Soil Security 4 (2021) 100011

Contents lists available at ScienceDirect

Soil Security

journal homepage: www.sciencedirect.com/journal/soil-security

A framework to assess changes in soil condition and capability over large areas

Mercedes Román Dobarco [a,*], Alex McBratney [a], Budiman Minasny [a], Brendan Malone [b]

[a] Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, 1 Central Avenue, Eveleigh, NSW 2015, Australia
[b] CSIRO Agriculture and Food, Clunies Ross Street, Black Mountain, ACT 2601, Australia

ARTICLE INFO

Keywords:
Soil condition
Soil capability
Soil change
Digital soil mapping

ABSTRACT

The assessment of changes in soil condition and capability requires the identification of a reference state specific to each soil class. This study develops a framework for mapping soil classes that can be used as a reference state. It identifies soil classes that should have undergone similar historic anthropedogenesis, and differentiate, within each class, zones that have been less affected by human activities. This approach could be used as a baseline for assessing contemporary soil change, as demonstrated in the state of New South Wales in Australia. First, we established soil classes with similar multimillennial natural pedogenesis and historic anthropedogenesis, called pedogenons. This was achieved by applying unsupervised classification (*k*-means) to a set of quantitative state variables, proxies of the soil-forming factors at the time of the European settlement in New South Wales (climate, relief, parent material, and estimated pre-1750s vegetation). Pedogenon classes were then stratified into subclasses (ranging from remnant pedogenons to different pedophenons) by combining information on native vegetation extent, status (remnant or cleared) and current land use (i.e., land use history). The stratification of 1000 pedogenon classes resulted in 5448 subclasses, ranging from remnant pedogenons (located in protected areas of intact native vegetation), quasi-remnant pedogenons (production with low intervention on remnant native vegetation), cleared, grazing, and cropping pedophenons. The median of the area proportion of the pedogenon that was still preserved as remnant vegetation was 5.3%. This quasi-remnant pedogenon or the less affected pedophenon could be used as reference state. Pedophenon grazing and cropping occupied larger areas, with mean values of 73 km^2 and 153 km^2, respectively. The application of this framework for assessing soil change is illustrated using legacy data of topsoil pH (5 – 15 cm) as one indicator of soil condition. The ability of the pedogenon and pedophenon subclasses for explaining the variation of three stable (total Si, total Al, clay) and three dynamic (bulk density, particulate organic carbon, pH) soil properties from agricultural soils. A generalised least squares model indicated that the effects of pedogenon, land use history and their interaction on topsoil pH were statistically significant ($p < 0.001$). Paired comparisons between pedogenon/pedophenon subclasses by pedogenon class were not statistically significant, although we observed the general trend: remnant pedogenon ≈ quasi-remnant pedogenon < pedophenon cleared ≈ pedophenon grazing < pedophenon cropping. Redundancy discriminant analysis indicated that pedogenons explained 40% of the variation of stable and dynamic soil properties, pedogenon/pedophenon subclasses explained 0.1% and the shared effect explained 18%, leaving 42% of unexplained variance. The effects of pedogenon/pedophenon subclasses on the location of group centroids were statistically significant only when dynamic soil properties were considered, but not for stable and dynamic soil properties. This framework can be integrated into a soil security assessment once the indicators of soil condition and capability are translated into soil functions and ecosystem services. Other potential applications include the design of soil monitoring sampling schemes and identifying thresholds of soil degradation.

Abbreviations: GLS, generalised least squares; PERMANOVA, permutational multivariate analysis of variance; PERMDISP, permutational analysis of multivariate dispersions; RDA, redundancy discriminant analysis; SCaRP, national soil carbon research program; SDGs, United Nations sustainable development goals; SOC, soil organic carbon; POC, particulate organic carbon; HOC, humic organic carbon; ROC, resistant organic carbon.

* Corresponding author.
E-mail address: mercedes.romandobarco@sydney.edu.au (M. Román Dobarco).

https://doi.org/10.1016/j.soisec.2021.100011
Received 14 March 2021; Received in revised form 3 June 2021; Accepted 5 July 2021
Available online 9 July 2021

1. Introduction

Soil functions are essential for providing ecosystem services to society and achieving the United Nations Sustainable Development Goals (SDGs) (Keesstra et al., 2016). The links between soil functions, soil ecosystem services and SDGs are evident for −but not limited to− achieving food security and ending hunger (SDG2), ensuring health and promoting well-being (SDG3), ensuring clean water and sanitation (SDG6), taking action to combat climate change (SDG13) and sustaining the life on land (SDG15) (Bouma et al., 2019; Pozza and Field, 2020). The demand for goods and services by the growing world population and the current economic and socio-political system increase the anthropogenic pressure on soils and accelerate the rate of soil change (Berthe, 2019). Soil change is also affected by global change processes (e.g., land conversion and intensification, climate change, pollution), reducing pedodiversity (Lo Papa et al., 2011) and the soils' ability to supply ecosystem services. In this context, optimal soil management needs to maximize the supply of ecosystem services according to the soil capability, identify baselines for measuring changes in soil condition (Berthe, 2019; McBratney et al., 2019), and define strategies for maintaining soil multifunctionality across the landscape (Greiner et al., 2017). Detailed information and understanding of soil change can guide managers and stakeholders in the decision-making process and assist in forecasting the effects of management alternatives on soil condition (Tugel et al., 2005). Similarly, incipient soil degradation can be detected by identifying thresholds in the temporal evolution of soil properties associated with different management (Kuzyakov and Zamanian, 2019).

The concept of soil security connects the biophysical, economic, and social soil attributes with the ecosystem services and can be used for sustainable management (McBratney et al., 2014). The ability of a soil to deliver an ecosystem service is determined by its capability, which is limited by its capacity and modified by its condition (McBratney et al., 2019; Field, 2020). Capability refers to the inherent potential for performing soil functions ('What functions can this soil be expected to perform?') (McBratney et al., 2014). Soil capacity is determined by physicochemical properties that generally evolve over long pedogenetic time scales (centuries, multimillenia) and that are not readily changed by human forcings (e.g., CEC, soil texture), although presumably, stable soil properties can also change over a relatively short time periods (e.g., human-induced erosion can modify surface texture in less than 100 years (Lyles and Tatarko, 1986)). Soil condition concerns biological, chemical, and physical soil properties that change at a faster rate and reflect contemporary soil management (e.g., microbial biomass, soil organic carbon, pH, aggregates). The effects of land use change and management on soil condition and capability need to be assessed with respect to a reference state. Previous studies established thresholds for soil condition from soil monitoring data or repeated soil surveys (Cotching and Kidd, 2010; Yang et al., 2018; Dazzi and Lo Papa, 2019). Spatially explicit soil security assessments have evaluated soil condition setting target values by soil order and land use type for several soil physicochemical properties (Kidd et al., 2018; Yang et al., 2018). Dazzi and Lo Papa (2019) quantified soil condition with the Soil Potential Index (Mancini and Ronchetti, 1968) for different soil orders. They assessed the effect of anthropogenic activity on the soil security dimensions by comparing the temporal change in an area occupied by soil order within their study area. Alternatively, the baseline for soil condition may be identified at a lower taxonomical level from detailed soil maps (Rossiter and Bouma, 2018), and the effects of land use history on soil condition assessed with a space-for-time substitution.

Soils are human-natural bodies, often a "kind of polygenetic paleosol" that accrued features through processes that evolved widely over pedogenetic time (Cline, 1961; Richter and Yaalon, 2012). Hence, the reference state has to be defined for a specific temporal context. Yaalon and Yaron (1966) defined a theoretical framework according to which a soil in steady-state (S_N), resulting from the long-term effects of the soil-forming factors (*clorpt* (Jenny, 1941)) is the initial point for human-induced changes in soil processes and properties, i.e., metapedogenesis or anthropedogenesis. The properties of the new soil (S_{HN}) will depend on the intensity of the anthropedogenetic processes, initial soil properties, and the resistance and resilience of the initial soil S_N. Richter (2007) incorporated the temporal scale of soil formation into the conceptual model by Yaalon and Yaron (1966) by differentiating between multimillennial natural pedogenesis, the legacy of historic anthropedogenesis (i.e., time scale of centuries and millennia) and contemporary anthropedogenesis (i.e., decadal time scale). This model can be implemented with a digital soil mapping approach to identify the reference state specific for each soil class and assess changes in soil condition and capability caused by contemporary management. A soil map representing soil classes at a reference point in time can be overlayed with spatial information of contemporary human pressures on soils. The reference state for each soil class will be the one with the minimum degree of anthropogenic modification of soil properties.

In the context of the Australian continent, the historic anthropedogenesis comprises the land management and agricultural practices carried by the First Nations Australians over millennia (Gammage, 2011; Pascoe, 2014). These practices include modifying watercourses for pisciculture, applying controlled fire for managing the vegetation structure and composition, creating wildlife habitats, and agriculture with native crops (Gammage, 2011; Pascoe, 2014). As one of the societies with an extensive presence in a geographic region (Tobler et al., 2017; Bird et al., 2018), the management of landscape structure, vegetation dynamics and fire regime influenced pedogenetic processes and maintained the provision of ecosystem services. Land use practices after the European settlement in Australia, from 1788 onwards, modified the pressures on soils (Russell and Isbell, 1986). Clearing native forests and farming led to increased soil erosion, gully formation and increased alluvium deposits in valley bottoms in some areas (Gale and Haworth, 2005; Muñoz-Salinas et al., 2014). The species composition and structure of native forests likely shifted as a result of clearing of *Eucalyptus* forests followed by fire suppression (Lunt et al., 2006). In central New South Wales, former *Eucalyptus*-dominated woodlands may have evolved into either open savanna-like pastoral landscapes or dense *Callistris*-dominated forests (Lunt et al., 2006). The effects of contemporary agricultural activities like intensive cropping and grazing on dynamic soil properties are well documented: loss of soil organic carbon, increased bulk density, reduction of infiltration rates, increased soil erosion, modified nutrient availability and pH (Wilson et al., 2011; Yates et al., 2000).

The concepts of genoform and phenoform were conceived for distinguishing pedons within the same map unit and shared long-term pedogenesis that present substantial differences in properties and functions due to management (Droogers and Bouma, 1997; Rossiter and Bouma, 2018). This approach has been applied for investigating soil change and its relationship with soil type and management (Stevenson et al., 2015; Huang et al., 2018; Seaton et al., 2020) and requires the availability of detailed soil maps in which the soil survey groups polypedons with common historic anthropedogenesis. Some systems classify soils with the primary aim of describing current morphological features rather than following genetic criteria or accounting for pre-disturbance features (e.g., Australian Soil Classification (ASC), Isbell et al., 1997). Hence, information on the contemporary human disturbance is not provided explicitly at some level of the classification. Contemporary soil change may have modified soil attributes to the extent that polypedons with common long-term pedogenesis are assigned to different classes (Smeck and Balduff, 2002).

The main goal of this study is to design a digital soil mapping (DSM) approach that can serve for assessing changes in soil capability and condition due to contemporary land use history. We expand a DSM framework designed for defining groups of homogeneous quantitative state variables representing the soil-forming factors for a reference time, under the hypothesis that the dominant soil-forming processes within each group are similar, and therefore develop unique soil entities (i.e., pedogenons) (Román Dobarco et al., 2021). The specific objectives of this study are: (1) to map pedogenon classes defined for the time of the European settlement in New South Wales and further divide them into subclasses according to contemporary land use and vegetation status, (2) to test differences in soil pH, as a proxy for soil condition, by pedogenon, and (3) to assess the ability of pedogenon classes, pedogenons and pedophenon subclasses for explaining the variation in dynamic and stable soil properties in agricultural soils.

2. Methods

2.1. Framework for mapping pedogenons and pedophenons

The modelling framework consists of two steps. First, soil entities are defined by identifying groups of homogeneous soil-forming factors, in accordance with classic factor-based approaches (Dokuchaev, 1883; Jenny, 1941). Each group represents a historic soil system (S_{NH}) that results from long-term anthropedogenetic processes up to a reference time, i.e., pedogenon (Fig. 1) (Román Dobarco et al., 2021). Then, indicators of contemporary soil change are implemented with a look-up table for dividing the pedogenons into subclasses depending on the type and degree of intensity of human pressure on soils. These subclasses correspond to contemporary human-affected soil systems (S_{NHC}) (Fig. 1) (Richter, 2007).

Pedogenon classes are generated by applying unsupervised classification to a regionalised set of quantitative state variables. The environmental covariates are proxies of the soil-forming factors for a given reference time, e.g., European settlement in New South Wales. The characteristics of the soil-forming factors and processes vary widely during soil formation, which can span from millennia to millions of years (Richter and Yaalon, 2012). Hence, the pedogenon model can be generalized as:

$$pedogenon = f(s, cl,\ o,\ r,\ p,\ t)\, where\ t = pedogenetic\ time\ up\ to\ a\ reference\ time$$

where pedogenons are a function of the prior natural soil system (s), climate or paleoclimates (cl), organisms (o), relief (r) and parent material (p) acting from the origin of soil formation until a reference time (t). Soil attributes that inform of long-term pedogenetic processes may be included as covariates and information on historic land use. In this example, we did not differentiate sequences of natural pedogenetic and historic pedogenetic processes, but rather grouped them in a more static approach. We selected covariates that we considered relatively constant (relief, parent material) or assumed representative of the conditions at the time of the European settlement (estimated native vegetation, climate). In this study, a pedogenon is defined as follows:

$$pedogenon = f(cl_t,\ o_t,\ r_t,\ p_t)\ where\ t = reference\ time$$

The pedogenons constitute the historic soil system (S_{NH}) or starting point for assessing the effects of contemporary management. Next, pedogenon classes are divided into different subclasses depending on the type and degree of intensity of anthropogenic activities. These subclasses constitute the contemporary human-affected system (S_{NHC}) (Richter and Yaalon, 2012), and range from remnant pedogenons (the closest to the pre-European baseline) to several pedophenon types. Several data layers (e.g., native vegetation extent and status, land cover changes, land use) combined with a rule-based algorithm inform where it is likely to find the reference state or minimally-altered soils for each pedogenon class:

$$pedogenon = f(cl_t,\ o_t,\ r_t,\ p_t)_{o_t = native\ vegetation\ 1750} = \begin{cases} remnant\ pedogenon, & o = remnant\ native\ vegetation \\ pedophenon, & o \neq remnant\ native\ vegetation \end{cases}$$

