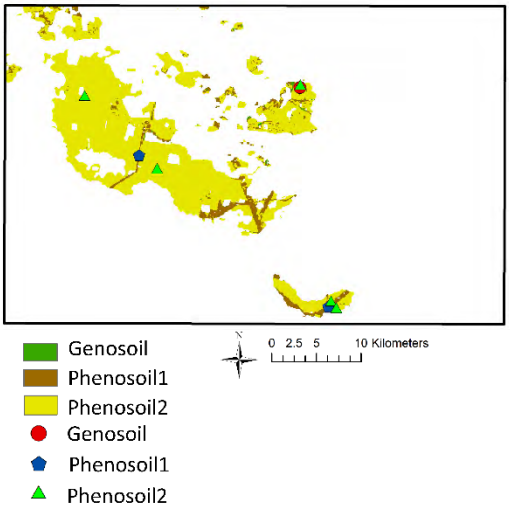
- **Parent material:** Alluvium
- **Topography:** Flat
- **Dominant Vegetation:** Native Grasslands
- **Genosoil (Geno) landuse:** Woodland
- **Phenosoil 2 (Pheno2) landuse:** Non-irrigated and Irrigated Cropping

**Soil**

- **Soil texture:** Heavy clay
- **Soil colour:** 7.5YR 3/2
- **Soil class:** Dark Brown Vertosol

| C | 0-30 (cm) | | 30-60 (cm) | | 60-100 (cm) | |
|---|---|---|---|---|---|---|
| | Geno | Pheno2 | Geno | Pheno2 | Geno | Pheno2 |
| Clay (%) | 61 | 61 | 66 | 66 | 64 | 64 |
| pH | 8.2 | 8.4 | 8.2 | 8.3 | 8.3 | 8.4 |
| CEC (mmol/kg) | 456.6 | 377.8 | 487.8 | 408.3 | 454.1 | 404.8 |
| OC (%) | 1.5 | 0.9 | 1.1 | 0.9 | 1.1 | 0.9 |

# Pedogenon D




Geno Pheno2

**Environment and landuse**

- **Parent material:** Alluvium
- **Topography:** Flat
- **Dominant Vegetation:** Native Grasslands
- **Genosoil (Geno) landuse :** Woodland
- **Phenosoil 1 (Pheno1) landuse :** Pasture
- **Phenosoil2 (Pheno2) landuse:** Irrigated Cropping

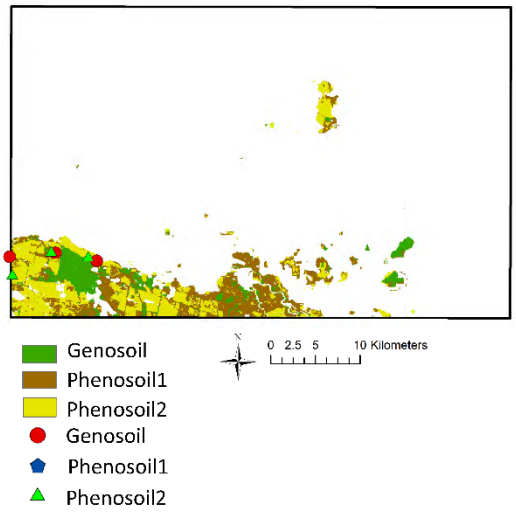
**Soil**

- **Soil texture:** Heavy clay
- **Soil colour:** 5YR 2.5/1
- **Soil class:** Black Vertosol

| D | 0-30 (cm) | | | 30-60 (cm) | | | 60-100 (cm) | | |
|---|---|---|---|---|---|---|---|---|---|
| | G | P1 | P2 | G | P1 | P2 | G | P1 | P2 |
| Clay (%) | 31 | 60 | 62 | 45 | 63 | 67 | 45 | 63 | 70 |
| pH | 7.4 | 7.9 | 8.3 | 7.4 | 8.0 | 8.3 | 7.8 | 8.3 | 8.4 |
| CEC (mmol/kg) | 225.4 | 453.1 | 443.4 | 335.3 | 419.9 | 455.9 | 317.2 | 409.0 | 479.0 |
| OC (%) | 1.2 | 1.4 | 0.8 | 1.7 | 1.0 | 0.7 | 0.9 | 0.9 | 0.7 |

G: Geno / P1: Pheno1 / P2: Pheno2

# Pedogenon E




Geno Pheno2

**Environment and landuse**

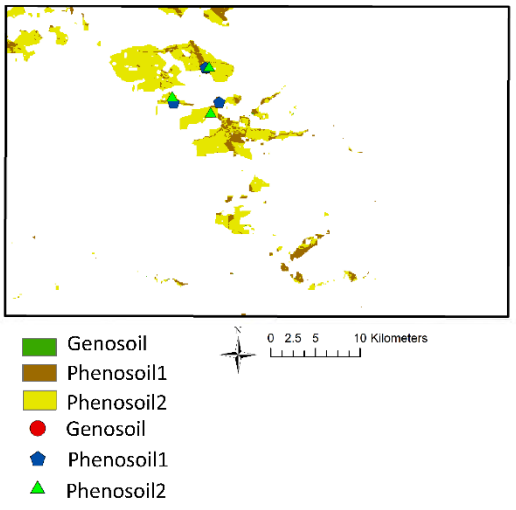
- **Parent material:** Sand Plain
- **Topography:** Flat
- **Dominant Vegetation:** Dry Sclerophyll Forests (Shrub/grass sub-formation or Shrub sub-formation)
- **Genosoil (Geno) landuse :** Woodland
- **Phenosoil2 (Pheno2) landuse:** Irrigated Cropping

**Soil**

- **Soil texture:** Sandy clay loam / Light clay
- **Soil colour:** 10R 3/6
- **Soil class:** Red Chromosol

| E | 0-30 (cm) | | 30-60 (cm) | | 60-100 (cm) | |
|---|---|---|---|---|---|---|
| | Geno | Pheno2 | Geno | Pheno2 | Geno | Pheno2 |
| Clay (%) | 13 | 19 | 17 | 25 | 32 | 40 |
| pH | 6.5 | 7.0 | 7.0 | 7.3 | 7.5 | 7.6 |
| CEC (mmol/kg) | 89.2 | 135.0 | 79.1 | 120.3 | 90.6 | 116.5 |
| OC (%) | 0.8 | 0.9 | 0.4 | 0.5 | 0.4 | 0.4 |

# Pedogenon F




Pheno1 Pheno2

**Environment and landuse**

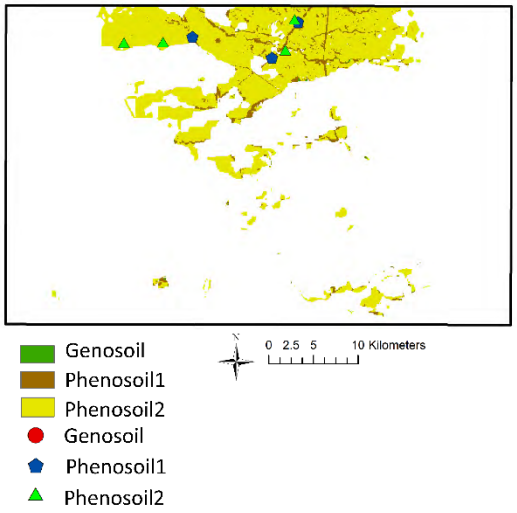
- **Parent material:** Sand Plain
- **Topography:** Flat
- **Dominant Vegetation:** Native Grasslands
- **Phenosoil1 (Pheno1) landuse :** Pasture
- **Phenosoil2 (Pheno2) landuse:** Non-irrigated Cropping

**Soil**

- **Soil texture:** Heavy clay
- **Soil colour:** 7.5YR 2.5/1
- **Soil class:** Black Vertosol

| F | 0-30 (cm) | | 30-60 (cm) | | 60-100 (cm) | |
|---|---|---|---|---|---|---|
| | Pheno1 | Pheno2 | Pheno1 | Pheno2 | Pheno1 | Pheno2 |
| Clay (%) | 53 | 50 | 57 | 56 | 58 | 57 |
| pH | 7.7 | 7.1 | 8.0 | 7.6 | 8.1 | 8.0 |
| CEC (mmol/kg) | 404.6 | 363.7 | 416.8 | 405.0 | 431.1 | 423.0 |
| OC (%) | 0.8 | 0.8 | 0.6 | 0.7 | 0.6 | 0.6 |

# Pedogenon G




Pheno1 Pheno2

**Environment and landuse**

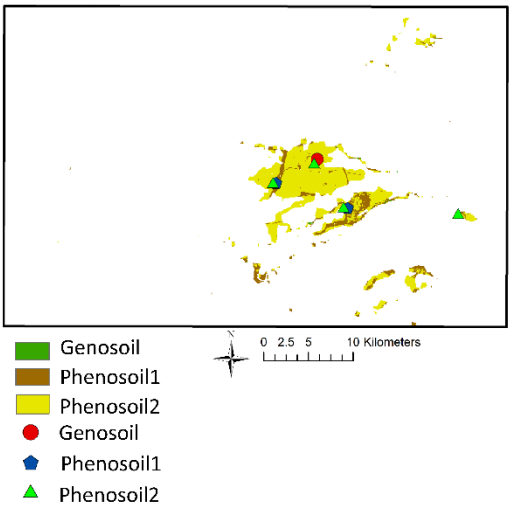
- **Parent material:** Alluvium
- **Topography:** Flat
- **Dominant Vegetation:** Native Grasslands
- **Phenosoil1 (Pheno1) landuse :** Pasture
- **Phenosoil2 (Pheno2) landuse:** Non-irrigated Cropping

**Soil**

- **Soil texture:** Heavy clay
- **Soil colour:** 5YR 2.5/1
- **Soil class:** Black Vertosol

| G | 0-30 (cm) | | 30-60 (cm) | | 60-100 (cm) | |
|---|---|---|---|---|---|---|
| | Pheno1 | Pheno2 | Pheno1 | Pheno2 | Pheno1 | Pheno2 |
| Clay (%) | 61 | 64 | 67 | 69 | 65 | 70 |
| pH | 8.0 | 8.5 | 8.6 | 8.2 | 8.6 | 8.3 |
| CEC (mmol/kg) | 414.3 | 532.7 | 447.3 | 544.3 | 448.9 | 567.9 |
| OC (%) | 1.1 | 0.9 | 0.7 | 0.8 | 0.7 | 0.8 |

# Pedogenon H




Geno Pheno2

**Environment and landuse**

- **Parent material:** Colluvium
- **Topography:** Flat
- **Dominant Vegetation:** Native Grasslands
- **Genosoil (Geno) landuse:** Woodland
- **Phenosoil1 (Pheno1) landuse:** Pasture
- **Phenosoil2 (Pheno2) landuse:** Non-irrigated Cropping

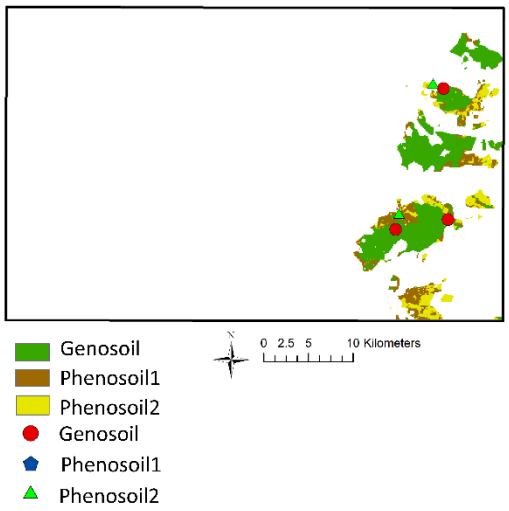
**Soil**

- **Soil texture:** Heavy clay
- **Soil colour:** 10YR 3/3
- **Soil class:** Dark Brown Vertosol

| H | 0-30 (cm) | | | 30-60 (cm) | | | 60-100 (cm) | | |
|---|---|---|---|---|---|---|---|---|---|
| | G | P1 | P2 | G | P1 | P2 | G | P1 | P2 |
| Clay (%) | 36 | 55 | 59 | 46 | 59 | 60 | 39 | 61 | 53 |
| pH | 7.2 | 8.1 | 8.5 | 8.0 | 8.5 | 8.6 | 8.0 | 8.6 | 8.8 |
| CEC (mmol/kg) | 236.6 | 465.4 | 453.2 | 303.2 | 462.4 | 454.7 | 270.3 | 477.3 | 416.4 |
| OC (%) | 1.8 | 1.4 | 1.2 | 1.5 | 1.2 | 1.0 | 1.2 | 1.1 | 1.1 |

G: Geno / P1: Pheno1 / P2: Pheno2

# Pedogenon I





**Environment and landuse**

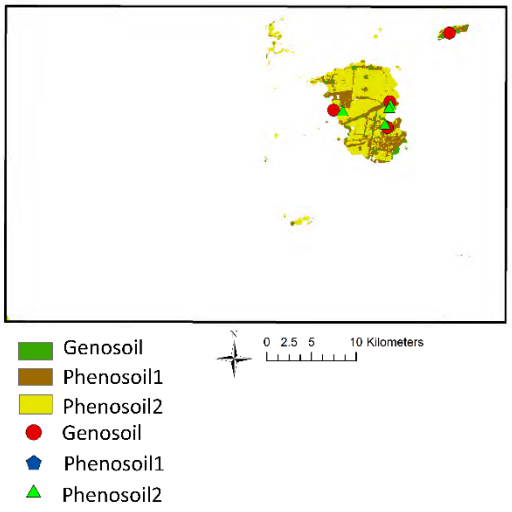
- **Parent material:** Pilliga Sandstone
- **Topography:** Gentle Slope
- **Dominant Vegetation:** Dry Sclerophyll Forests (Shrub/grass sub-formation or Shrub sub-formation)
- **Genosoil (Geno) landuse:** Woodland
- **Phenosoil2 (Pheno2) landuse:** Non-irrigated Cropping

**Soil**

- **Soil texture:** Sandy clay loam / Light clay
- **Soil colour:** 10YR 4/4
- **Soil class:** Dark Brown Tenosol

| I | 0-30 (cm) | | 30-60 (cm) | | 60-100 (cm) | |
|---|---|---|---|---|---|---|
| | Geno | Pheno2 | Geno | Pheno2 | Geno | Pheno2 |
| Clay (%) | 18 | 21 | 21 | 34 | 27 | 36 |
| pH | 6.5 | 6.9 | 6.9 | 7.2 | 7.4 | 7.5 |
| CEC (mmol/kg) | 140.2 | 168.6 | 138.6 | 202.7 | 151.5 | 218.0 |
| OC (%) | 1.4 | 0.7 | 1.1 | 0.5 | 0.8 | 0.5 |

# Pedogenon J





**Environment and landuse**

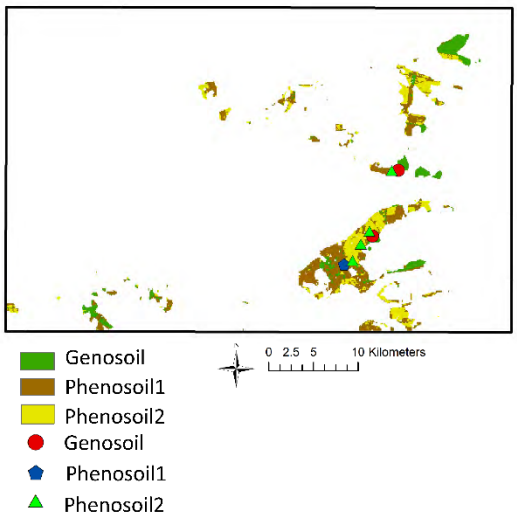
- **Parent material**: Keelindi Beds and Sand Plain
- **Topography:** Gentle Slope
- **Dominant Vegetation:** Native Grasslands and Western Vine Thickets
- **Genosoil (Geno) landuse:** Woodland
- **Phenosoil2 (Pheno2) landuse:** Non-irrigated Cropping

**Soil**

- **Soil texture:** Medium–heavy clay
- **Soil colour:** 5YR 2.5/1
- **Soil class:** Black Vertosol

| J | 0-30 (cm) | | 30-60 (cm) | | 60-100 (cm) | |
|---|---|---|---|---|---|---|
| | Geno | Pheno2 | Geno | Pheno2 | Geno | Pheno2 |
| Clay (%) | 37 | 49 | 44 | 55 | 44 | 56 |
| pH | 7.5 | 8.1 | 8.0 | 8.2 | 8.2 | 8.4 |
| CEC (mmol/kg) | 271.0 | 364.7 | 313.7 | 393.5 | 312.9 | 398.8 |
| OC (%) | 1.5 | 1.1 | 1.2 | 0.9 | 1.1 | 0.9 |

# Pedogenon K




Geno

Pheno2

**Environment and landuse**

- **Parent material:** Keelindi Beds
- **Topography:** Gentle Slope
- **Dominant Vegetation:** Native Grasslands and Dry Sclerophyll Forests (Shrub/grass sub-formation)
- **Genosoil (Geno) landuse:** Woodland
- **Phenosoil1 (Pheno1) landuse:** Pasture
- **Phenosoil2 (Pheno2) landuse:** Non-irrigated and Irrigated Cropping

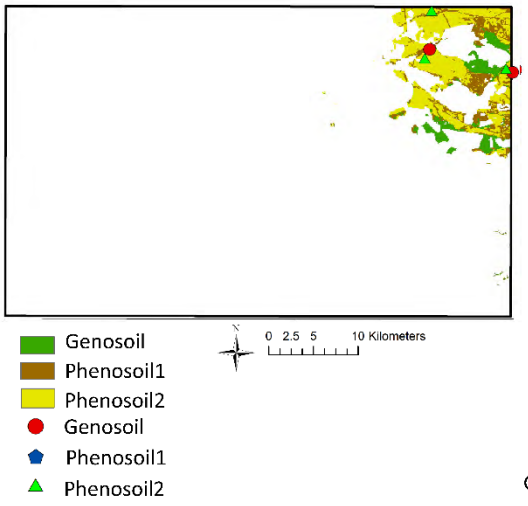
**Soil**

- **Soil texture:** Medium clay
- **Soil colour:** 5YR 2.5/2
- **Soil class:** Brown Vertosol

| K | 0-30 (cm) | | | 30-60 (cm) | | | 60-100 (cm) | | |
|---|---|---|---|---|---|---|---|---|---|
| | G | P1 | P2 | G | P1 | P2 | G | P1 | P2 |
| Clay (%) | 20 | 34 | 29 | 23 | 43 | 40 | 29 | 48 | 40 |
| pH | 6.6 | 7.6 | 7.4 | 7.3 | 8.6 | 7.9 | 7.9 | 8.8 | 7.9 |
| CEC (mmol/kg) | 146.4 | 273.9 | 199.5 | 148.9 | 326.3 | 237.0 | 190.2 | 341.5 | 240.0 |
| OC (%) | 1.1 | 1.7 | 1.0 | 0.5 | 1.0 | 0.8 | 0.5 | 0.8 | 0.8 |

G: Geno / P1: Pheno1 / P2: Pheno2

# Pedogenon L




Geno

Pheno2

**Environment and landuse**

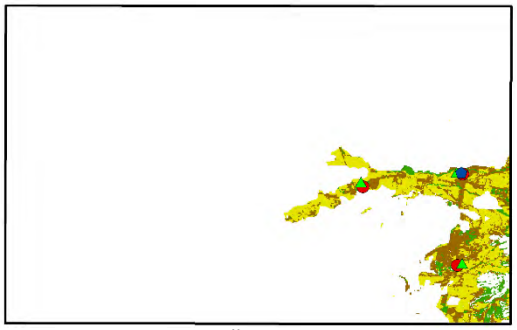
- **Parent material:** Colluvium
- **Topography:** Gentle Slope
- **Dominant Vegetation:** Native Grasslands and Dry Sclerophyll Forests (Shrub/grass sub-formation)
- **Genosoil (Geno) landuse:** Woodland
- **Phenosoil2 (Pheno2) landuse:** Non-irrigated Cropping

**Soil**

- **Soil texture:** Heavy clay
- **Soil colour:** 10YR 2.5/1
- **Soil class:** Black Vertosol

| L | 0-30 (cm) | | 30-60 (cm) | | 60-100 (cm) | |
|---|---|---|---|---|---|---|
| | Geno | Pheno2 | Geno | Pheno2 | Geno | Pheno2 |
| Clay (%) | 38 | 61 | 50 | 64 | 59 | 62 |
| pH | 7.7 | 8.5 | 8.4 | 8.4 | 8.8 | 8.4 |
| CEC (mmol/kg) | 275.3 | 454.9 | 350.5 | 468.3 | 401.3 | 487.9 |
| OC (%) | 1.5 | 1.2 | 1.2 | 1.0 | 1.0 | 1.0 |

# Pedogenon M




Geno

Pheno2

## Environment and landuse

- **Parent material:** Digby formation, Napperby formation – sandstone and Alluvium
- **Topography:** Gentle Slope
- **Dominant Vegetation:** Native Grasslands and Western Vine Thickets
- **Genosoil (Geno) landuse:** Woodland
- **Phenosoil1 (Pheno1) landuse:** Pasture
- **Phenosoil2 (Pheno2) landuse:** Non-irrigated Cropping

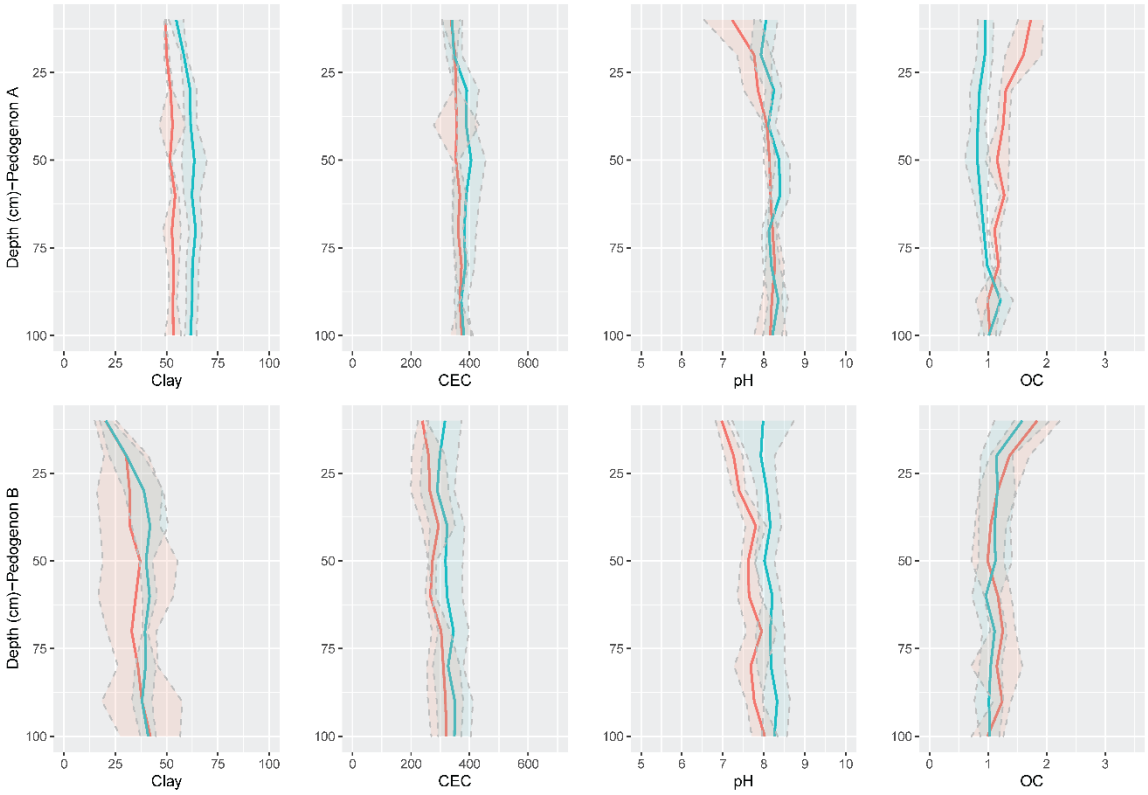
## Soil

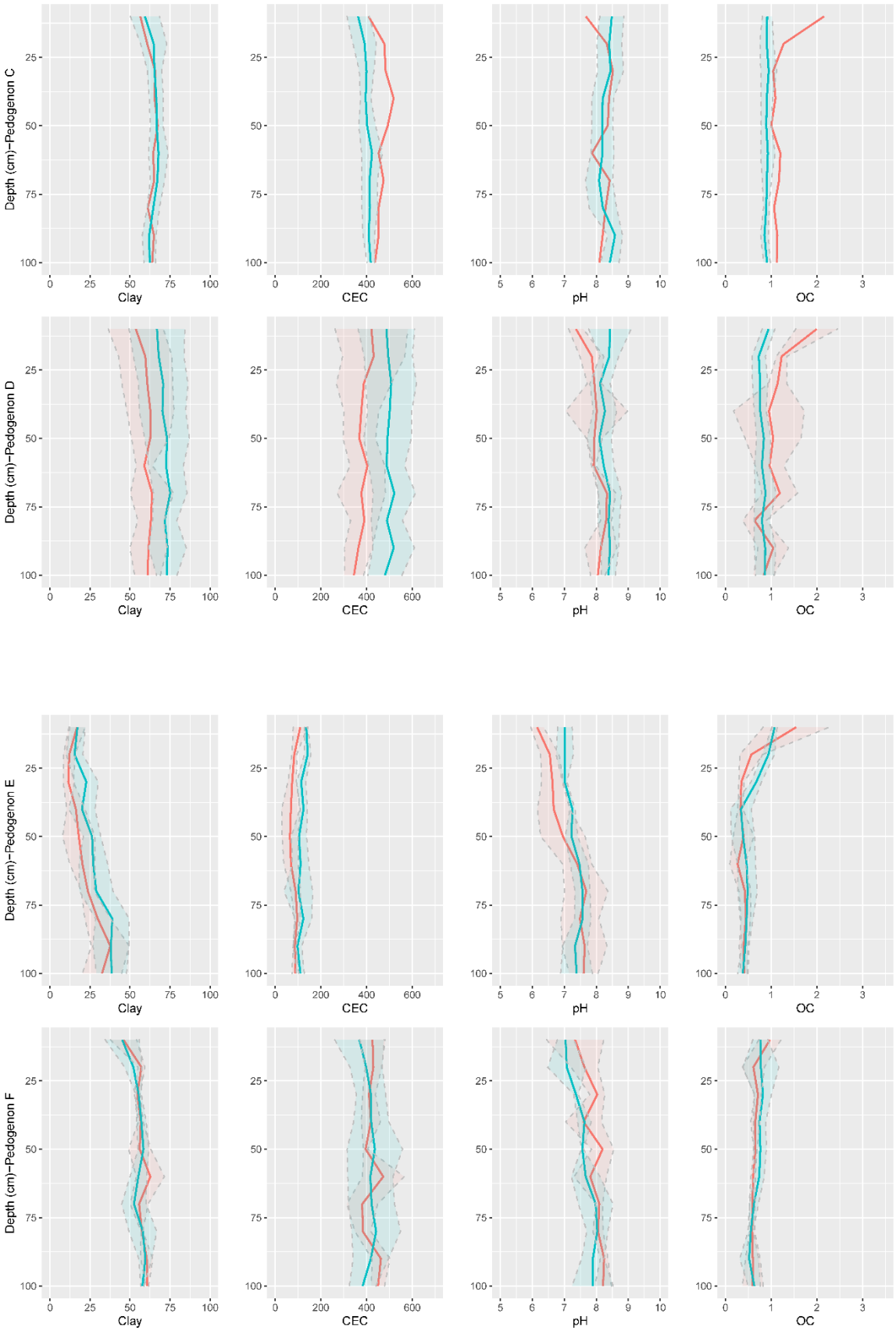
- **Soil texture:** Light medium clay
- **Soil colour:** 7.5YR 2.5/2
- **Soil class:** Dark brown Vertosol

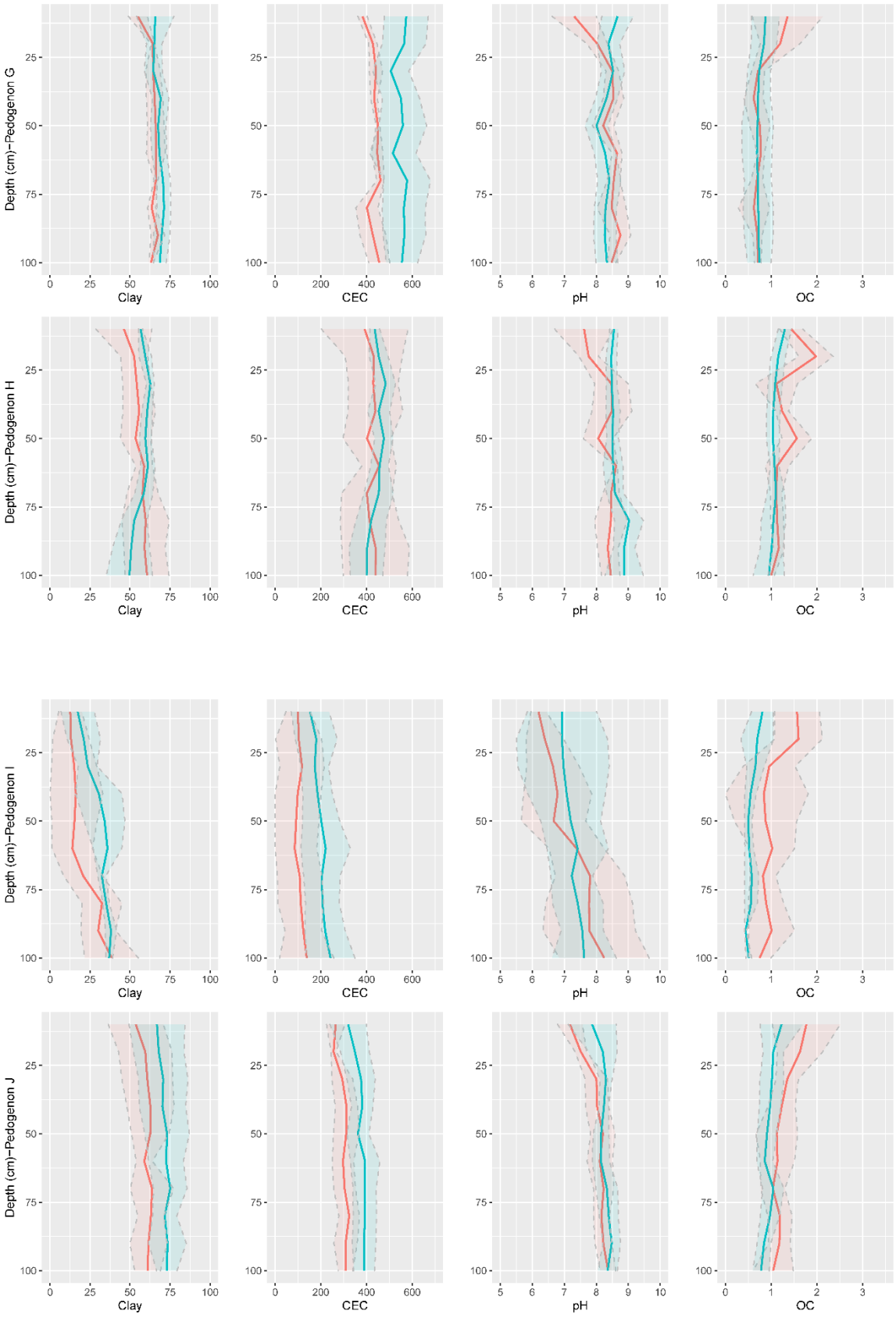
| M | 0-30 (cm) | | | 30-60 (cm) | | | 60-100 (cm) | | |
|---|---|---|---|---|---|---|---|---|---|
| | G | P1 | P2 | G | P1 | P2 | G | P1 | P2 |
| Clay (%) | 31 | 41 | 42 | 42 | 49 | 50 | 42 | 52 | 48 |
| pH | 6.9 | 6.9 | 7.6 | 7.7 | 8.0 | 7.9 | 8.0 | 8.1 | 8.1 |
| CEC (mmol/kg) | 193.2 | 305.2 | 295.4 | 269.4 | 355.1 | 355.6 | 296.2 | 366.7 | 355.3 |
| OC (%) | 1.5 | 2.6 | 1.6 | 1.2 | 1.4 | 1.3 | 1.2 | 1.1 | 1.2 |

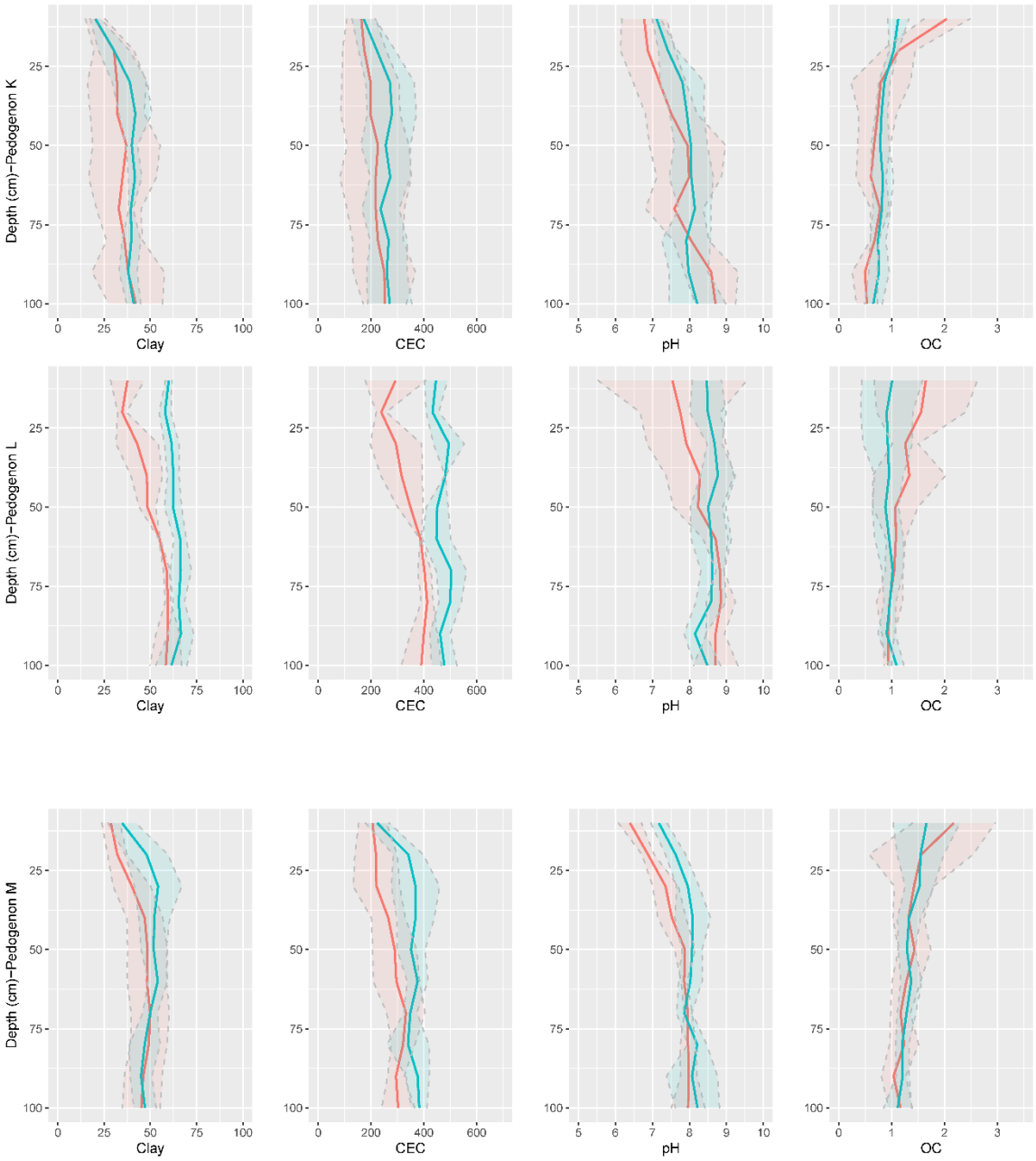
G: Geno / P1: Pheno1 / P2: Pheno2

**Appendix 3-2. Line graphs of soil profile properties for each pedogenon. The red lines represent mean values for genosoils &Phenosoils1, the blue lines represent phenosoils2. The shaded area represents 95% confidence interval of the mean of the data.**




Depth (cm)−Pedogenon C
Depth (cm)−Pedogenon D
Depth (cm)−Pedogenon E
Depth (cm)−Pedogenon F
Clay
CEC
pH
OC


Depth (cm)~Pedogenon G
Depth (cm)~Pedogenon H
Depth (cm)~Pedogenon I
Depth (cm)~Pedogenon J
Clay
CEC
pH
OC


Depth (cm)~Pedogenon K
Depth (cm)~Pedogenon L
Depth (cm)~Pedogenon M
Clay
CEC
pH
OC

# Chapter 4. Assessing soil organic carbon change in the Lower Namoi Valley, Australia using pedogenon sampling and digital soil mapping techniques

## 4.1. Abstract

Intensive agricultural activities have caused soil organic carbon (SOC) to decline in many parts of the world. However, there is a dearth of approaches that can spatially estimate the change of SOC due to land use change at a regional scale. In this study, we used the concept of pedogenons as the basis of soil carbon change estimation. Pedogenon is a conceptual soil class defined from a set of quantitative state variables representing the soil-forming factors for a given reference time. We recognise that pedogenons can be modified by human activities, forming phenosoils. This study identifies pedogenons representing SOC under native vegetation and the phenosoils representing SOC under intensive human management. We surveyed the lower Namoi Valley area, NSW, Australia (1700 $km^2$), comprising 13 pedogenons. Within each pedogenon, we sampled regions under native vegetation and areas under cropping. Ninety-nine sampling cores were collected from 0-1 m in woodlands and agricultural areas. The soil profile cores were scanned in 10 cm depth intervals (to 1 m) using mid-infrared spectroscopy to quantify SOC concentration. Using Digital Soil Mapping (DSM) techniques, the SOC data were used for mapping SOC every 10 cm down to 1 m using spatial layers (soil, organisms, topography and parent material and age). Sampling points under native vegetation were used to map SOC under the native state, and all data were used for mapping SOC current state. The neural networks model was utilised for making maps at 10 depth intervals simultaneously. By comparing the SOC maps at two states (native and current), we assessed SOC change from 0-1 m. The results show that the SOC loss in irrigated cropping areas was the largest (average: 3.60 t C $ha^{-1}$, 0-30 cm), followed by non-irrigated cropping (3.32 t C $ha^{-1}$), pasture (1.92 t C $ha^{-1}$), and woodland (0.81 t C $ha^{-1}$). Overall the area had a loss of 1,737 Gg C due to cropping and pasture establishment. The understanding of SOC change can provide information on areas of SOC loss threat. The digital maps could be used to inform farm management and environmental planning.

## 4.2. Introduction

Soil carbon is central to soil functions, influencing crop production, nutrient and water cycles, climate, and above- and below-ground biodiversity (McBratney et al., 2014). As the largest terrestrial store of carbon, soil has two to three times more carbon than the atmosphere and plants. Thus, a slight increase in carbon in the soil could significantly mitigate greenhouse gas emissions and enhance ecosystem services (Lal, 2013). It is known that increasing SOC improves soil physical, chemical and biological conditions, thus enhancing crop productivity (Kopittke et al., 2021). However, intensive agricultural activities have caused SOC decline in many parts of the world. Most global soils under agricultural practices are depleted of their initial carbon stock by 25%−75%, in the range of 10−30 Mg $ha^{-1}$ (Lal, 2013). A study by Padarian et al. (2022) indicated that global land use change in the past two decades caused an annual loss of about 1.9 Pg C per year. In Australia, estimates have shown that areas that have been cultivated for more than ten years lose about half of their SOC content (Luo et al., 2010, Post and Kwon, 2000).

This decrease in the amount of carbon in the soil has a direct negative consequence on the ability of soils to produce crops, cycle the nutrients, protect the soil biodiversity, regulate water and climate and other ecosystem services (Kopittke et al., 2021). The continued cumulative effect of soil degradation could have devastating consequences on soil functions (Evans et al., 2020). Unfortunately, there is a lack of an approach that can quantify and monitor the change of SOC at a regional scale. Inherent variation of SOC is a function of soil forming factors, and thus there is also a lack of reference SOC content that allows us to quantify how much SOC had been lost and how much SOC can be gained through SOC sequestration approaches.

Monitoring soil change has several ways (Filippi et al., 2016). Most studies on quantifying soil change require resampling, meaning a site needs to be revisited after some time. The

disadvantage of this approach is the high cost and effort involved. Another commonly used approach is the "over the fence" comparison or space-for-time substitution. This approach compares soil conditions under natural vegetation or less disturbed conditions with soils under cropping or agricultural activities. Filippi et al. (2016) mentioned that using space-for-time substitution to observe soil change is useful, especially when the time of land use change is known. Many studies use this approach to assess how human activities have modified soil over time (Cattle et al., 1994, Tye et al., 2013). However, this method only provides a pairwise comparison over a field or limited area and cannot be extended to the mapping of soil change over a region.

To address the spatial limitation of the space-for-time substitution method, this paper uses the concept of pedogenons proposed by Román Dobarco et al. (2021) to assess soil change and also set a reference state for SOC. The definition of pedogenon is "a conceptual soil taxon defined from a set of quantitative state variables representing the soil-forming factors for a given reference time". Within a pedogenon, we can distinguish genosoil and phenosoil (adapted from the concept of genoform and phenoform by Rossiter and Bouma (2018)). The genosoil is the dominant genetic soil type within a map unit, produced by long-term pedogenesis, and the phenosoil is a persistent variant of the genosoil modified by human activities (e.g., agricultural, industrial or mining practices). And thus, sampling the remnant genosoil and phenosoils of a particular pedogenon will allow us to map and quantify soil change. Furthermore, soil change can be quantified based on the type of land use or the intensity of anthropogenic disturbance.

We derived pedogenon maps for the Edgeroi region in New South Wales, Australia from Chapter 3 (Figure 4-1). Further, we designed a sampling scheme that captures soil variation under natural conditions and intensive agricultural activities. In this paper, we extend the data

collected from the pedogenons to quantify and map SOC change over the area. Hence, the aims of this study are:

i) To map SOC content (under natural or less disturbed conditions and the current state) using pedogenons and the digital soil mapping approach,

ii) To calculate how much soil carbon content and stock have changed as a function of pedogenon and land use due to agricultural practices.

iii) To set SOC targets for each of the pedogenons.

This study will provide a framework to quantify SOC change and also establish a benchmark or a reference state for SOC stock.

## 4.3. Materials and Methods

We first detail the research area and then describe the methodology of measuring SOC content using MIR spectroscopy. Subsequently, we detail the digital techniques for mapping soil under native vegetation and current conditions, mapping SOC content and stock followed by SOC change.

### 4.3.1. Description of the study area

The study area is in the Edgeroi district in New South Wales (NSW), Australia (1700 $km^2$). The region's climate contrasts hot and dry in the summer and cold in the winter. During summer, the average maximum temperature starts at about 30 °C and peaks at 40 °C. The average minimum temperature during winter is 3 °C. The mean annual precipitation is about 712 mm (Ward, 1999). The region has been intensively cropped for more than 200 years since European settlers came in the early 1800s (Green et al., 2011, Narrabri Shire Council, 2021). Because

grazing was the predominant agricultural activity in the past, vast regions were used as stock routes. Wheat, sorghum, and irrigated cotton farming were introduced in the 1960s (McGarry et al., 1989). Few remaining national parks are relatively unaffected by human activities for an extended period. The study area has undergone several land use changes in the last 200 years.

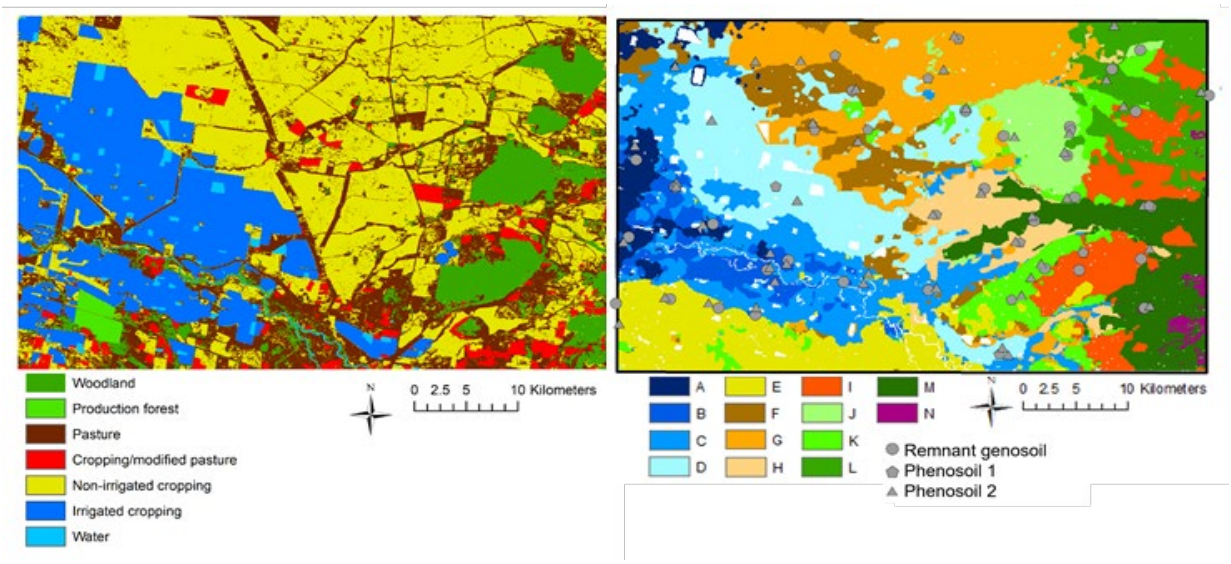



**Figure 4-1. Left: the map of land use or cover, and there are seven broad land use/cover classes in the area. Right: the map of fourteen pedogenons (Pedogenon A to N) and sampling points of genosoil and phenosoil.**

### 4.3.2. Pedogenon map

The Pedogenon map of the study area (Figure 4-1) was created from Chapter 3 based on the method of Román Dobarco et al. (2021). Fourteen pedogenons were determined in the area using an unsupervised classification method, i.e. k -means clustering of soil forming factors represented as spatial layers (soil, climate, organisms, topography, parent material and age). The resolution of the map is 90 m. The detail of each pedogenon is described in Chapter 3. The pedogenon map was validated with the five soil properties of each pedogenon (pH, clay, sand, cation exchange capacity, and organic carbon) and multivariate statistical analyses (redundancy discriminant analysis (RDA) and variance partitioning).

### 4.3.3. Soil sampling and MIR spectra acquisition

Sampling points were designated on the pedogenons based on land use data (Figure 4-1). First, for each pedogenon, we located genosoil sampling points in large patches of woodland areas, and then phenosoil points were randomly allocated in different types of agricultural lands or the land disturbed by anthropogenic activities. In total, soil cores to a depth of 1m were collected from 89 sampling points in woodland and farming areas. Extra samples (10 sampling points) were collected in the woodland area to ensure an even spatial coverage of the area. One pedogenon (pedogenon N, in Chapter 3) only occupied a small area on a high elevation and thus could not be sampled. Thus, this study used data from 13 pedogenons.

Soil samples were air-dried in the laboratory, parted every 10 cm, ground and sieved (< 2 mm). The samples for calibration data were sent to an external laboratory for analysis. Total carbon was analysed with the combustion method using Elementar analysis, and inorganic carbon (calcium carbonate) was analysed using a titration method. SOC was calculated as the difference between total and inorganic carbon.

The soil samples were scanned using a MIR diffuse reflectance spectrometer. Approximately 15 g of soil samples were further ground using a ball mill. Two replicates of each sample were prepared. The samples were loaded into Al microtiter plates (A752-96, Bruker Optik GmbH, Ettlingen, Germany) using a micro spatula to fill the 6 mm diameter wells and level the soil. The first two wells of every plate were filled with potassium bromide (KBr) used for the standard and blank, respectively, and the first sample was filled from the third well. Demyan et al. (2012) and Rasche et al. (2013) recommend maintaining the plate's temperature at 32 ˚C overnight to remove moisture in soil samples before scanning using the MIR spectrometer. The prepared plates for scanning were stored in a desiccator with dried silica gel for one day (humidity inside the desiccator, about 15-20 %). Spectral measurement was conducted using a high throughput screening device (HTS-XT) attached to a Tensor 37 spectrometer (Bruker Optik GmbH, Ettlingen, Germany). The detector was a mid-band mercury-cadmium-telluride detector which was cooled with liquid nitrogen. The wavenumber range of MIR spectra measured is 4000–400 $cm^{-1}$. The average reflectance of 60 scans per soil sample was transformed to absorbance and recorded using the Optics software.

### 4.3.4. Analysing and calibrating the spectra data for SOC prediction

The MIR spectra were used to predict SOC content. Four pre-processing methods for the spectra were used in this study. The first one is trimming. The spectrum edges (<600 and >3995 $cm^{-1}$) were excluded due to the high noise-to-signal ratio. After that, the region of atmospheric carbon dioxide ($CO_2$) (2300 to 2400 $cm^{-1}$) was removed because MIR spectroscopy instrument is susceptible to the concentration of $CO_2$ inside the instrument (Mirzaeitalarposhti et al., 2017). Then, the Savitzky-Golay filtering method was applied to smooth spectra. SNV (Standard Normal Variate) was used to normalise the spectra. All processes were conducted in R statistical software with the *spectroscopy* package (Campbell et al., 2016).

A total of 151 samples (89 topsoils (0-15 cm) and 62 subsoils (30-40 and 70-80 cm from the soil cores)) were used to calibrate a model relating the MIR spectra with SOC content. The cubist model was used for modelling, using the package Cubist (Kuhn et al., 2014).

A 10-fold cross-validation was employed to evaluate the accuracy of the SOC predictions. This 10-fold cross-validation was repeated 10 times. The order of the data was randomly shuffled before creating the folds. The result of validation data from each fold was accumulated and plotted for testing the goodness of fit (Figure 4-2).

### 4.3.5. Creating SOC and SOC change maps using DSM techniques

The spectral calibration model predicted SOC content at 10 cm intervals from the surface to 1 m depth for the 99 sampling points. Subsequently, maps of SOC concentration at every 10 cm depth interval from surface to 1 m depth were created using the DSM approach. The neural networks model was utilised to establish quantitative relationships between the SOC content and 13 environmental covariates representing the factors soil (*s*), organisms (*o*), relief (*r*), parent material (*p*) and age (*a*) from the *scorpan* model (McBratney et al., 2003) (Table 4-1). Non-normally distributed covariates were log-transformed (i.e., DEM and Gamma K/Th). The vegetation and parent material maps were converted to continuous variables for the clustering requirement. The method of conversion was well explained in Chapter 3. Water areas, including irrigation channels or dams, were removed from the map.

Two soil condition maps were produced: estimated SOC assuming natural vegetation across the study area and SOC under the current condition. For SOC under natural vegetation, only sampling points from the genosoil areas that were under natural vegetation were used (47 sampling points) as training data. For mapping the current SOC concentration, all sampling points were used.

The covariates representing *s, r and p* factors for both maps were the same except for the vegetation (*o*) covariates. For mapping SOC under natural conditions, the map of vegetation pre-European settlement (1788) was used as the covariate in the organism factor. This dataset was derived from Eco Logical Australia (2013), representing soil under natural conditions with the least human disturbance.

The map of current SOC content was created using the *s, r and p* factors and the current vegetation data representing *o*. The current vegetation map was derived from the Department of Planning and Environment (2018). The vegetation map contains information about plant community types for the Namoi and Border Rivers Gwydir catchments. A hybrid classification method (visual interpretation and spatial modelling) was utilised to combine the features and produce this map.

To create the SOC maps of 10 depths simultaneously, we used the neural networks model with multiple outputs, called the multilayer perceptron (MLP) model (Manaswi et al., 2018, Padarian et al., 2019). The model was implemented in JMP statistical software. As the neural network model only accept continuous data, categorical data were first converted to continuous data. Then, the categorical data were coded into binary variables (presence/absence for a particular class), and the principal component analysis was performed on the coded variables to produce continuous data.

For the modelling, the input layer has 13 nodes corresponding to the 13 environmental covariates, one hidden layer, and an output layer has 10 nodes which correspond to the prediction of 10 depth intervals (every 10 cm). The optimal number of nodes and hidden layers were chosen after trial and error. Here we used one hidden layer and the tanh activation function.

The neural networks training was executed 50 times using the bootstrapping method to calculate the model's validation and prediction uncertainty, resulting in 50 maps per depth.

These results were then averaged into one final map. The bootstrap uses sampling with a replacement on the data, creating training datasets size n = 47 (under native vegetation) and n = 99 (current). The data that were not selected during the sampling (approximately 1/3 of the data) were used for validation.

The change in SOC was calculated using a model-based approach, with the difference between the DSM-predicted current SOC ($C_{current}$) and SOC under native vegetation ($C_{native}$):

$$\Delta SOC = C_{current} - C_{native}$$

The uncertainty of the prediction was calculated from the variance of the two DSM predictions:

$$var(C_{current} - C_{native}) = var(C_{current}) + var(C_{native}) - 2cov(C_{current}, C_{native})$$

Where $var$ is variance, and $cov$ is covariance.

**Table 4-1. Environmental covariates were used to create both SOC maps (current and before intensive agriculture).**

| Soil-forming factors | Covariate name | Units | Reference |
|---|---|---|---|
| **Soil** | Gamma K | | Wilford and Kroll (2020) |
| | Gamma Th | | Wilford and Kroll (2020) |
| | Gamma K/Th | | Wilford and Kroll (2020) |
| | Gamma dose | | Wilford and Kroll (2020) |
| **Organisms** | Native Vegetation pre-European settlement (3 PCs, explained 82% variation) (For SOC map Native intensive agriculture) | | Eco Logical Australia (2013) |
| | Current vegetation (3 PCs, explained 82% variation) (For current SOC map ) | | Department of Planning and Environment (2018) |
| **Topography** | Altitude from Digital elevation model (DEM) | m | Gallant et al. (2009) |
| | Slope | % | Gallant et al. (2009) |
| | Topographic wetness index (TWI) | | Quinn et al. (1991) |
| **Parent material & age** | Parent material (3 PCs, explained 46% variation) | | From Chapter 2 |

### 4.3.6. Calculating SOC stock value

For calculating the SOC stock, bulk density data were used. Bulk density data were measured from the soil corer, where the soil was divided every 20 cm. The SOC stock (Mg $ha^{-1}$ or ton $ha^{-1}$) was calculated using the data of SOC, bulk density and soil thickness:

$$\text{SOC stock (Mg ha}^{-1}\text{)} = \text{SOC (kg kg}^{-1}\text{)} \times \text{bulk density (Mg m}^{-3}\text{)} \times \text{soil thickness(m)} \times 10{,}000 \text{ (ha m}^{-2}\text{)}$$

As there is a minimal coarse fraction in the study area, there is no need to correct for it.

## 4.4. Results

First, we present the result of predicting SOC using MIR spectra, and then the spatial prediction of SOC content. Moreover, the three maps of SOC (SOC under natural vegetation, current SOC and SOC change) are presented and discussed. We then detailed the spatial pattern of SOC and SOC stock change from different land uses. Finally, we calculate total SOC stock change by different land uses and discuss the overall trend of SOC change.

### 4.4.1. Predicting soil organic carbon with MIR spectra and cubist model

Soil organic carbon was predicted using MIR spectra and the cubist tree model. The results of the 10-fold cross-validation (averaged from the 10 repetitions) are shown in Figure 4-2. The $R^2$ of the prediction is 0.80, with a root mean squared error (RMSE) of 3.07 g $kg^{-1}$. The model was able to predict SOC accurately and concurred with previous studies, which found that MIR spectra can accurately predict SOC (Ng et al., 2019, Santos et al., 2020, Siebielec et al., 2004).

The spectral model predicting SOC was applied to the whole dataset (observation at every 10 cm depth), and the summary statistics of the predicted SOC values are displayed in Table 4-2. The average and standard deviation SOC values by thirteen pedogenons (pedogenon A to M) and three land uses (woodland, pasture, and non-irrigated and irrigated cropping) were calculated. The results show that the SOC content on the topsoil (0-30 cm) was, as expected, larger than other depths (30-60 and 60-100 cm). SOC content in the woodland area was highest (mean 0-30 cm: 12.47 g $kg^{-1}$) followed by soils under pasture (mean 0-30 cm: 10.37 g $kg^{-1}$) and the least SOC in irrigated and non-irrigated cropping regions (mean 0-30 cm: 7.46 g $kg^{-1}$).

Woodlands in pedogenon M had the highest SOC content for all depth intervals (mean all depth: 11.46 g $kg^{-1}$). Pedogenon M is dominated by dark brown vertosol according to the Australian Soil Classification, ASC (Isbell, 2016) or vertisol in USDA Soil Taxonomy. The pedogenon with the smallest SOC content was pedogenon E (mean all depth: 3.12 g $kg^{-1}$), where the soil was a red chromosol in the ASC or alfisol according to Soil Taxonomy.

For pasture, the highest SOC content was in pedogenon M (mean all depth: 9.01 g $kg^{-1}$). In contrast, pedogenon K had the lowest SOC content, 4.01 g $kg^{-1}$ where the soil is a brown vertosol.

The highest SOC in cropping areas (non-irrigated and irrigated cropping areas) was pedogenon M (mean all depths: 9.51 g $kg^{-1}$). Conversely, the SOC content in pedogenon E had the lowest value (mean all depth: 2.18 g $kg^{-1}$), where the soil was a red chromosol.

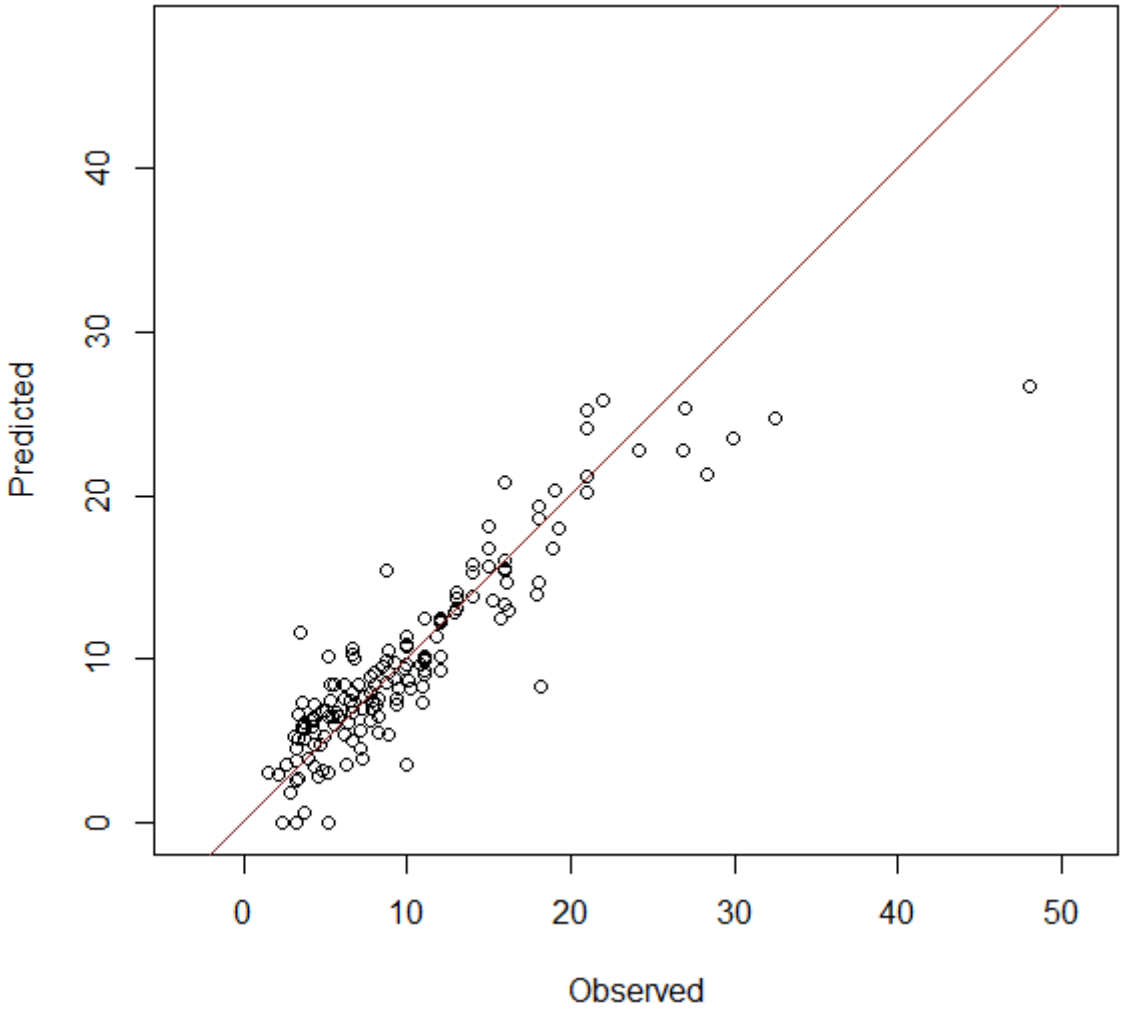



**Figure 4-2. Comparing data from the cross-validation statistics and actual data from the laboratory. The $R^2$ value is 0.80, and The RMSE value is 3.07 g $kg^{-1}$. The red line is a trendline.**

**Table 4-2. Mean and standard deviation values of SOC (g kg-1) by thirteen pedogenons (pedogenon A to M) and three land uses. The letters (W, P and C) represent woodland, pasture, and non-irrigated and irrigated cropping, respectively.**

| Pedogenon | Land use | 0-30cm | | 30-60cm | | 60-100cm | |
|---|---|---|---|---|---|---|---|
| | | Mean | SD | Mean | SD | Mean | SD |
| A | C | 8.55 | 2.07 | 6.11 | 1.02 | 5.38 | 1.41 |
| A | W | 14.04 | 5.17 | 8.73 | 1.84 | 6.88 | 2.35 |
| B | C | 9.73 | 3.47 | 6.44 | 2.41 | 5.26 | 2.27 |
| B | W | 14.12 | 6.99 | 8.87 | 2.25 | 6.80 | 2.97 |
| C | C | 6.88 | 1.57 | 5.29 | 1.12 | 4.44 | 1.34 |
| C | W | 9.49 | 5.39 | 5.74 | 0.72 | 4.15 | 0.86 |
| D | C | 6.85 | 1.31 | 6.46 | 2.27 | 4.73 | 2.02 |
| D | P | 11.97 | 8.71 | 5.33 | 0.88 | 4.85 | 0.37 |
| D | W | 8.86 | 3.64 | 6.72 | 1.90 | 2.68 | 2.05 |
| E | C | 4.52 | 2.80 | 1.08 | 0.49 | 0.95 | 0.60 |
| E | W | 6.66 | 6.26 | 1.32 | 0.62 | 1.39 | 0.65 |
| F | C | 5.29 | 2.16 | 3.36 | 0.79 | 1.88 | 0.71 |
| F | P | 8.22 | 3.71 | 4.57 | 0.91 | 3.21 | 0.68 |
| G | C | 5.31 | 1.26 | 4.28 | 1.41 | 4.07 | 1.70 |
| G | P | 9.11 | 4.36 | 5.00 | 0.31 | 4.22 | 0.56 |
| H | C | 7.44 | 2.50 | 5.91 | 2.47 | 5.94 | 3.97 |
| H | P | 10.58 | 5.30 | 6.53 | 0.57 | 4.38 | 1.06 |
| H | W | 10.17 | 1.61 | 6.43 | 1.12 | 4.54 | 0.66 |
| I | C | 5.80 | 2.56 | 3.30 | 1.34 | 2.56 | 1.78 |
| I | W | 20.27 | 11.16 | 7.26 | 5.49 | 2.97 | 2.52 |
| J | C | 7.96 | 2.57 | 5.22 | 0.72 | 3.49 | 0.92 |
| J | P | 11.39 | 4.04 | 7.00 | 0.91 | 4.39 | 1.95 |
| J | W | 14.30 | 6.94 | 7.79 | 3.17 | 5.35 | 2.65 |
| K | C | 6.41 | 3.52 | 2.68 | 1.75 | 1.86 | 1.88 |
| K | P | 8.73 | 3.06 | 2.50 | 1.40 | 0.80 | 0.14 |
| K | W | 9.33 | 4.99 | 1.91 | 0.92 | 1.52 | 0.86 |
| L | C | 9.68 | 3.41 | 8.32 | 1.24 | 6.32 | 1.72 |
| L | W | 15.52 | 6.84 | 10.00 | 1.96 | 7.73 | 2.30 |
| M | C | 12.61 | 2.94 | 8.84 | 1.06 | 7.08 | 1.78 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **M** | **P** | 12.60 | 2.71 | 8.90 | 1.18 | 5.53 | 1.43 |
| **M** | **W** | 14.46 | 6.41 | 10.68 | 2.67 | 9.25 | 2.15 |

### 4.4.2. Maps of SOC content and SOC change

The SOC contents at 10 cm depth intervals were spatialised using DSM techniques utilising the neural networks model. To improve the accuracy of the model, different numbers of hidden layers and nodes were trialled using cross-validation and models with the highest $R^2$ values across all depths were selected. One hidden layer and three nodes show the highest $R^2$ pattern in all depths.