For clarity, we may refer to the factor pedogenon or pedophenon

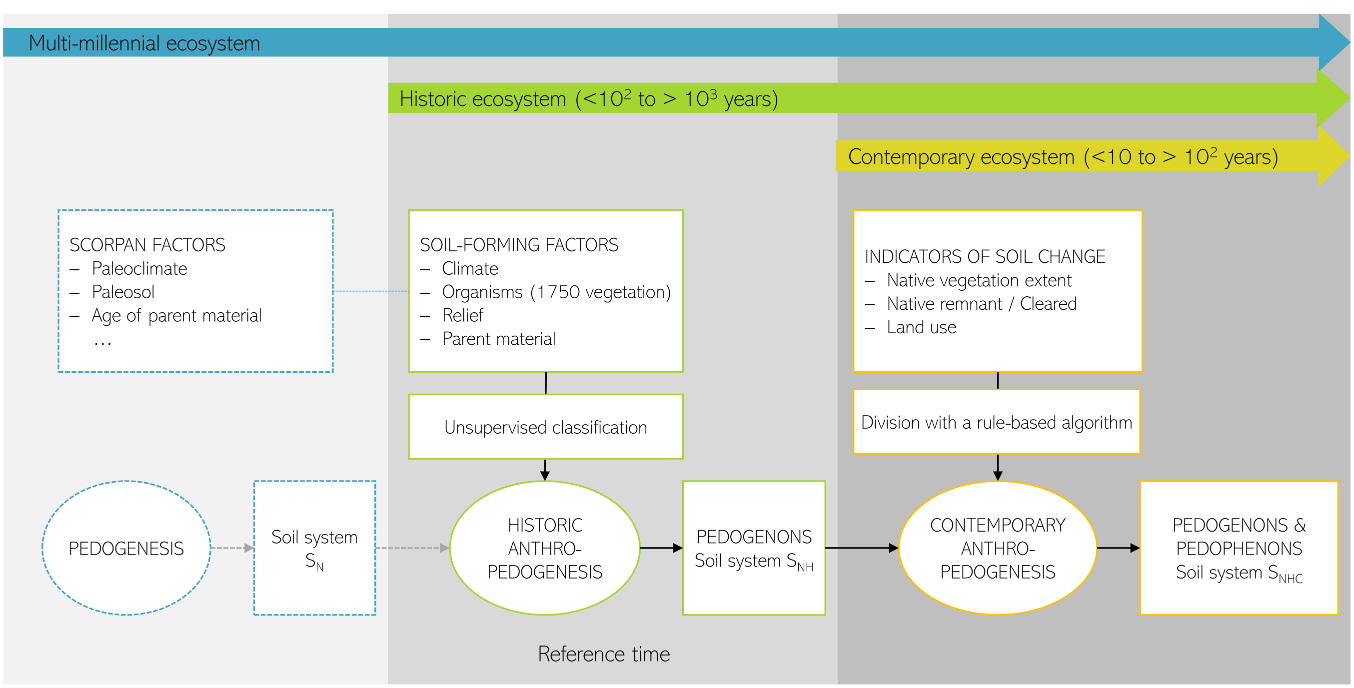

Fig. 1. Diagram of the modelling framework for mapping pedogenons and pedophenons. The framework of soil change proposed by Richter (2007) distinguishes three time scales: the multimillennial natural soil system (S_N), the historic soil system (S_{NH}) affected by natural and historic anthropedogenetic processes (which in the case of Australia comprise several millennia), and the contemporary human-affected system (S_{NHC}). The pedogenons are related to the historic soil system and constitute the baseline for assessing the effects of contemporary management on soil properties. This diagram is adapted from Richter and Yaalon (2012), Richter (2007) and Yaalon and Yaron (1966).

subclasses, either as pedogenon/pedophenon subclasses or land use history, to differentiate it from the factor pedogenon class. The pedogenon subclasses are populated with soil data. Information on stable and dynamic soil properties by fixed depths, and horizon and profile features can be used to calculate distance metrics either for a depth interval or the entire profile (Carré and Jacobson, 2009) between observations from different pedogenon/pedophenon subclasses. The distances between pedogenon/pedophenon subclasses (defined in soil properties) indicate whether recent management and land use have modified soil condition. Multivariate statistical analyses that compare distances and dispersion within- and between-groups were used for this objective.

2.2. Digital pedogenon mapping

The modelling framework for digital pedogenon mapping applied to New South Wales, the characteristics of the pedogenon classes and their spatial distribution are explained in detail in Román Dobarco et al. (2021). Here, we provide a brief description of the modelling process.

We selected 25 continuous environmental covariates (Table 1), proxies of the soil-forming factors that would have remained relatively constant (e.g., relief and parent material) or that we assume representative of the conditions prior 1750 (e.g., vegetation). Due to the lack of accurate estimates and limited spatial coverage for the pre-industrialized climate data in Australia (Ashcroft et al., 2014), we used nine bioclimatic indices calculated from the ANUCLIM 6.1 (Xu and Hutchinson, 2011) 30-year average climate surfaces (1975–2005). Four gamma-ray spectrometry variables and a weathering intensity index (Wilford, 2012) informed on the geochemistry, mineralogy and degree of weathering of the regolith and bedrock. Five covariates derived from a 3-second digital elevation model produced from the Shuttle Radar Topographic Mission (SRTM) (Farr and Kobrick, 2000) described the relief and geomorphology. The categories of the estimated vegetation at the time of the European settlement (pre-1750 vegetation) (National Vegetation Information System V5.1 © 2018) were grouped into 15 classes (Fig. 2), transformed into binary variables, and subjected to principal component analysis. We kept the first seven principal components that collectively retained 50% of the variation which amply discriminated the major vegetation groups. Including available estimates of paleoclimate data would have raised the matter of whether applying a sequential workflow (grouping covariates by geological period) or weighing the climate data depending on the duration of the period they represent. Hence, we excluded paleoclimate from this exercise for simplicity, but we will consider it in future studies. We trialled several combinations of covariates in preliminary analyses since the output of the unsupervised classification can be susceptible to covariate selection (Román Dobarco et al., 2021). The 25 environmental covariates were centred, standardized, and sampled at 259,000 locations in a 1.6 km × 1.6 km grid.

We created 1000 pedogenon classes applying the *k*-means algorithm (Hartigan and Wong, 1979) to the dataset of 25 environmental covariates. *K*-means is a non-hierarchical clustering method, efficient for very large datasets of numerical data. The algorithm searches the partition of *k* clusters from dataset *X* that minimized the within-cluster sum of squared errors, i.e., the sum of squared distances to the cluster centroids (Han et al., 2012). Before clustering, the environmental dataset *X* was rescaled by applying the inverse of the Cholesky transformation of the variance-covariance matrix (Wicklin, 2012). The Euclidean distance calculated on the rescaled dataset *Y* is equivalent to the Mahalanobis distance calculated in *X* (Wicklin, 2012). This transformation allows for correlations among variables to be considered in the partitioning process. The clustering process was repeated 10 times with the *k*-means++ initialization (Arthur and Vassilvitskii, 2007) and a maximum of 5000 iterations. The *k*-means algorithm was implemented with the *Kmeans_rccp* function of the ClusterR package (Mouselimis,

Table 1
Covariates used to describe clorpt (Jenny, 1941) or scorpan (McBratney et al., 2003) factors and generate pedogenon classes. P: parent material; S: soil; T: time; R: relief; Cl: climate; O: organisms.

Covariate	Description	Clorpt factor	Original raster resolution (m)	Reference
PTA	Annual precipitation (mm)	Cl	270	Xu and Hutchinson (2011) Williams et al. (2012)
PTS1MP	Precipitation: ratio of annual contrast in regional rainfall conditions between summer and winter solstice conditions.	Cl	270	Xu and Hutchinson (2011) Williams et al. (2012)
PTS2MP	Precipitation: ratio of annual contrast in regional rainfall conditions between spring and autumn equinox conditions.	Cl	270	Xu and Hutchinson (2011) Williams et al. (2012)
TNM	Minimum temperature (annual mean) (°C)	Cl	270	Xu and Hutchinson (2011) Williams et al. (2012)
TXX	Maximum temperature (monthly maximum) (°C)	Cl	270	Xu and Hutchinson (2011) Williams et al. (2012)
TNI	Minimum temperature (monthly minimum) (°C)	Cl	270	Xu and Hutchinson (2011) Williams et al. (2012)
TRX	Maximum monthly mean diurnal temperature range (°C). high variation in temperature conditions (inland or continental locations).	Cl	270	Xu and Hutchinson (2011) Williams et al. (2012)
TRI	Minimum monthly mean diurnal temperature range (°C). Consistent temperature conditions (coastal locations).	Cl	270	Xu and Hutchinson (2011) Williams et al. (2012)
RSM	Short-wave solar radiation - annual mean (MJ/m2/day)	Cl	90	Wilson and Gallant (2000)
K	Radiometrics: filtered K element concentrations (%)	S, P	100	Minty (2019a); Geoscience Australia (2019)
Th	Radiometrics: filtered Th element concentrations (ppm)	S, P	100	Minty (2019b); Geoscience Australia (2019)
Th/K	Radiometrics: Ratio Th/K derived from the filtered Th and K grids	S, P	100	Minty (2019c); Geoscience Australia (2019)
WII	Weathering intensity index	P, T	100	Wilford (2012)
Elevation	SRTM-derived 3 S Smoothed Digital Elevation Model	R	90	Gallant et al. (2009)
Slope	Slope (%)	R	90	Gallant et al. (2009)
TWI	Topographic wetness index	R	90	Quinn et al., (1991)
MRVBF	Multi-resolution valley bottom flatness index	R	90	Gallant and Dowling (2003)
MRRTF	Multi-resolution ridge top flatness index	R	90	Gallant and Dowling (2003)
MVG	Estimated pre-1750 major vegetation groups (MVGs). First 7 principal components retaining 50% of the variation of the MVGs represented as binary variables (presence/absence).	O	100	National Vegetation Information System V5.1 (2018)

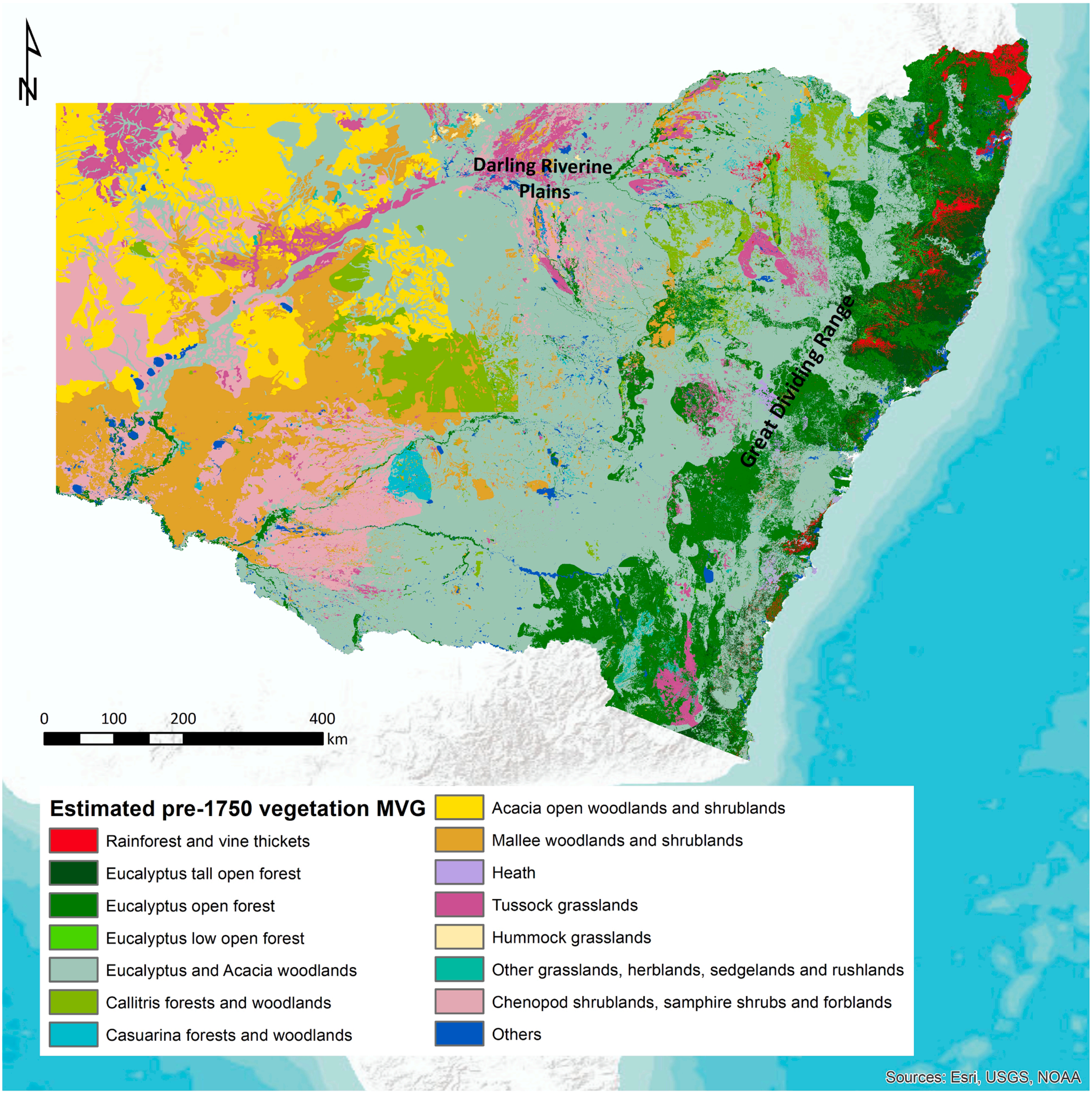

Fig. 2. Estimated pre-1750 native vegetation (National Vegetation Information System V5.1, 2018) reclassified into 15 major classes used for pedogenon mapping.

2021). Pedogenon classes were mapped at 90 m resolution assigning each pixel to its closest cluster centroid after rescaling with the inverse Cholesky transformation.

The number of clusters was selected based on three criteria: the elbow method, pedodiversity-area equations (Guo et al., 2003), and visual examination of the spatial patterns of pedogenon maps at smaller study areas where soil properties have been well studied in the past. These areas included the Namoi-Edgeroi district (Triantafilis and McBratney, 1993; Ward, 1999), the Hunter Valley region (Malone et al., 2014) and Nowley farm (Stockmann et al., 2016).

We hypothesized that the variation and complexity in combinations of environmental variables, proxies of soil-forming factors, would follow a power function, $S = cA^z$ (S is richness, A is area and c and z are constants) (Guo et al., 2003). The richness-area relationships observed at continental and subregional scales for established soil classification systems (Guo et al., 2003; Minasny et al., 2010) should be intrinsic to pedogenon classes. We used equations developed by Guo et al. (2003) for the conterminous USA for the family level of Soil Taxonomy (Soil Survey Staff, 2010) to approximate the number of desired pedogenon classes for New South Wales. This taxonomic level is above the soil series, but it is considered a useful general group for management. For the area of mainland New South Wales (801,137 km^2) the number of classes at the taxonomic level of family would be around 1040 (Román Dobarco et al., 2021). Further information on selecting the optimal number of pedogenon classes, and a discussion on the relevance of the number of classes, representation of the different soil-forming factors and selection of environmental covariates on the output pedogenon maps can be found in Román Dobarco et al. (2021). The ratio between cluster sum of

Table 2
Look-up table for classifying pedogenon or pedophenon subclasses.

Native vegetation extent	Native vegetation status	Land Use class	CLUM category	Pedogenon / pedophenon subclass
Native	Native remnant	110:117	1.1 Nature conservation	Pedogenon Remnant
		120:125,	1.2 Managed Resource Protection	
		130,131,133	1.3 Other Minimal Use	
		610,611,614,	6.1 Lake (conservation and saline)	
		630,631,	6.3 River conservation	
		650,651,	6.5 Marsh/wetland	
		660,661	6.6 Estuary/coastal	
		132	1.3.2 Stock route	Quasi-remnant pedogenon
		200:222	2. Production from relatively natural environments	
		314,	3.1.4 Environmental forest plantation	
		414,	4.1.4 Irrigated environmental forest plantation	
		612,	6.1.2 Lake production	
		632,	6.3.2 River production	
		652	6.5.2 Marsh production	
		662	6.6.2 Estuarine production	
Native	Native cleared	110:117, 120:125, 130,131,133, 200:222, 314,414, 610:614, 630:632, 650:652, 660,661		Pedophenon Cleared
Native or non-native	Remnant or cleared	310:313,	3.1 Dry forestry	Pedophenon Forestry
		410:413	4.1 Irrigation forestry	
		320:325	3.2 Dry Grazing	Pedophenon Grazing
		420:424	4.2 Irrigated Grazing	
		330:353,	3.3 Cropping	Pedophenon Cropping
		360:365,	3.4 Perennial horticulture	
		430:465	3.5 Seasonal horticulture	
		542	3.6 Land in transition	
			4.3 Irrigated Cropping	
			5.4.2 Rural residential with agriculture	
		500:595	5. Intensive Uses (mining, industrial, urban, greenhouses, livestock facilities, etc.)	Excluded from analysis
		613	6.1.3 Lake intensive	
		620:623	6.2 Reservoir/dam	
		633	6.3.3 River intensive	
		640:643	6.4 Channel/aqueduct	
		134	1.3.4 Rehabilitation	

squared errors to total sum of squared errors (between-CSS / total-SS) was used as an indicator of clustering quality. When the ratio between-CSS / total-SS is close to 1, it is indicative that the observations follow a pattern in the clustering (Mouselimis, 2021).

Hierarchical agglomerative clustering was applied to investigate the similarities and hierarchical organization of the pedogenon classes. The pedogenon centroids were treated as individuals and the clustering was performed with Ward's method. The optimal number of branches or pedogenon families was selected with the Silhouette and the Dunn indices, setting 30 as the maximum number of clusters. The visualization of the 1000 pedogenon classes was determined by trial and error, assigning a different colour palette to different dendrogram branches.

2.3. *Stratification into pedogenon/pedophenon subclasses with a rule-based algorithm*

A categorical map distinguishing areas with the different expected degree of anthropogenic modifications of soil properties (i.e., land use history) was generated combining information on native vegetation extent, status (cleared or intact), and current land use. The map was produced at 90 m resolution, although the resolution of the input layers ranged between 5 m and 250 m. The maps used to delineate pedogenon and pedophenon subclasses answered the following questions:

- *Is the vegetation potentially native?* The NSW Native Vegetation Extent 5 m Raster v1.2 (NSW Office of Environment and Heritage (OEH) 2019) determined the extent of native vegetation. This layer was developed under the NSW State Vegetation Type Map program and differentiates tree cover, woodland matrix, candidate native grasslands, forestry plantations, non-native areas, and water bodies. The mapping method used high-resolution and SPOT5 satellite imagery for mapping individual trees (Fisher et al., 2017). Its currency is for 2011–2018. It assumes that all tree cover different from plantations are native. Areas with grassland vegetation and no signs of agricultural management are classified as candidate native grasslands.
- *Is it remnant or secondary native vegetation?* A drawback of using current vegetation extent and structure is that the anthropogenic degradation of natural ecosystems might have taken place prior to the availability of remote sensing imagery. Land cover classification may not distinguish vegetation structures resulting from human activities (Harwood et al., 2016). For example, it may be difficult to determine whether a native open woodland or native grassland results from ecosystem degradation following forest clearing or reflective of natural environmental conditions. Hence, we included the map by Keith and Simpson (2008) on native vegetation at 250 m resolution in the analysis. This binary map discriminated between remnant native vegetation (intact grasslands and woody vegetation) and cleared vegetation (including non-native and secondary grasslands of native vegetation). The map was produced combining the best available information from 46 vegetation maps with currency ranging between 1966 and 2005, although the available information

on native grasslands was limited at the time (Keith and Simpson 2008).

- *Can the current land use affect soil functions substantially?* Current land use is an indicator of the intensity of human pressures on soils. The Catchment Scale Land Use of Australia dataset (CLUM) at 50 m (ABARES, 2019) was reclassified into five categories to differentiate between conservation areas, production from relatively natural environments, forestry, grazing, and cropping. Intensive uses (mining, industrial, greenhouse horticulture, livestock production facilities, etc.), urban areas, and water bodies were excluded from the analysis.

The rules for classification are presented as a look-up table (Table 2). For an area to be considered a remnant pedogenon it was required that the vegetation was potentially native, intact from clearing and that current land use has a small or negligible impact on soil condition (i.e., conservation). The next pedogenon subclass along the gradient of anthropogenic pressure (quasi-remnant pedogenon) was designed for areas where land use was close-to-nature management. Areas of potentially native vegetation that have been cleared (Keith and Simpson, 2008), and where current land use has a relatively small effect on soil condition, were designated as a pedophenon with a low degree of modification (i.e., Pedophenon Cleared). This class may be relevant for grasslands resulting from degradation of woodlands and forests due to previous land use management (grazing, fire management, etc.). Land uses indicative of severe modification on native vegetation and dynamic soil properties (e.g., cropping, grazing on managed pastures) override the information on native vegetation extent and status. Water bodies and intensive land uses (mining, industrial, etc.) were excluded from this analysis since the focus of this study is on agricultural activities.

2.4. Population of pedogenon and pedophenon subclasses with legacy soil data

Soil pH was chosen as an indicator of soil condition because it is a dynamic property that is affected by management. Legacy data were accessed with the Soil Data Federator (http://esoil.io/TERNLandscapes/SoilDataFederatoR/R/help/index.html). This web API is managed by CSIRO and compiles soil data from different institutions and government agencies throughout Australia. We selected soil pH measured with 1:5 soil/0.01 M calcium chloride extract (Rayment and Lyons, 2011). After cleaning and checking the quality of the data, mass-preserving splines (Bishop et al., 1999) were applied to estimate pH values by the *GlobalSoilMap* depth intervals (Arrouays et al., 2014). The subset of 5047 observations for the 5–15 cm depth interval was selected for assessing changes in soil condition.

Data from the national Soil Carbon Research Program (SCaRP) (Baldock et al., 2013b) constituted a second dataset. SCaRP was designed to quantify carbon stocks across combinations of land uses and management practices in the main agricultural regions of Australia. The main land uses were cropping and grazing in managed pastures or native grasslands. Hence, forests and rangelands outside the main agricultural regions are not represented in the dataset. Soil samples were mainly collected at 0–10, 10–20, and 20–30 cm depth from 10 sampling points in 25 m × 25 m plots within a paddock and combined in a composite sample (Sanderman et al., 2011). A minimum of three additional samples per depth interval was collected for measuring bulk density. The samples were air dried at 40 °C and sieved with a 2 mm mesh. At the laboratory total organic carbon was determined with a LECO CN analyser (LECO Corporation, MI, USA). Soil organic carbon (SOC) fractions were determined for a representative subset of samples with a combination of size-density fractionation and solid-state ^{13}C nuclear magnetic resonance spectroscopy (Baldock et al., 2013a). The fractionation scheme separated particulate organic carbon (POC), mineral-associated SOC or humic organic carbon (HOC), and resistant organic carbon (ROC) or charcoal. These fractions were related with carbon pools of different turnover time of the Roth C model (Skjemstad et al., 2004). Diffuse reflectance mid-infrared spectra were acquired for all samples with a Thermo Nicolet 6700 FTIR spectrometer (Thermo Fisher Scientific Inc., MA, USA) over the range 8000–400 cm^{-1} at 8 cm^{-1} resolution. SOC fraction contents of all samples were determined with partial least squares regression models calibrated with MIR spectra and fractionation data (Baldock et al., 2013a). Gravimetric clay content, soil pH, total nitrogen, and total concentration of Al, Si, and Fe were determined with MIR predictive models (Janik et al., 1995; Janik and Skjemstad, 1995).

2.5. Statistical analyses

We calculated summary statistics on the distribution of pedogenons into subclasses as total area and percentages of their respective pedogenon of origin. We applied a generalised least squares (GLS) model to test the effect of pedogenon, land use history (pedogenon/pedophenon subclass) and their interaction on topsoil pH (5–15 cm depth). Only

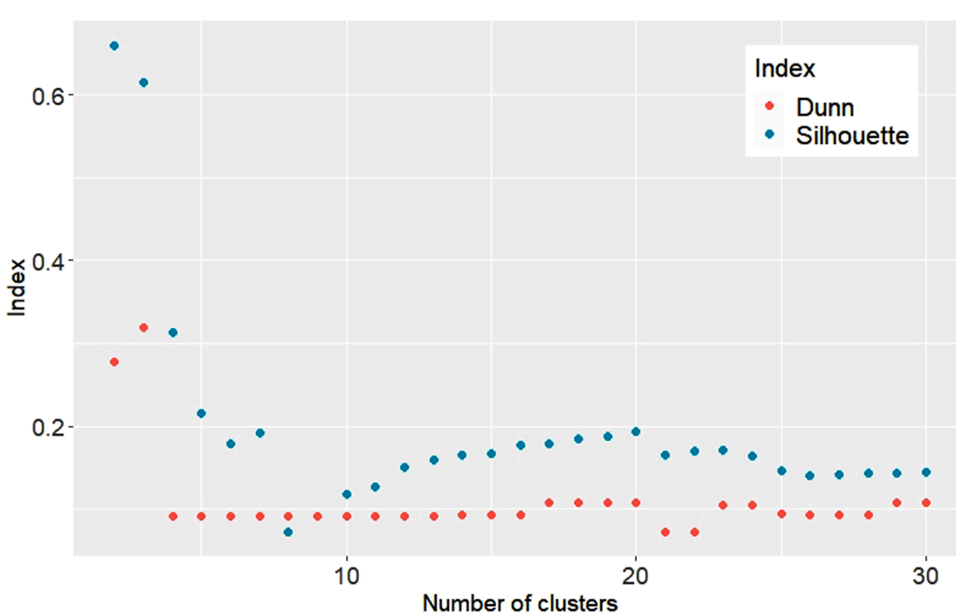

Fig. 3. Silhouette and Dunn indices for different number of clusters applied in the hierarchical clustering of the pedogenon centroids.

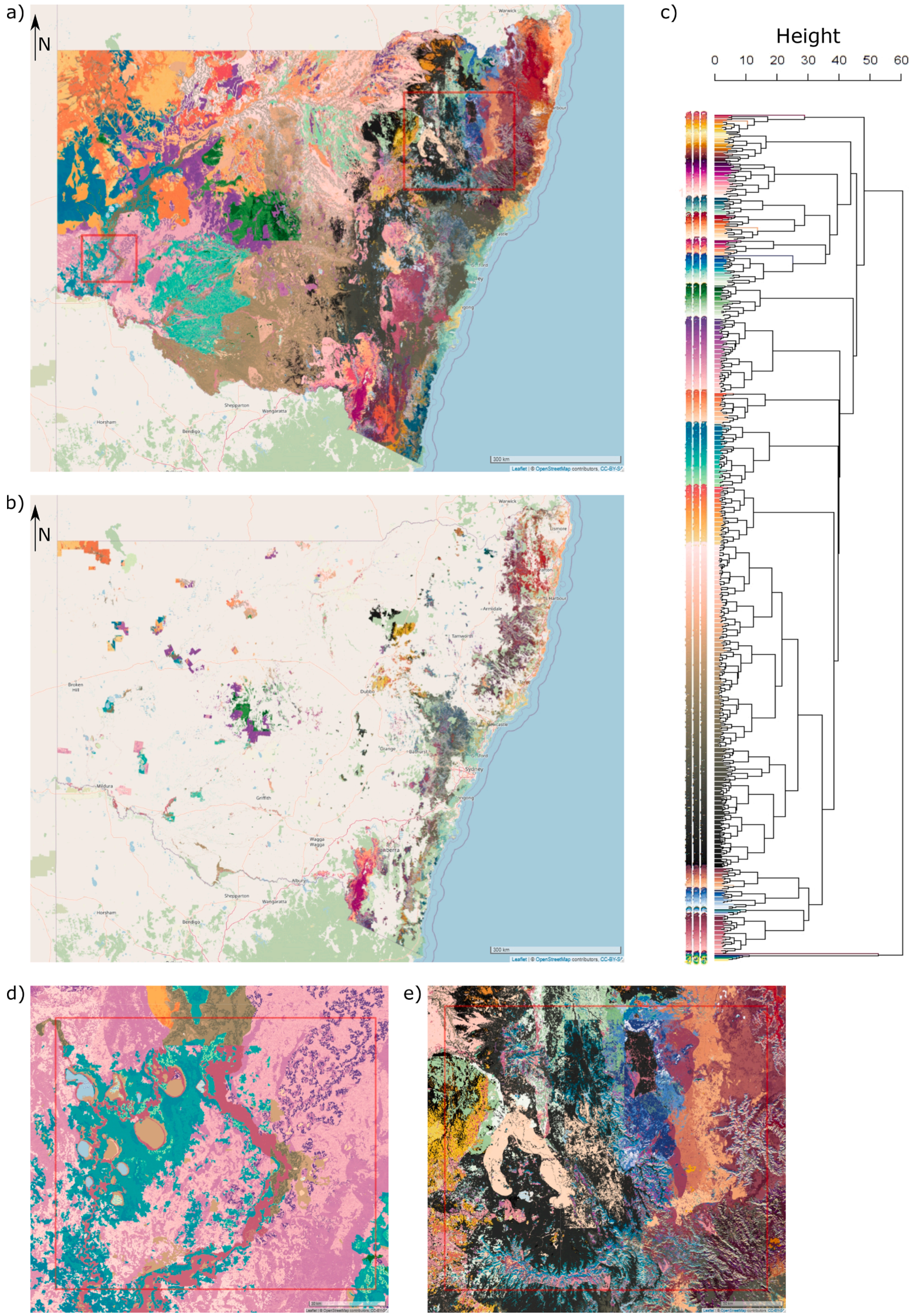

Fig. 4. (a) Pedogenon classes for New South Wales and organization into different branches following hierarchical clustering. (b) Spatial distribution of remnant pedogenons. The colour designates the pedogenon of origin following the hierarchical dendrogram.

pedogenon/pedophenon subclasses with at least 5 observations were included in the model. GLS models can deal with variance heterogeneity, and hence we fitted the variance structure by pedogenon class (Zuur et al., 2009). Since not all pedogenon/pedophenon subclass levels existed in every pedogenon, we created a new factor designating their combinations. We tested the significance of the interaction term with the likelihood-ratio test, i.e., comparing nested models (with and without interaction) fitted with maximum likelihood (Zuur et al., 2009). Paired comparisons between estimated means were calculated across all pedogenon and pedophenon subclasses by pedogenon class with Tukey's