Models with the optimal hyperparameters were trained to predict SOC under native vegetation and current state from 0 – 1 m at 10 cm intervals. Table 4-3 shows the accuracy of the models from the bootstrap validation. The average $R^2$ of all depths (under native vegetation and current) are 0.31. However, the pattern of $R^2$ is different with different depths. The $R^2$ under native vegetation shows more variation than the current one, and the $R^2$ of topsoil under native vegetation is generally lower than subsoil). The maximum value of $R^2$ from both SOC maps was 0.45, while 0.18 was the minimum value.

**Table 4-3. the accuracy ($R^2$) of SOC maps (under native vegetation and current condition) using multiple outputs neural networks models validated from the our-of-bag bootstrap samples.**

| Depth (cm) | Under native vegetation | Current |
|---|---|---|
| **0-10** | 0.18 | 0.31 |
| **10-20** | 0.31 | 0.42 |
| **20-30** | 0.19 | 0.30 |
| **30-40** | 0.23 | 0.30 |
| **40-50** | 0.27 | 0.30 |
| **50-60** | 0.34 | 0.31 |
| **60-70** | 0.38 | 0.34 |
| **70-80** | 0.40 | 0.27 |
| **80-90** | 0.45 | 0.32 |
| **90-100** | 0.38 | 0.27 |

Using the calibrated DSM models, maps of SOC content under native vegetation and current states were generated. Subsequently, the difference map (SOC current – SOC native) for each depth interval was calculated. Figures 3 and 4 show an example of maps of SOC at two depths (Figure 4-3: 0-10 cm and Figure 4-4: 30-40 cm), and maps for every 10 cm depth interval (down to 1 m) are given in Appendix 4-1.

The map of SOC under native vegetation estimates the soil condition without intensive agricultural practice. As expected, SOC content under native vegetation was higher than the current content at all depths. Nevertheless, significant differences in SOC with land uses were only observed for depths <30 cm due to intensive cultivation on the topsoil.

Under native vegetation, the SOC content of the surface soils (0-10 cm) (Figure 4-3) in the whole study area was above 10 g $kg^{-1}$, with a mean of 18.3 g $kg^{-1}$ (standard deviation (SD) 3.23 g $kg^{-1}$) and an interquartile range/IQR (16.05 g $kg^{-1}$ to 19.10 g $kg^{-1}$). The map of current SOC shows that mean surface soils (0-10 cm) under pasture was 15.27 g $kg^{-1}$ (IQR = 5.11 g $kg^{-1}$), dryland cropping was 12.39 g $kg^{-1}$ (IQR = 4.18 g $kg^{-1}$), while under irrigated cropping dropped to 10.17 g $kg^{-1}$ (IQR = 1.74 g $kg^{-1}$). As shown in Figure 4-3, many cropping areas had topsoil OC content below 10 g $kg^{-1}$ (about 67 % of the cropping area). The topsoil change map (difference in current SOC and under native vegetation) shows that SOC content in most cropping areas had decreased by -5.83 g $kg^{-1}$ (IQR=1.17 g $kg^{-1}$).

Figure 4 shows SOC content of the subsoil (30-40 cm). The mean SOC value under native vegetation was 6.80 g $kg^{-1}$ (SD 1.67 g $kg^{-1}$). In contrast, SOC of the current condition had a similar mean of 6.13 g $kg^{-1}$ (SD 1.82 g $kg^{-1}$). The mean values of SOC by land use were 7.45, 6.57, 5.89, 5.51 g $kg^{-1,}$ and SD values were 0.26, 0.16, 0.12 and 0.11 g $kg^{-1}$ (order of land use: woodland, pasture, non-irrigated, cropping and irrigated cropping). Their IQR values were 2.77, 3.38, 2.83 and 1.07 g $kg^{-1}$. There was a minimal change of SOC in the subsoil.

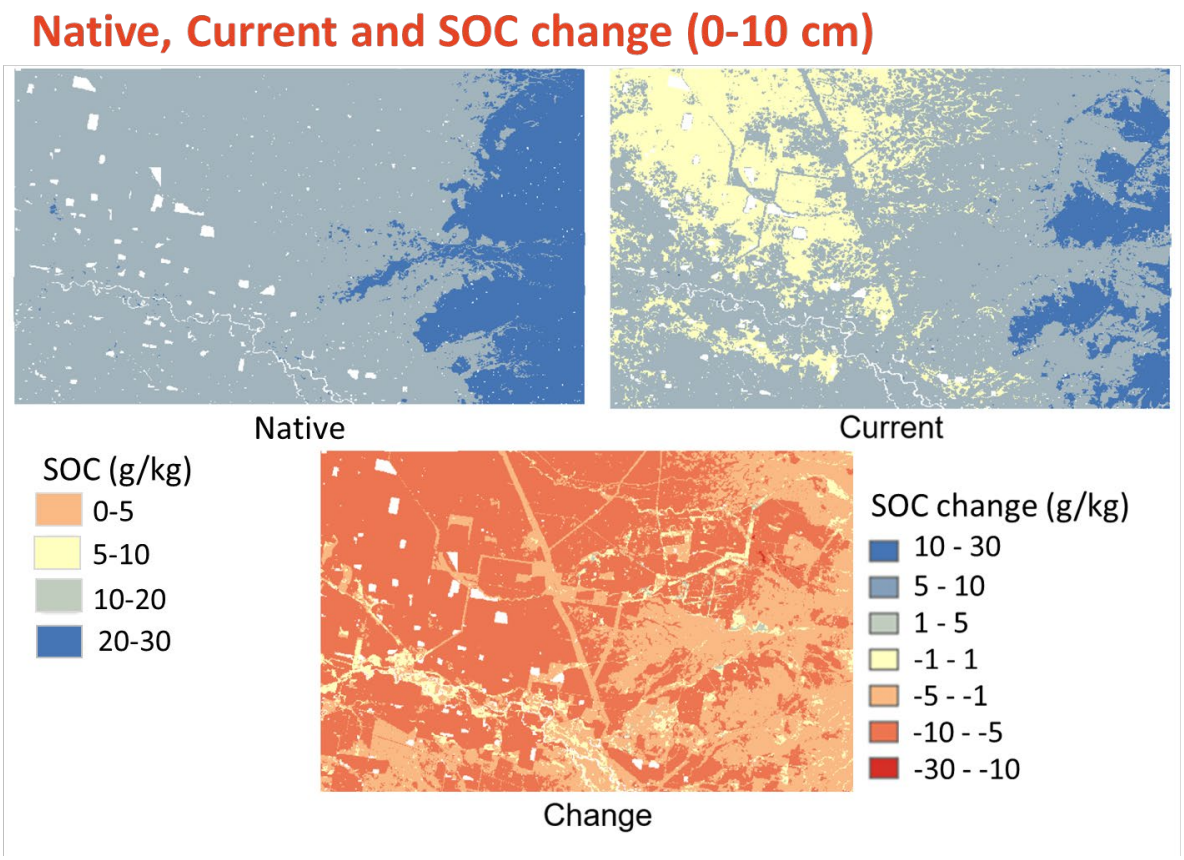



**Figure 4-3. SOC under Native, Current, and Change maps at 0 – 10 cm depth interval. Native: the map of SOC content under native vegetation. Current: the map of current SOC content. Change: the map of SOC change (Current – Native).**

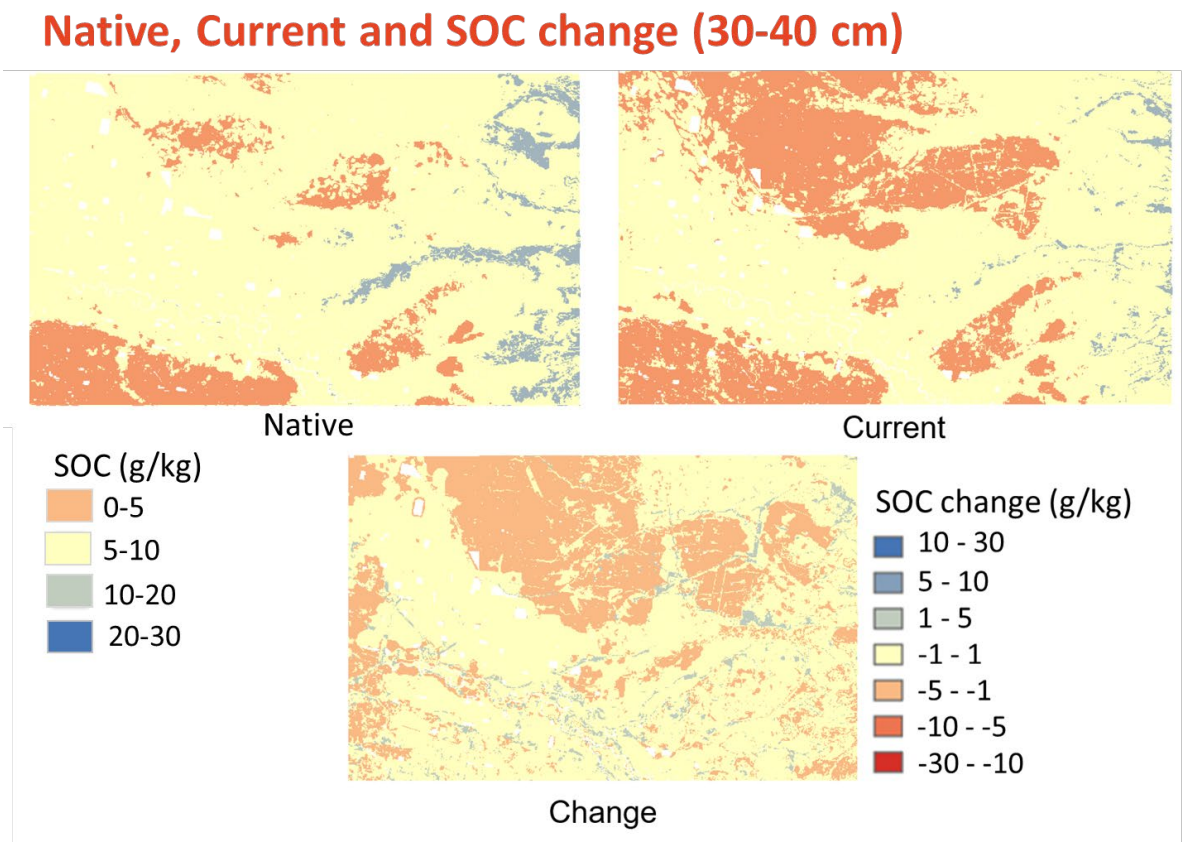



**Figure 4-4. SOC under Native, Current, and Change maps at 30 – 40 cm depth interval. Native: the map of SOC content under native vegetation. Current: the map of current SOC content. Change: the map of SOC change (Current – Native).**

### 4.4.3. SOC content change

The mean values of current SOC and SOC change by different land uses were calculated for each 10 cm depth interval map. Figure 4-5A shows the mean values of the current SOC contents by four land uses. As expected, SOC is highest in woodland, followed by pasture, dryland, and irrigated cropping. Figure 4-5B shows the SOC change calculated from the map of soils under current conditions and native vegetation. SOC change in woodland was not included because the SOC change was insignificant. Overall, as discussed in the previous section, soils under irrigated cropping had the most significant SOC loss, with content at 0-10 cm having a decrease of 38% (compared to SOC under native vegetation), followed by dryland cropping (30% loss) and pasture (19% loss). Additionally, SOC loss was greatest in the surface soils and decreased exponentially with depth.

The loss of SOC in the pasture was mainly in the top 30 cm (average 1.82 g $kg^{-1}$ or 13% loss compared to soils under native vegetation). Under dryland agriculture, the loss of SOC was determined down to 50 cm depth (representing a total of 95% loss compared to soils under native vegetation). There was an average loss of 2.33 g $kg^{-1}$, or 18% decrease compared to soils under native vegetation (0-50 cm). Under irrigated agriculture, the loss of SOC was also down to 50 cm depth (representing 96% total loss compared to soils under native vegetation). There was an average loss of 2.49 g $kg^{-1}$, or 19% decrease compared to soils under native vegetation (0-50 cm).

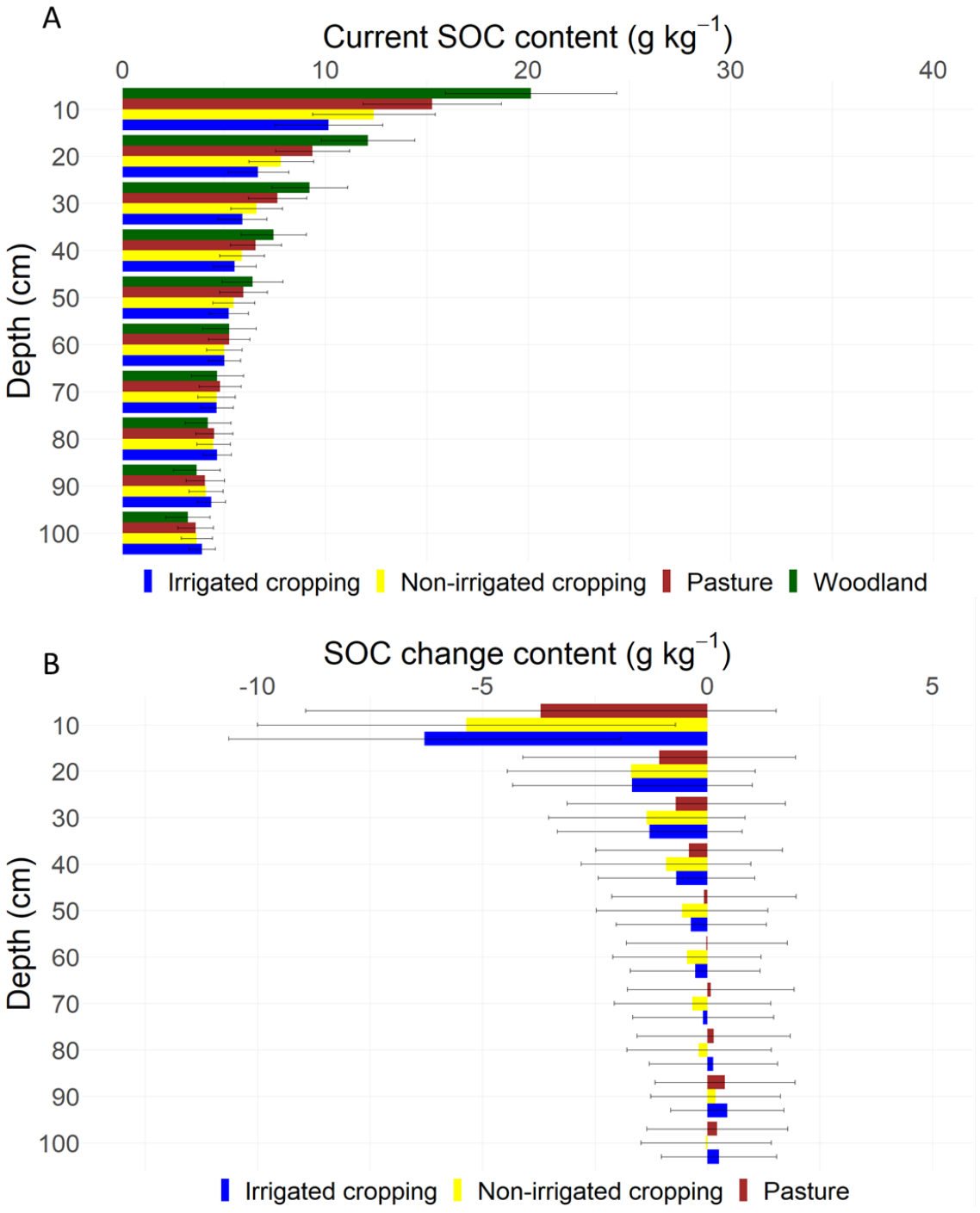



**Figure 4-5. Current SOC content and SOC change (unit: g kg$^{-1}$) for every 10 cm depth until 1 m. The green, brown, yellow and blue bars represent woodland, pasture, non-irrigated cropping and irrigated cropping, respectively. The black lines are standard deviation values.**

### 4.4.4. SOC stock change by pedogenon and land use

To understand the pattern and magnitude of change, SOC stock change was calculated by pedogenons and land use by unit area (t C $ha^{-1}$ ) and for their total area (Gg C). All pedogenons show a decrease in SOC stock content with cropping of at least 5 t C $ha^{-1}$ (Figure 4-6). In relative terms, pedogenon A and J had a larger SOC loss compared to other pedogenons under agriculture, with values ranging between 17-21 t C $ha^{-1}$. In contrast, pedogenon M had the smallest loss, between 4-6 t C $ha^{-1}$. The highest SOC stock loss under the pasture area was also pedogenon A and J (10.19 and 14.98 t C $ha^{-1}$, respectively ), while the lowest one was under pedogenon B (1.54 t C $ha^{-1}$). Few pedogenons show a small positive SOC stock change at deeper depth. However, the uncertainty of prediction at these depths was quite large.

Pedogenon A consists of black vertosol, and the main land use in the area was irrigated cropping, which produced cotton intensively. Additionally, pasture in pedogenon A showed a decline in SOC of approximately 10 t C $ha^{-1}$. The soil class of Pedogenon J is also a black vertosol, and the main land use is non-irrigated cropping.

Figure 4-6 shows the total SOC stock change (in Gg C) for each land use by pedogenon class. Pedogenon G, a black vertosol that occupied a large area of 21,462 ha, showed the largest carbon stock loss (-302.86 Gg C (SD: 127.68 Gg C) under irrigated cropping). Although the SOC change (in terms of t C $ha^{-1}$) was relatively significant under irrigated cropping area in pedogenon H (-8.38 t C $ha^{-1}$) and M (-6.00 t C $ha^{-1}$), due to their relatively small area (< 3 ha), the SOC stock change was small. Similarly, some pedogenons with moderate to significant SOC loss per unit area under dryland cropping, like pedogenon I (-12.31 t C $ha^{-1}$) and K (-13.77 t C $ha^{-1}$), resulted in small total SOC stock loss.

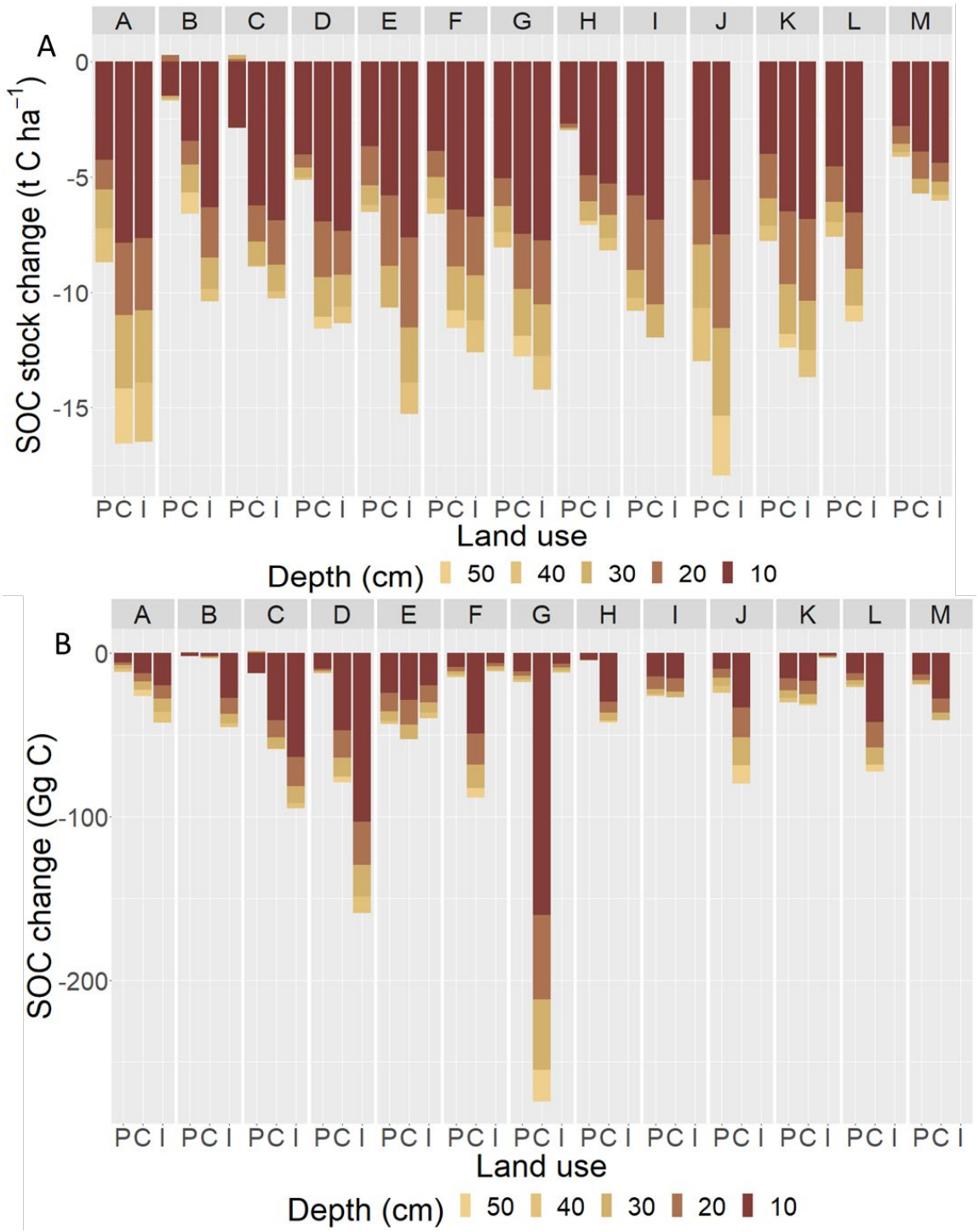



**Figure 4-6. A: SOC stock change (t C ha$^{-1}$) and B: SOC change (Gg C) (SOC stock change (t C/ha) multiplied by area (ha)) of pedogenon and land use. The letters (P, C and I) represent pasture, non-irrigated, and irrigated cropping, respectively. The depth of the soil profile is every 10 cm down to 50 cm.**

### 4.4.5. Total SOC loss by land uses.

Finally, we can calculate the total SOC loss due to different agricultural activities in the Edgeroi area, as described in Figure 4-7.

Non-irrigated cropping area (total area: 79,070 ha) shows the largest carbon stock loss (967.91 Gg, SD: 169.20 Gg C), and the subsequent largest loss was irrigated cropping (419.83 Gg C, SD: 88.62 Gg C ) (35,284 ha). The smallest SOC loss among the three agricultural land uses was in the pasture areas (38,989 ha) (238.49 Gg C, SD: 87.42 Gg C). The SOC change in the woodland area was negligible, so it is not included in Figure 4-7.

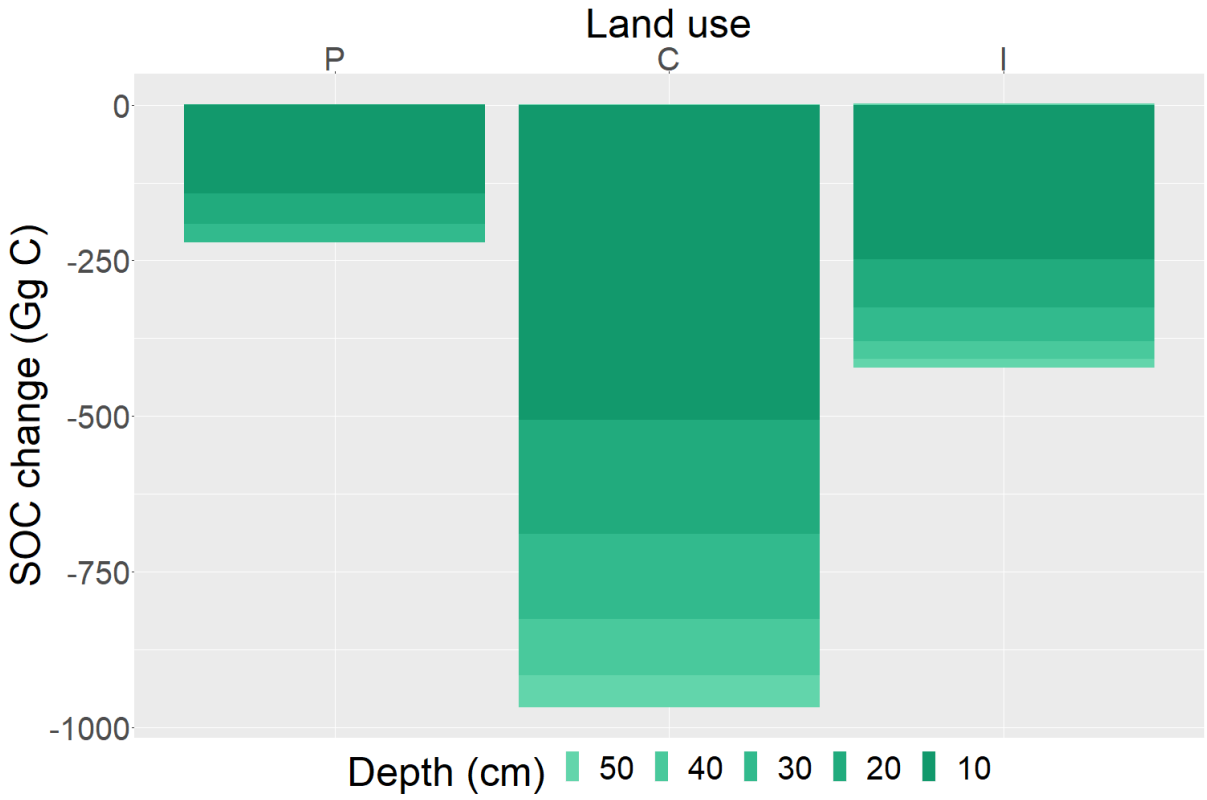



**Figure 4-7. the overall SOC stock loss (Gg C by land use). The letters (P, C and I) represent pasture, non-irrigated, and irrigated cropping, respectively. The depth of the soil profile is separated into five depths.**

### 4.4.6. SOC targets for each of the pedogenons.

The amount of SOC that can be regained on cropping soils via improved management practices for each pedogenon can be calculated. The SOC stock target for topsoil (0-30 cm) was calculated by averaging three datas (0-10, 10-20 and 20–30 cm) and subtracting the current SOC stock under cropping (non-irrigated and irrigated cropping areas) from SOC stock in native vegetation areas. The results of the average value by each land use and pedogenon is given in Appendix 4-2.

Figure 4-8 shows the SOC stock target for each of the pedogenons, and they can be grouped into four groups. The first group included pedogenon J and A, which had a high SOC stock loss. The target SOC stock was 5 t C $ha^{-1}$, where the soil class of both pedogenons is a black vertosol and the dominant native vegetation before cropping practice is native grassland.

The following group (pedogenon G, K, I, E and L) shows the second-highest SOC loss with an average SOC target of 4 t C $ha^{-1}$. These pedogenons are diverse (pedogenon G and L: black vertosol, pedogenon K: brown vertosol, pedogenon I: dark brown tenosol and pedogenon E: red chromosol), but the native vegetation under all five pedogenons before human activities is dry sclerophyll forest except pedogenon G (native grassland). These pedogenons are closely located to each other at the top right side of the study area except pedogenon E, located at the bottom left side.

Pedogenon F, D, C and M represent the third group, the third highest SOC loss, with an average value was 3 t C $ha^{-1}$. The soil type of the four pedogenons is vertosol, and all their parent material is alluvium, located at the top left side of the study area, except pedogenon M (at the

bottom right). The last group explains the lowest SOC loss (pedogenon H and B), averaging 1.5 t C $ha^{-1}$. Their main vegetation is native grassland; their soil type is dark brown vertosol.

With the data on SOC stock change, it is possible to estimate possible SOC target values. Improving the farming management approaches where pedogenons A, G, J, K, I, and L should be the focus area with a potential of 4-5 t C $ha^{-1}$.

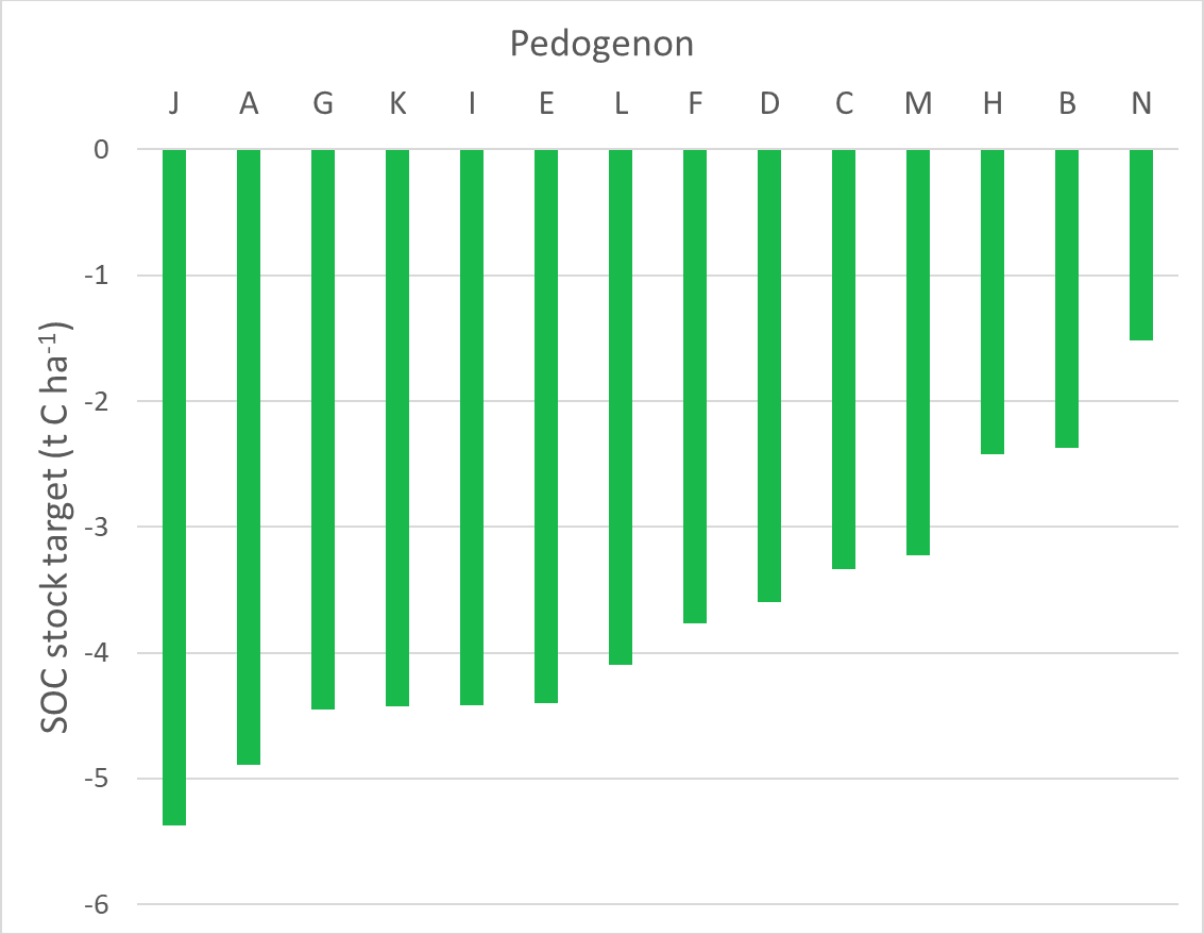



**Figure 4-8. SOC stock target (t C ha-1) by pedogenons for cropping soils.**

## 4.5. Discussion

This section discusses the efforts to improve the uncertainties in the steps of modelling and mapping. We also discuss SOC change by pedogenons and land use.