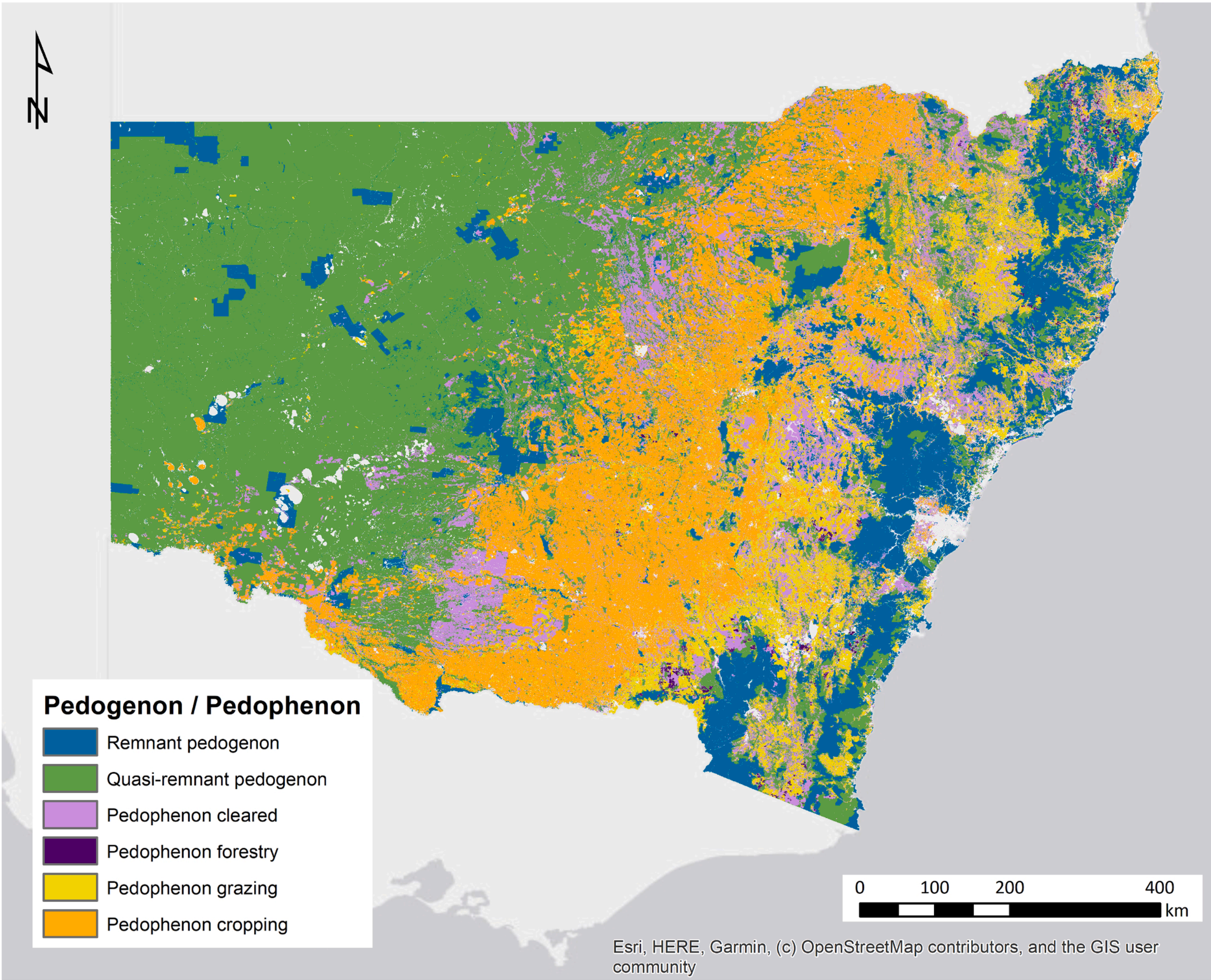

Fig. 5. Distribution of six categories for pedogenons and pedophenons in New South Wales.

Table 3
Summary statistics of the area occupied by different pedogenon and pedophenon subclasses (km^2).

Pedogenon / Pedophenon subclass	Distinct Pedogenons classes	Minimum Area	Q25 Area	Median Area	Mean Area	Q75 Area	Maximum Area
Remnant pedogenon	997	0.007	9.8	38.5	101.8	130.7	976.7
Quasi-remnant pedogenon	995	0.299	54.2	158.3	360.4	578	2827.6
Pedophenon Cleared	987	0.007	25.7	73.9	114.3	163.7	1042.7
Pedophenon Forestry	604	0.007	0.1	0.4	3.8	1.8	134.9
Pedophenon Grazing	950	0.007	5.3	25.1	72.8	79.6	881
Pedophenon Cropping	915	0.007	2.1	32.5	152.6	230.9	1244.7

honest significance test. Paired comparisons between estimated means were performed aggregated by pedogenon and pedophenon classes.

We applied redundancy discriminant analysis (RDA) to identify patterns of variation in the multivariate SCaRP data that could be associated to pedogenons and pedogenon/pedophenon classes. We selected three stable (clay, total Si, total Al) and three dynamic (bulk density, pH, POC) soil properties. The variance inflation factor smaller than 3 indicated acceptable collinearity. POC is considered a proxy of the active C pool and very responsive to management, and hence we chose it over SOC, although both were highly correlated. We also performed RDA to compare the ability of soil order (ASC, Isbell et al., 1997) and higher-level taxa pedogenon class, i.e., branch or family, to explain the variation of soil properties because both variables had a number of factor levels in the same order of magnitude. We tested the suitability of the rule-based algorithm by performing the RDA with the three source layers (native vegetation extent, status and land use) as explanatory variables and only the classification pedogenon/pedophenon as explanatory variable. RDA is a linear canonical ordination method that can be considered as a constrained version of principal component analysis (Zuur et al., 2007). The canonical axes are built from linear combinations of the response variables and the explanatory variables (Legendre and Legendre, 2012; Borcard et al., 2018).

Distance metrics calculated with a multivariate set of soil properties can inform on substantial differences in condition and capability between pedogenon and pedophenon groups, by comparing within- and between-group distances and the location of the group centroids in the multivariate space. A permutational multivariate analysis of variance (PERMANOVA) (Anderson 2001) was applied to the subset of stable and dynamic soil properties using the Mahalanobis distance as dissimilarity metric. PERMANOVA is a multivariate variance partitioning method

Table 4
Summary statistics of the contribution (%) that the pedogenon/pedophenon subclasses represent of their respective pedogenon class.

Pedogenon / Pedophenon subclass	Minimum	Q25	Median	Mean	Q75	Maximum
Remnant pedogenon	0.001	1.3	5.3	17.1	21.4	98.7
Quasi-remnant pedogenon	0.337	9	26.2	39.4	73.5	99.8
Pedophenon Cleared	0.001	4.1	11.9	14.9	23.1	85.3
Pedophenon Forestry	0	0	0.1	0.5	0.3	14.6
Pedophenon Grazing	0	0.8	3.6	9.8	12.5	77
Pedophenon Cropping	0.001	0.3	3.9	19.2	32.5	90.6

with p-values obtained with distribution-free permutation techniques (Anderson and Walsh 2013). PERMANOVA tests the null hypothesis of equivalence in the position (location) of group centroids in the space of dissimilarity measure, under the assumption of exchangeability of sample units among the groups (Anderson and Walsh 2013). We tested the effect of pedogenon class, land use history (pedogenon/pedophenon subclass) and their interaction on the location of group centroids with 9999 permutations. PERMANOVA makes no assumptions in the distributions of the variables or the dissimilarity metrics, and is very robust to heterogeneity for balanced designs but not unbalanced designs, and insensitive to the correlation (shape) among groups (Anderson and Walsh 2013). Then we performed multilevel pairwise comparisons and calculated the Bonferroni-corrected *p*-values between pedogenon/pedophenon subclasses (combinations of pedogenon and land use history) to test if the effect of management was significant within a pedogenon class. A test for homogeneity of multivariate dispersion (PERMDISP) was also done for each factor using the Mahalanobis distance as dissimilarity metric and 9999 permutations. PERMDISP compares within-group spread among groups with the average distances from individual observations to their group centroid (Anderson, 2017). We made pairwise

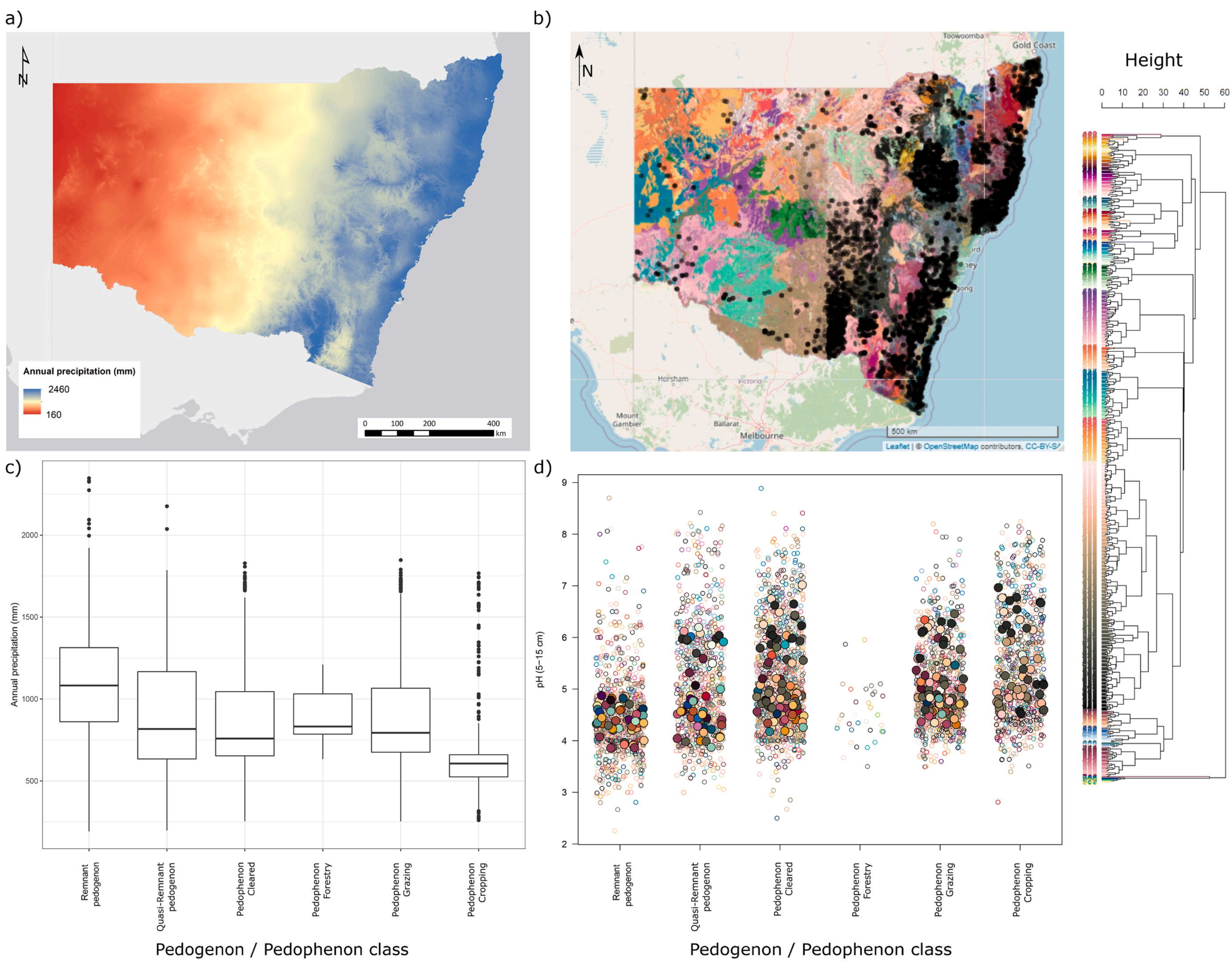

Fig. 6. (a) Annual precipitation (mm) in New South Wales (NSW), (b) Location of pH data (5–15 cm) across pedogenon classes in NSW ($N = 5047$) and hierarchical clustering of pedogenon classes, with a colour palette by pedogenon branch, (c) Boxplot of annual precipitation by pedogenon/ pedophenon class. The lower and upper whiskers represent respectively the ranges between the first (Q1) and third (Q3) quantiles to the smallest value at most Q1 - 1.5 * IQR and to the largest values no greater than Q3 + 1.5 * IQR (IQR is the inter-quartile range, or Q3 – Q1), and (d) soil pH (5–15 cm) observations by pedogenon and pedophenon class. In open circles individual soil pH observations, in closed circles, mean pH for those subclasses with at least 5 observations by pedogenon-pedophenon subclass (3224 observations across 176 pedogenons). The colour indicates the pedogenon class with the same colour scheme as in the hierarchical dendrogram.

comparisons between different pedogenon/pedophenon classes by pedogenon. The GLS model was fitted with the *nlme* package (reference), pairwise mean comparisons with the package *emmeans* (Lenth, 2020), RDA, PERMANOVA and PERMDISP analyses were performed with the *vegan* package in R (Oksanen et al., 2019) and the pairwise.adonis function (Martinez Arbizu, 2020).

3. Results

3.1. Pedogenon mapping and stratification into pedogenon and pedophenon subclasses

The k-means algorithm set for $k = 1000$ produced 997 pedogenon classes for NSW (three centroids had no observations assigned to them) with a mean area of 799 km^2, ranging between 4 and 2842 km^2. The ratio between-CSS / total-SS was 0.85, indicating a relatively good clustering pattern. The Silhouette index indicated that 2 was the optimal number of clusters. The index decreased from 2 to 7 clusters and increased progressively until reaching a local maximum at 20 clusters. The Dunn index indicated that 3 was the optimal number of clusters. The Dunn index had very small values between 4 and 16 clusters and increased slightly for 17–20 clusters (Fig. 3). Hence, the 997 pedogenon classes were grouped into 20 branches or families according to the hierarchical dendrogram (Fig. 4.a). The pedogenon families distributed along the Great Dividing Range and towards the coast followed the gradient in relief and climatic conditions and generally had a smaller size than those towards the west of the Great Dividing Range (Román Dobarco et al., 2021). In central NSW, a large pedogenon family was distributed along the north-south direction, where the predominant estimated pre-1750 vegetation were *Eucalyptus* and *Acacia* woodlands (Fig. 2). The influence of the estimated pre-1750 vegetation on the pedogenon classification was apparent for several pedogenon branches, e.g., Casuarina forests or Chenopod shrublands (Figs. 4 and 2).

The distribution of the pedogenons among pedogenon and pedophenon subclasses indicated that 46% of the area in NSW corresponded to areas of native remnant vegetation dedicated to productive uses, mainly grazing, classified as quasi-remnant pedogenon, whereas remnant pedogenon represented 13% of NSW. The predominant pedophenon was pedophenon cropping (18%) followed by pedophenon cleared (14%) and pedophenon grazing (9%), while pedophenon forestry occupied only 0.3% of NSW (Fig. 5).

The stratification of the 997 pedogenons by land use history produced a total of 5448 subclasses. For each pedogenon there was at least one grid cell ($\approx$ 0.007 km^2) classified as a remnant pedogenon (Table 3). Remnant pedogenon subclasses represented on average, 17% of their respective pedogenon class, ranging between 0.001 and 99% (Table 4 and Fig. 3.b). The median of the proportion of the pedogenon of origin preserved as a remnant pedogenon was 5.3% (Table 4). The 995 quasi-remnant pedogenon subclasses occupied larger areas, with a mean of 360 km^2 in comparison to the 102 km^2 of remnant pedogenon (Table 3). Quasi-remnant pedogenon constituted on average 39% of their pedogenon of origin, with a median of 26% (Table 4). The next pedophenon along the gradient of anthropogenic pressure (pedophenon cleared) were present in 987 pedogenons and had a mean area of 114 km^2. The pedophenons occupied by forestry were present in only 604 pedogenons and had a mean area of 3.4 km^2, although the maximum value was 134

Table 5
Analysis of variance of the generalised least squares (GLS) model parameters and maximum likelihood test of the interaction term.