### 4.5.1. Modelling SOC at multiple depths using the neural networks model

The neural network model was utilised for predicting SOC contents from the soil surface until the depth of 1 m at 10 cm intervals. We initially utilised the cubist regression tree models to create separate models for each depth interval. However, creating independent models for each depth resulted in discontinuity of the prediction as the regression tree models used different covariates for each depth interval. For example, gamma data was heavily used in predicting the 0-10 cm model, while the topography covariates strongly affected the SOC maps of 10-20 cm. As we know, SOC content is highest in topsoil and decreases exponentially with depth (Angst et al., 2018); creating individual depth maps makes it difficult to observe the SOC trends with depth. Thus, a single model predicting SOC at multiple depths is preferable. One option is including soil depth as a covariate in the machine learning models (e.g., (Roudier et al., 2020)). However, Ma et al. (2021) concluded that care must be taken when using depth as a covariate in tree-based models, which could create discontinuous depth trends. As we regularly observed SOC every 10 cm depth interval, we used the neural networks model (MLP model) with multiple outputs to resolve the problem (Padarian et al., 2019). A single MLP model predicted SOC content at 10 different depth intervals and can continuously model the depth trend data.

An example is shown in Figure 4-9, where the MLP model predicts SOC in woodland and cropping areas of pedogenon F. As in all machine learning models, the MLP model requires tuning the number of hidden layers and nodes. The tuning needs to ensure that the model does not overfit but produces accurate predictions.

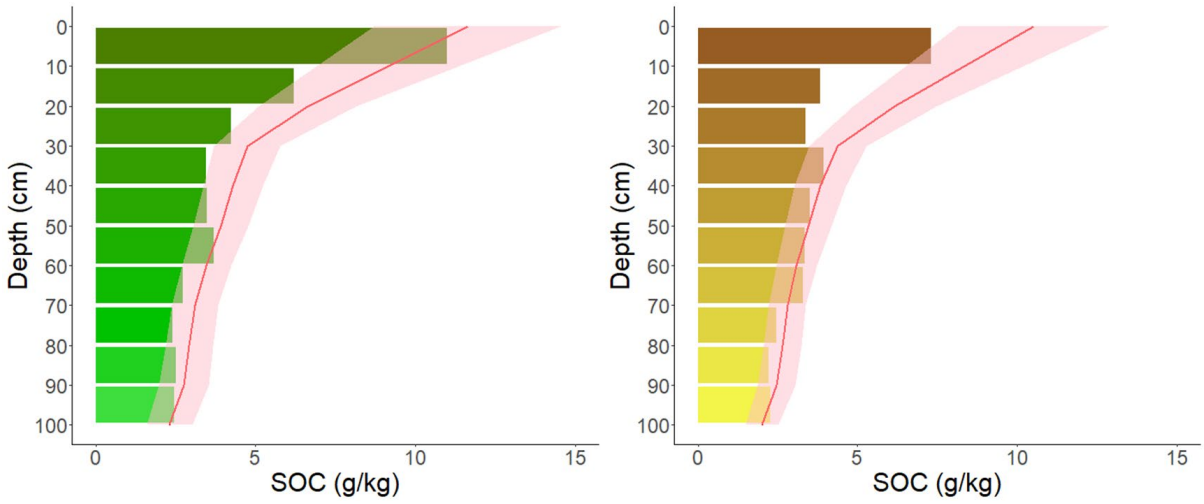



**Figure 4-9. Bar graph of observations with depth and neural network predictions (lines). The data of both graphs were derived from pedogenon F. The green bar group (left side of the figure) represents the woodland area, while the brown bar graph (right side of the figure) represents the cropping area.**

Here we model SOC under native vegetation and current conditions separately to calculate soil change. This approach was based on sampling from each pedogenon, i.e., soil type formed by the same soil forming factor for a given reference time. After that, data collected under natural vegetation or less disturbed conditions was compared with soils under cropping or agricultural activities in each pedogenon with a space-for-time substitution approach as it is often applied to investigate soil change during landform development (Li et al., 2022). The pedogenon concept supported this study to find similar soil types under the same process of soil formation, and the concept of space for time substitution was used to segregate the different types of land use based on intensive human activity. Combining the two concepts allowed us to calculate SOC content under native vegetation and the current condition.

We acknowledge the high uncertainty of prediction due to the limited time and resources, which limits the number of genosoil samples and also the number of samples for the large area. We assume the genosoil samples (under current native vegetation) have not changed since European settlement. We know that SOC stock is affected by climate change. In addition, soils under native vegetation are also not evenly distributed in the geographical and covariate space, hence the high uncertainty in prediction. Nevertheless, the method proposed in this study shows the possibility of investigating SOC change to recreate soil conditions under native vegetation. This new method addresses the limitation of the current DSM method of space for time substitution where a DSM model was created using current observations and to predict soil condition under the native condition, only the land use in the calibrated model was substituted (e.g. (Li et al., 2022)).

### 4.5.2. SOC change as a function of pedogenons and land use

Our results show that cropping on all pedogenons resulted in the loss of SOC compared to soils under native vegetation. We can derive this conclusion by sampling from pedogenons. The soil properties from Chapter 3 were analysed (clay, sand, cation exchange capacity, pH, and organic carbon) at every 10 cm depth increment for the pedogenons in this study area. They demonstrate that pedogenon significantly explained the variation in soil properties. They also show that soil properties on genosoils had twice the variation of phenosoils. The SOC maps produced in this study reaffirmed this trend for SOC. Under native vegetation, SOC content of the surface soils (0-30 cm) had a mean of 12.10 g $kg^{-1}$ (variance = 0.81 $g^2$ $kg^{-2}$). The map of current SOC shows that mean surface soils under pasture were 10.76 g $kg^{-1}$ (variance = 0.57 $g^2$ $kg^{-2}$), dryland cropping was 8.93 g $kg^{-1}$ (variance = 0.44 $g^2$ $kg^{-2}$), while under irrigated cropping dropped to 7.58 g $kg^{-1}$ (variance = 0.36 $g^2$ $kg^{-2}$). These results indicate that agricultural activities reduced the SOC and homogenised the variation of surface SOC.

The study area has been converted to agricultural areas since European settlers came in the early 1800s. Initially, the area was used for grazing, followed by wheat, sorghum, and irrigated cotton farming. Based on the Historical Aerial Photography (from Geoscience Australia), we can visually assess the land use and starting cultivation period. This information can be used to support and interpret the results of SOC change.

Figure 4-10 shows some of the pedogenon A and J areas at different periods. Both areas were already actively cultivated in the 1950s. Until recently, these areas continuously produced wheat and cotton. SOC change (loss) was greatest in irrigated cropping, followed by dryland cropping and pasture. This is mainly reflected in pedogenon A, which had the second largest SOC content loss among other pedogenons (Figure 4-6). The main land use in pedognon A is irrigated cotton, which has been grown for over 60 years (Figure 4-10A and B). The modification of SOC content under the cotton crop areas can be attributed to various effects

(soil tillage intensity, the fallow phase period, etc.). The large decrease in SOC in irrigated cotton was confirmed by the study of Rabbi et al. (2014), who found that irrigated cotton land use in South-eastern Australia showed a negative correlation between SOC content and its fractions. Conteh et al. (1997) found that SOC on clay topsoils (vertosols) under irrigated cotton in the Namoi valley had decreased by 37-48 % compared to nearby reference sites (uncropped).

The SOC loss in the non-irrigated cropping area (pedogenon J) was high compared to other pedogenons. The main crop produced in this area is wheat, and the SOC change can vary significantly due to different management of farming practices. The key management factors for SOC change in wheat fields are the intensity of tillage, removal of stubbles or plant residues and application of nitrogen fertilisers (Wang and Dalal, 2015). So et al. (1999) informed that reduced tillage had half the impact on SOC loss than conventional tillage. The management of stubble can change SOC loss. Burning stubbles had more SOC loss than retaining stubble, with more than two times SOC carbon loss (Chan et al., 1998). The crop rotation can also have a significant impact on SOC loss. Continuous cropping (wheat only) shows substantial SOC loss compared to rotation cropping (So et al., 1999). The high carbon loss in pedogenon J could be due to unsustainable management. Further studies should investigate this pedogenon to minimise SOC loss in the cropping area.

On the other hand, the SOC change of pedogenon M is the smallest. This area is located on steep slopes compared to other sites and was not under intensive agricultural practices.

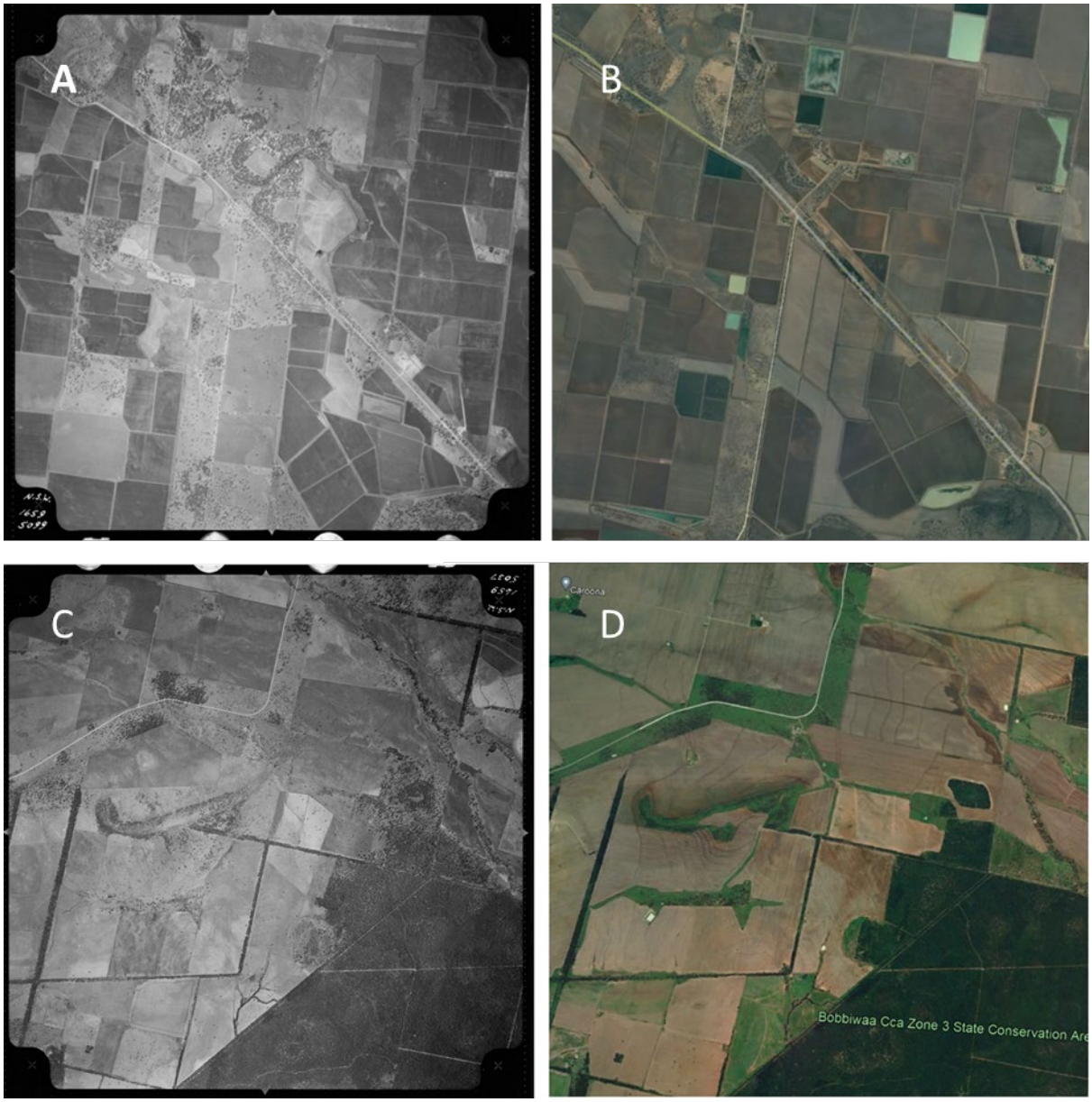



**Figure 4-10. Aerial imagery from Historical Aerial Photography of NSW and recent satellite imageries of the study area. The top two pictures (A and B) are a part of pedogenon A. Picture A was taken in 1969, and picture B was taken in 2018. Pictures C and D are part of pedogenon J, taken in 1959 and (D) in 2021. Both areas had been used for cropping for more than 60 years.**

## 4.6. Conclusion

This paper investigated SOC change (between the SOC content before intensive agriculture and the content of the current stage) using the space for time substitution method and pedogenon concept.

The SOC content at every 10 cm depth interval was predicted using MIR spectra, and the content of SOC was mapped using the predicted SOC content and soil forming factors (soil, organisms, topography and parent material and age). The results conclude that the highest carbon loss was in irrigated cropping areas, followed by non-irrigated cropping, pasture, and woodland. Pedogenon A and J show the highest SOC loss, while pedogenon M shows the lowest. Based on the SOC stock change result, the possible SOC target value can be calculated.

In addition, further investigations can be considered to advance the methodology. With the result of this research and an extra survey of farming management methods from farmers, it is possible to find out which management method can maximise yield and minimise SOC loss. Incorporating farmers' management systems (e.g., the intensity of tillage, the period of fallow, crop rotation, the type and amount of fertiliser application, etc.) can elucidate the cause of SOC change.

The change in SOC stock after 200 years (since European settlement) was calculated from this study, and with this data, it is possible to provide a warning before reaching the point at which yield may rapidly decrease due to soil degradation.

## 4.7. References


ANGST, G., MESSINGER, J., GREINER, M., HÄUSLER, W., HERTEL, D., KIRFEL, K., KÖGEL-KNABNER, I., LEUSCHNER, C., RETHEMEYER, J. & MUELLER, C. W. 2018. Soil organic carbon stocks in topsoil and subsoil controlled by parent material, carbon input in the rhizosphere, and microbial-derived compounds. *Soil Biology and Biochemistry,* 122**,** 19-30.

CAMPBELL, S., MALONE, B., MINASNY, B., NELSON, M. & FAJARDO, M. 2016. *spectroscopy: Functions for Visible and Near Infrared data manipulation* [Online]. Available: https://github.com/Soilsecuritylab/spectroscopy_package [Accessed].

CATTLE, S., KOPPI, A. & MCBRATNEY, A. 1994. The effect of cultivation on the properties of a Rhodoxeralf from the wheat/sheep belt of New South Wales. *Geoderma,* 63**,** 215-225.

CHAN, K., HEENAN, D., GRABSKI, A. & SO, H. 1998. Soil organic carbon sequestration and changes in soil quality under conservation tillage on lighter-textured soils in Australia. *Conservation tillage: can it assist in mitigating the greenhouse gas problem*.

CONTEH, A., BLAIR, G., LEFROY, R. & MACLEOD, D. 1997. Soil organic carbon changes in cracking clay soils under cotton production as studied by carbon fractionation. *Australian Journal of Agricultural Research,* 48**,** 1049-1058.

DEPARTMENT OF PLANNING AND ENVIRONMENT 2018. State Vegetation Type Map: Border Rivers Gwydir / Namoi Region Version 2.0. VIS_ID 4467.

ECO LOGICAL AUSTRALIA 2013. Refinement of vegetation mapping in the Namoi Catchment: Extant and pre-European. Armadale NSW.

EVANS, D., QUINTON, J. N., DAVIES, J. A., ZHAO, J. & GOVERS, G. 2020. Soil lifespans and how they can be extended by land use and management change. 15**,** 0940b2.

FILIPPI, P., MINASNY, B., CATTLE, S. R. & BISHOP, T. F. A. 2016. Chapter Four - Monitoring and Modeling Soil Change: The Influence of Human Activity and Climatic Shifts on Aspects of Soil Spatiotemporally. *In:* SPARKS, D. L. (ed.) *Advances in Agronomy.* Academic Press.

GREEN, D., PETROVIC, J., BURRELL, M. & MOSS, P. 2011. *Water resources and management overview: Namoi catchment*.

ISBELL, R. 2016. *The Australian soil classification*, CSIRO publishing.

KOPITTKE, P. M., BERHE, A. A., CARRILLO, Y., CAVAGNARO, T. R., CHEN, D., CHEN, Q.-L., ROMÁN DOBARCO, M., DIJKSTRA, F. A., FIELD, D. J. & GRUNDY, M. J. 2021. Ensuring planetary survival: the centrality of organic carbon in balancing the multifunctional nature of soils. *Critical Reviews in Environmental Science and Technology***,** 1-17.

KUHN, M., WESTON, S., KEEFER, C., COULTER, N. & QUINLAN, R. 2014. Cubist: rule-and instance-based regression modeling. *R package version 0.0,* 13.

LAL, R. 2013. Intensive agriculture and the soil carbon pool. *Journal of Crop Improvement,* 27**,** 735-751.

LI, N., ZHOU, S. & MARGENOT, A. J. 2022. From prairie to crop: Spatiotemporal dynamics of surface soil organic carbon stocks over 167 years in Illinois, USA. *Science of The Total Environment***,** 159038.

LUO, Z., WANG, E. & SUN, O. J. J. G. 2010. Soil carbon change and its responses to agricultural practices in Australian agro-ecosystems: a review and synthesis. 155**,** 211-223.

MA, Y., MINASNY, B., MCBRATNEY, A., POGGIO, L. & FAJARDO, M. 2021. Predicting soil properties in 3D: Should depth be a covariate? *Geoderma,* 383**,** 114794.

MANASWI, N. K., MANASWI, N. K. & JOHN, S. 2018. *Deep learning with applications using python*, Springer.

MCBRATNEY, A. B., SANTOS, M. M. & MINASNY, B. 2003. On digital soil mapping. *Geoderma,* 117**,** 3-52.

MCGARRY, D., WARD, W. T., MCBRATNEY, A. B. & SOILS, C. D. O. 1989. *Soil Studies in the Lower Namoi Valley: Methods and Data. 1, The Edgeroi Data Set*, CSIRO Division of Soils.

MIRZAEITALARPOSHTI, R., DEMYAN, M. S., RASCHE, F., CADISCH, G. & MÜLLER, T. 2017. Mid-infrared spectroscopy to support regional-scale digital soil mapping on selected croplands of South-West Germany. *CATENA,* 149**,** 283-293.

NARRABRI SHIRE COUNCIL. 2021. *Historical Walks* [Online]. [Accessed].

NG, W., MINASNY, B., MONTAZEROLGHAEM, M., PADARIAN, J., FERGUSON, R., BAILEY, S. & MCBRATNEY, A. B. 2019. Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra. *Geoderma,* 352**,** 251-267.

PADARIAN, J., MINASNY, B. & MCBRATNEY, A. B. 2019. Using deep learning for digital soil mapping. *Soil,* 5**,** 79-89.

PADARIAN, J., STOCKMANN, U., MINASNY, B. & MCBRATNEY, A. 2022. Monitoring changes in global soil organic carbon stocks from space. *Remote Sensing of Environment,* 281**,** 113260.

POST, W. M. & KWON, K. C. 2000. Soil carbon sequestration and land-use change: processes and potential. 6**,** 317-327.

RABBI, S. F., TIGHE, M., COWIE, A., WILSON, B. R., SCHWENKE, G., MCLEOD, M., BADGERY, W. & BALDOCK, J. 2014. The relationships between land uses, soil management practices, and soil carbon fractions in South Eastern Australia. *Agriculture, Ecosystems & Environment,* 197**,** 41-52.

ROMÁN DOBARCO, M., MCBRATNEY, A., MINASNY, B. & MALONE, B. 2021. A modelling framework for pedogenon mapping. *Geoderma,* 393**,** 115012.

ROSSITER, D. G. & BOUMA, J. 2018. A new look at soil phenoforms – Definition, identification, mapping. *Geoderma,* 314**,** 113-121.

ROUDIER, P., BURGE, O. R., RICHARDSON, S. J., MCCARTHY, J. K., GREALISH, G. J. & AUSSEIL, A.-G. 2020. National scale 3D mapping of soil pH using a data augmentation approach. *Remote Sensing,* 12**,** 2872.

SANTOS, U. J. D., DEMATTÊ, J. A. D. M., MENEZES, R. S. C., DOTTO, A. C., GUIMARÃES, C. C. B., ALVES, B. J. R., PRIMO, D. C. & SAMPAIO, E. V. D. S. B. 2020. Predicting carbon and nitrogen by visible near-infrared (Vis-NIR) and mid-infrared (MIR) spectroscopy in soils of Northeast Brazil. *Geoderma Regional,* 23**,** e00333.

SIEBIELEC, G., MCCARTY, G. W., STUCZYNSKI, T. I. & REEVES III, J. B. 2004. Near- and mid-infrared diffuse reflectance spectroscopy for measuring soil metal content. *Journal of Environmental Quality,* 33**,** 2056-2069.

SO, H., DALAL, R., CHAN, K., MENZIES, N. & FREEBAIRN, D. Potential of conservation tillage to reduce carbon dioxide emission in Australian soils. Sustaining the Global Farm. Selected paper for the 10 th International Soil Conservation Organization meeting held May, 1999. 24-29.

TYE, A., ROBINSON, D. & LARK, R. 2013. Gradual and anthropogenic soil change for fertility and carbon on marginal sandy soils. *Geoderma,* 207**,** 35-48.

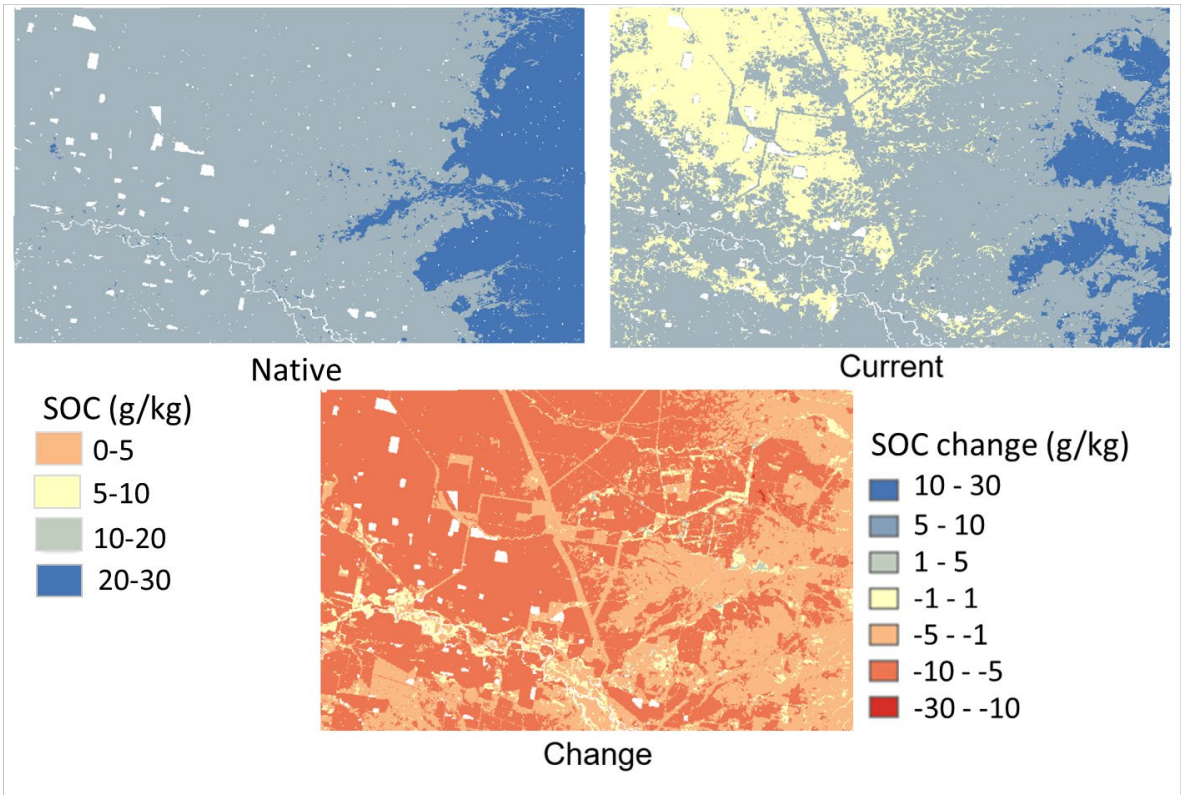
WANG, W. & DALAL, R. C. 2015. Nitrogen management is the key for low-emission wheat production in Australia: A life cycle perspective. *European Journal of Agronomy,* 66**,** 74-82.

WARD, W. 1999. Soils and landscapes near Narrabri and Edgeroi, NSW, with data analysis and using fuzzy k-means.

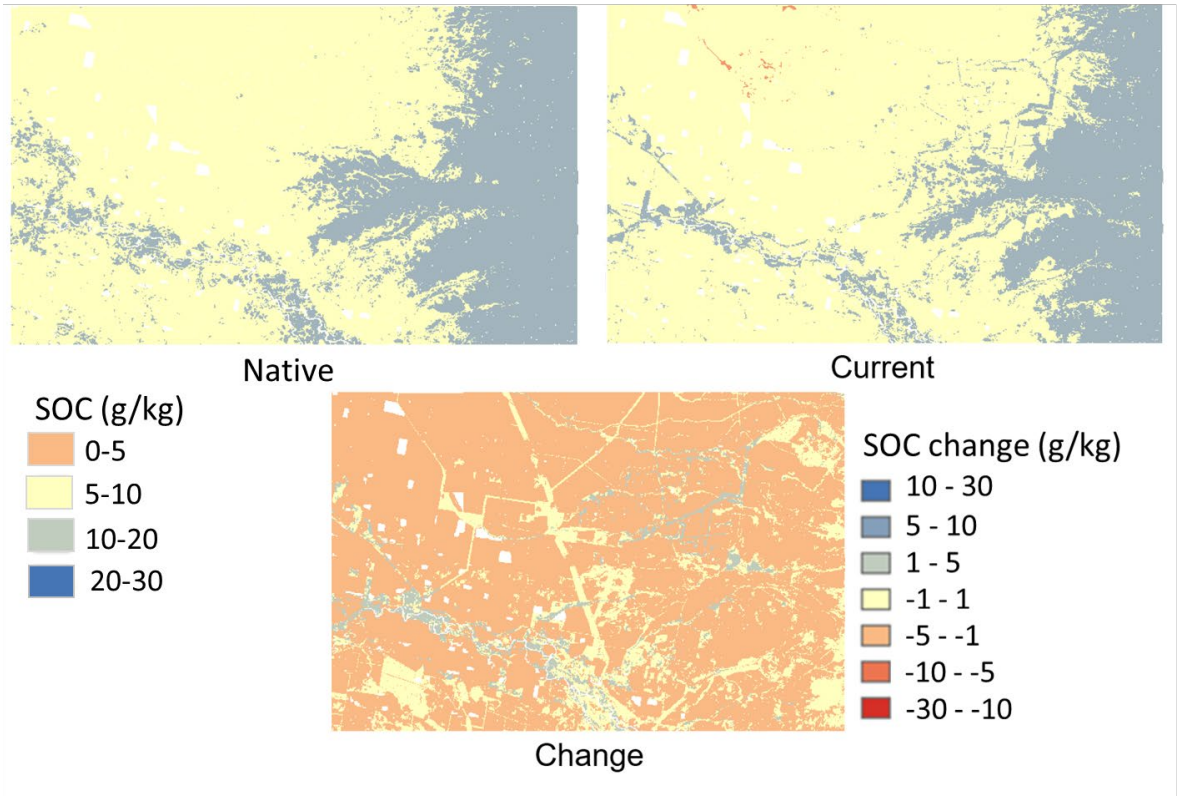
## 4.8. Appendices

**Appendix 4-1. SOC under Native, Current, and Change maps. Native: the map of SOC content under native vegetation. Current: the map of current SOC content. Change: the map of SOC change (Current – Native).**

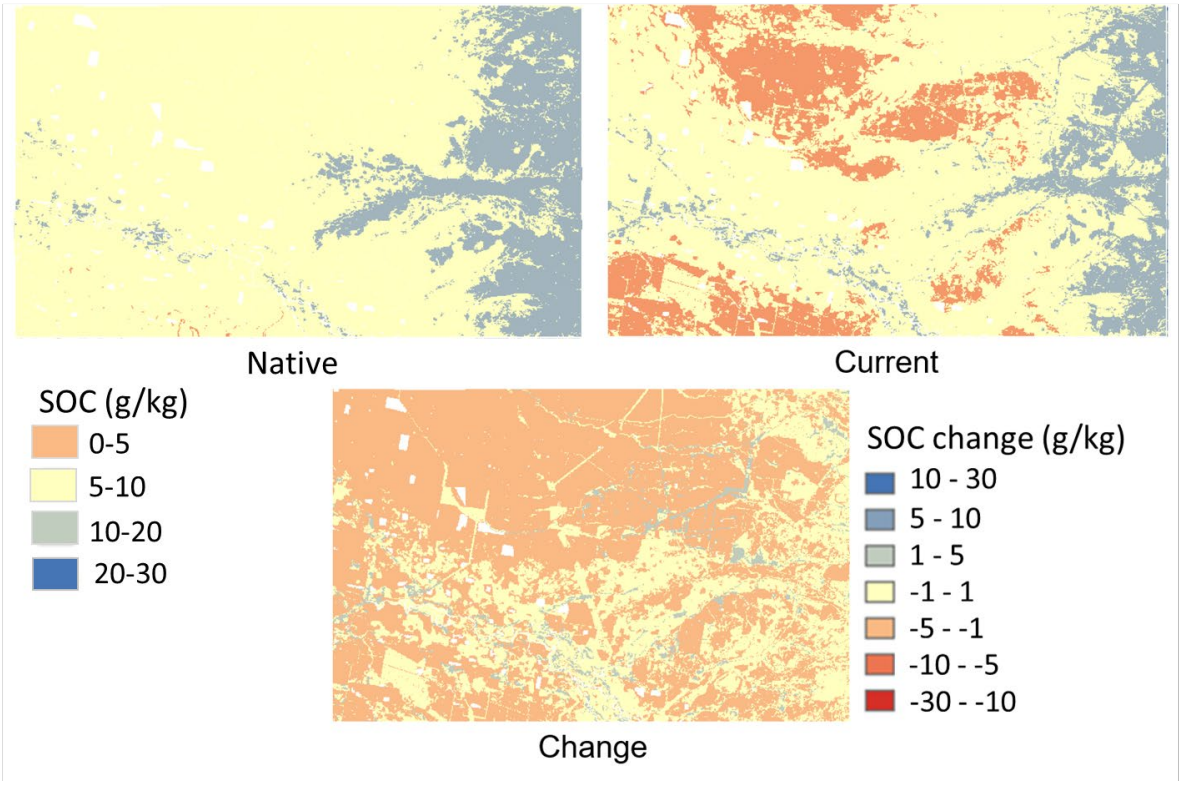
## Native, Current and SOC change (0-10 cm)



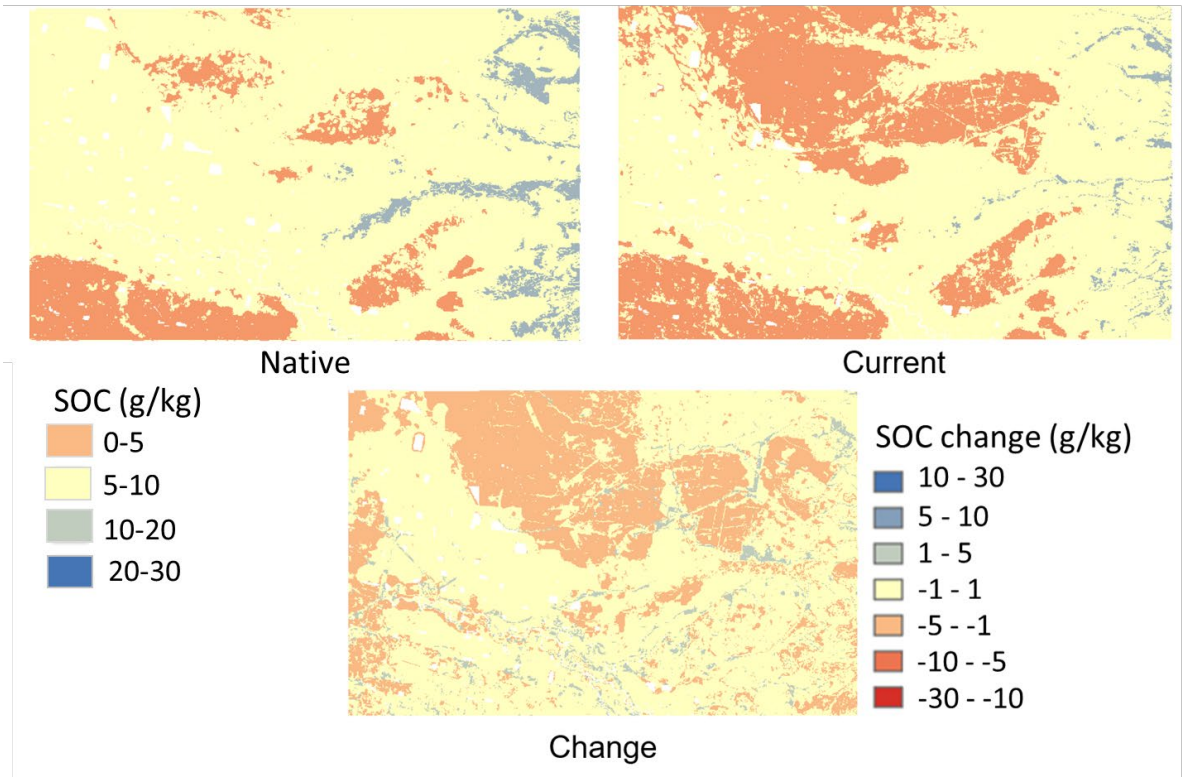

## Native, Current and SOC change (10-20 cm)



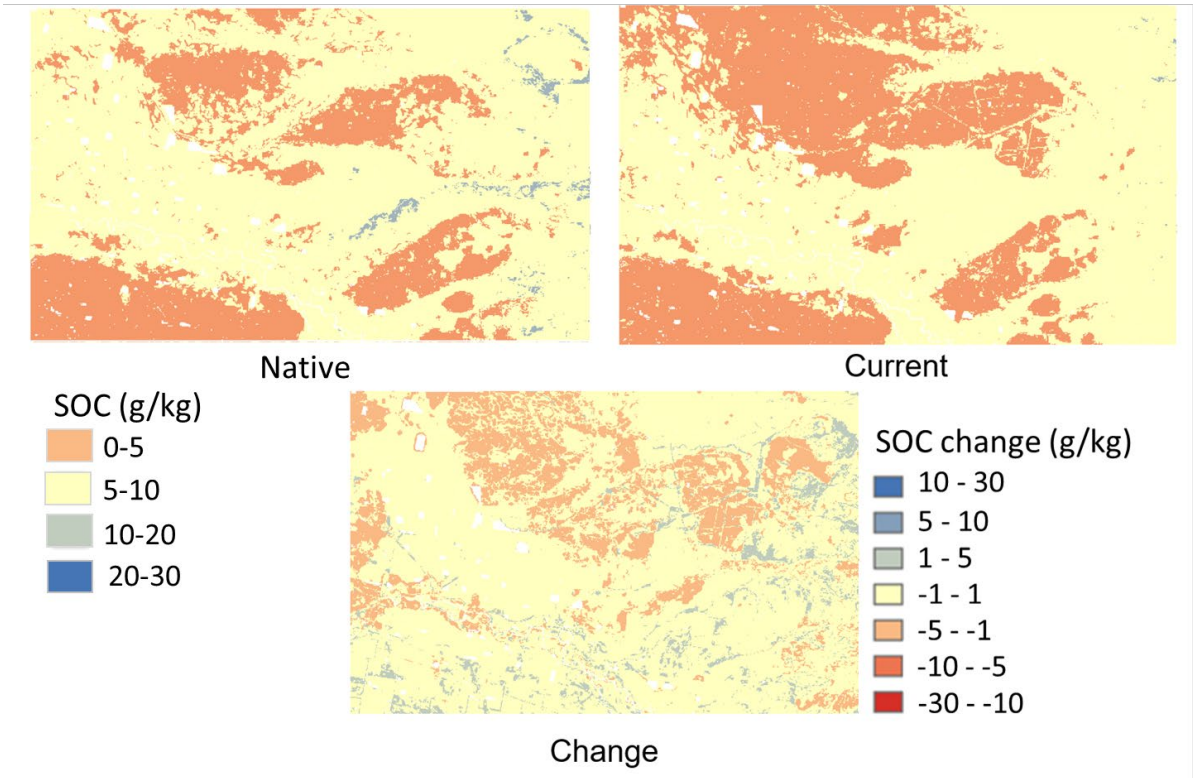
## Native, Current and SOC change (20-30 cm)



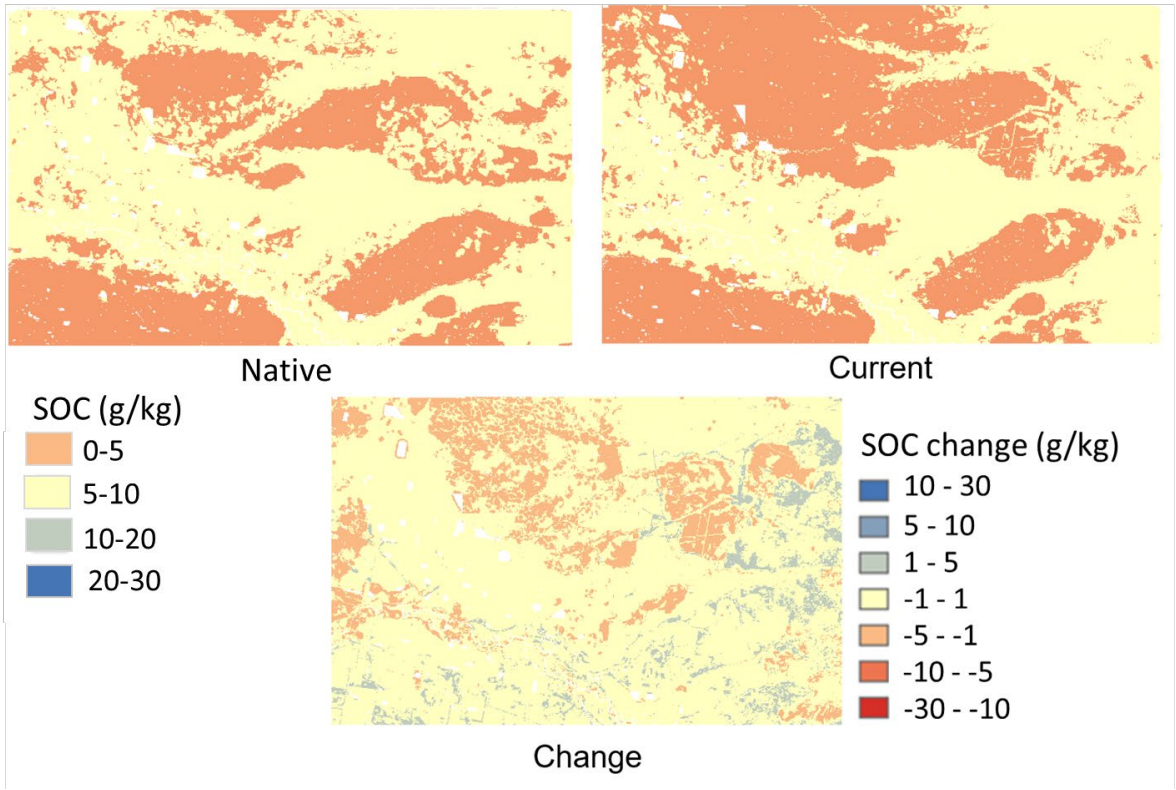

## Native, Current and SOC change (30-40 cm)



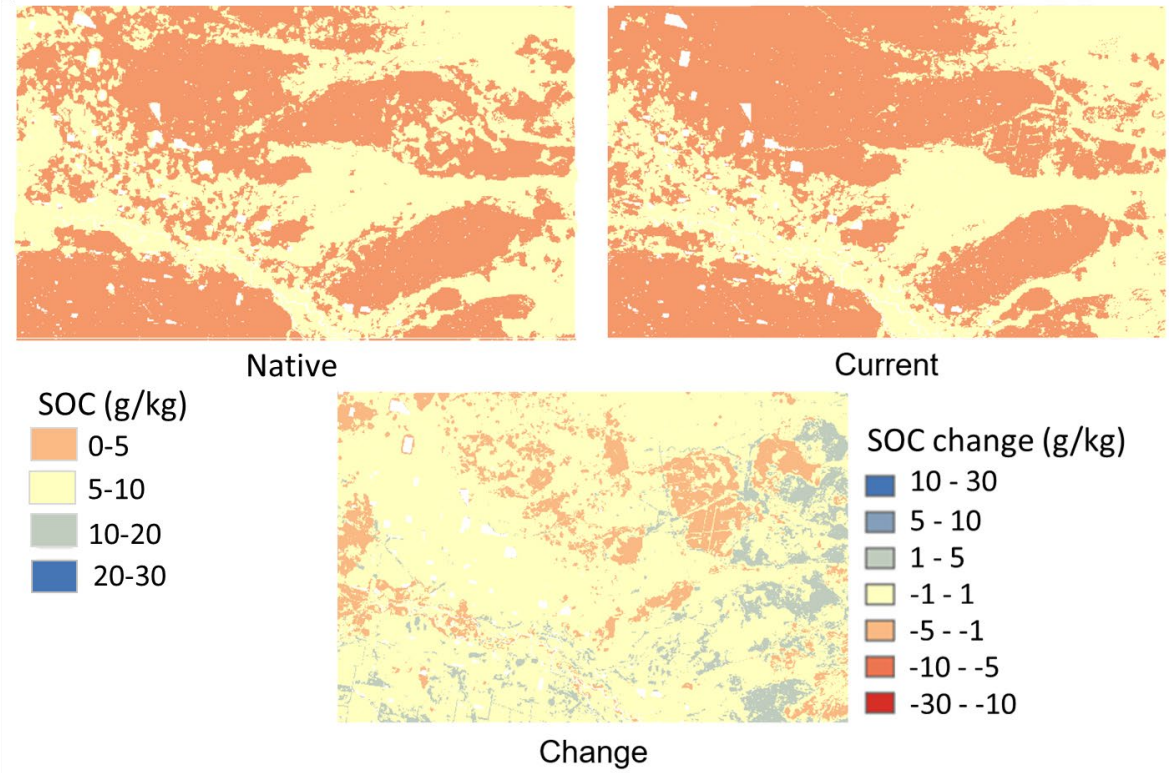
## Native, Current and SOC change (40-50 cm)



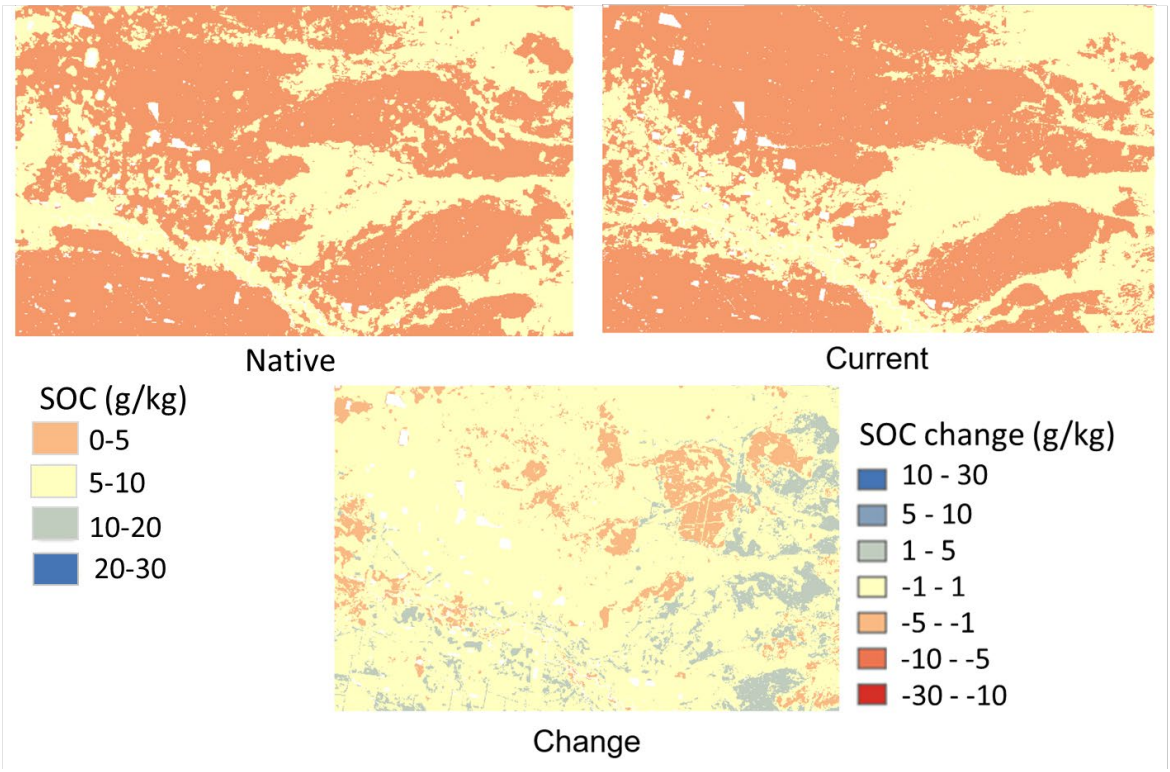

## Native, Current and SOC change (50-60 cm)



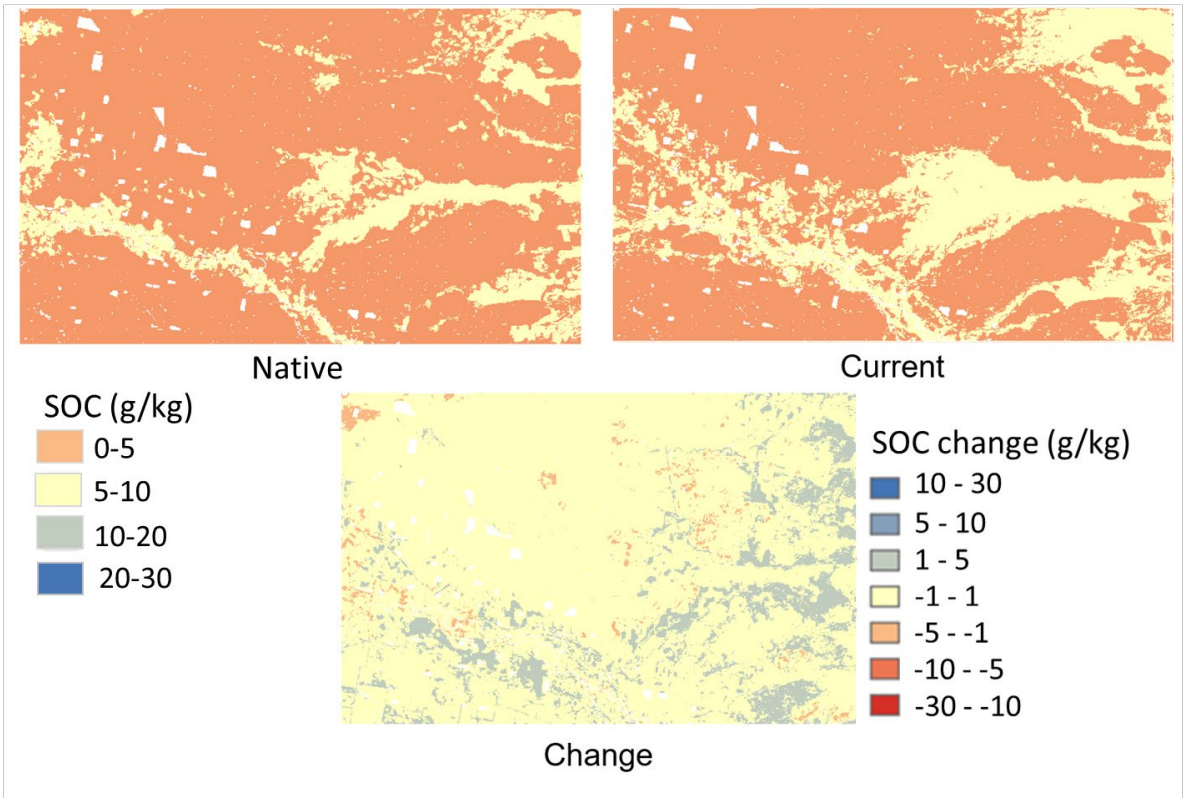
## Native, Current and SOC change (60-70 cm)



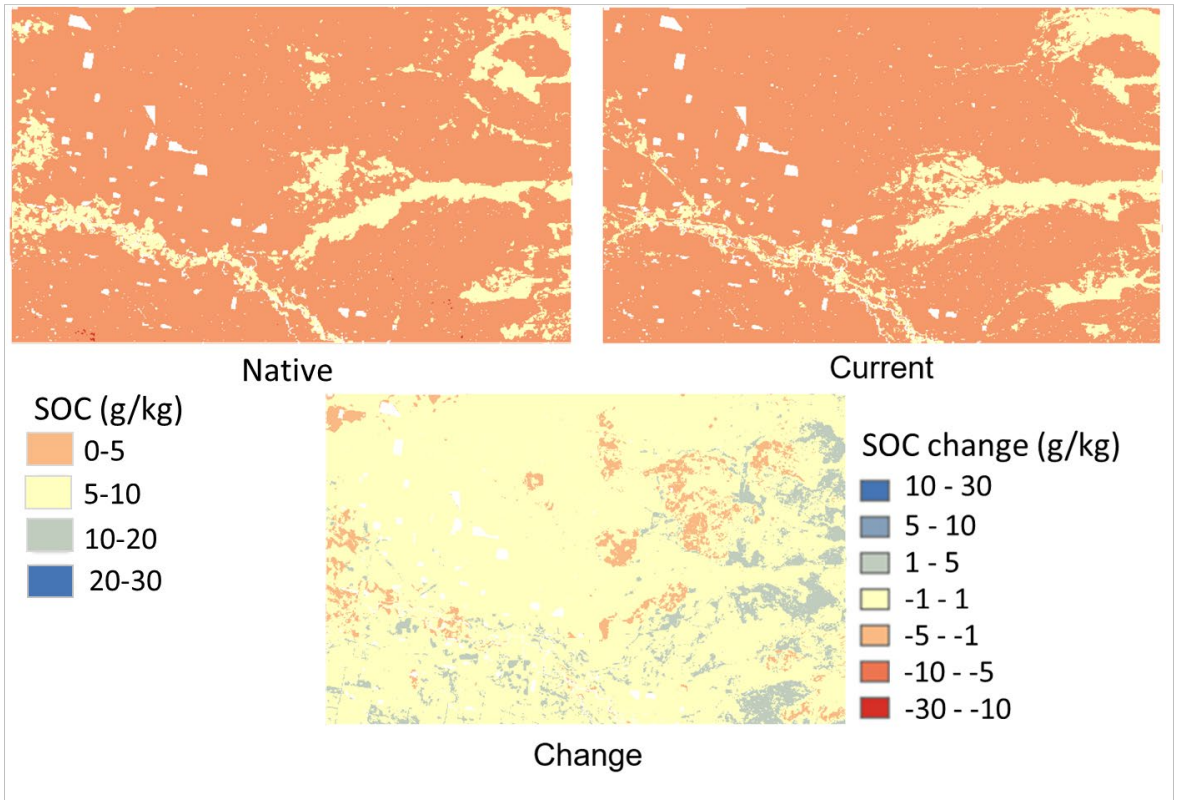

## Native, Current and SOC change (70-80 cm)



## Native, Current and SOC change (80-90 cm)




## Native, Current and SOC change (90-100 cm)



**Appendix 4-2. Results of average value by each land use and pedogenon.**

| Pedogenon | Depth | Native | Pasture | Irrigated_cropping | Non_irrigated_cropping |
|---|---|---|---|---|---|
| **A** | 10 | 21.76 | 17.82 | 14.13 | 13.59 |
| | 20 | 13.15 | 12.36 | 10.06 | 9.45 |
| | 30 | 12.17 | 10.89 | 9.10 | 8.51 |
| **B** | 10 | 21.58 | 20.89 | 14.88 | 18.84 |
| | 20 | 13.29 | 14.51 | 10.64 | 13.16 |
| | 30 | 11.47 | 12.60 | 9.52 | 11.44 |
| **C** | 10 | 20.83 | 18.53 | 13.51 | 14.66 |
| | 20 | 11.92 | 12.71 | 9.57 | 10.32 |
| | 30 | 10.40 | 11.18 | 8.80 | 9.43 |
| **D** | 10 | 20.56 | 16.88 | 12.87 | 14.18 |
| | 20 | 11.05 | 10.80 | 8.94 | 8.95 |
| | 30 | 9.52 | 9.22 | 8.29 | 7.45 |
| **E** | 10 | 22.45 | 18.73 | 14.53 | 16.60 |
| | 20 | 12.89 | 11.39 | 8.50 | 9.74 |
| | 30 | 9.16 | 8.69 | 6.13 | 7.12 |
| **F** | 10 | 20.67 | 16.97 | 13.65 | 14.23 |
| | 20 | 11.42 | 10.55 | 8.75 | 8.89 |
| | 30 | 9.24 | 8.64 | 7.30 | 7.24 |
| **G** | 10 | 20.96 | 15.96 | 12.57 | 13.51 |
| | 20 | 11.63 | 10.47 | 8.25 | 9.25 |
| | 30 | 10.15 | 9.03 | 7.00 | 8.16 |
| **H** | 10 | 22.55 | 20.00 | 17.35 | 17.57 |
| | 20 | 14.12 | 14.11 | 12.81 | 12.96 |
| | 30 | 12.63 | 12.45 | 11.61 | 11.77 |
| **I** | 10 | 32.06 | 25.57 | | 24.63 |
| | 20 | 19.90 | 16.21 | | 15.77 |
| | 30 | 13.82 | 12.26 | | 12.12 |
| **J** | 10 | 23.72 | 19.09 | | 15.94 |
| | 20 | 14.34 | 12.00 | | 10.00 |
| | 30 | 11.86 | 9.51 | | 7.84 |
| **K** | 10 | 22.96 | 18.45 | 15.60 | 16.27 |
| | 20 | 13.73 | 11.45 | 9.80 | 10.42 |
| | 30 | 10.43 | 8.81 | 7.39 | 8.19 |
| **L** | 10 | 27.69 | 23.60 | | 20.36 |
| | 20 | 17.55 | 16.47 | | 14.48 |
| | 30 | 14.55 | 14.01 | | 12.67 |
| **M** | 10 | 26.37 | 23.74 | 18.86 | 22.02 |
| | 20 | 17.43 | 16.79 | 14.17 | 15.88 |
| | 30 | 14.57 | 14.24 | 12.77 | 13.71 |
| **N** | 10 | 29.43 | 27.56 | | 26.67 |
| | 20 | 19.88 | 19.18 | | 18.71 |
| | 30 | 15.53 | 15.13 | | 14.91 |

# Chapter 5. Assessing the capacity for mineral-associated organic carbon storage in croplands of New South Wales, Australia

## 5.1. Abstract

Soil carbon sequestration entails the capture of carbon from the atmosphere and storing it in the soil in a stable form. Soil organic carbon (SOC) is diverse; thus, investigating SOC fraction relevant to climate change mitigation is more important than total SOC. This study quantified mineral-associated organic carbon (MAOC) capacity and condition in Edgeroi, New South Wales, Australia. The MAOC capacity refers to the maximum OC the soil can hold, while the condition refers to the current MAOC content. The difference between MAOC capacity and condition is considered as the SOC sequestration potential. The MAOC capacity and condition for the study area was mapped using two methods. The first approach is using pedogenon mapping, where soil samples in the study area (n = 99) were collected under native vegetation (genosoils) and cropping (phenosoils). MAOC capacity was mapped based on genosoil samples and MAOC condition was mapped using samples from phenosoils using spatial covariates and random forest models. Different land uses impacted MAOC, generally, MAOC concentration in genosoils was always higher than in phenosoils. The second approach spatially predicted median and upper quantile MAOC content using quantile general additive models based on a regional dataset. Mapping results from both methods show that MAOC content was higher in the woodland area, whereas MAOC in the agricultural area was decreased. There was an inconsistency between the two approaches in the resulting SOC sequestration potential map. We conclude that MAOC prediction using the pedogenon mapping with local data produced more reliable results. The results of this study can be connected with two dimensions of soil security (capacity and condition), and these two dimensions can be supported for decision-making related to other dimensions (capital, connectivity and codification).

## 5.2. Introduction

Soil stores the largest amount of carbon in the terrestrial ecosystem. Carbon in a stable form or associated with minerals is less likely to be released back into the atmosphere, which helps to lower the concentration of atmospheric $CO_2$. By exchanging carbon with the atmosphere through processes like plant growth and microbial activities, soil also contributes to the global carbon cycle. Understanding and managing these processes makes it possible to influence the amount of carbon stored in the soil. Overall, soil carbon is an essential factor in climate change mitigation, and there are numerous ways in which soil management practices can be used to enhance soil carbon sequestration and storage. In addition, increasing the content of SOC improves soil fertility, which supports soil and food security and global mitigation of $CO_2$ emissions (Paul, 2016, McBratney et al., 2014).

However, SOC is very heterogeneous in nature, so it is critical to investigate relevant SOC fractions rather than bulk SOC (Lehmann and Kleber, 2015). Many papers use different fractionation approaches to separate SOC into fractions (Six et al., 2000, Sollins et al., 2006). The SOC fractions are characterised by fundamentally different processes of formation, stabilisation, chemistry and protection against decomposition (Xiao et al., 2022). In particular, the stable form of SOC is more relevant for climate change mitigation than bulk SOC. Investigating the distribution of SOC fractions can support monitoring and modelling carbon sequestration (Chenu et al., 2019). There are many types of SOC fractions, but in recent years, there has been a consensus on two operational fractions: mineral-associated organic carbon (MAOC) and particulate organic carbon (POC). MAOC refers to organic carbon associated with fine mineral soil particles (silt- and clay-sized minerals) by chemical bonding or physical occlusion, which may confer protection from fast decomposition. In contrast, POC is largely made up of lightweight fragments that are relatively undecomposed (Lavallee et al., 2020, Kleber et al., 2015). The MAOC and POC contents are controlled by several biotic (vegetation

and soil microbial community) (Cotrufo et al., 2019b) and abiotic factors (climate, soil physicochemical properties) (Doetterl et al., 2015). In Australia, the pyrogenic OC (PyOC) is another form of stable carbon generally found in the landscape due to its long history of burning.

According to Hassink (1997) and Six et al. (2002), the soil mineral matrix has a maximum capacity for protecting SOC through physical and chemical stabilisation mechanisms, i.e., C (carbon) saturation, which is often estimated with its texture and dominant mineralogy. MAOC corresponds to the SOC fraction that saturates (Lavallee et al., 2020, Cotrufo et al., 2019a), and the C saturation concept has been corroborated experimentally (Stewart et al., 2007).

MAOC can indicate soil capacity for the function of carbon storage, as it can be determined from inherent soil properties (e.g., particle size distribution). The soil capacity is represented by soil attributes that determine the performance of soil functions but are not readily modified by human activities (e.g., soil clay content) (McBratney et al., 2019). Soil capacity is one of the biophysical dimensions of soil security together with soil condition. Soil condition refers to the current state of the soil and is characterised by soil attributes modified by human activities at a temporal scale relevant to management and affecting soil function (McBratney et al., 2019). Soil capability is an emerging concept that measures the ability of soil to perform a function (e.g., storage of carbon), which is limited by its capacity and modified by its condition (capability = capacity + condition). The mean residence time of POC is much shorter (from more than 10 years to decades) than MAOC (from decades to centuries) (Keiluweit et al., 2015, Kögel-Knabner et al., 2008, von Lützow et al., 2007), and is more sensitive to management than MAOC (Rocci et al., 2021). Therefore, POC has been suggested as an indicator of soil condition and to assess the effects of management on soil quality. However, the comparison of MAOC levels between different land uses, or MAOC change as a result of management, can also be used as an indicator of soil condition.

Soil carbon saturation implies a theoretical maximum concentration of SOC attainable by the soil. The C saturation can also be referred to as the upper limit of soil carbon stabilisation or the protected soil carbon (Baisden et al., 2013). The C saturation can be determined by the fine fraction (the clay and silt contents and the different types of mineralogy), climate and land use. Many approaches for quantifying the C saturation capacity range from multiple linear regression to data-driven models. Hassink (1997) used linear regression methods with data from uncultivated grasslands' soil samples to fit C saturation equations, which were tested in different soil types and regions under temperate and tropical climates. Hassink's equation was applied to estimate the C saturation capacity under agricultural soils (Carter et al., 2003, Sparrow et al., 2006, Zhao et al., 2006, Wiesmeier et al., 2014). Hassink (1997) also illustrated the strong relationship between the maximum carbon and the proportion of silt and clay particles (fine fraction). However, SOC measurements in Australia were entirely below the values estimated by Hassink's model. The possible reason for this result is low precipitation and high temperatures in dry areas leading to low primary productivity. In addition, Six et al. (2002) expanded the analysis of Hassink (1997). They also used linear regression with different types of mineralogy (1:1 clays and 2:1 clays) and land use (forest, grassland and cultivated area) as predictor variables. Wiesmeier et al. (2015) estimated the C saturation in semi-arid grasslands in northern China and estimated the SOC sequestration potential in different land uses (grazing land, arable land and eroded areas). SOC sequestration potential was calculated by the difference between the SOC saturation and current SOC (Guillaume et al., 2022):

$$\text{SOC sequestration potential} = C_{\text{Saturation}} - C_{Current}.$$

Feng et al. (2013) used the boundary line analysis for estimating C saturation. This analysis is a regression method that fits a function to the upper portion response of SOC content as a function of fine soil fraction (<20 μm). Alvarez and Berhongaray (2021) used a quantile regression, which estimates the conditional quantile of the response variable. They set the 0.75

quantiles of SOC to estimate the C saturation capacity from clay + silt (<20 μm) content, soil depth, land use, mean annual temperature and precipitation. They mentioned that when the quantile regression was fitted with only one predictor (i.e. soil fine fraction), using the 0.9 quantile for the upper boundary of estimates is common. However, for multiple covariates, not all can attain their maximum values in the dataset, and the 0.90 quantiles can provide unrealistic values of C saturation capacity. Thus, the 0.75 quantiles provided a reasonable estimate of saturation capacity. Similarly, Padarian et al. (2022) estimated global SOC sequestration potential on croplands by comparing the median and upper quantile prediction (q0.75-q0.50).

Chen et al. (2018) estimated the SOC sequestration potential in France with Hassink's equation using clay+silt as an explanatory variable (Hassink, 1997). The sequestration potential in France was recalculated by Chen et al. (2019) on arable soils using a data-driven approach. They first produced carbon-landscape zones by clustering covariates as proxies of the environmental control of SOC storage (climatic decomposition index (SOC decomposition), soil clay content (SOC protection from decomposition) and net primary production (C input)). The maximum SOC stocks of arable soils (top and subsoil) were predicted for each carbon-landscape zone with four quantiles (0.80, 0.85, 0.90 and 0.95). The difference between the current SOC stocks and the SOC storage potential under the four quantiles was calculated. Even though SOC associated with the fine fraction does not represent the total SOC, their estimate of SOC sequestration potential was close to the result of 0.9 percentile. The study concluded that maps of SOC storage potential produced using the data-driven approach show better qualitative agreement than the approach using Hassink's equation.

These past studies only use SOC concentration or stock to define SOC sequestration potential, while as described above, a more relevant measure of carbon storage in soil is the MAOC. Thus, this paper aims to map MAOC capacity and condition. We define the difference between MAOC capacity and condition as SOC sequestration potential.

In addition to the quantile regression approach, this paper proposes a new approach to measure SOC sequestration potential using a data-driven approach by comparing genosoils and phenosoils within a digital pedogenon map. Pedogenons defined regions of homogeneous soil forming factors for a given reference time, assuming that the inherent soil properties of each pedogenon were developed similarly (Román Dobarco et al., 2021b). The definition of genosoil is the dominant genetic soil type within a map unit created by long-term pedogenesis, and the phenosoil is a persistent variant of the genosoil changed by anthropogenic activities (Rossiter and Bouma, 2018). Few studies used these concepts to study soil change (Huang et al., 2018, Román Dobarco et al., 2021a). Genosoils serve as reference soils to assess changes in condition and determine capacity (within a soil class or a soil map unit).

Similarly, the phenosoils inform on current condition after management or land use change. The unique properties in each pedogenon would have a different capacity for soil carbon storage. The level of C saturation can be set by pedogenon classes (genosoil). Hence, comparing genosoil and phenosoil within a pedogenon class can be used to estimating SOC sequestration potential.