GLS model	df	F-value	p-value
pH = Pedogenon + Land use history variance weights = varIdent(form = ~ 1 | Pedogenon)			
Intercept	1	261,008.4	<0.0001
Pedogenon	175	24.5	< 0.0001
Land use history	4	11.8	< 0.0001

GLS model (interaction term)	df	AIC	BIC	logLik	L.Ratio	p-value
pH = Pedogenon x Land use history variance weights = varIdent(form = ~ 1 | Pedogenon)	487	6476.8	9436.9	−2751.4		
pH = Pedogenon + Land use history variance weights = varIdent(form = ~ 1 | Pedogenon)	356	6413.2	8577.1	−2850.6	198.5	0.0001

Table 6
Multiple comparisons of mean pH by pedogenon/pedophenon subclasses. The diagonal presents the mean pH value by class. The lower triangle the difference between estimated means and the upper triangle the p-values.

Pedogenon/Pedophenon	Remnant pedogenon	Quasi-remnant pedogenon	Pedophenon Cleared	Pedophenon Grazing	Pedophenon Cropping
Remnant pedogenon	4.79	0.88	<0.0001	<0.0001	<0.0001
Quasi-remnant pedogenon	−0.04	4.83	0.0004	<0.0001	<0.0001
Pedophenon Cleared	−0.21	−0.17	5.00	0.60	0.27
Pedophenon Grazing	−0.26	−0.22	−0.05	5.05	0.87
Pedophenon Cropping	−0.31	−0.27	0.10	−0.05	5.11

Table 7
Summary statistics of pairwise comparisons of estimated mean pH by pedogenon/pedophenon subclasses by pedogenon class. N indicates the number of pedogenon classes where the pairwise comparison by subclass could be made.

Pedogenon/ Pedophenon	Pedogenon/Pedophenon	Minimum	25th percentile	Mean	75th percentile	Maximum	Standard deviation	N
Remnant Pedogenon	Quasi-remnant pedogenon	−0.331	−0.201	−0.029	0.124	0.369	0.216	22
Remnant Pedogenon	Pedophenon Cleared	−1.791	−0.45	−0.216	0.045	0.51	0.456	26
Remnant Pedogenon	Pedophenon Grazing	−0.577	−0.512	−0.297	−0.094	0.057	0.232	19
Remnant Pedogenon	Pedophenon Cropping	−1.238	−1.141	−1.045	−0.948	−0.851	0.274	2
Quasi-remnant pedogenon	Pedophenon Cleared	−0.804	−0.412	−0.155	0.102	0.507	0.367	29
Quasi-remnant pedogenon	Pedophenon Grazing	−0.747	−0.454	−0.256	−0.077	0.253	0.291	17
Quasi-remnant pedogenon	Pedophenon Cropping	−0.994	−0.364	−0.255	−0.152	0.535	0.433	8
Pedophenon Cleared	Pedophenon Grazing	−0.57	−0.114	0.055	0.162	1.032	0.311	48
Pedophenon Cleared	Pedophenon Cropping	−0.79	−0.27	−0.036	0.249	0.538	0.354	21
Pedophenon Grazing	Pedophenon Cropping	−0.804	−0.52	−0.298	−0.104	0.258	0.31	15

km^2. Thus, the proportion that pedophenon forestry occupied of their respective pedogenons was on average 0.5%. Pedophenons grazing and cropping occupied larger areas, with mean values of 73 km^2 and 153 km^2, respectively. The subclasses dedicated to cropping constituted in general a larger proportion of the pedogenons than grazing, with mean contributions of 19% and 10%, respectively (Table 4).

3.2. Soil pH as an indicator in soil condition. Effects of pedogenon and pedophenon subclasses

The soil pH dataset consisted of 5047 observations distributed in 577 pedogenon classes, while the remaining 420 pedogenon classes did not have any pH observation. The pH data were mainly located in the agricultural areas and the coastal fringe (Fig. 6.b). The subset with at least 5 observations by pedogenon/pedophenon subclass (combination of the factors pedogenon and land use history) resulted in 3224 observations across 176 pedogenon classes. We will also refer to the latter as land use history to differentiate between the pedogenon class and the pedogenon/pedophenon subclass levels. The GLS model indicated that the effects of pedogenon, land use history and their interaction were statistically significant ($p < 0.001$) (Table 5). Paired comparisons of mean pH aggregated by pedogenon/pedophenon level suggested that topsoil pH did not differ between remnant pedogenon and quasi-remnant pedogenon. Mean pH of both pedogenon classes differed from pedophenons. Estimated pH means did not differ between pedophenon classes, although it followed the trend remnant pedogenon < quasi-remnant pedogenon < pedophenon cleared < pedophenon grazing < pedophenon cropping (Table 6). Paired mean comparisons between pedogenon/pedophenon subclasses within the same pedogenon class were not statistically significant. However, the differences between estimated means had a maximum absolute value of 1.8 (Table 7). Most of the differences in mean pH (absolute values) that were in the upper 25th percentile = 0.4 (irrespective of the *p*-value) corresponded to pedophenon cleared and remnant pedogenon.

3.3. Variation in stable and dynamic soil properties from the SCaRP dataset explained by pedogenon classes and land use history

The number of SCaRP observations per pedogenon class ranged between 1 and 51 with a mean of 3.6 observations/pedogenon for the top 0–10 cm. From the total of 1444 observations, only 18 were located in soils classified as remnant pedogenon and 76 in quasi-remnant pedogenon. Most of the soil samples were located in cropping and grazing pedophenons (Table 8).

3.3.1. RDA on stable and dynamic soil properties

The subset of the SCaRP dataset that had at least 5 observations per pedogenon/pedophenon subclass consisted of 599 observations across 69 different pedogenon classes. The lack of SCaRP observations in most pedogenon classes is explained by the purpose of the SCaRP programme, focused on characterizing the SOC stocks of agricultural soils across Australia. The RDA axes explained 63% of the variation, leaving 37% of unexplained variance. The first and second axes of the RDA explained respectively 67% and 18% of the constrained variance (Table 9). A permutation test indicated that the global RDA model and the first four RDA axis were statistically significant ($p = 0.001$). The RDA distance triplot (scaling 1) did not discriminate pedogenon and pedophenon classes completely. Pedophenons grazing had higher POC content than pedophenons cropping, which had higher pH, clay, and bulk density (Table 8 and Fig. 7.b). Sites from pedogenons that belonged to the same family (similar colours) were closer in the RDA distance triplot, with the exception of some pedogenon classes that had a high dispersion. Variance partitioning of the RDA using the adjusted R^2 (Borcard et al., 1992; Zuur et al., 2007; Legendre and Legendre, 2012) indicated that the pure pedogenon effect was equal to 40% of the variation and the pure pedogenon/pedophenon effect (land use history or management) was 0.1% of the total variation. The shared amount of variation was 18% and it was not possible to differentiate between them due to some collinearity between pedogenon and pedogenon/pedophenon subclasses. These results indicate that pedogenon class (and hence environmental variables) explained most of the variation of soil properties. The shared effect of land use history and pedogenon suggests that the occurrence of some land uses mediates the variation explained by management where the environmental conditions and intrinsic soil properties (pedogenons and their capability) make them more suitable. The adjusted R^2 was 58% and the residual variation was 42%. RDA on dynamic soil properties explained a similar amount of variation (63%) and partition of variance between pedogenon classes, and pedogenon/pedophenon subclasses.

Table 8
Summary statistics of soil properties from the SCaRP dataset (0–10 cm) per pedogenon/pedophenon class ($N = 1444$), mean annual temperature (MAT) and mean annual precipitation (MAP). BD: bulk density; SOC: soil organic carbon; POC: particulate organic carbon; HOC: humic organic carbon; TN: total nitrogen; Si: total silica; Fe: total iron; Al: total aluminium.

Pedogenon/ Pedophenon	N	MAT (°C)	MAP (mm)	BD (g soil cm^{-3})	SOC (mg C g^{-1} soil)	POC (mg C g^{-1} soil)	HOC (mg C g^{-1} soil)
Remnant pedogenon	18	14.3 ± 1.8	815 ± 248	1.32 ± 0.12	25.78 ± 7.82	4.59 ± 2.07	13.79 ± 3.81
Quasi-remnant pedogenon	76	17.0 ± 2.5	495 ± 252	1.38 ± 0.18	14.15 ± 8.17	2.48 ± 1.56	7.61 ± 4.52
Pedophenon Cleared	213	15.6 ± 2.8	672 ± 175	1.30 ± 0.18	21.81 ± 11.97	4.30 ± 3.24	11.55 ± 6.20
Pedophenon Grazing	464	14.4 ± 2.2	746 ± 177	1.28 ± 0.18	24.70 ± 11.25	5.41 ± 2.93	12.86 ± 6.04
Pedophenon Cropping	673	16.9 ± 1.8	572 ± 129	1.34 ± 0.20	14.78 ± 6.66	2.11 ± 1.57	8.60 ± 3.75

Pedogenon/ Pedophenon	TN (mg N g^{-1} soil)	pH	Clay (mg g^{-1} soil)	Si (mg g^{-1} soil)	Fe (mg g^{-1} soil)	Al (mg g^{-1} soil)
Remnant pedogenon	1.60 ± 0.48	5.35 ± 1.08	293.84 ± 131.21	754.04 ± 58.61	19.71 ± 19.21	90.91 ± 36.18
Quasi-remnant pedogenon	1.06 ± 0.48	6.54 ± 1.12	444.07 ± 191.05	731.96 ± 99.45	38.53 ± 26.21	98.87 ± 31.39
Pedophenon Cleared	1.61 ± 0.69	5.90 ± 1.10	424.95 ± 194.97	714.17 ± 121.53	42.61 ± 37.64	96.80 ± 34.62
Pedophenon Grazing	1.91 ± 0.72	5.24 ± 0.72	320.96 ± 144.14	746.47 ± 114.06	34.40 ± 35.82	84.39 ± 32.58
Pedophenon Cropping	1.17 ± 0.42	6.13 ± 0.94	468.39 ± 168.62	660.84 ± 194.16	50.22 ± 36.78	92.91 ± 39.68

Table 9
Redundancy discriminant analysis with topsoil data (0–10 cm) from SCaRP sites with at least 5 observations per pedogenons and pedogenon/pedophenon subclass ($N = 599$). The explanatory variables were pedogenon class and pedogenon/pedophenon.

Axis	λ	λ as%	λ as cumulative%	λ as% of sum of all canonical eigenvalues	λ as cumulative% of sum of all canonical eigenvalues
RDA 1	2.53	42.2	41.5	66.8	66.8
RDA 2	0.69	11.6	53.9	18.4	85.2
RDA 3	0.22	3.7	57.6	5.8	91.0
RDA 4	0.17	2.9	60.4	4.5	95.5

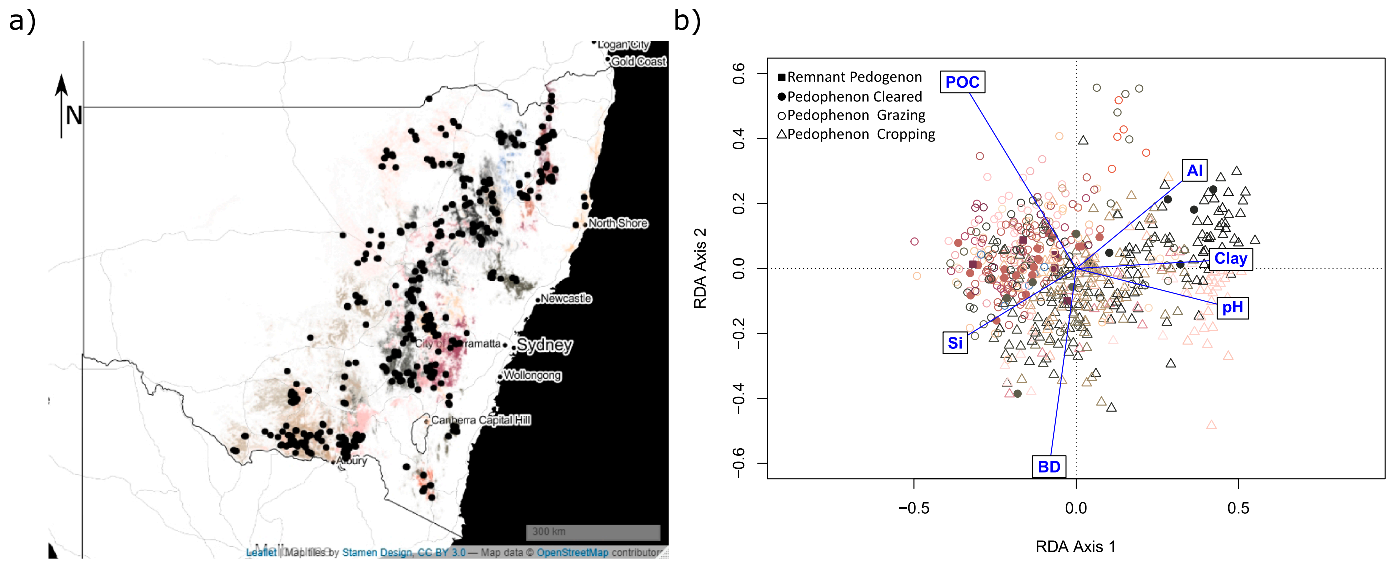

Fig. 7. (a) Location of SCaRP samples (black circles) and pedogenon classes. The subset of data selected the pedogenon and pedogenon/pedophenon subclasses with at least 5 observations per combination ($N = 599$) from 69 different pedogenon classes. The colours follow the dendrogram from Fig. 3, (b) RDA distance triplot (scaling 1) on SCaRP (0–10 cm) samples with pedogenon and pedogenon/pedophenon classes as explanatory variables and six soil properties (clay, total Si, total Al, POC, pH and bulk density) as response variables.