This study aimed to investigate the SOC sequestration potential, or storage deficit in the MAOC fraction, with two different approaches: i) using the comparison between genosoil and phenosoil for assessing MAOC capacity and condition, and ii) developing a regional C saturation model (i.e., the upper limit of the C concentration in the fraction < 53 μm (MAOC)) for south-eastern Australia with general additive model quantile regression. We calculated the difference between two values (MAOC capacity and current MAOC) to estimate the SOC sequestration potential in the fine fraction. We use the concept of SOC sequestration potential as an indicator of capacity for the function of carbon storage with the conceptual framework of soil security.

## 5.3. Materials and Methods

### 5.3.1. Description of the study area

The study area is in the Edgeroi district in New South Wales (NSW), Australia; the size of the area is 1700 $km^2$ (Figure 5-1). The region has been intensively cropped for more than 200 years (Green et al., 2011, Narrabri Shire Council, 2021). Extensive expanses were used as pastures and stock routes when grazing was the primary agricultural activity. Wheat, sorghum, and irrigated cotton were common broadacre crops in the 1960s (McGarry et al., 1989). Soils in national parks have been relatively unaffected by human activities.

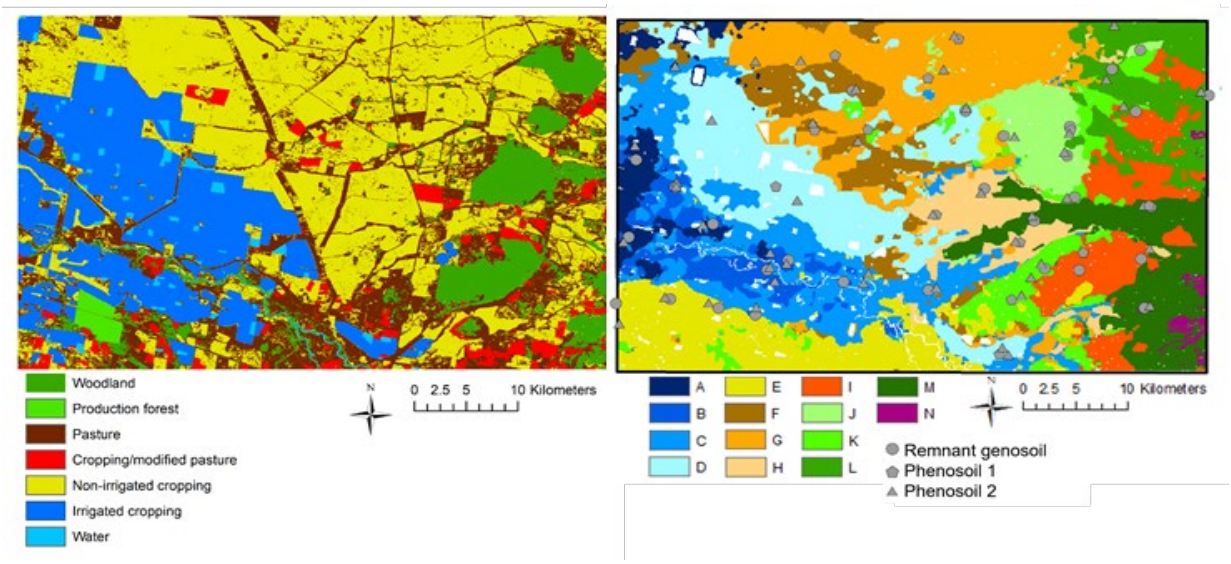



**Figure 5-1. Left: the map of land use or cover, and there are seven broad land use/cover classes in the area. Right: the map of fourteen pedogenons (Pedogenon A to N) and sampling points of genosoil and phenosoil.**

### 5.3.2. Pedogenon map

A pedogenon map was generated, as described in Chapter 3 following the approach of Román Dobarco et al. (2021b). Pedogenons are classes defined within areas with homogeneous soil-forming factors representing soil classes for a given reference time (Román Dobarco et al., 2021b). Fourteen pedogenons (pedogenons A to N, Figure 4-1) were established in the study area using an unsupervised classification method, i.e., k-means clustering of spatial layers representing soil forming factors (climate, soil, topography, organisms, parent material and age). The characteristics of each pedogenon are explained in Chapter 3. This pedogenon map has been validated with five soil attributes of each pedogenon (clay, sand, pH, organic carbon and cation exchange capacity) and multivariate statistical analyses (redundancy discriminant analysis (RDA) and variance partitioning). With the map of pedogenons, genosoils and phenosoils could be distinguished by overlaying land use information. Genosoils are soil classes that have resulted from long-term natural pedogenesis, or soil classes used in detailed soil mapping (dominant class within a soil map unit or at the level of soil series or equivalent) (Huang et al., 2018, Rossiter and Bouma, 2018). In practice, genosoils are found in areas human activities have not intensely influenced. On the other hand, phenosoils are soil variants of genosoils due to past land use history and management, mainly cropping practices, and have seen altered soil functions (Rossiter and Bouma, 2018). Based on the genosoil and phenosoil delimitation, the sampling points were decided and collected for further studies (Figure 4-1). The criteria for defining genosoil and phenosoil areas is explained in Chapter 3.

### 5.3.3. Calculating carbon fractions using MIR spectra and mapping MAOC change in condition.

Two datasets were used to map MAOC in our study area: a national dataset and a local dataset.

#### 5.3.3.1. National SOC and SOC fractions data

The national dataset was from the Soil Carbon Research Programme (SCaRP) (Baldock et al., 2013a) where soil samples from agricultural soils were collected across Australia to assess its stock. Topsoil samples (0-30 cm) were collected from 4,526 sampling sites (Baldock et al., 2013a). Figure 2 shows the soil sampling locations of the SCaRP across Australia. The soil samples were allocated to soil order according to the Australian Soil Classification system (Isbell, 2002) and were dominated mainly by Chromosols, Dermosols, Sodosols, Tenosols, and Vertosols. More information about the SCaRP data can be found in Baldock et al. (2013a).

Within the ScaRP dataset, a subset of 287 sample points representing the range of SOC content was selected and measured for their SOC content and SOC factions (MAOC, POC and pyrogenic OC (PyOC)) using the method described in Baldock et al. (2013b). In brief, air-dried (≤ 2 mm) soil samples were dispersed and separated into coarse (>50 μm) and fine (<50 μm) fractions using wet sieving. The organic C concentrations of the coarse and fine fractions were analysed with dry combustion. The PyOC was estimated using solid-state NMR (nuclear magnetic resonance spectroscopy). POC and MAOC contents (mg C g soil$^{-1}$) were calculated by subtracting the proportion of PyOC in each fraction (Baldock et al., 2013b).

The measured C-fraction samples were scanned using a diffuse reflectance mid-infrared (MIR) spectrometer. Soil samples were dried, sieved and fine ground. Spectral reflectance of the samples was measured using a high throughput screening device (HTS-XT) attached to a Tensor 37 spectrometer (Bruker Optik GmbH, Ettlingen, Germany) between 4000–400 cm$^{-1}$. Each soil sample was averaged with 60 times scanned data, and the averaged reflectance spectra were converted to absorbance.

Calibration models were built to predict MAOC concentration and other SOC factions from mid-infrared (MIR) spectra of the 287 soil samples. The SOC concentrations data were square-

root transformed to minimise the generation of negative predicted values and provide the best linear relationship (Baldock et al., 2013a). Moreover, the purpose of a square-root transformation was to produce normally distributed residuals.

Partial least squares (PLS) regression was used to produce a spectral model for predicting MAOC content. To enhance model performance, several common spectrum pre-processing methods were utilised. These include trimming spectral region with a high signal-to-noise ratio (<600 nm and >3997 nm) and $CO_2$ peak (2300 nm to 2400 nm). The spectra were then smoothed and a standard normal variate transformation (SNV) was applied (Wadoux et al. (2021). The optimal number of components of the PLS model was selected using a ten-fold cross-validation approach (plsr function in R, (Mevik and Wehrens, 2007)), setting 30 components as the maximum number of components. To assess the accuracy of the model, the spectra data were split randomly into two sets: a training set for calibrating a model (80% of all data) and a test set for evaluating the model performance (20%). This process was repeated ten times, and the accuracy metrics were averaged.

#### 5.3.3.2. Edgeroi SOC and SOC fractions data

MIR data of the 99 sites collected in the study area (Edgeroi) were acquired from the surface down to 30 cm depth at 10 cm depth intervals. The calibrated PLS model was used to predict the MAOC content on these samples. All information regarding the data collection in the field and scanning soil samples using MIR spectroscopy is described in Chapter 4. Before the prediction, the data was pre-processed with the same techniques applied to the calibration data. As no SOC fractions were analysed in Edgeroi, a step for assessing the model accuracy was comparing MAOC with organic carbon (SOC) content and POC+MAOC+PyOC with SOC content. Baldock et al. (2013a) found a high correlation coefficient between OC and MAOC

of 0.98. The SOC content of the Edgeroi dataset was predicted with MIR spectra and locally measured OC content using the cubist model as described in Chapter 4.

### 5.3.4. Creating MAOC capability and condition maps

Maps of MAOC (averaged for 0-30 cm) for genosoils and phenosoils were created using the digital soil mapping (DSM) approach, as detailed in Chapter 4. A machine learning model was utilised to establish a quantitative relationship between MAOC and 12 environmental covariates representing the factors soil (s), organisms (o), relief (I), parent material (p) and age (a) from the *SCORPAN* model (McBratney et al., 2003). The covariates are listed in Table 5-1. The covariate of DEM that was not normally distributed was log-transformed.

Two soil maps were created: the first one represents MAOC capacity, assuming natural vegetation across the study area, i.e., pre-European settlement conditions (genosoil MAOC), whereas the second map represents MAOC condition (phenosoil+genosoil MAOC). For mapping MAOC capacity, only 47 sampling points located at genosoil areas (currently under natural vegetation) were used as training data. For producing the MAOC condition map, all sites (99 points) were used.

The covariates representing s, r and p factors for creating both maps were the same except for the vegetation (o) covariate. For mapping MAOC capacity, a map of the estimated vegetation pre-European settlement (Eco Logical Australia, 2013) was used as the covariate representing the organism factor. The MAOC condition map was produced using the s, r and p factors and the current vegetation data representing o. Chapter 4 provides further detailed information about two different vegetation maps. The spatial models were fitted with random forest models with 500 trees.

**Table 5-1. Environmental covariates are used to create MAOC maps (MAOC capacity and condition).**

| Soil-forming factors | Covariate name | Units | Reference |
|---|---|---|---|
| **Soil** | Gamma K | | Wilford and Kroll (2020) |
| | Gamma Th | | Wilford and Kroll (2020) |
| | Gamma dose | | Wilford and Kroll (2020) |
| | Sand | % | From Chapter 2 |
| **Organisms** | Vegetation Native European settlement (3 PCs, explained 82% variation) (For genosoil MAOC) | | Eco Logical Australia (2013) |
| | Current vegetation (3 PCs, explained 82% variation) (For phenosoil MAOC) | | Department of Planning and Environment (2018) |
| **Topography** | Altitude from Digital elevation model (DEM) | m | Gallant et al. (2009) |
| | Slope | % | Gallant et al. (2009) |
| **Parent material & age** | Parent material (3 PCs, explained 46% of the total variation) | | From Chapter 2 |

The difference between MAOC capacity and condition maps represents SOC sequestration potential:

$$\Delta MAOC = MAOC_{capacity} - MAOC_{condition}$$

### 5.3.5. Mapping regional SOC sequestration potential using quantile regression

The second approach for estimating the capacity and condition for MAOC storage consisted in developing a regional model. The capacity was modelled with an upper quartile generalised additive model, qgam (Wood and Wood, 2015) and the condition with a generalised additive model, gam (Hastie, 2006).

#### 5.3.5.1. Subsetting calibration data from SCaRP dataset

As described above, soil data from the SCaRP dataset (Baldock et al., 2013a) were extracted to create regional MAOC models. A buffer was created in the east of Australia for extracting the soil sampling points of the SCaRP dataset with similar land use and pedoclimatic regional conditions to the study area. The main land use of the study area is cropping, and the buffer area was decided based on the information on the grain growing region in Australia (Armstrong et al., 2019, Robertson et al., 2012). Both the study area and the buffer have a similar land use history. Figure 5-2 shows the SCaRP data locations extracted within the buffer (1607 sites).

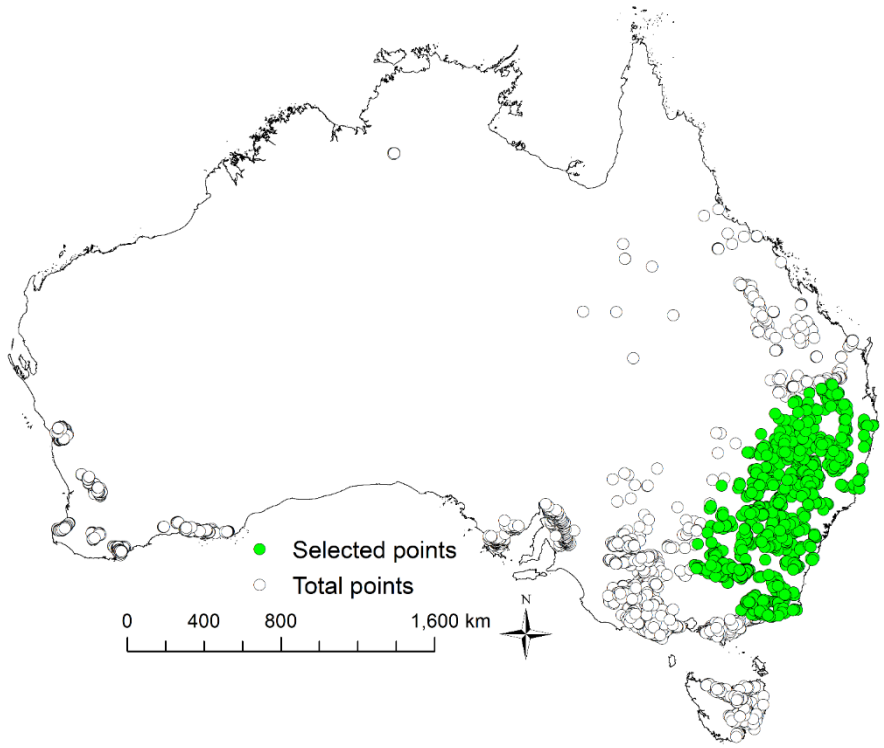



**Figure 5-2. SCaRp soil points and the selected regional points for the study area.**

#### 5.3.5.2. Finding gam model structure with manual backward selection

A generalised additive model (gam) is a flexible nonlinear regression model that can be fitted efficiently using the approximate Bayesian methods (Fasiolo et al., 2020, Miller, 2019). gam offers a middle ground between a simple model (linear regression) and more complex machine learning models (including neural networks) (Brenning, 2008). This model can fit complex, nonlinear relationships and make a good prediction. However, unlike black box machine learning models, inferential statistics can be done on gam, and the structure of models can be interpretable. The data were fitted with smooths, or splines, which are functions that can take on a wide variety of shapes. The gam model was fitted using the gam() function from the mgcv package (Hastie, 2006). The independent variables were wrapped with s() function, which is a smooth function to make a nonlinear regression model. The smoothing basis from the function of s() is a form of the thin plate regression spline (Wood, 2017).

The current MAOC concentration was predicted with the gam model, whereas the upper limit or capacity for MAOC was predicted with the upper quartile (q = 0.75) regression using the qgam model. The model of qgam was proposed as a novel framework for fitting additive quantile regression models by (Fasiolo et al., 2021). This model provided well-calibrated inference about the conditional quantile and automatic prediction of smoothing parameters. This model maintains equivalent numerical efficiency and stability. Moreover, this method is statistically rigorous and computationally efficient (Bissiri et al., 2016) (Wood et al., 2016).

The optimal predictive variables of the gam and qgam (q = 0.75) models were selected by combining several model-building strategies. The initial candidate variables were 103 variables representing the SCORPAN factors, including climate, parent material, relief, soil, organisms and time. These covariates are available from TERN Landscapes Cloud Optimised Geotiff Datastor (Searle et al., 2022). The MAOC data, which was not normally distributed, were log-transformed. The first step for selecting the model structure was identifying potential variables

from each SCORPAN factor using exploratory data analysis. Correlation plots were produced by grouping the variables by soil-forming factor. The top three or four variables with the highest correlation coefficients with MAOC by soil-forming factors were selected.

Moreover, the suitability of the potential variables was confirmed with an extra step. A univariate gam model was fitted for each predictor variable and compared with the null model using AIC (Akaike information criterion). However, there is one fixed variable, that is the fine_fraction (% of clay + silt (< 20 μm)). This variable was included following the approach from Hassink (1997) and Six et al. (2002). Note that the silt fraction used in Australia is 2-20 μm.

Next, a gam model was fitted using all candidate variables selected in the previous step (full model). The suitability of the model structure and the accuracy of the model were assessed with 10-fold cross-validation. The model performance statistics consisted of the average AIC, root mean square error (RMSE), mean error (ME), $R^2$ and concordance correlation coefficient (CCC) from 10-fold cross-validation statistics. In addition, the statistics of the gam models were also assessed, i.e., normality of residuals and lack of concurvity of predictor variables. Concurvity is a measure similar to the concept of collinearity from a linear model. Concurvity values are between 0 and 1, with 0 indicating no collinearity and 1 indicating high collinearity. If two covariates are highly correlated, it is difficult to fit the model as either one can affect the output and result in poorly fitted models with large confidence intervals (He et al., 2006). If the necessary conditions of the model were not met, a manual backward selection was performed. Several gam ("leave one variable out") models were built by eliminating each potential variable from the full model, and the concurvity was calculated. A threshold concurvity value of 0.8 was used to decide if a variable has high collinearity.

The full and each of the "leave one variable out" models was compared with cross-validation statistics. This step was repeated successively until an optimal gam model was found. The final model had lower AIC, higher $R^2$ and lower RMSE compared with models built with fewer variables. A qgam model for the 0.75 quantiles was built with the same variable structure as the optimal gam model except for the variable describing the soil-forming factor of *organisms*. The vegetation variable was excluded because it is indicative of the current land use and management, i.e., the current *condition* of the soil for storing MAOC. However, we preferred to estimate the *capacity* of the soil to store MAOC based on soil, parent material, and climatic factors, as it is outside the scope of this paper to represent or simulate different estates of potential land uses and vegetation.

#### 5.3.5.3. Mapping MAOC capacity and condition using the quantile regression models

A second set of MAOC maps (0-30 cm) for the Edgeroi area was created: qgam regional model (q = 0.75) represents MAOC capacity, and gam represents MAOC condition. After stacking all covariates in the study area, the waterbodies were removed. The prediction of MAOC was calculated throughout the study area and back-transformed into the original scale.

## 5.4. Results

### 5.4.1. SOC fractions distribution

#### 5.4.1.1. Assessing the accuracy of the PLS model

The results of the model accuracy for SOC content (SOC) and each SOC fraction (MAOC, POC and PyOC) ($R^2$ and RMSE) from the selected SCaRP dataset are shown in Appendix 5-1. The PLS model had a high accuracy with $R^2 = 0.91$ and RMSE = 2.49 mg C g $soil^{-1}$ for MAOC and $R^2 = 0.86$ for POC and PyOC.

To assess the accuracy of the SCaRP PLS model on the Edgeroi dataset, we calculated the linear correlation coefficient between predicted MAOC, the sum of predicted total fractions (MAOC, POC and PyOC) and measured OC (OC_edgeroi) from Chapter 4 (Figure 5-3). The correlation coefficients between OC_edgeroi and MAOC and total fractions were 0.79 and 0.78, respectively. The high correlation coefficients suggest that the PLS model predicted the MAOC and other fractions quite well.

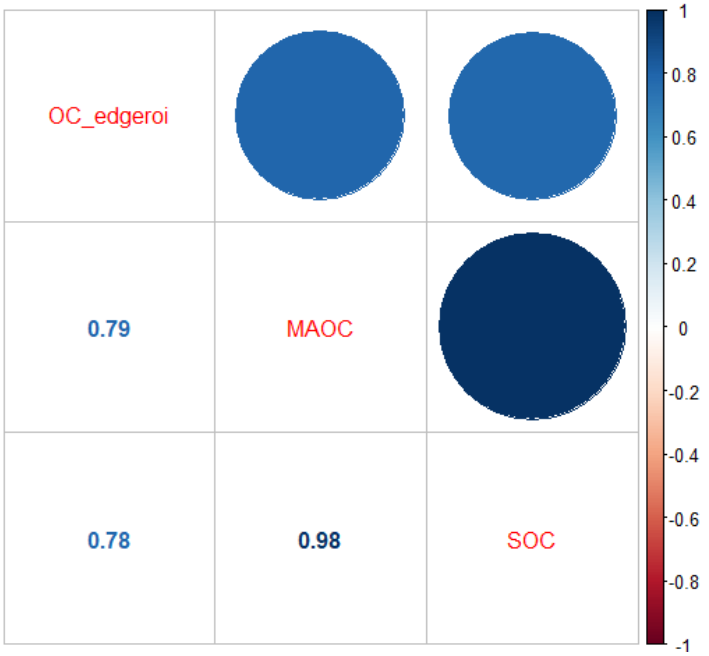



**Figure 5-3. the correlation of total SOC content (MAOC+POC+ROC), MAOC and OC_edgeroi (from Chapter 4).**

#### 5.4.1.2. MAOC and POC for each land use and pedogenon

MAOC and POC content for the 99 soil cores was predicted using MIR spectra and the PLS model and was summarised by land use and pedogenon class. The study area had seven land uses (Figure 4-1), grouped into four major land uses: woodland, pasture, non-irrigated cropping and irrigated cropping. Figure 5-4 shows the boxplot of MAOC content by the four land uses. The mean value of all MAOC data is 4.52 mg C g soil$^{-1}$ (standard deviation (SD): 4.21 mg C g soil$^{-1}$), and an interquartile range (IQR) (2.05 mg C g soil$^{-1}$ to 5.20 mg C g soil$^{-1}$).

The result shows that the highest MAOC content was in woodland areas, followed by pasture, non-irrigated and irrigated cropping. The mean value of MAOC in woodland was 7.10 mg C g soil$^{-1}$ (SD: 5.68 mg C g soil$^{-1}$ and IQR: 4.46 mg C g soil$^{-1}$). Pasture had the second highest MAOC content (mean: 3.42 mg C g soil$^{-1}$, SD: 2.05 mg C g soil$^{-1}$ and IQR: 2.77 mg C g soil$^{-1}$). The next was irrigated cropping (mean: 2.93 mg C g soil$^{-1}$, SD:1.47 mg C g soil$^{-1}$ and IQR: 1.44 mg C g soil$^{-1}$). The lowest MAOC was in non-irrigated cropping (mean: 2.77 mg C g soil$^{-1}$, SD: 1.36 mg C g soil$^{-1}$ and IQR: 1.86 mg C g soil$^{-1}$). The highest POC mean value was from woodland (1.91 mg C g soil$^{-1}$), and the mean of the other three land uses was similar (between 0.26 and 0.31 mg C g soil$^{-1}$).

The boxplot of MAOC for each pedogenon is shown in Figure 5-5A. The highest mean value of MAOC was found in pedogenon I (mean: 8.59 mg C g soil$^{-1}$, SD: 7.54 mg C g soil$^{-1}$, IQR: 10.10 mg C g soil$^{-1}$)s. While the lowest one was in pedogenon G (mean: 1.96 mg C g soil$^{-1}$, SD: 0.95 mg C g soil$^{-1}$, IQR: 1.18 mg C g soil$^{-1}$). The boxplot of MAOC by genosoil and phenosoil for each pedogenon is shown in Figure 5-5B, where G, P1, and P2 represent genosoil, phenosoil 1 and phenosoil 2, respectively. The area of genosoil was under the woodland, which was not disturbed by human activities. Phenosoil 1 was under the pasture area, while phenosoil 2 was under intense agricultural practices. Figure 5-5B confirms that MAOC values of genosoil

from all pedogenon were higher than the phenosoil. In this case, genosoil and phenosoil values are similar (pedogenon F).

Surprisingly, the POC content from seven pedogenons (pedogenon A, B, C, G, H, L and M) had POC content in genosoils that was lower than the phenosoils (Figure 5-6). In contrast, the POC content of genosoils from four pedogenons (pedogenon H, I, J and K) was much higher than other pedogenons.

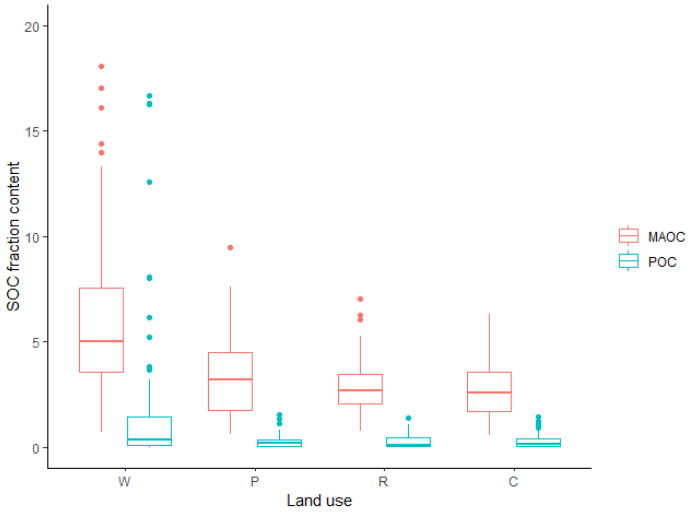



**Figure 5-4. MAOC and POC values are grouped by four land uses, W: woodland, P: pasture, R: irrigated cropping and C: non-irrigated cropping. The unit of both fractions is mg C g soil$^{-1}$.**

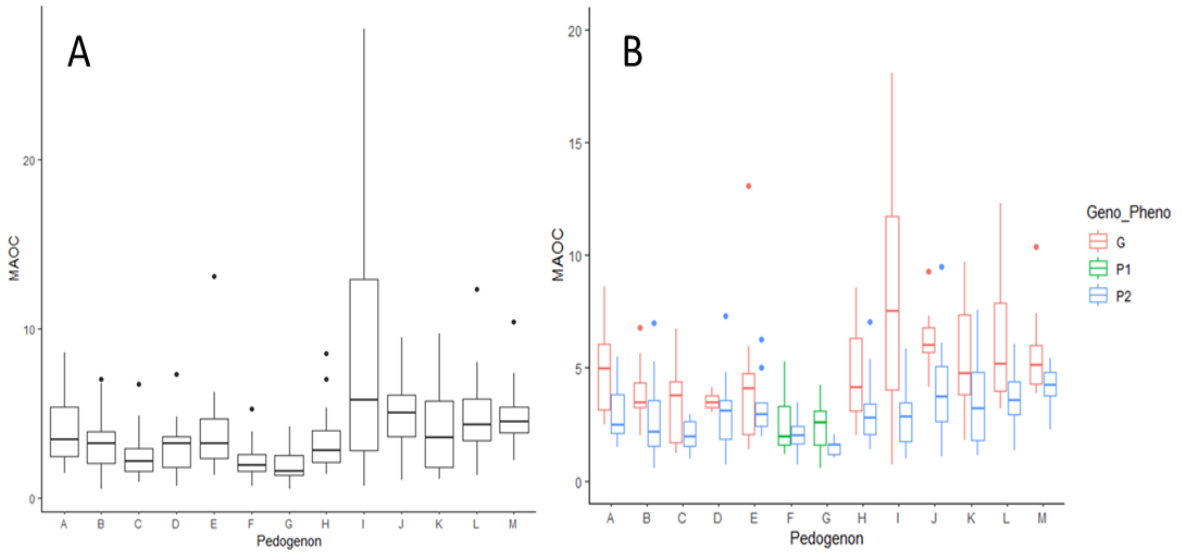



**Figure 5-5. (A) MAOC content was grouped by 13 pedogenons (A to M) and (B) MAOC grouped by pedogenons and genosoils and phenosoils (G: genosoil, P1: phenosoil 1 and P2: phenosoil 2). The unit for MAOC is mg C g soil$^{-1}$.**

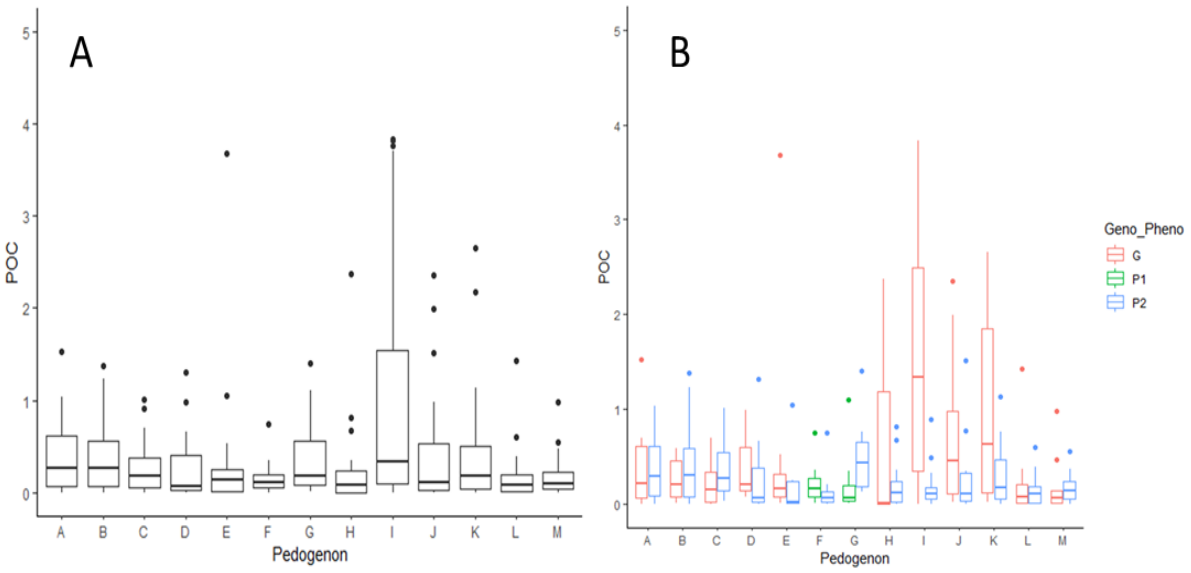



**Figure 5-6. (A) POC by 13 pedogenons (pedogenon A to M) and (B) POC by 13 pedogenons and genosoils and phenosoils (G: genosoil, P1: phenosoil 1 and P2: phenosoil 2) (B). The unit of POC is mg C g soil$^{-1}$.**

#### 5.4.1.3. MAOC capacity and condition using a pedogenon mapping approach

The MAOC (0-30 cm depth) was spatialised using DSM techniques utilising the random forest model. The calibrated DSM models produced maps of MAOC capacity, condition, and difference. Figure 5-7 shows the three maps of MAOC (0-30 cm). The map of MAOC capacity estimates what the MAOC content would have been without intensive agricultural practice. The genosoil MAOC (Figure 5-7A) for the whole study area had a mean of 5.00 mg C g $soil^{-1}$ (SD: 2.19 mg C g $soil^{-1}$). The map of the MAOC condition refers to the current MAOC content. The mean value of MAOC condition in the whole study area (Figure 5-7B) shows 3.92 mg C g $soil^{-1}$ (SD: 2.12 mg C g $soil^{-1}$ and IQR: 1.87 mg C g $soil^{-1}$). The MAOC difference for the study area (Figure 5-7C) had a mean of 1.07 mg C g soil-1 and SD of 0.83 mg C g $soil^{-1}$.

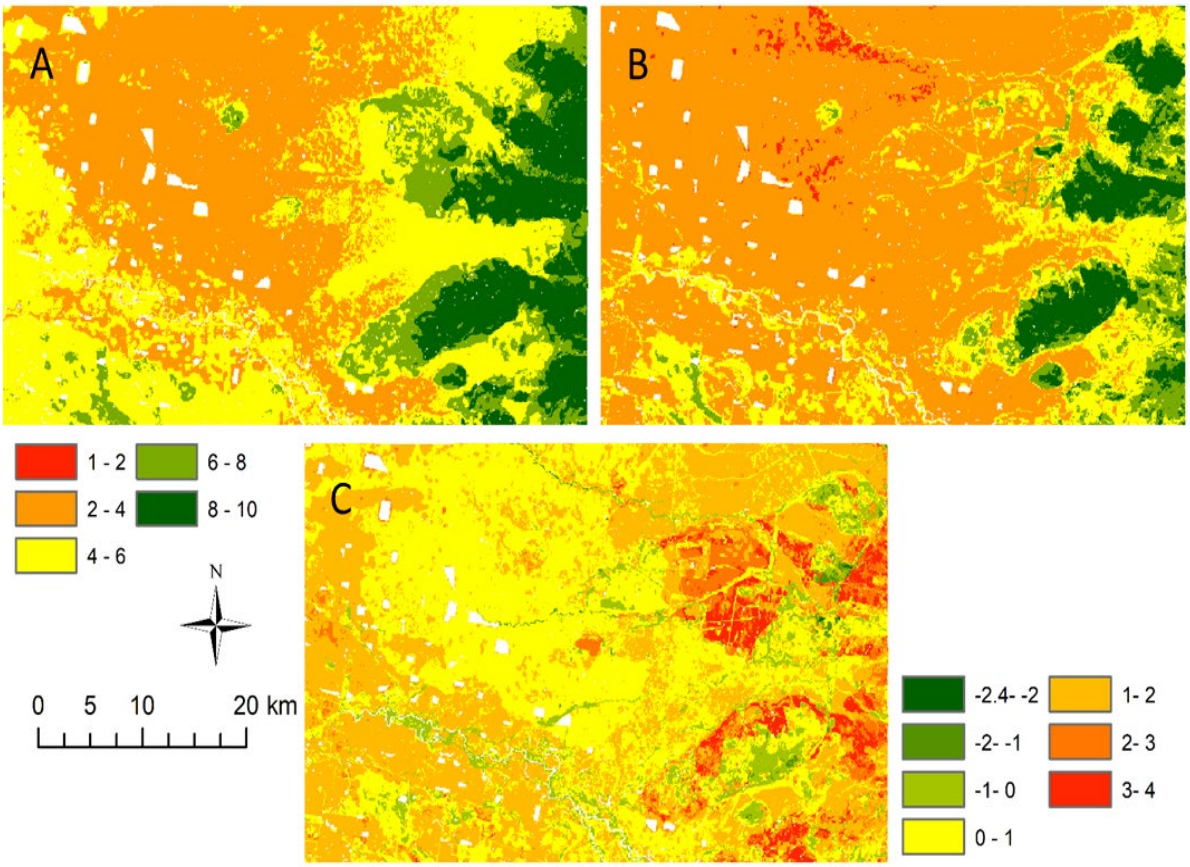



**Figure 5-7. MAOC capacity (top left: A), MAOC condition (top right: B) and MAOC difference (capacity – condition)(bottom middle: C) at 0 - 30 cm produced using the pedogenon mapping approach. The unit of MAOC is mg C g soil$^{-1}$.**

### 5.4.2. Mapping SOC sequestration potential using quantile regression

#### 5.4.2.1. Finding optimal covariates for gam and qgam model structure

The optimal covariates from predicting MAOC at a regional scale were selected from 103 covariates representing the SCORPAN factors. Three or four covariates with the highest correlation coefficients with MAOC were selected from each soil-forming factor. This results in a full model with 11 covariates, including two soil, four climate, one parent material, two organisms and two relief covariates. From the full model (11 covariates), the optimal structure for gam and qgam model (final model) was found with several iterations of removing ineffective variables using cross-validation statistics with a concurvity threshold of 0.8. Table 3 shows the iterative backward selection of covariates from 11 covariates to the final model containing 8 covariates (Table 5-3).

**Table 5-2. Result of model performance statistics of the gam from the backward selection of covariates.**

| Model | AIC | BIC | Adjusted $R^2$ (calibration) | ME | RMSE | $R^2$ | CCC |
|---|---|---|---|---|---|---|---|
| **Full model** | 7146 | 7473 | 0.52 | -0.04 | 2.93 | 0.53 | 0.68 |
| **No FC** | 7144 | 7462 | 0.52 | -0.04 | 2.93 | 0.53 | 0.69 |
| **No FC and pH** | 7241 | 7537 | 0.49 | -0.03 | 3.01 | 0.51 | 0.67 |
| **No FC, pH and TWI (Final model)** | 7238 | 7517 | 0.49 | -0.03 | 3.02 | 0.51 | 0.66 |

**Table 5-3. Optimal covariates for gam and qgam models.**

| Soil-forming factors | Name | Explanation |
|---|---|---|
| **Soil** | Fine_fraction | Fine fraction (% particles < 20 μm) |
| **Climate** | Clim_ADM | Mean annual aridity index (precipitation / potential evaporation) |
| | Clim_EPX | Maximum monthly potential evaporation (mm) |
| | Clim_RSM | Short-wave solar radiation - annual mean (SRAD data) |
| | Clim_TRA | Annual temperature range (max - min) (°C) |
| **Parent Material** | PM_ratio_Th_K | Gamma ray - Thorium to Potassium ratio |
| **Organisms** | NDVI_mean_Q2 | Landsat 5 long-term average NDVI (April-June) 1986-2011 |
| **Relief** | Relief_roughness | Landscape Roughness |

#### 5.4.2.2. Maps of MAOC using regional quantile regression models

The MAOC (0-30 cm depth) for the Edgeroi area was spatialised using the regional gam model. The map of MAOC using gam represents the mean or current MAOC condition (q = 0.5), while the map of MAOC using quantile regression (q=0.75) represents MAOC capacity. With two maps, the SOC sequestration potential map was calculated ($MAOC_{q0.5}$ - $MAOC_{q0.75}$). All three maps are shown in Figure 5-8.

The map of $MAOC_{q0.5}$ had a mean value of 5.47 mg C g $soil^{-1}$ (SD: 0.98 mg C g $soil^{-1}$ and IQR: 4.82 mg C g $soil^{-1}$ to 5.95 mg C g $soil^{-1}$). The mean value of $MAOC_{q0.75}$ was 6.98 mg C g $soil^{-1}$ (SD: 0.81 mg C g $soil^{-1}$ and IQR: 1.17 mg C g $soil^{-1}$). The SOC sequestration potential had a mean of 1.51 mg C g $soil^{-1}$ and SD of 0.93 mg C g $soil^{-1}$.

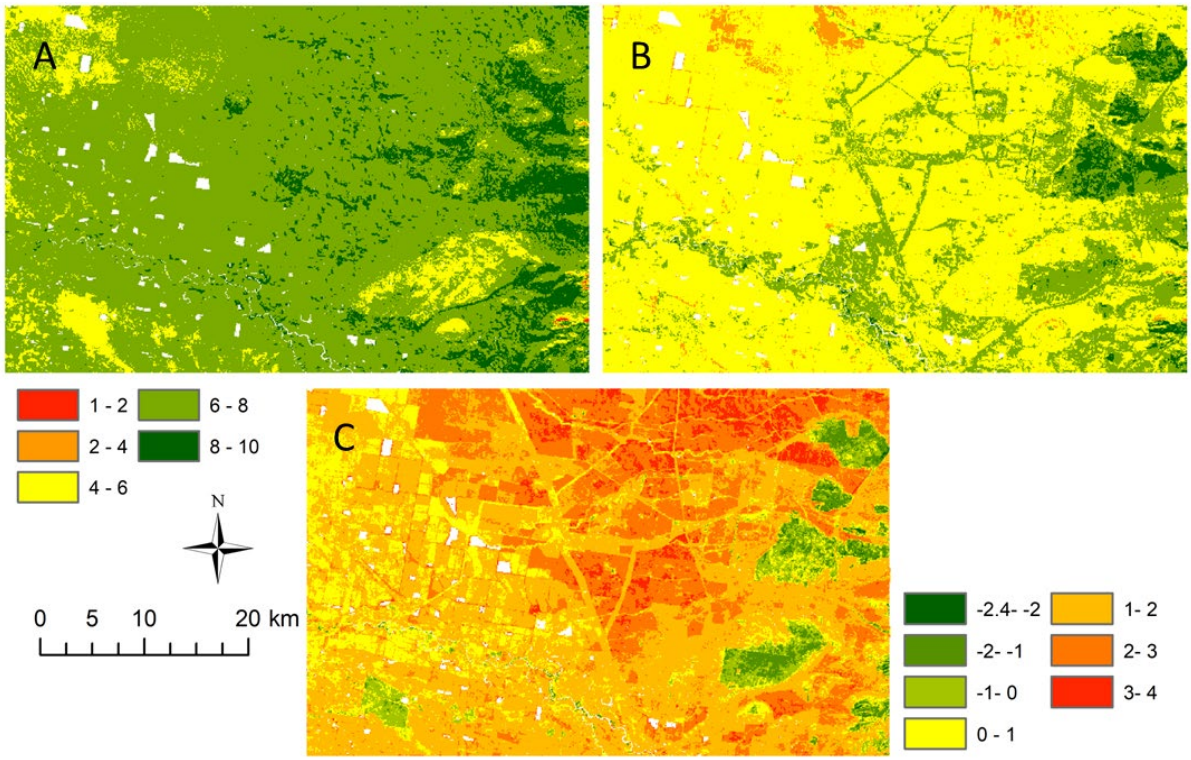



**Figure 5-8. MAOC capacity estimated with qgam (q = 0.75) (top left: A), Current MAOC condition predicted using a gam model (top right: B) and SOC sequestration potential maps (0.75 – 0.50)(bottom middle: C) at 0 - 30 cm. The unit of MAOC is mg C g soil-1.**

## 5.5. Discussion

### 5.5.1. MAOC capacity and condition using pedogenon mapping approach

#### 5.5.1.1. MAOC content by land use and pedogenons

The predicted SOC fraction values were grouped by different land use and pedogenons. The results (Figure 5-4) show that MAOC content was much higher than the POC. This is in accordance with past studies. Briedis et al. (2012) explained that in agricultural soils of Brazil, the dominant SOC fraction found is MAOC. Román Dobarco et al. (2022) found that for soils in Australia, the largest contribution of SOC fractions to total SOC is MAOC (59 % ± 17.5 %), while the smallest one is POC (13 % ± 11.1 %).

Figure 5-4 further shows the MAOC and POC values based on four different land use. Both SOC fractions under woodland areas were the highest, and the subsequent highest were under pasture. MAOC contents under non-irrigated and irrigated cropping areas were similar. Other studies also found that the MAOC is much higher in the woodland area than cropping area (Sainepo et al., 2018, Toru and Kibret, 2019). Chapter 4 shows that the highest SOC content was in woodland, followed by pasture, irrigated cropping, and non-irrigated cropping areas. However, this study shows that MAOC and POC were not significantly different between irrigated cropping and non-irrigated cropping areas. The main reason for SOC loss by land use is the depletion of MAOC fraction, and its content is impacted by root exudates, tillage and microbial biomass (Modak et al., 2019, Ramesh et al., 2019, Wuaden et al., 2020).

The MAOC value under phenosoil was generally lower than genosoil, as land uses highly impact MAOC. However, the POC trend was not the same as the MAOC. Some pedogenons show that genosoil POC was higher than phenosoil, while other pedogenons showed the opposite. POC is dynamic and can rapidly change due to the decomposition of plant debris that is plentiful on the surface soil (Cambardella and Elliott, 1992).

### 5.5.2. SOC sequestration potential using SCORPAN factors and quantile regression.

#### 5.5.2.1. The physical interpretation of the covariates

Eight covariates were selected for the gam and qgam models, including one soil, four climate, one parent material, one organism and one relief factors. The fine fraction was included as the covariate because it is essential for the physicochemical stabilisation mechanisms of MAOC (Alvarez and Berhongaray, 2021, Xiao et al., 2022). Sanderman et al. (2021) mentioned that MAOC content is correlated with climate, and the temperature, precipitation and solar radiation strongly relate to the MAOC. Moreover, mineralogy and surface geochemistry information is useful for predicting MAOC because the concentration and ratios of gamma radiometric indicate the lithology and degree of weathering (Román Dobarco et al., 2022).

Based on the interpretable covariate response function (Appendix 5-2), the spline on the response of MAOC to fine fraction shows that the fine fraction increases with increasing MAOC content. The response of NDVI (vegetation covariate) shows a similar trend to the fine fraction. These trends agree with the common understanding of the relationship between MAOC and fine fraction (Paterson et al., 2020) and vegetation (Solomon et al., 2000). In constast, climate variables on aridity (ADM), solar radiation (RSM) and annual temperature range (TRA) shows a decreasing trend, i.e. increase dryness result in decrease MAOC. Soil minerals (Th/K) ratio shows a gradual increase response at small values but constant afterward.

### 5.5.3. Comparing C sequestration potential maps from two different approaches

Two methods were used to predict MAOC capacity, condition and potential. The first approach predicted MAOC content using pedogenon map (genosoil and phenosoil) approach with local

data (Figure 5-7). The second approach used regional data with climatic, vegetation, and soil physicochemical variables and quantile regression models (Figure 5-8).

Figure 5-9 shows the distribution of the MAOC values from the maps created by pedogenon and regional quantile regression approaches. The MAOC capacity and condition values created by the pedogenon approach were lower than the quantile regression approach, with a difference of about 1.89 mg C g $soil^{-1}$ for capacity and 1.54 mg C g $soil^{-1}$ for condition.

The C sequestration potential (capacity – condition) values for the quantile regression model were higher than the pedogenon map but insignificant. As the regional model does not include local data, we can conclude that the quantile regression model overpredicted the MAOC in Edgeroi.

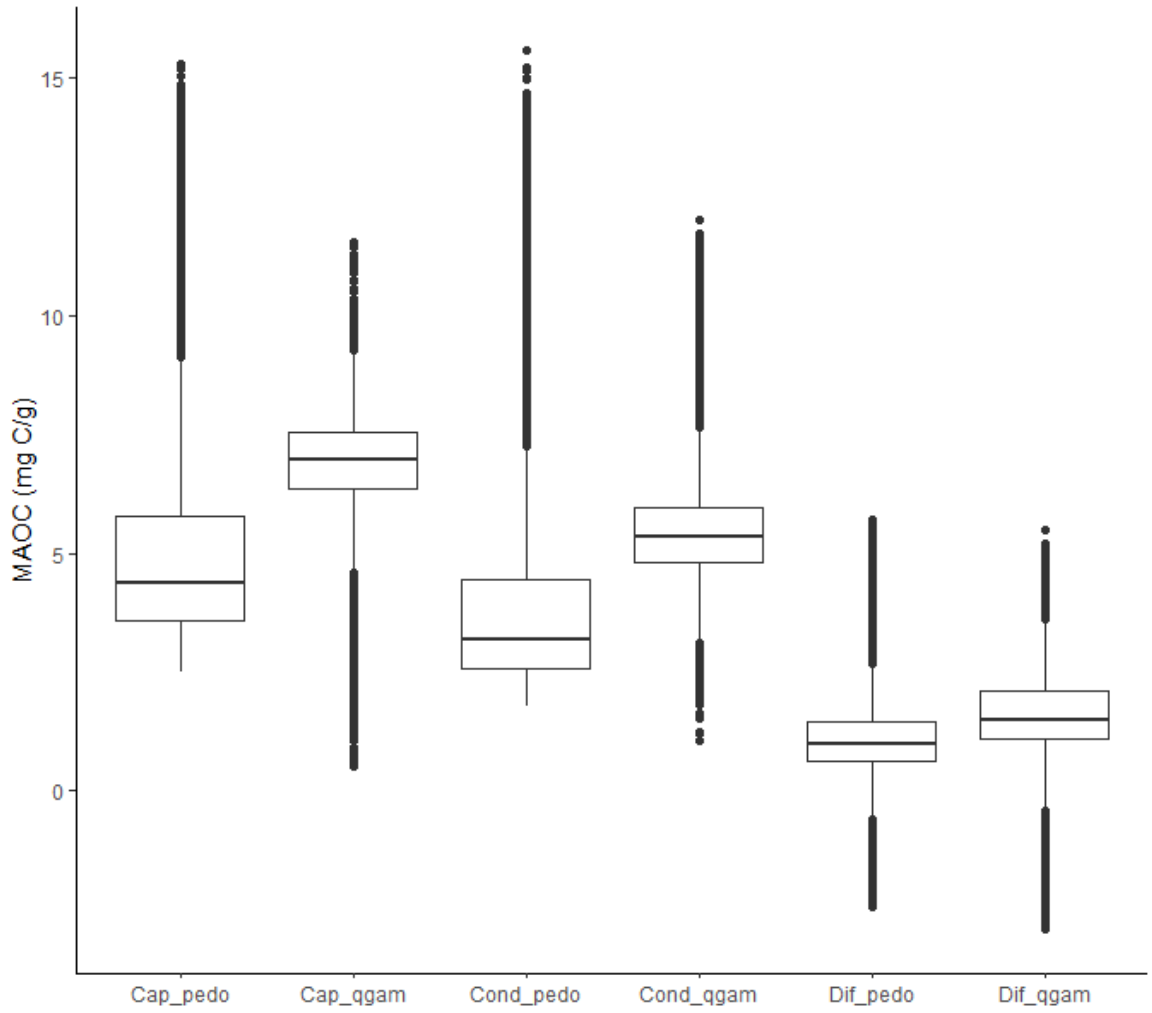



**Figure 5-9. A comparison of pedogenon (pedo) and quantile regression (qgam) approaches in predicting MAOC capacity (cap), condition (cond), and difference (dif) in Edgeroi.**

Figure 5-10 further compares observed MAOC in Edgeroi and their predicted values using the regional gam model. The regional model slightly overpredicted the median MAOC in the area but had a small range of values, and it also did not predict high values of observed MAOC. It is important to note that the SCaRP data were collected from agricultural soils and thus may not have the information to predict soils under woodland and native vegetation.

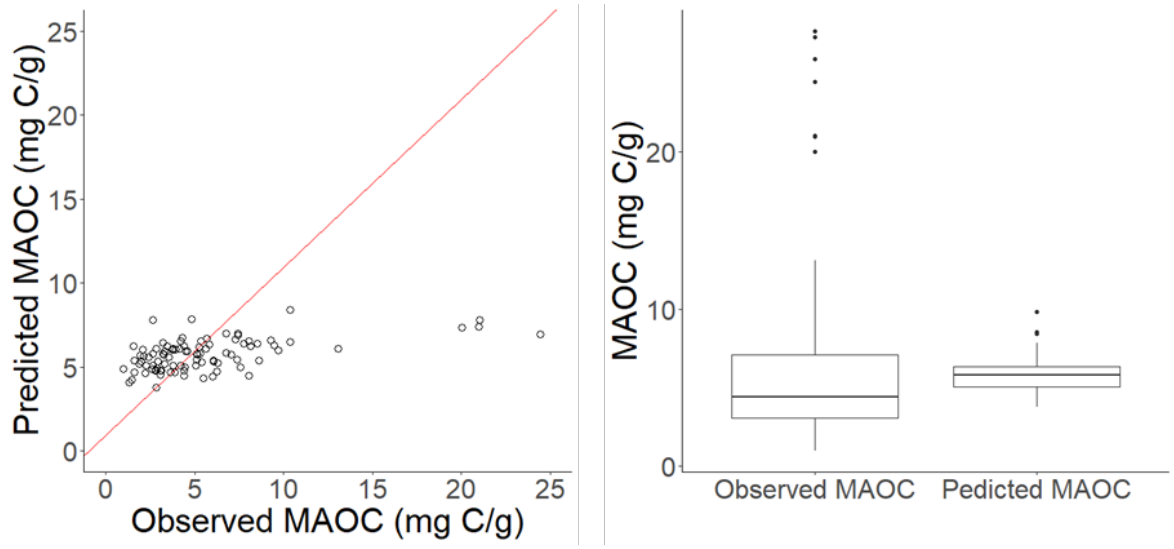



**Figure 5-10. A comparison of observed MAOC in Edgeroi and regional gam model predicted MAOC.**

#### 5.5.3.1. Comparing the pattern of maps created by two approaches

The pattern of C sequestration potential created by the two approaches was contrasted. The map created by pedogenon model showed that the value of MAOC change was much larger in the irrigated area than in the non-irrigated cropping area (Figure 5-7). This result is in line with the findings in Chapter 4; irrigated cropping area in Edgeroi shows high carbon loss due to intensive agriculture practices, including the short fallow phase and the intensive soil tillage. The main crop under the irrigated cropping area is cotton. Rabbi et al. (2014) confirmed that the significant decrease in SOC is due to the production of irrigated cotton, and SOC content and fractions were correlated negatively with the land use of irrigated cotton. Past studies have shown that SOC on clay soil under irrigated cotton in the Namoi Valley had decreased by 37% to 48% compared to uncropped areas (Conteh et al., 1997). Hence, the irrigated cropping practice impacted the stabilisation of soil C into MAOC. The pedogenon approach also shows that MAOC change was highest in pedogenon J (Figure 4-1, pedogenon map). This result is also in accordance with Chapter 4, where the loss of SOC content in pedogenon J is the highest among other pedogenons.

However, the regional quantile regression model result shows the opposite pattern. The MAOC content in non-irrigated cropping areas was higher than in irrigated cropping areas (Figure 5-8). This implies that the regional model did not capture the MAOC distribution correctly.

The C sequestration potential map generated by the pedogenon approach (Figure 5-7) recognises the floodplain areas where the values of MAOC difference were negative, indicating SOC sequestration ($C_{phenosoil} > C_{genosoil}$), while the positive $\Delta MAOC$ represents SOC loss ($C_{genosoil} > C_{phenosoil}$). This result indicates that under favourable conditions, MAOC can be increased over time. The value of MAOC change in condition can be affected by different vegetation densities, structure and composition, and intensity of human activities between the cropping area and floodplain area (Dalal et al., 2021).

Compared to the pedogenon approach, the map produced by the quantile regression mainly shows the pattern of cultivation land. The boundary of farming land can be observed in Figure 5-8, and the woodland area was clearly separated from other land use areas. This pattern is possibly affected by the covariate of the vegetation factor (NDVI Q2), which was included in the gam model to estimate current MAOC.

We concluded that the first approach (pedogenon mapping) produced a more reliable soil C sequestration potential map than the regional quantile regression model. There are several reasons. The first reason is that the pedogenon approach used local data to create a local DSM model, and the quantile regression approach built a regional model and applied it at a local scale. Hence the regional model cannot correctly capture the relationships between scorpan factors and MAOC content, and the result can have high uncertainty. Moreover, in the first approach, MAOC content was estimated from samples collected in the study area even though the number of samples was only 99. In contrast, the calibration data of the second approach consisted of 1607 sampling points extracted from a regional dataset. However, only 16 sample points out of 1607 were located in the study area on agricultural soils and were not evenly spread, and these few sampling points could not be representative of the study area. In addition, the covariate data on particle size fraction (clay and silt) from the regional model was estimated from DSM, not from the direct measurement, which could be a source of uncertainty in the model.

There are a few limitations in this study. In particular, the carbon fractions (MAOC) were predicted from a national (SCaRP) soil spectra library, which does not include soils under woodland and native vegetation. And as discussed above, the lack of samples in woodland areas may have impacted the inaccuracy of the regional model. As in Chapter 3, the maps of MAOC capacity are constrained by the limited number of genosoils.

### 5.5.4. Relationship with soil security

As described in the introduction section, the mapping of MAOC is related to two dimensions of soil security (capacity and condition). The capacity is characterised by soil attributes that determine the performance of soil functions but are not impacted by human activities. In contrast, the condition is the current state of the soil and is represented by soil attributes that are changed by human activities. In this study, we are mapping the capacity and condition of the soil to store carbon. Several factors can influence the capacity of soil to store carbon, including the type of soil, minerals, vegetation and climate. While soil has an inherent capacity to store carbon (mineralogy), it is affected by the environment (climate, water availability) and the type of vegetation. The two approaches used different methods to calculate the capacity and condition of MAOC.

Most studies on SOC potential mapping only used bulk SOC content, while this study focuses on the stable form SOC, which is more representative of stable form C storage in soil. The results show that MAOC on cropping soils was generally lower than its counterpart soil under woodland and grassland. This suggests that long-term cultivation has resulted in the loss of stabilised C and an opportunity to gain a more stable C with improved management. Nevertheless, the soil samples collected in this study (genosoils) defined the reference, and we term as capacity, and there could be a higher capacity of the soil to store carbon under different vegetation or conditions. For example, the MAOC difference map (Condition – Capacity) (Figure 5-7) shows positive values in forested and wetland areas, indicating SOC sequestration.

The data on capacity and condition can be used as an indicator for assessing soil security and also applied for measuring the effect of other dimensions (capital, connectivity and codification). For example, soil carbon sequestration is an essential aspect of natural capital, the value of natural resources and ecosystem services that are provided by nature. Soil carbon sequestration is an important ecosystem service because it stores C in the soil and can help in

mitigating climate change. Soil carbon is an important factor affecting soil's health and crop productivity, which in turn can have important implications for the general public. Maintaining healthy soil through SOC sequestration is, therefore, crucial for the well-being of people and the environment. Soil carbon connects to the public through its role in crop production and mitigating climate change. Maintaining healthy soil and increasing soil carbon levels through sustainable land management practices can benefit the public.

Finally, governments worldwide are now implementing policies that aim to mitigate climate change by promoting practices that increase soil carbon sequestration. The Carbon Farming Initiative (CFI) programme, launched by the Australian government, attempts to persuade farmers and land managers to adopt techniques that boost soil carbon retention and lower glasshouse gas emissions. Farmers and land managers who adopt measures that boost soil carbon sequestration or lower glasshouse gas emissions from livestock are given financial incentives by the CFI. These methods include adopting reduced tillage, crop rotation, rotational grazing systems, and restoring degraded ecosystems. Farmers and land managers can produce carbon credits by boosting SOC that can be traded on the carbon market. The money from the sale of carbon credits can be used to pay for the costs of putting the practices in place and bring in extra money for the landowners or farmers. Adopting practices that increase soil carbon sequestration and reduce greenhouse gas emissions cannot be applied as a blanket approach. This study's capacity and condition maps can help target areas with large sequestration potential and reduce investment risk.

## 5.6. Conclusion

SOC fraction change in south-eastern Australia was investigated with two approaches. The first approach compares genosoil and phenosoil for assessing MAOC. The other method developed a regional model to predict MAOC capacity and condition using a quantile gam approach. The change in MAOC content is heavily affected by land use, especially agricultural activities. The MAOC content in each pedogenon has a notable trend, but the MAOC value in genosoil is always higher than in phenosoil. The mapping result shows that MAOC was sequestrated in the woodland area, and MAOC was decreased in the agriculture area. The results can be linked with capacity and condition dimensions from soil security, and the capacity and condition data can support calculating the impact of other dimensions (capital, connectivity and codification).

## 5.7. References


ALVAREZ, R. & BERHONGARAY, G. 2021. Soil organic carbon sequestration potential of Pampean soils: comparing methods and estimation for surface and deep layers. Soil Research, 59, 346-358.

ARMSTRONG, R. N., POTGIETER, A. B., MARES, D. J., MRVA, K., BRIDER, J. & HAMMER, G. L. 2019. An integrated framework for predicting the risk of experiencing temperature conditions that may trigger late-maturity alpha-amylase in wheat across Australia. Crop and Pasture Science, 71, 1-11.

BAISDEN, W. T., PARFITT, R. L., ROSS, C., SCHIPPER, L. A. & CANESSA, S. 2013. Evaluating 50 years of time-series soil radiocarbon data: towards routine calculation of robust C residence times. Biogeochemistry, 112, 129-137.

BALDOCK, J., HAWKE, B., SANDERMAN, J. & MACDONALD, L. 2013a. Predicting contents of carbon and its component fractions in Australian soils from diffuse reflectance mid-infrared spectra. Soil Research, 51, 577-595.

BALDOCK, J., SANDERMAN, J., MACDONALD, L., PUCCINI, A., HAWKE, B., SZARVAS, S. & MCGOWAN, J. 2013b. Quantifying the allocation of soil organic carbon to biologically significant fractions. Soil Research, 51, 561-576.

BISSIRI, P. G., HOLMES, C. C. & WALKER, S. G. 2016. A general framework for updating belief distributions. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 78, 1103-1130.

BRENNING, A. 2008. Statistical geocomputing combining R and SAGA: The example of landslide susceptibility analysis with generalized additive models. Hamburger Beiträge zur Physischen Geographie und Landschaftsökologie, 19, 410.

BRIEDIS, C., DE MORAES SÁ, J. C., CAIRES, E. F., DE FÁTIMA NAVARRO, J., INAGAKI, T. M., BOER, A., NETO, C. Q., DE OLIVEIRA FERREIRA, A., CANALLI, L. B. & DOS SANTOS, J. B. 2012. Soil organic matter pools and carbon-protection mechanisms in aggregate classes influenced by surface liming in a no-till system. Geoderma, 170, 80-88.

CAMBARDELLA, C. A. & ELLIOTT, E. 1992. Particulate soil organic-matter changes across a grassland cultivation sequence. Soil science society of America journal, 56, 777-783.

CARTER, M., ANGERS, D., GREGORICH, E. & BOLINDER, M. 2003. Characterizing organic matter retention for surface soils in eastern Canada using density and particle size fractions. Canadian Journal of Soil Science, 83, 11-23.

CHEN, S., ARROUAYS, D., ANGERS, D. A., CHENU, C., BARRÉ, P., MARTIN, M. P., SABY, N. P. & WALTER, C. 2019. National estimation of soil organic carbon storage potential for arable soils: A data-driven approach coupled with carbon-landscape zones. Science of the Total Environment, 666, 355-367.

CHEN, S., MARTIN, M. P., SABY, N. P., WALTER, C., ANGERS, D. A. & ARROUAYS, D. 2018. Fine resolution map of top-and subsoil carbon sequestration potential in France. Science of the Total Environment, 630, 389-400.

CHENU, C., ANGERS, D. A., BARRÉ, P., DERRIEN, D., ARROUAYS, D. & BALESDENT, J. 2019. Increasing organic stocks in agricultural soils: Knowledge gaps and potential innovations. Soil and Tillage Research, 188, 41-52.

CONTEH, A., BLAIR, G., LEFROY, R. & MACLEOD, D. 1997. Soil organic carbon changes in cracking clay soils under cotton production as studied by carbon fractionation. Australian Journal of Agricultural Research, 48, 1049-1058.

COTRUFO, M. F., RANALLI, M. G., HADDIX, M. L., SIX, J. & LUGATO, E. 2019a. Soil carbon storage informed by particulate and mineral-associated organic matter. Nature Geoscience, 12, 989-+.

COTRUFO, M. F., RANALLI, M. G., HADDIX, M. L., SIX, J. & LUGATO, E. 2019b. Soil carbon storage informed by particulate and mineral-associated organic matter. Nature Geoscience, 12, 989-994.

DALAL, R. C., THORNTON, C. M., ALLEN, D. E., OWENS, J. S. & KOPITTKE, P. M. 2021. Long-term land use change in Australia from native forest decreases all fractions of soil organic

carbon, including resistant organic carbon, for cropping but not sown pasture. Agriculture, Ecosystems & Environment, 311, 107326.
DEPARTMENT OF PLANNING AND ENVIRONMENT 2018. State Vegetation Type Map: Border Rivers Gwydir / Namoi Region Version 2.0. VIS_ID 4467.
DOETTERL, S., STEVENS, A., SIX, J., MERCKX, R., VAN OOST, K., CASANOVA PINTO, M., CASANOVA-KATNY, A., MUÑOZ, C., BOUDIN, M. & ZAGAL VENEGAS, E. 2015. Soil carbon storage controlled by interactions between geochemistry and climate. Nature Geoscience, 8, 780-783.
ECO LOGICAL AUSTRALIA 2013. Refinement of vegetation mapping in the Namoi Catchment: Extant and pre-European. Armadale NSW.
FASIOLO, M., WOOD, S. N., ZAFFRAN, M., NEDELLEC, R. & GOUDE, Y. 2020. qgam: Bayesian non-parametric quantile regression modelling in R. arXiv preprint arXiv:2007.03303.
FASIOLO, M., WOOD, S. N., ZAFFRAN, M., NEDELLEC, R. & GOUDE, Y. 2021. Fast calibrated additive quantile regression. Journal of the American Statistical Association, 116, 1402-1412.
FENG, W., PLANTE, A. F. & SIX, J. 2013. Improving estimates of maximal organic carbon stabilization by fine soil particles. Biogeochemistry, 112, 81-93.
GALLANT, J., WILSON, N., PK, T., DOWLING, T. & READ, A. 2009. 3 second SRTM Derived Digital Elevation Model (DEM) Version 1.0. Canberra: Geoscience Australia.
GREEN, D., PETROVIC, J., BURRELL, M. & MOSS, P. 2011. Water resources and management overview: Namoi catchment.
GUILLAUME, T., MAKOWSKI, D., LIBOHOVA, Z., BRAGAZZA, L., SALLAKU, F. & SINAJ, S. 2022. Soil organic carbon saturation in cropland-grassland systems: Storage potential and soil quality. Geoderma, 406, 115529.
HASSINK, J. 1997. The capacity of soils to preserve organic C and N by their association with clay and silt particles. Plant and soil, 191, 77-87.
HASTIE, T. 2006. Gam: Generalized Additive Models, R Package, Version 0.98. 2006.
HE, S., MAZUMDAR, S. & ARENA, V. C. 2006. A comparative study of the use of GAM and GLM in air pollution research. Environmetrics: The official journal of the International Environmetrics Society, 17, 81-93.
HUANG, J., MCBRATNEY, A. B., MALONE, B. P. & FIELD, D. J. 2018. Mapping the transition from pre-European settlement to contemporary soil conditions in the Lower Hunter Valley, Australia. Geoderma, 329, 27-42.
ISBELL, R. 2002. The Australian Soil Classification., Rev. edn (CSIRO Publishing: Melbourne).
KEILUWEIT, M., BOUGOURE, J. J., NICO, P. S., PETT-RIDGE, J., WEBER, P. K. & KLEBER, M. 2015. Mineral protection of soil carbon counteracted by root exudates. Nature Climate Change, 5, 588-595.
KLEBER, M., EUSTERHUES, K., KEILUWEIT, M., MIKUTTA, C., MIKUTTA, R. & NICO, P. S. 2015. Mineral–organic associations: formation, properties, and relevance in soil environments. Advances in agronomy, 130, 1-140.
KÖGEL-KNABNER, I., GUGGENBERGER, G., KLEBER, M., KANDELER, E., KALBITZ, K., SCHEU, S., EUSTERHUES, K. & LEINWEBER, P. 2008. Organo-mineral associations in temperate soils: Integrating biology, mineralogy, and organic matter chemistry. Journal of Plant Nutrition and Soil Science, 171, 61-82.
LAVALLEE, J. M., SOONG, J. L. & COTRUFO, M. F. 2020. Conceptualizing soil organic matter into particulate and mineral-associated forms to address global change in the 21st century. Global Change Biology, 26, 261-273.
LEHMANN, J. & KLEBER, M. 2015. The contentious nature of soil organic matter. Nature, 528, 60-68.
MCBRATNEY, A., FIELD, D. J. & KOCH, A. 2014. The dimensions of soil security. Geoderma, 213, 203-213.
MCBRATNEY, A. B., FIELD, D., MORGAN, C. L. & HUANG, J. 2019. On soil capability, capacity, and condition. Sustainability, 11, 3350.
MCBRATNEY, A. B., SANTOS, M. M. & MINASNY, B. 2003. On digital soil mapping. Geoderma, 117, 3-52.