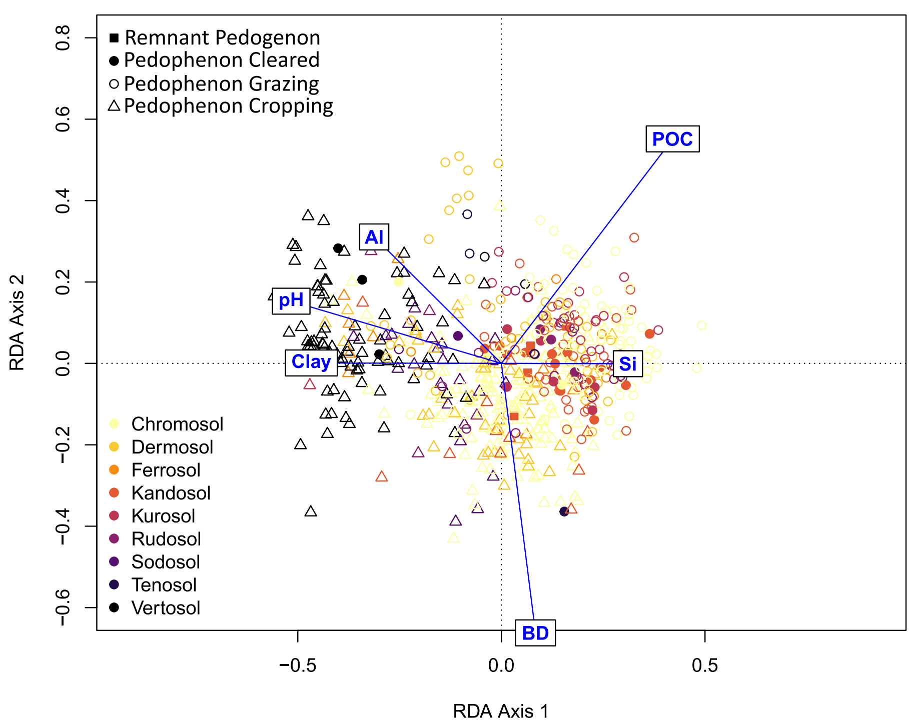

Fig. 8. RDA site conditional triplot (scaling 1) with soil order and pedogenon/pedophenon as explanatory variables and six soil properties (clay, total Si, total Al, POC, pH and bulk density) as response variables.

Fig. 8

Soil order (ASC, Isbell et al., 1997) and land use history (pedogenon/pedophenon classes) explained 39% of the variation of the response variables (clay, total Si, total Al, bulk density, pH, POC) and the adjusted R^2 was 37% (Fig. 7). The global model and the first three RDA axis were statistically significant following a permutation test. Variance partitioning indicated that the pure soil order effect corresponded to 19% of the variation, the pure pedogenon/pedophenon effect 8% and the shared effect 10% of the variation. Compared to the RDA with pedogenon as explanatory variable, land use and management explained more variation of soil properties. On the other hand, when the RDA included as explanatory variables the branch (pedogenon family) and pedogenon/pedophenon class, the constrained variance decreased to 23% of the variation and the adjusted R^2 to 22%, suggesting that this aggregation of pedogenon classes into higher-level taxa failed to explain the variation of stable and dynamic soil properties. The pure branch effect explained 3% of the variation and the pure pedogenon/pedophenon effect 7% of the total variation. The shared effect was 12 with

Table 10
PERMANOVA on six soil properties (clay, total Si, total Al, POC, pH, bulk density) of the SCaRP subset with at least 5 observations per subclass (9999 permutations). For clarity, we refer to the factor pedogenon/pedophenon subclass as land use history.

Factor	Df	SS	R^2	F	*p*-value
Pedogenon ($k = 1000$)	68	337.76	0.56	10.12	0.0001
Land use history	3	2.65	0.004	1.08	0.1416
Pedogenon: Land use history	4	0.98	0.002	0.50	0.7373
Residual	523	256.61	0.43		
Total	598	598	1		
Soil order	8	114.86	0.19	19.19	0.0001
Land use history	3	19.57	0.03	8.72	0.0001
Soil order: Land use history	14	34.94	0.06	3.34	0.0001
Residual	523	428.63	0.72		
Total	598	598.00	1		
Pedogenon branch	5	69.04	0.12	16.19	0.0001
Land use history	3	22.83	0.03	8.92	0.0001
Pedogenon branch: Land use history	3	5.50	0.01	2.15	0.0973
Residual	587	500.64	0.84		
Total	598	598.00	1		

Table 11
PERMANOVA on three dynamic soil properties (POC, pH, bulk density) of the SCaRP subset with at least 5 observations per subclass (9999 permutations). For clarity, we refer to the factor pedogenon/pedophenon subclass as land use history.

Factor	Df	SS	R^2	F	*p*-value
Pedogenon ($k = 1000$)	68	948.11	0.53	8.87	0.0001
Land use history	3	17.99	0.01	3.81	0.0002
Pedogenon: Land use history	4	5.34	0.003	0.85	0.6
Residual	523	822.55	0.46		
Total	598	1794.00	1		
Soil order	8	413.11	0.23	26.10	0.0001
Land use history	3	171.97	0.10	28.97	0.0001
Soil order: Land use history	14	75.21	0.04	2.72	0.0001
Residual	523	1133.71	0.63		
Total	598	1794.00	1		
Pedogenon branch	5	204.39	0.11	16.74	0.0001
Land use history	3	123.43	0.07	16.86	0.0001
Pedogenon branch: Land use history	3	33.37	0.02	4.56	0.0001
Residual	587	1432.81	0.80		
Total	598	1794.00	1		

78% of residual variance. Similarly, when only dynamic soil properties were included as response variables, the RDA with pedogenon branch had an adjusted R^2 of 26%. In this case, the shared effect explained 14% of the variance, the pure pedogenon/pedophenon effect 9% and the pure branch effect explained just 2% of the variation.

The RDA with land use history variables (vegetation extent, status, land use) had an adjusted R^2 of 22% on stable and dynamic soil properties and 27% on dynamic soil properties. When the pedogenon/pedophenon classification was the only explanatory variable the adjusted R^2 was 19% for stable and dynamic soil properties and 24% for dynamic soil properties.

3.3.2. Permanova on stable and dynamic soil properties

The PERMANOVA test was applied to the same SCaRP subset as in Section 3.3.1. The centroids of the pedogenon classes ($n = 69$) differed in location according to the PERMANOVA test ($p = 0.001$) (Table 11). The pedogenon/pedophenon classes and their interaction with pedogenon class did not have a significant effect on the location of the class centroids. Pairwise comparisons found statistically significant differences between the pedophenon grazing ($n = 11$) and pedophenon cropping ($n = 7$) of one pedogenon class. The permutation test for homogeneity of multivariate dispersions indicated that the pedogenon classes and the pedogenon/pedophenon groups had non-homogeneous dispersions.

Soil order, pedogenon/pedophenon class and their interaction had a significant effect on the location of the class centroids ($p = 0.0001$). The differences in the dispersion of observation to the class centroids by soil order were statistically significant. The dispersion was higher in Ferrosols, Dermosols and Tenosols, and smaller for Vertosols, Kurosols, Kandosols and Sodosols. These differences in the dispersion may be related with the choice of response variables and the characteristics of the soil orders (e.g., high clay content in Vertosols) or the higher diversity within some soil orders (e.g., Dermosols, Tenosols) (ASC, Isbell et al., 1997). The differences among centroids by pedogenon branch and pedogenon/pedophenon class were statistically significant, but they showed non-homogeneous dispersion ($p = 0.04$).

3.3.3. Permanova on dynamic soil properties

Pedogenon and pedogenon/pedophenon classes (land use history) had significant effects on the location of group centroids when the PERMANOVA test was applied only to dynamic soil properties (Table 10). Pairwise comparisons between pedogenons/pedophenons within the same pedogenon class (7 pedogenons) indicated differences in group centroids for five pedogenon classes. The effects of soil order, land use history and their interaction on centroid location were statistically significant ($p = 0.0001$). Similarly, the effects of pedogenon branch, land use history, and interaction were statistically significant although the amount of variance explained was very small. The PERMDISP test indicated non-homogeneous dispersion among pedogenon, pedogenon/pedophenon classes, soil order and pedogenon branch ($p = 0.0001$).

4. Discussion

4.1. Identification of reference state, current limitations and potential applications

The main goal of this paper was to design a DSM framework for identifying areas that can serve as a reference state for assessing changes in soil condition and capability. The DSM framework was inspired by the concepts of genoform and phenoform (Rossiter and Bouma, 2018) and the conceptual model of soil change developed by Yaalon and Yaron (1966) and Richter (2007). For each pedogenon class, there was at least one grid cell (≈ 0.007 km^2) classified as remnant pedogenon, with a median contribution to the area of their respective pedogenon of 5.3%. These results may suggest that every pedophenon could be compared with its correspondent remnant pedogenon. However, the reference state for some classes should be quasi-remnant pedogenon or cleared pedophenon subclass, to have a sufficiently large area for taking a representative sample and incorporate the spatial variability of soil properties within the pedogenon subclass. Small patches of native remnant vegetation, including isolated trees, were classified as remnant or quasi-remnant pedogenons. Isolated trees have suffered a decline in abundance and density since the 1960s (Ozolins et al., 2001), which likely has affected soil condition due to differences in soil microclimate and biogeochemical cycling linked to a decrease in organic matter input (Eldridge and Wong, 2005). Hence, using these native vegetation patches as reference state for studying soil change should be done with caution, although in some areas they constitute the last fragments of native vegetation.

Digital pedogenon mapping was designed as a first-order method for mapping soil classes with similar long-term pedogenesis and historic anthropedogenesis. Subsequently, we expanded the approach for dividing them into subclasses with different land use history. Pedogenon classes are produced under the assumption that groups with homogeneous state variables representing the soil-forming factors for a given reference time, undergo similar dominant soil-forming processes over pedogenetic time, and thus have similar soil properties (Román Dobarco et al., 2021). The viability of the derived pedogenon and pedophenon maps for assessing changes in soil condition and capability relies on this hypothesis. The RDA and PERMANOVA analyses attributed most of the explained variance to the pedogenon classes, supporting the hypothesis that classes defined by homogeneous state variables present differences in stable and dynamic soil properties. However, the set of soil properties analysed in this study and previously (Román Dobarco et al., 2021) are not sufficient for proving links between pedogenon classes and long-term pedogenesis. Thus, this hypothesis cannot be accepted yet. Secondly, the results of the statistical analyses indicate that the stratification into pedogenon and pedophenon subclasses was not able to detect changes in soil properties caused by contemporary management. Since the effects of management and land use history have been confirmed across different areas of New South Wales, this suggests that the stratification with the rule-based algorithm is incorrect (see Section 4.2). Another possible explanation is that the legacy soil data may not be representative of pedogenon and pedophenon subclasses within a study area. Errors in the pre-European vegetation mapping, land use (history) allocation, and space and time biases in the sampling negatively impacted the analysis and that is why we might not detect differences even though these might exist. A way to remediate these limitations would be to use the mapping in local areas: the first step would be a local recalibration/updating to improve the mapping followed by new dedicated (stratified random) sampling to test change hypotheses and estimate the amount of change.

The environmental factors (pedogenon class) and the interaction with land use history (e.g., shared effect of the interaction in RDA analyses) explained between 50 and 60% of the variation of soil properties whereas land use history (pedogenon and pedophenon subclasses) alone explained a minimal proportion of the variance (< 10%). However, land use history categories (pedogenons and pedophenons) aggregated across pedogenon classes showed differences in mean pH (Table 6). This may be an example of the Simpson's paradox (Sprenger and Weinberger, 2021) or be linked to the predominance of land uses in the biophysical settings more suitable for their purpose (e.g., cropping at low slopes, more fertile soils, and suitable climatic conditions). Finally, the aggregation of pedogenon classes into higher-level taxa with hierarchical clustering did not successfully define groups that explained the variation of soil properties. This suggests that the aggregation into higher-level taxa should be done with a different method (e.g., weighing the centroids with the number of observations belonging to each class, clustering method, distance), or if we are interested in a smaller number of pedogenon classes a smaller k should be directly set. Hence, future work should develop methods for evaluating the correspondence between pedogenon classes and long-term pedogenesis, refine the selection of covariates and the optimal number of classes for a study area, and investigate how different clustering methods and objective functions influence the resulting maps.

Pedogenon mapping is appropriate for large extents with limited soil observations or when soil classification systems do not explicitly regard genetic criteria. At smaller spatial scales (landscape, catchment) where detailed soil maps consider genetic pathways in the classification, or for soil classification systems that account for recent disturbances, the genoform and phenoform approach (Rossiter and Bouma, 2018) may be more appropriate. The results of the PERMANOVA and RDA analyses with soil order as explanatory variable (Tables 9 and 10, Fig. 7) suggest that mapping soil great groups at a reference time (Huang et al., 2018) are valid for detecting the effect of contemporary management on stable and dynamic soil properties. These approaches, however, are challenging to implement at a larger extent.

This framework has the potential to be integrated into a holistic soil security assessment (McBratney et al., 2014; Field, 2020). This requires to translate the indicators of soil condition and capability (e.g., SOC, particle size distribution, pH) into soil functions and ecosystem services (e.g., carbon sequestration, food production, water and nutrient cycling, and storage) (Schulte et al., 2014; Greiner et al., 2017; Bouma et al., 2019; Ellili-Bargaoui et al., 2021). Another potential application of this framework is the design of sampling strategies and monitoring soil change. The pedogenons and their subclasses can be used as strata for distributing soil sampling points. The number of pedogenon classes and stratification into pedogenons/pedophenons can be modified to meet the available resources, e.g., creating maps with a smaller *k* or merging pedogenon classes into higher-level taxa with the hierarchical dendrogram.

4.2. Stratification of pedogenons into subclasses by land use history

In this study, we used a simple classification of pedogenons/pedophenons based on three sources of information (native vegetation extent, status, and current land use). These classes had a significant effect on the centroids of groups of dynamic soil properties although they explained a very small fraction of the variation of soil stable and dynamic properties. Hence, this suggests that the rule-based algorithm failed to identify meaningful classes in terms of soil change. However, it is also likely that the data sources did not provide enough information on land use history, because the difference in adjusted R^2 of RDA analyses performed with the three layers separately or reclassified into pedogenon/pedophenon classes was < 4%. This top-down approach did not include detailed site information on management practices and intensity (e.g., livestock density), time since land use change, time since clearing, etc. For example, we assumed that grazing in relatively natural environments could be considered a low-intensity pressure, but there were no significant differences between quasi-remnant pedogenons and pedophenon grazing. Detailed information on land use history and management practices, aggregated into classes or used as independent variables, could improve the classification of pedogenons and pedophenons. Long-term time series of satellite imagery should be used to inform on the changes in land cover/land use for as long as there are available images. A possible improvement of the classification of pedophenons would be to create a composite index of cumulative soil anthropogenic pressures, similar to the index of cumulative human modification developed by Kennedy et al. (2019) for terrestrial lands or the index on threats for soil biodiversity developed by Gardi et al. (2013). However, this index should describe the degree of intensity of anthropogenic activities by the main type of activity (e.g., varying degree of pressure for cropping), because the type of activity greatly affects the trajectory of soil change.

4.3. Trajectories of soil properties

Kuzyakov and Zamanian (2019) proposed that whereas natural pedogenesis causes the diversification of soils, agropedogenesis (i.e., agricultural practices are the dominant factor of soil formation) leads to narrowing and convergence of soil properties. The reduction in the ranges of soil properties and their final convergence is driven by fostering a single ecosystem function (crop production) while other soil functions (e.g., biodiversity pool, climate regulation) are reduced. The rates of change would differ for different soil properties, as a function of climate, biophysical conditions and land use intensity (Kuzyakov and Zamanian 2019; Richter 2007). The interactions between the type of management and pedogenon on topsoil pH partly support this hypothesis. Topsoil pH followed the trend remnant pedogenon $\approx$ quasi-remnant pedogenon $\leq$ pedophenon cleared $\approx$ pedophenon grazing $\leq$ pedophenon cropping, but it is not possible to ascribe it only to the effect of management (e.g., liming). Rather, the data suggests that differences in pH between pedophenons may be linked to the occurrence of different land uses in soils with different soil properties and suitability. Across the agricultural belt of NSW, cropping is generally located in areas with lower precipitation and more alkaline soils. In contrast, grazing is located in a broader range of environments, including areas with higher precipitation. Vast areas of remnant pedogenons were located in the Great Dividing Range and towards the coast, with higher humidity. However, pairwise comparisons by pedogenon class also suggested that pH increased from remnant pedogenons to pedophenons, while pedophenon grazing may have lower pH than pedophenon cleared and cropping (Table 7). Wilson et al. (2011) observed lower pH of surface soil layers in managed pastures from northwest NSW compared to woodlands, native grasslands, and croplands despite periodic liming. The acidification was associated with nitrogen leaching from legume roots (Slattery et al., 1999; Lockwood et al., 2003; Wilson et al., 2011). The pH data in the SCaRP dataset contradict in part this trend (Table 8). Pedophenon grazing had lower pH than pedophenon cleared and cropping. However, the pH at quasi-remnant pedogenons was higher than among the rest of subclasses. A possible explanation is the lower sample number compared to pedophenon subclasses, or that the quasi-remnant pedogenons sampled during the SCaRP may have been located in more alkaline, drier rangeland areas.

The convergence of dynamic soil properties from different soil orders (interpreted as genoforms) driven by land use history was observed for several land uses (i.e., phenoforms: native forests, pastures, crops) in New Zealand (Stevenson et al., 2015). Conversely, different land use management led to distinct functional soil classes within a soil order (Stevenson et al., 2015). The strong influence of land use management on soil condition was reflected in the correlation between soil functional classes (defined with cluster analysis on soil physicochemical properties) and different habitats rather than with historic pedogenetic classes (Seaton et al., 2020). We hypothesized that the response of a historic soil system to different anthropogenic activities, in magnitude and direction, would differ among pedogenon classes depending on their intrinsic resistance and resilience. The trajectory of dynamic soil properties would also vary depending on the type of land management practice, its intensity and duration. The results could neither support nor reject this hypothesis. The effect of management was pertinent when we considered dynamic soil properties only, as indicated by the RDA and the PERMANOVA analyses on the SCaRP dataset. Differences in group location between remnant pedogenon/pedophenons of the same pedogenon class suggest that it was possible to detect the effect of management and land use history on indicators of soil condition. However, there were almost no changes in stable and dynamic soil properties, indicators of soil capability, among pedophenons of the same pedogenon. We did not observe clear patterns in trajectories of dynamic soil properties, or their centroids in the multivariate space, by pedogenon classes (data not shown). The shared effect of pedogenon and remnant pedogenon/pedophenon explaining the variation of soil properties again indicates the occurrence of land uses in pedogenons that are more suitable due to environmental and soil characteristics. Rabbi et al. (2014) analysed the SCaRP dataset and found that environmental variables and soil properties explained 42% of the variation of SOC fractions in NSW while land use and management practices explained 9.2% of the variation. Higher POC stock under pasture than under cropping was likely caused by disruption of aggregates and enhanced decomposition of POC under cropping and higher carbon inputs (shoot residues, rhizodeposition) under pastures (Rabbi et al., 2014). The negative correlation between soil pH and POC stocks was linked to lower pH under pastures than under cropping.

Thresholds of soil degradation specific to different soil classes can be identified with phase diagrams (Kuzyakov and Zamanian, 2019) or ratios of key soil properties (Prout et al., 2020). Alternatively, we can calculate distance metrics with multivariate indicators of soil condition and capability and compare between- and within-group distances along the gradient from remnant to cropping pedogenon subclasses. For example, the only pedogenon class that had statistically significant differences between the centroids (defined with stable and dynamic soil properties) of subclasses grazing and cropping had average within-group Mahalanobis distances of 1.14 and 0.68 and an average between-group Mahalanobis distance of 1.22. We could not illustrate this application across the whole sequence of anthropogenic pressure because no pedogenon class had SCaRP data in more than two subclasses. In addition, the datasets used here were targeted for agricultural soils or composed of legacy soil data, not intended to characterize soil change. Future studies will apply this method at local scale and collect soil samples with the aim of assessing soil change and investigate trajectories of soil properties as a response to management.

5. Conclusions

- We developed a top-down framework for detecting soil change that can be applied to large areas.
- This framework gives a new and valuable stratification of the landscape.
- It integrates a theoretical model of soil change with a digital soil mapping approach.
- Detailed land use history is an essential factor for detecting soil change.

5.1. Future work

- The methodology for defining pedogenon classes needs to be improved for including the past conditions in soil formation (e.g., paleoclimates, paleosols, etc.), optimizing the selection of covariates and the number of classes.
- Integrate the indicators of soil capability and condition into a holistic soil security assessment.
- Adapt this framework to human-natural systems with a long history (millennial) of agricultural use and extensive anthropogenic pressure.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to thank Ross Searle of CSIRO for his help with the Soil Data Federator and his suggestions on the covariate selection. The authors acknowledge the Terrestrial Ecosystem Research Network (TERN), an Australian Government NCRIS-enabled research infrastructure project, for facilitating and supporting this research.

Code availability

The scripts for implementing the modelling framework and visualize pedogenons and dendrograms, and the stratification into subclasses are available at https://github.com/MercedesRD/GenoPheno. There is a reproducible example for pedogenon mapping, and the script with the analyses for this paper.

References

ABARES, 2019. Catchment Scale Land use of Australia – Update December 2018. ABARES, Canberra. https://doi.org/10.25814/5c7728700fd2a. March CC BY 4.0.
Anderson, M.J., 2001. A new method for non-parametric multivariate analysis of variance. Aust. Ecol. 26 (1), 32–46. https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x.
Anderson, M.J., 2017. Permutational multivariate analysis of variance (PERMANOVA). In: Balakrishnan, N., Colton, T., Everitt, B., Piegorsch, W., Ruggeri, F., Teugels, J.L. (Eds.), Wiley StatsRef: Statistics Reference Online. Wiley. https://doi.org/10.1002/9781118445112.stat07841.
Anderson, M.J., Walsh, D.C.I., 2013. PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: what null hypothesis are you testing? Ecol. Monogr. 83 (4), 557–574. https://doi.org/10.1890/12-2010.1.
Arthur, D., Vassilvitskii, S., 2007. K-means++: the advantages of careful seeding. In: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '07). USA. Society for Industrial and Applied Mathematics, pp. 1027–1035. https://doi.org/10.1145/1283383.1283494.

Arrouays, D., McBratney, A.B., Minasny, B., Hempel, J.W., Heuvelink, G.B.M., MacMillan, R.A., Hartemink, A.E., Lagacherie, P., Mckenzie, N.J., 2014. The GlobalSoilMap project specifications. In: Arrouays, D., Mckenzie, N.J., Hempel, J., Richer de Forges, A., McBratney, A. (Eds.), Global Soil Map: Basis of the Global Spatial Soil Information System. CRC Press, London, pp. 9–12.
Ashcroft, L., Gergis, J., Karoly, D.J., 2014. A historical climate dataset for southeastern Australia, 1788–1859. Geosci. Data J. 1, 158–178. https://doi.org/10.1002/gdj3.19.
Baldock, J.A., Hawke, B., Sanderman, J., Macdonald, L.M., 2013a. Predicting contents of carbon and its component fractions in Australian soils from diffuse reflectance mid-infrared spectra. Soil Res. 51, 577–595. https://doi.org/10.1071/SR13077.
Baldock, J., Sanderman, J., Macdonald, L., Allen, D., Cowie, A., Dalal, R., Davy, M., Doyle, R., Herrmann, T., Murphy, D., Robertson, F., 2013b. Australian soil carbon research program. v2. CSIRO. Data Collection. 10.25919/5ddfd6888d4e5.
Berthe, A.A., 2019. Chapter 3 - drivers of soil change. In: Busse, M., Giardina, C.P., Morris, D.M., Page-Dumroese, D.S. (Eds.), Global Change and Forest Soils. Developments in Soil Science (36). Elsevier, pp. 27–42. https://doi.org/10.1016/B978-0-444-63998-1.00003-3.
Bird, M.I., Beaman, R.J., Condie, S.A., Cooper, A., Ulm, S., Veth, P., 2018. Palaeogeography and voyage modeling indicates early human colonization of Australia was likely from Timor-Roti. Quaternary Sci. Rev. 191, 431–439. https://doi.org/10.1016/j.quascirev.2018.04.027.
Bishop, T.F.A., McBratney, A.B., Laslett, G.M., 1999. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. Geoderma 91 (1–2), 27–45. https://doi.org/10.1016/S0016-7061(99)00003-8.
Borcard, D., Legendre, P., Drapeau, P., 1992. Partialling out the spatial component of ecological variation. Ecol 73, 1045–1055. https://doi.org/10.2307/1940179.
Borcard, D., Gillet, F., Legendre, P., 2018. Canonical Ordination, in: Numerical Ecology with R. Use R!, 2nd ed. Springer, Cham. https://doi.org/10.1007/978-3-319-71404-2_6.
Bouma, J., Montanarella, L., Evanylo, G., 2019. The challenge for the soil science community to contribute to the implementation of the UN sustainable development goals. Soil Use Manag. 35 (4), 538–546. https://doi.org/10.1111/sum.12518.
Carré, F., Jacobson, M., 2009. Numerical classification of soil profile data using distance metrics. Geoderma 148, 336–345. https://doi.org/10.1016/j.geoderma.2008.11.008.
Cline, M.G., 1961. The changing model of soil. Soil Sci. Soc. Am. J. 25 (6), 442–446. https://doi.org/10.2136/sssaj1961.03615995002500060009x.
Cotching, W., Kidd, D., 2010. Soil quality evaluation and the interaction with land use and soil order in Tasmania, Australia. Agric. Ecosyst. Environ. 137 (3), 358–366. https://doi.org/10.1016/j.agee.2010.03.006.
Dazzi, C, Lo Papa, G, 2019. Soil genetic erosion: New conceptual developments in soil security. International Soil and Water Conservation Research 7 (4), 317–324. https://doi.org/10.1016/j.iswcr.2019.08.001.
Dokuchaev, V.V., 1883. Russian Chernozem. Selected Works of V.V. Dokuchaev. Volume I (Translated in 1967). Israel Program for Scientific Translations. U.S. Department of Agriculture, Washington DC.
Droogers, P., Bouma, J., 1997. Soil survey input in exploratory modeling of sustainable soil management practices. Soil Sci. Soc. Am. J. 61, 1704–1710. https://doi.org/10.2136/sssaj1997.03615995006100060023x.
Eldridge, D.J., Wong, V.N.L., 2005. Clumped and isolated trees influence soil nutrient levels in an Australian temperate box woodland. Plant Soil 270, 331–342. https://doi.org/10.1007/s11104-004-1774-2.
Ellili-Bargaoui, Y., Walter, C., Lemercier, B., Michot, D., 2021. Assessment of six soil ecosystem services by coupling simulation modelling and field measurement of soil properties. Ecol. Indic. 121, 107211 https://doi.org/10.1016/j.ecolind.2020.107211.
Farr, T.G., Kobrick, M., 2000. Shuttle radar topography mission produces a wealth of data. Eos Trans. AGU 81 (48), 583–585. https://doi.org/10.1029/EO081i048p00583.
Field, D., 2020. Sustaining agri-food systems framed using soil security and education. Int. J. Agric. Nat. Resour. 47 (3), 152–165. https://doi.org/10.7764/ijanr.v47i3.2289.
Fisher, A., Danaher, T., Gill, T., 2017. Mapping trees in high resolution imagery across large areas using locally variable thresholds guided by medium resolution tree maps. Int. J. Appl. Earth Obs. Geoinf. 58, 86–96. https://doi.org/10.1016/j.jag.2017.02.004.
Gale, S.J., Haworth, R.J., 2005. Catchment-wide soil loss from pre-agricultural times to the present: transport-and supply-limitation of erosion. Geomorphology 68 (3–4), 314–333. https://doi.org/10.1016/j.geomorph.2004.10.008.
Gallant, J.C., Dowling, T.I., 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. Water Resour. Res. 39, 1347. https://doi.org/10.1029/2002WR001426, 12.
Gallant, J., Wilson, N., Tickle, P.K., Dowling, T., Read, A., 2009. 3s SRTM Derived Digital Elevation Model (DEM) Version 1.0. Record 1.0. Geoscience Australia, Canberra. http://pid.geoscience.gov.au/dataset/ga/69888.
Gammage, B., 2011. The Biggest Estate on Earth: How Aborigines Made Australia. Allen & Unwin, Crows Nest, NSW, Australia.
Gardi, C., Jeffery, S., Saltelli, A., 2013. An estimate of potential threats levels to soil biodiversity in EU. Glob. Change Biol. 19 (5), 1538–1548. https://doi.org/10.1111/gcb.12159.
Greiner, L., Keller, A., Grêt-Regamey, A., Papritz, A., 2017. Soil function assessment: review of methods for quantifying the contributions of soils to ecosystem services. Land Use Policy 69, 224–237. https://doi.org/10.1016/j.landusepol.2017.06.025.
Guo, Y.Y., Gong, P., Amundson, R., 2003. Pedodiversity in the United States of America. Geoderma 117 (1–2), 99–115. https://doi.org/10.1016/S0016-7061(03)00137-X.
Han, J., Kamber, M., Pei, J., 2012. 10-cluster analysis: basic concepts and methods. In: Han, J., Kamber, M., Pei, J. (Eds.), Data Mining, 3rd ed. Morgan Kaufmann, pp. 443–495. https://doi.org/10.1016/B978-0-12-381479-1.00010-1.
Harwood, T.D., Donohue, R.J., Williams, K.J., Ferrier, S., McVicar, T.R., Newell, G., White, M., 2016. Habitat Condition Assessment System: a new way to assess the condition of natural habitats for terrestrial biodiversity across whole regions using remote sensing data. Methods Ecol. Evol. 7 (9), 1050–1059. https://doi.org/10.1111/2041-210X.12579.
Hartigan, J.A., Wong, M.A., 1979. A K-means clustering algorithm. J. R. Stat. Soc. Series C Appl. Stat. 28, 100–108. https://doi.org/10.2307/2346830.
Huang, J.Y., 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. https://doi.org/10.1016/j.geoderma.2018.05.016.
Isbell, R.F., McDonald, W.S., Ashton, L.J., 1997. Concepts and Rationale of the Australian Soil Classification. CSIRO Land and Water.
Janik, L., Skjemstad, J., 1995. Characterization and analysis of soils using midinfrared partial least-squares. 2. Correlations with some laboratory data. Aust. J. Soil Res. 33, 637–650. https://doi.org/10.1071/SR9950637.
Janik, L., Skjemstad, J., Raven, M., 1995. Characterization and analysis of soils using mid-infrared partial least-squares. 1. Correlations with XRF-determined major-element composition. Aust. J. Soil Res. 33, 621–636. https://doi.org/10.1071/SR9950621.
Jenny, H., 1941. Factors of Soil Formation: A System of Quantitative Pedology. Dover Publications, New York.
Keesstra, S.D., Bouma, J., Wallinga, J., Tittonell, P., Smith, P., Cerdà, A., Montanarella, L., Quinton, J.N., Pachepsky, Y., van der Putten, W.H., Bardgett, R.D., Moolenaar, S., Mol, G., Jansen, B., Fresco, L.O., 2016. The significance of soils and soil science towards realization of the United Nations Sustainable Development Goals. Soil 2, 111–128. https://doi.org/10.5194/soil-2-111-2016, 2016.
Keith, D.A., Simpson, C.C., 2008. A protocol for assessment and integration of vegetation maps, with an application to spatial data sets from south-eastern Australia. Austral Ecol. 33 (6), 761–774. https://doi.org/10.1111/j.1442-9993.2008.01844.x.
Kennedy, C.M., Oakleaf, J.R., Theobald, D.M., Baruch-Mordo, S., Kiesecker, J., 2019. Managing the middle: a shift in conservation priorities based on the global human modification gradient. Glob. Change Biol. 25, 811–826. https://doi.org/10.1111/gcb.14549.
Kidd, D., Field, D., McBratney, A., Webb, M., 2018. A preliminary spatial quantification of the soil security dimensions for Tasmania. Geoderma 322, 184–200. https://doi.org/10.1016/j.geoderma.2018.02.018.
Kuzyakov, Y., Zamanian, K., 2019. Reviews and syntheses: agropedogenesis - humankind as the sixth soil-forming factor and attractors of agricultural soil degradation. Biogeosciences 16 (24), 4783–4803. https://doi.org/10.5194/bg-16-4783-2019.
Legendre, P., Legendre, L.F., 2012. Canonical analysis. Numerical Ecology, 3rd ed. Elsevier, Amsterdam, The Netherlands.
Lenth, R.V., 2020. Emmeans: estimated Marginal Means, aka Least-Squares Means. R package version 1.5.3. https://CRAN.R-project.org/package=emmeans.
Lockwood, P.V., Wilson, B.R., Daniel, H., Jones, M., 2003. Soil Acidification and Natural Resource management: Directions For the Future. University of New England, Armidale, NSW, Australia.
Lo Papa, G., Palermo, V., Dazzi, C., 2011. Is land-use change a cause of loss of pedodiversity? The case of the Mazzarrone study area, Sicily. Geomorphol. 135 (3–4), 332–342. https://doi.org/10.1016/j.geomorph.2011.02.015.
Lunt, I.D., Jones, N., Spooner, P.G., Petrow, M., 2006. Effects of European colonization on indigenous ecosystems: post-settlement changes in tree stand structures in Eucalyptus-Callitris woodlands in central New South Wales, Australia. J. Biogeogr. 33 (6), 1102–1115. https://doi.org/10.1111/j.1365-2699.2006.01484.x.
Lyles, L., Tatarko, J., 1986. Wind erosion effects on soil texture and organic matter. J. Soil Water Conserv. 41 (3), 191–193.
Malone, B.P., Hughes, P., McBratney, A.B., Minasny, B., 2014. A model for the identification of terrons in the Lower Hunter Valley, Australia. Geoderma Reg. 1, 31–47. https://doi.org/10.1016/j.geodrs.2014.08.001.
Mancini, F., Ronchetti, G., 1968. Carta Della Potenzialità dei Suoli d'Italia (con note illustrative). Comitato per La Carta Dei Suoli. Tipografia Coppini, Firenze, p. 37 (in Italian).
Martinez Arbizu, P., 2020. pairwiseAdonis: pairwise multilevel comparison using adonis. R package version 0.4. Retrieved from https://github.com/pmartinezarbizu/pairwiseAdonis on February 8, 2021.
McBratney, A., Field, D.J., Koch, A., 2014. The dimensions of soil security. Geoderma 213, 203–213. https://doi.org/10.1016/j.geoderma.2013.08.013.
McBratney, A.B., Field, D., Morgan, C.L.S., Huang, J.Y., 2019. On soil capability, capacity, and condition. Sustainability 11, 3350. https://doi.org/10.3390/su11123350.
Minasny, B., McBratney, A.B., Hartemink, A.E., 2010. Global pedodiversity, taxonomic distance, and the World Reference Base. Geoderma 155, 132–139. https://doi.org/10.1016/j.geoderma.2009.04.024.
Minty, B.R.S., 2019a. Radiometric Grid of Australia (Radmap) v4 2019 Filtered Pct Potassium Grid. Geoscience Australia, Canberra. https://doi.org/10.26186/5dd48d628f4f6 http://dx.doi.org/.
Minty, B.R.S., 2019b. Radiometric Grid of Australia (Radmap) v4 2019 Filtered ppm Thorium. Geoscience Australia, Canberra. https://doi.org/10.26186/5dd48e3eb6367.
Minty, B.R.S., 2019c. Radiometric Grid of Australia (Radmap) v4 2019 Ratio Uranium Over Thorium. Geoscience Australia, Canberra. https://doi.org/10.26186/5dd4a63603704.