MCGARRY, D., WARD, W. T., MCBRATNEY, A. B. & SOILS, C. D. O. 1989. Soil Studies in the Lower Namoi Valley: Methods and Data. 1, The Edgeroi Data Set, CSIRO Division of Soils.
MEVIK, B.-H. & WEHRENS, R. 2007. The pls package: principal component and partial least squares regression in R. Journal of Statistical software, 18, 1-23.
MILLER, D. L. 2019. Bayesian views of generalized additive modelling. arXiv preprint arXiv:1902.01330.
MODAK, K., GHOSH, A., BHATTACHARYYA, R., BISWAS, D. R., DAS, T. K., DAS, S. & SINGH, G. 2019. Response of oxidative stability of aggregate-associated soil organic carbon and deep soil carbon sequestration to zero-tillage in subtropical India. Soil and Tillage Research, 195, 104370.
NARRABRI SHIRE COUNCIL. 2021. Historical Walks [Online]. [Accessed].
PADARIAN, J., MINASNY, B., MCBRATNEY, A. & SMITH, P. 2022. Soil carbon sequestration potential in global croplands. PeerJ, 10, e13740.
PATERSON, K. C., CLOY, J. M., REES, R. M., BAGGS, E. M., MARTINEAU, H., FORNARA, D., MACDONALD, A. J. & BUCKINGHAM, S. 2020. Estimating maximum mineral associated organic carbon in UK grasslands. Biogeosciences Discuss, 2020, 1-28.
PAUL, E. A. 2016. The nature and dynamics of soil organic matter: Plant inputs, microbial transformations, and organic matter stabilization. Soil Biology and Biochemistry, 98, 109-126.
RABBI, S. F., TIGHE, M., COWIE, A., WILSON, B. R., SCHWENKE, G., MCLEOD, M., BADGERY, W. & BALDOCK, J. 2014. The relationships between land uses, soil management practices, and soil carbon fractions in South Eastern Australia. Agriculture, Ecosystems & Environment, 197, 41-52.
RAMESH, T., BOLAN, N. S., KIRKHAM, M. B., WIJESEKARA, H., KANCHIKERIMATH, M., RAO, C. S., SANDEEP, S., RINKLEBE, J., OK, Y. S. & CHOUDHURY, B. U. 2019. Soil organic carbon dynamics: Impact of land use changes and management practices: A review. Advances in agronomy, 156, 1-107.
ROBERTSON, M., LLEWELLYN, R., MANDEL, R., LAWES, R., BRAMLEY, R., SWIFT, L., METZ, N. & O'CALLAGHAN, C. 2012. Adoption of variable rate fertiliser application in the Australian grains industry: status, issues and prospects. Precision agriculture, 13, 181-199.
ROCCI, K. S., LAVALLEE, J. M., STEWART, C. E. & COTRUFO, M. F. 2021. Soil organic carbon response to global environmental change depends on its distribution between mineral-associated and particulate organic matter: A meta-analysis. Science of The Total Environment, 793, 148569.
ROMÁN DOBARCO, M., MCBRATNEY, A., MINASNY, B. & MALONE, B. 2021a. A framework to assess changes in soil condition and capability over large areas. Soil Security, 4, 100011.
ROMÁN DOBARCO, M., MCBRATNEY, A., MINASNY, B. & MALONE, B. 2021b. A modelling framework for pedogenon mapping. Geoderma, 393, 115012.
ROMÁN DOBARCO, M., WADOUX, A. M., MALONE, B., MINASNY, B., MCBRATNEY, A. B. & SEARLE, R. 2022. Mapping soil organic carbon fractions for Australia, their stocks and uncertainty. Biogeosciences Discussions, 1-38.
ROSSITER, D. G. & BOUMA, J. 2018. A new look at soil phenoforms – Definition, identification, mapping. Geoderma, 314, 113-121.
SAINEPO, B. M., GACHENE, C. K. & KARUMA, A. 2018. Assessment of soil organic carbon fractions and carbon management index under different land use types in Olesharo Catchment, Narok County, Kenya. Carbon balance and management, 13, 1-9.
SANDERMAN, J., BALDOCK, J. A., DANGAL, S. R., LUDWIG, S., POTTER, S., RIVARD, C. & SAVAGE, K. 2021. Soil organic carbon fractions in the Great Plains of the United States: an application of mid-infrared spectroscopy. Biogeochemistry, 156, 97-114.
SEARLE, R., MALONE, B., WILFORD, J., AUSTIN, J., WARE, C., WEBB, M., ROMAN DOBARCO, M. & VAN NIEL, T. 2022. TERN Digital Soil Mapping Raster Covariate Stacks. CSIRO.
SIX, J., CONANT, R. T., PAUL, E. A. & PAUSTIAN, K. 2002. Stabilization mechanisms of soil organic matter: implications for C-saturation of soils. Plant and soil, 241, 155-176.

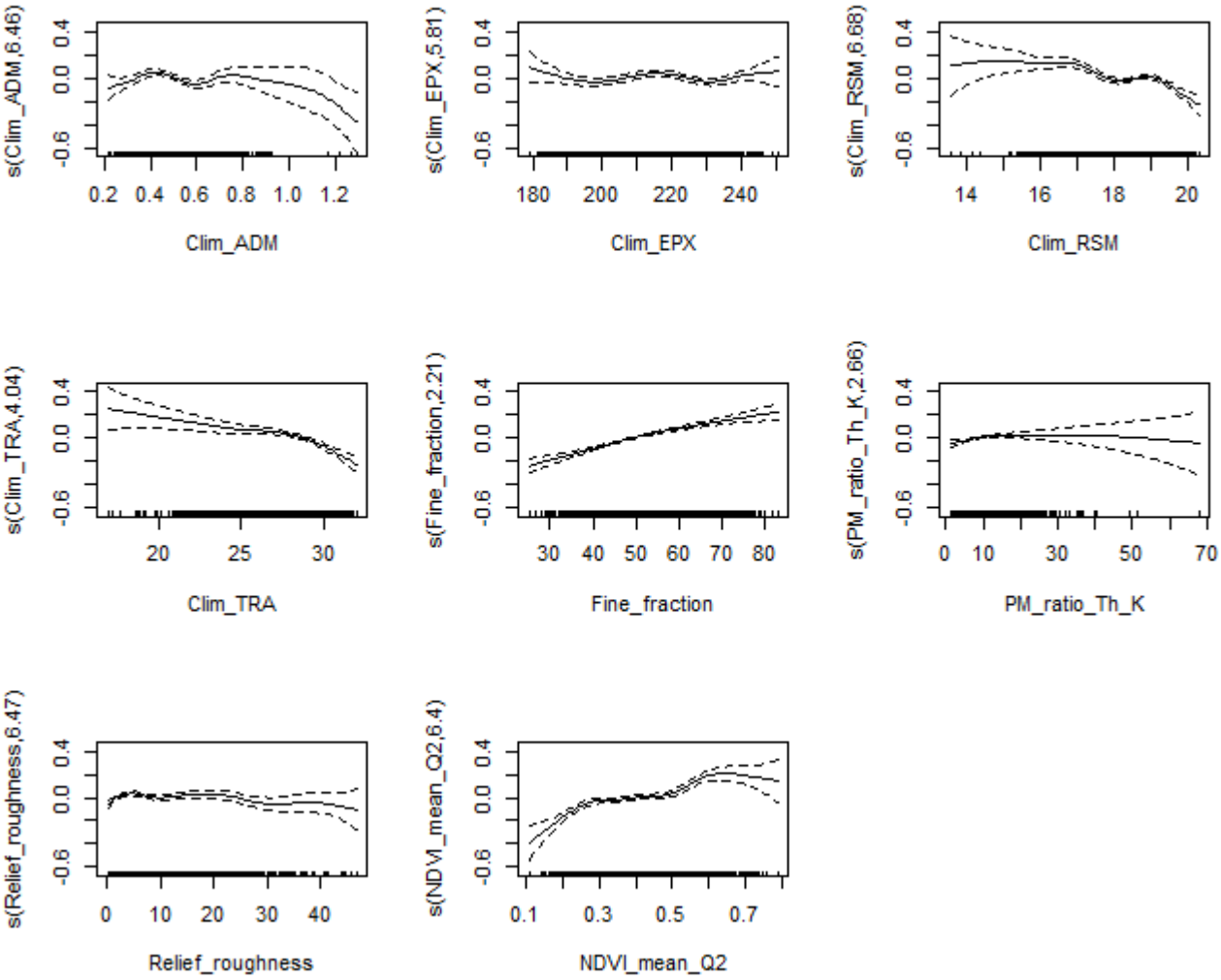
SIX, J., ELLIOTT, E. T. & PAUSTIAN, K. 2000. Soil macroaggregate turnover and microaggregate formation: a mechanism for C sequestration under no-tillage agriculture. Soil Biology and Biochemistry, 32, 2099-2103.
SOLLINS, P., SWANSTON, C., KLEBER, M., FILLEY, T., KRAMER, M., CROW, S., CALDWELL, B. A., LAJTHA, K. & BOWDEN, R. 2006. Organic C and N stabilization in a forest soil: evidence from sequential density fractionation. Soil Biology and Biochemistry, 38, 3313-3324.
SOLOMON, D., LEHMANN, J. & ZECH, W. 2000. Land use effects on soil organic matter properties of chromic luvisols in semi-arid northern Tanzania: carbon, nitrogen, lignin and carbohydrates. Agriculture, Ecosystems & Environment, 78, 203-213.
SPARROW, L., BELBIN, K. & DOYLE, R. 2006. Organic carbon in the silt+ clay fraction of Tasmanian soils. Soil use and management, 22, 219-220.
STEWART, C. E., PAUSTIAN, K., CONANT, R. T., PLANTE, A. F. & SIX, J. 2007. Soil carbon saturation: concept, evidence and evaluation. Biogeochemistry, 86, 19-31.
TORU, T. & KIBRET, K. 2019. Carbon stock under major land use/land cover types of Hades sub-watershed, eastern Ethiopia. Carbon balance and management, 14, 1-14.
VON LÜTZOW, M., KÖGEL-KNABNER, I., EKSCHMITT, K., FLESSA, H., GUGGENBERGER, G., MATZNER, E. & MARSCHNER, B. 2007. SOM fractionation methods: relevance to functional pools and to stabilization mechanisms. Soil Biology and Biochemistry, 39, 2183-2207.
WADOUX, A., MALONE, B., MINASNY, B., FAJARDO, M. & MCBRATNEY, A. B. 2021. Soil Spectral Inference with R.'.
WIESMEIER, M., HÜBNER, R., SPÖRLEIN, P., GEUß, U., HANGEN, E., REISCHL, A., SCHILLING, B., VON LÜTZOW, M. & KÖGEL-KNABNER, I. 2014. Carbon sequestration potential of soils in southeast Germany derived from stable soil organic carbon saturation. Global change biology, 20, 653-665.
WIESMEIER, M., MUNRO, S., BARTHOLD, F., STEFFENS, M., SCHAD, P. & KÖGEL-KNABNER, I. 2015. Carbon storage capacity of semi-arid grassland soils and sequestration potentials in northern China. Global Change Biology, 21, 3836-3845.
WILFORD, J. & KROLL, A. 2020. Complete Radiometric Grid of Australia (Radmap) v4 2019 with modelled infill. Canberra: Geoscience Australia.
WOOD, S. 2017. Generalized additive models: An introduction with R 2 An Introduction with R. Generalized Additive Models, 10, 9781315370279.
WOOD, S. & WOOD, M. S. 2015. Package 'mgcv'. R package version, 1, 729.
WOOD, S. N., PYA, N. & SÄFKEN, B. 2016. Smoothing parameter and model selection for general smooth models. Journal of the American Statistical Association, 111, 1548-1563.
WUADEN, C. R., NICOLOSO, R. S., BARROS, E. C. & GRAVE, R. A. 2020. Early adoption of no-till mitigates soil organic carbon and nitrogen losses due to land use change. Soil and Tillage Research, 204, 104728.
XIAO, Y., XUE, J., ZHANG, X., WANG, N., HONG, Y., JIANG, Y., ZHOU, Y., TENG, H., HU, B. & LUGATO, E. 2022. Improving pedotransfer functions for predicting soil mineral associated organic carbon by ensemble machine learning. Geoderma, 428, 116208.
ZHAO, L., SUN, Y., ZHANG, X., YANG, X. & DRURY, C. 2006. Soil organic carbon in clay and silt sized particles in Chinese mollisols: relationship to the predicted capacity. Geoderma, 132, 315-323.

## 5.8. Appendices

**Appendix 5-1. Results of model accuracy for the prediction of SOC content and fractions (MAOC, POC and PyOC) using MIR on the SCaRP dataset.**

| | $R^2$ | RMSE (mg C/g) |
|---|---|---|
| **SOC** | 0.95 | 4.06 |
| **POC** | 0.86 | 1.85 |
| **MAOC** | 0.91 | 2.49 |
| **PyOC** | 0.86 | 1.46 |

**Appendix 5-2. The response surface splines for each variable on the gam prediction of MAOC.**



# Chapter 6. Overall Discussion and Conclusions

## 6.1. Soil change

Humans have impacted the Australian continent and its soil for more than 50,000 years. However, the impact from humans has been most significant in the last 200 years, following European settlement. These include land use change, agricultural practices, urban development, pollution and climate change. As a result, widespread degradation of cropping soils, including soil erosion, acidification, SOC decline, and nutrient imbalances, have been reported. We are aware that soil has changed considerably, but we have not precisely quantified the extent and location of that change. Thus, there is a need for a new way of mapping where and how soil has changed and where soil functions have been compromised. This information will allow us to identify areas under threat to degradation and understand how soil responds to climate and land use changes. Furthermore, this knowledge will allow us to recommend interventions that protect our soil resources to maintain the production of food, fibre and fresh water and sustain the biodiversity and overall ecosystem protection.

This thesis addresses this challenge by designing a method to map how much soil has changed and applied it in the Edgeroi area of New South Wales.

## 6.2. Pedogenon mapping

This thesis created a *local* pedogenon map in Edgeroi, New South Wales, Australia, following the framework by Román Dobarco et al. (2021b). Pedogenon mapping was proposed by Román Dobarco et al. (2021b) as a way to stratify the landscape into soil units. They created soil groups in New South Wales by applying unsupervised classification (k-means) to quantitative variables representing soil-forming factors at a reference time. Applying the pedogenon mapping in NSW resulted in pedogenon classes that can explain 40% of the variation in stable soil properties (e.g. texture). In a subsequent study, Román Dobarco et al. (2021a) estimated soil change in NSW with subclasses of the pedogenon map (remnant pedogenon (genosoil) and pedophenon (phenosoil)). The subclasses were stratified using the combined information, including the condition of the land, current land use and the location of native vegetation. Comparing these pedophenons under agricultural influences and pedogenons that are still under native vegetation, they detect trends of soil acidification and SOC losses in topsoils.

This study further expands on the concept by applying it and testing it at a regional scale in Edgeroi, New South Wales. Some key findings from this topic include:

- Generation of optimal covariates representing the scorpan factors which could not be done at a state level. In particular, Chapter 2 shows an approach to generating a parent material map that contains information on both parent material and the age of the landscape.
- Creation of local pedogenon classes using statistical approaches. Using local scorpan layers, Chapter 3 describes the method for creating pedogenon and identifying the optimal number of classes. Based on the land use map, within a pedogenon, genosoils and phenosoils were derived.
- Design of a soil sampling strategy to capture genosoils and phenosoil within a pedogenon.

- Soil sampling to collect 1 m soil cores and analysing the soils every 10 cm depth using NIR and MIR spectroscopy. This rapid soil analysis technique allows the investigation of soil properties up to 1 m.
- Validation of the pedogenon classes using multivariate statistical analyses (PCA and RDA) of soil profile properties (clay, sand, cation exchange capacity, pH, and organic carbon). Variance partitioning analysis confirmed that a large proportion of the variation of the data is explained by pedogenon (49 %) as opposed to landuse (5 %).
- Pedogenons act as soil series in a local soil map. The qualitative description of the pedogenons demonstrates that each pedogenon originated from distinct parent material, occupied specific geographic locations, and produced distinctive soil profiles.

## 6.3. Mapping Soil Change

This thesis aims to find effective ways to map and quantify soil change over a region using the pedogenon mapping approach. Current methods for mapping soil change require repeated surveys, which will be costly and time-consuming. The space-for-time substitution method offers another estimation of soil change, however the comparison is usually made locally, so-called "over the fence" approach (Filippi et al., 2016). An extension of this method is called the space-for-time extrapolation method. This involves fitting a spatial model using *scorpan* factors on existing soil data. This fitted model can be used for backcasting and forecasting. For example, replacing the "*o*" factor under cropland with native vegetation or woodland would allow the calculation of the original SOC content before cultivation. Alternatively, replacing the current "c" factor with future climate projection allows forecasting SOC under future climate change scenarios. Such an approach has its assumption of immediate equilibrium of all factors. It may be valid under certain circumstances if the samples expand all scorpan factors. However, if the soil samples collected were collected with a preference in agricultural soils, such extrapolation over native vegetation for the whole area is invalid.

This thesis addresses the current model limitations and fills this knowledge gap by proposing pedogenon mapping to sample, estimate and map soil change by spatialising the "space for time" idea. Based on the samples collected in Chapter 4, SOC condition maps under native vegetation and current conditions was created using the DSM approach. The first map is the map of SOC content under native vegetation, produced using only genosoil sampling points, and the second map is the current SOC condition map, produced with both genosoil and phenosoil points. Thus the difference between the two maps shows the SOC change since agricultural establishment in the area.

Some key findings from this topic include:

- Agricultural activities homogenised the variation of soil profiles, especially SOC. Natural pedogenesis produced soils with high levels of variability, but decades of agricultural methods had changed the soil qualities to make them less variable. The variability of soil properties under agriculture was cut in half compared to the soils under native vegetation.
- Different land uses varied SOC change. The biggest SOC loss was in soils under irrigated cropping, with a 38% decrease in content at the 0–10 cm depth (relative to SOC under natural vegetation), followed by dryland cropping (30% loss) and pasture (19% loss).
- SOC loss reduced dramatically with depth and was greatest in the surface soils. Compared to soils with natural vegetation, the top 30 cm of pasture experienced the greatest loss of SOC. The loss of SOC under dryland and irrigated agriculture was measured to a depth of 50 cm.
- SOC change under each pedogenon was variable due to different soil types, the intensity of agricultural practices and vegetation types. The highest SOC loss was under pedogenon A and J. The soil type under both pedogenons is black vertosol. Pedogenon A, located west of the study area, was under irrigated cropping, while pedogenon J, east of the study area, was under non-irrigated cropping. The lowest SOC loss was under pedogenon M, located in the southeast of the study area. Its soil type is dark brown vertosol, and its landuse is non-irrigated cropping.

## 6.4. SOC Sequestration potential

SOC can be lost or accumulated due to land management and climate change. Soil can be a carbon store for climate mitigation purposes, and this thesis demonstrates the use of pedogenon mapping created in Chapter 3 for mapping an important soil function: storing carbon.

Several methods can be used to determine the potential of soil to sequester carbon. Some of these methods include: models and simulations such as RothC or Century to estimate the potential for carbon sequestration in a given soil type based on factors such as climate, vegetation, and land use, field trials to measure carbon sequestration that occurs in a given soil under different management practices or conditions, and meta analysis based on published results for given soils under different management practices or conditions.

It is important to consider the specific characteristics of the soil and the local environment in order to estimate the potential for carbon sequestration accurately. This study particularly looks at the stable form of carbon, mineral-associated organic carbon (MAOC), and organic carbon bound to the surfaces of mineral particles. MAOC is typically more stable and resistant to decomposition than other forms of organic carbon found in soil. MAOC is a critical component of soil, as it can contribute to soil fertility and play a role in the carbon cycle. Several factors can influence the amount of MAOC in soil, including the type and composition of the minerals present, fine fraction, and soil pH. However, the amount of MAOC in the soil is further limited by climate and vegetation and can be affected by human activities, such as land use and management practices.

This thesis used the pedogenon mapping concept to predict MAOC potential and condition. MAOC capacity was mapped based on genosoil samples and MAOC condition was mapped using samples from phenosoils using spatial covariates and random forest models. Different land uses impacted MAOC; generally, MAOC concentration in genosoils was always higher

than in phenosoils. The results show the potential of SOC sequestration on cropping soils of Edgeroi.

## 6.5. Relationship with Soil Security

Soil security is concerned with the maintenance and improvement of the global soil resource to produce food, fibre and fresh water, contribute to energy and climate sustainability, and maintain the biodiversity and the overall protection of the ecosystem (McBratney et al., 2014).

Soil security has five dimensions: capability, condition, capital, connectivity & codification. The pedogenon classes measure soil capacity and conditions, two biophysical measures of soil security dimensions (Román Dobarco et al., 2021b, Román Dobarco et al., 2021a). The capacity of soils is a measure of soil properties that determine the performance of soil function but are not modified by anthropogenic activities. The soil condition is the current soil state and corresponds to soil properties, which are changed by anthropogenic activities.

The concept of capacity and conditions represented by genosoils and phenosoils allow us to quantify soil security. Previous concepts, such as soil quality, usually measure the quality of soil to perform a function based on a standard set of criteria. However, soils have different capacities to perform functions. Thus, setting the genosoil as a reference for a soil type (pedogenon) will allow a better comparison.

Additionally, the capacity and condition can improve soil connectivity to the public in several ways, such as agricultural productivity, climate change mitigation and human health.

Soil carbon is an essential component of natural capital, the stock of natural resources and ecosystem services that provide value to society, as soil carbon provides a range of ecosystem services that provide value to the community.

Finally, soil carbon is a fundamental component of the Carbon Farming Initiative (CFI), a program in Australia that aims to promote the adoption of practices that can reduce greenhouse gas emissions and sequester carbon in the landscape. The CFI is governed by the Carbon

Credits (Carbon Farming Initiative) Act 2011 and the Carbon Credits (Carbon Farming Initiative) Regulations 2011.

## 6.6. Assumptions and Limitations

### 6.6.1. Assumption

This study has a few assumptions about using pedogenons for investigating soil change. The first assumption is that other soil forming factors, except organism factors, have not changed significantly over the past 200 years. Most area in the study area is flat except for the west of the study area (mountain area), so there may be no topography change over the years. However, agricultural activities may cause soil water and wind erosion and change the soil and landscape. Indeed data from Chapter 3 shows that topsoil clay content in phenosoils was consistently higher compared to genosoils. This implies that the original topsoil under cropping may have been eroded; thus, the current topsoil would have higher clay content.

The second assumption is that climate change does not affect SOC content. The same climate factor was applied to both SOC content under native vegetation and current SOC content. In this study, the intensity of agropedogenesis is much higher than natural pedogenesis. While climate change impacts SOC, its impact could be small compared to human activities (Brevik, 2013, Kuzyakov and Zamanian, 2019).

The last assumption is that the genosoil samples (under current native vegetation) have not changed since European settlement. However, collecting land history data in the area for the past 100-200 years is impossible. Current land use in most genosoil areas is densely packed woodland, implying that land use has not changed for at least 100 years (Chapter 3). It is also known that some areas may have been replanted.

### 6.6.2. Limitations

There are a few limitations in this study. The first limitation is that the pre-European vegetation map was created with modelling approaches (Eco Logical Australia, 2013), so the prediction

result is highly uncertain and can be biased. Fortunately, the pedogenon map produced using the vegetation map showed meaningful information and supported the study of soil change. However, when covariates are created by modelling, it is imperative to calculate the model accuracy and understand the covariate uncertainty effects (Gu et al., 2010).

The other limitation is that the number of genosoil is limited, with only 47 sample points. The location of sampling points was decided based on the genosoil and phenosoil or map of land use. Hence, this sampling is not optimal for mapping as genosoil sampling points do not cover the whole study area. Moreover, the number of sampling points is small due to limited time and resources. These two limitations impacted the mapping result, which has high prediction uncertainty.

The selection of remnant genosoil sites could be biased and not representative of the remnant pedogenon. It is known that land managers selected the best soil for agricultural lands, and thus, not all pedogenons may have remnant genosoils, or the soils under woodland may not be representative of the area. To accurately study soil changes resulting from human activities, it is recommended to select study areas that have limited or absent remnant genosoil areas, such as woodlands. Instead, areas that have been less disturbed by agricultural practices should be classified as phenosoil 1. This classification will distinguish them from remnant genosoil (woodlands) and phenosoil 2 (agricultural land) in the area. Calculating soil changes can be achieved by determining the difference between phenosoil 1 and phenosoil 2. This approach can help researchers more accurately quantify the impact of human activities on soil properties and develop appropriate soil management strategies to ensure sustainable land use.

## 6.7. Recommendation to improve the method of the thesis for future works

The combined approaches from this thesis can be further developed for assessing soil change which can be implemented from regional to continental scales. For future works, a few recommendations exist to produce a more robust pedogenon map at different scales and better investigate soil change.

The first recommendation is selecting optimal environmental covariates. More diverse and relevant covariates based on the soil forming factors will produce more accurate and robust pedogenon maps. The large number and types of covariates could positively affect the creation of a calibration model for mapping due to the high probability of selecting the best covariates, but there is also a problem of redundancy and irrelevant covariates. Hence, selecting the covariates requires specific soil science knowledge and statistical approaches. Several statistical methods are already used to find optimal covariates for producing accurate soil maps. For example, Poggio et al. (2013) implemented two approaches (general additive models and regression trees) to select a subset of the covariates providing a calibration model which shows high performance.

Moreover, specific machine learning methods have been designed to find optimal covariates. Many studies have made efforts to find optimal covariates for soil mapping (Zeraatpisheh et al., 2022, Behrens et al., 2010, Lacoste et al., 2016). Hence, creating a better pedogenon map requires the selection of optimal covariates.

The second recommendation is on the sampling design and the number and location of soil sampling points. Soil samples should cover the whole area, and the limitation of the current design is that the sampling points did not cover the whole study area. Hence, the design of the sampling point with the information of the pedogenon classes and land use should be

considered carefully due to the high cost of each soil sampling trip. Many papers also explained and concluded that the design of a soil survey scheme is critical in the first step of the digital soil mapping technique (Domburg et al., 1994, De Gruijter et al., 2006). However, in the field, one could be challenged with restricted access to specific sites, and thus the issue of sample alternatives should be considered.

The third recommendation is to collect information on land use history. The limitation of using satellite time series information is that the human-induced land degradation might have taken place prior to the availability of remote sensing imagery. When available, historical land use maps and aerial photographs will inform on land use changes that might have taken place before the availability of remote sensing imagery. The history of land use, crop types, and how long and often they were cultivated are essential to consider when mapping such diverse areas. This information with the land use data can support deciding the soil sampling points in each pedogenon class. The decision of sampling points in this study was based only pedogenon and land use data. With extra information on land use history, more areas of phenosoils can be stratified and represent different degrees of human influences. Moreover, the information on land use history can be essential data for data interpretation. For example, Chapter 3 shows that certain pedogenons have large SOC change, but without detailed land use history data, it is challenging to find the specific reason.

The fourth recommendation is the combination of pedogenon mapping and process-based modelling for soil function evaluation. Process-based models can be utilised to capture the unique interactions among climate, soil, and land management and their impacts on soil functions, thus helping to quantify the value of soils. Scenario modelling can also be used to forecast soil functions under different cropping systems, management, and climates across pedogenons.

## 6.8. References


BEHRENS, T., ZHU, A.-X., SCHMIDT, K. & SCHOLTEN, T. 2010. Multi-scale digital terrain analysis and feature selection for digital soil mapping. Geoderma, 155, 175-185.

BREVIK, E. C. 2013. The potential impact of climate change on soil properties and processes and corresponding influence on food security. Agriculture, 3, 398-417.

DE GRUIJTER, J., BRUS, D. J., BIERKENS, M. F. & KNOTTERS, M. 2006. Sampling for natural resource monitoring, Springer.

DOMBURG, P., DE GRUIJTER, J. & BRUS, D. 1994. A structured approach to designing soil survey schemes with prediction of sampling error from variograms. Geoderma, 62, 151-164.

ECO LOGICAL AUSTRALIA 2013. Refinement of vegetation mapping in the Namoi Catchment: Extant and pre-European. Armadale NSW.

FILIPPI, P., MINASNY, B., CATTLE, S. & BISHOP, T. 2016. Monitoring and modeling soil change: the influence of human activity and climatic shifts on aspects of soil spatiotemporally. Advances in agronomy, 139, 153-214.

GU, H., KENNEY, T. & ZHU, M. 2010. Partial generalized additive models: an information-theoretic approach for dealing with concurvity and selecting variables. Journal of Computational and Graphical Statistics, 19, 531-551.

KUZYAKOV, Y. & ZAMANIAN, K. 2019. Reviews and syntheses: Agropedogenesis–humankind as the sixth soil-forming factor and attractors of agricultural soil degradation. Biogeosciences, 16, 4783-4803.

LACOSTE, M., MULDER, V. L., RICHER-DE-FORGES, A. C., MARTIN, M. P. & ARROUAYS, D. 2016. Evaluating large-extent spatial modeling approaches: A case study for soil depth for France. Geoderma Regional, 7, 137-152.

MCBRATNEY, A., FIELD, D. J. & KOCH, A. 2014. The dimensions of soil security. Geoderma, 213, 203-213.

POGGIO, L., GIMONA, A. & BREWER, M. J. 2013. Regional scale mapping of soil properties and their uncertainty with a large number of satellite-derived covariates. Geoderma, 209, 1-14.

ROMÁN DOBARCO, M., MCBRATNEY, A., MINASNY, B. & MALONE, B. 2021a. A framework to assess changes in soil condition and capability over large areas. Soil Security, 4, 100011.

ROMÁN DOBARCO, M., MCBRATNEY, A., MINASNY, B. & MALONE, B. 2021b. A modelling framework for pedogenon mapping. Geoderma, 393, 115012.

ZERAATPISHEH, M., GAROSI, Y., OWLIAIE, H. R., AYOUBI, S., TAGHIZADEH-MEHRJARDI, R., SCHOLTEN, T. & XU, M. 2022. Improving the spatial prediction of soil organic carbon using environmental covariates selection: A comparison of a group of environmental covariates. Catena, 208, 105723.