Mouselimis, L., 2021. ClusterR: Gaussian Mixture Models, K-Means, Mini-Batch-Kmeans, K-Medoids and Affinity Propagation Clustering. R package version 1.2.5. https://CRAN.R-project.org/package=ClusterR.

Muñoz-Salinas, E., Bishop, P., Sanderson, D., Kinnaird, T., 2014. Using OSL to assess hypotheses related to the impacts of land use change with the early nineteenth century arrival of Europeans in southeastern Australia: an exploratory case study from Grabben Gullen Creek. N. S. W. Earth Surf. Process. Landf. 39 (12), 1576–1586. https://doi.org/10.1002/esp.3542.

National Vegetation Information System V5.1 ©, 2018. Australia - Pre-1750 Major Vegetation Groups - NVIS Version 5.1 (Albers 100m Analysis Product). Australian Government Department of Agriculture, Water and the Environment.

NSW Office of Environment and Heritage (OEH), 2019. NSW Native Vegetation Extent 5m Raster v1.2. NSW Department of Planning, Industry and Environment, Sydney, Australia. Retrieved from. https://data.nsw.gov.au/data/dataset/nsw-native-vegetation-extent-5m-raster-v1-0.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H., 2019. vegan: community ecology package.

Ozolins, A., Brack, C., Freudenberger, D., 2001. Abundance and decline of isolated trees in the agricultural landscapes of central New South Wales, Australia. Pac. Conserv. Biol. 7, 195–203. https://doi.org/10.1071/PC010195.

Pascoe, B., 2014. Dark Emu: Aboriginal Australia and the Birth of Agriculture. Magabala Books, Broome, Australia.

Pozza, L.E., Field, D.J., 2020. The science of soil Security and food security. Soil Secur. 1, 100002 https://doi.org/10.1016/j.soisec.2020.100002.

Prout, J.M., Shepherd, K.D., McGrath, S.P., Kirk, G.J.D., Haefele, S.M., 2020. What is a good level of soil organic matter? An index based on organic carbon to clay ratio. Eur. J. Soil Sci. 1–11. https://doi.org/10.1111/ejss.13012.

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. Agric. Ecosys. Environ. 197, 41–52. https://doi.org/10.1016/j.agee.2014.06.020.

Rayment, G.E., Lyons, D.J., 2011. Soil Chemical methods: Australasia (Vol. 3). CSIRO publishing.

Richter, D.D., 2007. Humanity's transformation of Earth's soil: pedology's new frontier. Soil Sci. 172 (12), 957–967. https://doi.org/10.1097/ss.0b013e3181586bb7.

Richter, D.D., Yaalon, D.H., 2012. The changing model of soil" revisited. Soil Sci. Soc. Am. J. 76, 766–778. https://doi.org/10.2136/sssaj2011.0407.

Román Dobarco, M., McBratney, A., Minasny, B., Malone, B., 2021. A modelling framework for pedogenon mapping. Geoderma 393, 115012. https://doi.org/10.1016/j.geoderma.2021.115012.

Rossiter, D.G., Bouma, J., 2018. A new look at soil phenoforms - definition, identification, mapping. Geoderma 314, 113–121. https://doi.org/10.1016/j.geoderma.2017.11.002.

Russell, J.S., Isbell, R.F., 1986. Australian Soils: the Human Impact. University of Queensland Press, Brisbane, Australia.

Sanderman, J., Baldock, J., Hawke, B., Macdonald, L., Puccini, A., Szarvas, S., 2011. National Soil Carbon Research Programme: Field and Laboratory Methodologies. CSIRO. http://hdl.handle.net/102.100.100/101994?index=1.

Schulte, R.P., Creamer, R.E., Donnellan, T., Farrelly, N., Fealy, R., O'Donoghue, C., O'huallachain, D., 2014. Functional land management: a framework for managing soil-based ecosystem services for the sustainable intensification of agriculture. Environ. Sci. Policy 38, 45–58. https://doi.org/10.1016/j.envsci.2013.10.002.

Seaton, F.M., Barrett, G., Burden, A., Creer, S., Fitos, E., Garbutt, A., Griffiths, R.I., Henrys, P., Jones, D.L., Keenan, P., Keith, A., Lebron, I., Maskell, L., Pereira, M.G., Reinsch, S., Smart, S.M., Williams, B., Emmett, B.A., Robinson, D.A., 2020. Soil health cluster analysis based on national monitoring of soil indicators. Eur. J. Soil Sci. 1–16. https://doi.org/10.1111/ejss.12958.

Skjemstad, J.O., Spouncer, L.R., Cowie, B., Swift, R.S., 2004. Calibration of the Rothamsted organic carbon turnover model (RothC ver. 26.3), using measurable soil organic carbon pools. Aust. J. Soil Res. 42, 79–88. https://doi.org/10.1071/SR03013.

Slattery, W.J., Conyers, M.K., Aitken, R.I., 1999. Soil pH, manganese and lime requirement. In: Peverill, K.I., Sparrow, L.A., Reuter, D.J. (Eds.), Soil Analysis: an Interpretation Manual. CSIRO, Collingwood, Vic, Australia, pp. 103–128.

Smeck, N.E., Balduff, D., 2002. Contrasting approaches for the classification of eroded soils in the USA. In: Proceedings of the Paper no. 616 in Transactions of the 17th World Congress of Soil Science: Confronting New Realities in the 21st Century. Bangkok, Thailand.

Soil Survey Staff, 2010. Keys to Soil Taxonomy. United States Department of Agriculture, Soil Conservation Service, Washington, DC.

Sprenger, J., Weinberger, N., 2021. Simpson's Paradox. In: Zalta, E.N. (Ed.), The Stanford Encyclopedia of Philosophy (Summer 2021 Edition). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2021/entries/paradox-simpson/. (accessed on May 22nd, 2021).

Stevenson, B.A., McNeill, S., Hewitt, A.E., 2015. Characterising soil quality clusters in relation to land use and soil order in New Zealand: an application of the phenoform concept. Geoderma 239, 135–142. https://doi.org/10.1016/j.catena.2011.06.004.

Stockmann, U., Cattle, S.R., Minasny, B., McBratney, A.B., 2016. Utilizing portable X-ray fluorescence spectrometry for in-field investigation of pedogenesis. Catena 139, 220–231. https://doi.org/10.1016/j.catena.2016.01.00.

Tobler, R., Rohrlach, A., Soubrier, J., Bover, P., Llamas, B., Tuke, J., Bean, N., Abdullah-Highfold, A., Agius, S., O'Donoghue, A., O'Loughlin, I., Sutton, P., Zilio, F., Walshe, K., Williams, A.N., Turney, C.S.M., Williams, M., Richards, S.M., Mitchell, R. J., Kowal, E., Stephen, J.R., Williams, L., Haak, W., Cooper, A., 2017. Aboriginal mitogenomes reveal 50,000 years of regionalism in Australia. Nature 544, 180–184. https://doi.org/10.1038/nature21416.

Triantafilis, J., McBratney, A.B., 1993. Application of Continuous Methods of Soil Classification and Land Suitability Assessment in the Lower Namoi Valley. Canberra, ACT: CSIRO Division of Soils. https://doi.org/10.25919/5c9522aec8bad.

Tugel, A.J., Herrick, J.E., Brown, J.R., Mausbach, M.J., Puckett, W., Hipple, K., 2005. Soil change, soil survey, and natural resources decision making. Soil Sci. Soc. Am. J. 69, 738–747. https://doi.org/10.2136/sssaj2004.0163.

Ward, W.T., 1999. Soils and landscapes near Narrabri and Edgeroi, NSW, with data analysis and using fuzzy k-means. CSIRO Land and Water Technical Report No.:22/99. http://hdl.handle.net/102.100.100/213385?index=1.

Wicklin, R., 2012. What is mahalanobis distance? https://blogs.sas.com/content/iml/2012/02/15/what-is-mahalanobis-distance.html (accesed 4 March 2020).

Wilford, J., 2012. A weathering intensity index for the Australian continent using airborne gamma-ray spectrometry and digital terrain analysis. Geoderma 183, 124–142. https://doi.org/10.1016/j.geoderma.2010.12.022.

Wilson, B.R., Koen, T.B., Barnes, P., Ghosh, S., King, D., 2011. Soil carbon and related soil properties along a soil type and land-use intensity gradient, New South Wales, Australia. Soil Use Manag. 27 (4), 437–447. https://doi.org/10.1111/j.1475-2743.2011.00357.x.

Williams, K.J., Belbin, L., Austin, M.P., Stein, J.L., Ferrier, S., 2012. Which environmental variables should I use in my biodiversity model? Int. J. Geogr. Inf. Sci. 26, 2009–2047. https://doi.org/10.1080/13658816.2012.698015.

Wilson, J.P., Gallant, J.C., 2000. Secondary topographic attributes. In: Wilson, J.P., Gallant, J.C. (Eds.), Terrain Analysis: Principles and Applications. John Wiley & Sons, New York, pp. 87–131.

Xu, T., Hutchinson, M.F., 2011. ANUCLIM Version 6.1 User Guide. Fenner School of Environment and Society, The Australian National University.

Yaalon, D.H., Yaron, B., 1966. Framework for man-made soil changes - an outline of metapedogenesis. Soil Sci. 102 (4), 272–277.

Yang, R.M., Minasny, B., Ma, Y.X., Field, D., McBratney, A., Wu, C.F., 2018. A preliminary soil security assessment of agricultural land in middle-eastern China. Soil Use Manag. 34 (4), 584–596. https://doi.org/10.1111/sum.12463.

Yates, C.J., Norton, D.A., Hobbs, R.J., 2000. Grazing effects on plant cover, soil and microclimate in fragmented woodlands in south-western Australia: implications for restoration. Austral Ecol. 25 (1), 36–47. https://doi.org/10.1046/j.1442-9993.2000.01030.x.

Zuur, A.F., Ieno, E.N., Smith, G.M., 2007. Principal Component Analysis and Redundancy analysis, in: Analysing Ecological Data. Statistics for Biology and Health. Springer, New York, NY. https://doi.org/10.1007/978-0-387-45972-1_12.

Zuur, A.F., Ieno, E.N., Walker, N., Saveliev, A.A., Smith, G.M., 2009. Dealing With Heterogeneity, in: Mixed Effects Models and Extensions in Ecology with R. Statistics for Biology and Health. Springer, New York, NY. https://doi.org/10.1007/978-0-387-87458-64.