Forest Monitoring and Improvement Program

Project 4: Baselines, drivers and trends in soil health and stability

DPIE (Science) and University of Sydney, 30 September 2020

Jonathan Gray[1], Brian Jenkins[1], Thomas Bishop[2], Sabastine Ugbaje[3]

[1] NSW Department of Planning, Industry and Environment, Division of Science, Economics and Insights

[2] School of Life & Environmental Science, The University of Sydney

[3] Sydney Informatics Hub, The University of Sydney

Update 1: Description of overall methodology including modelling approaches

1.1 Overview

The overall strategy being adopted by the joint project team is to apply differing but complementary assessment approaches to derive results and insights into soil condition baseline levels, their driving influences and their trends.

The assessment approaches range from empirical to machine learning:

(i)   purely empirical approach, based on existing soil sampling and survey data, eg, NSW soil landscape mapping program and the NSW 2008-09 Monitoring Evaluation and Reporting (MER)

(ii)   a pragmatic but robust digital soil modelling and mapping (DSMM) approach, mainly employing multiple linear regression (MLR) but possibly random forest (RF) techniques;

(iii)   a data cube approach that combines machine learning techniques with covariates that vary in space, time and space and time to allow predictions through space and time of soil condition.

We will assess the results and similarities between the different approaches on an ongoing basis during their implementation and upon completion to derive final conclusions and recommendations. Together, the results from the different approaches will provide an indication of levels of consistency and uncertainty of the final combined products.

We will attempt to derive models and maps of soil condition relating to individual indicators (properties) for past, present and future temporal frames. We then combine these indicators into a single index of overall soil condition. The proposed key indicators to be assessed include: soil organic carbon, pH (acidity/alkalinity), soil nutrients (macro and micro), soil erosion (hillslope water) and soil structure/compaction. Other indicators such as soil biodiversity may also be explored. These indicators are consistent with the Montreal Process.

For combining the different indicators into a single condition index, we will either use an existing scheme or develop a new scheme suitable for this program based on our own experience and knowledge. Several possible schemes were outlined in *Soil health and stability monitoring in forests, Background and review of potential indicators* (Soil and Landscape Assessment Team, 2020, advanced

draft as submitted to NRC, 21.9.2020). These include the 2008-09 MER scheme (Chapman et al. 2011, OEH 2014), and those used for North American forest soil health (Amacher *et al.* 2007, Lawrence *et al.* 2016).

Soil health has been defined in several ways. We will attempt to address soil health in terms of two broad concepts:

(i)   condition relative to natural (pre-European) state; the greater the difference from natural soil condition the poorer the soil condition

(ii)   production potential – or soil quality, considering soil fertility and constraints (not climate), the greater the productive potential of the soil, the better the soil condition.

We propose to assess soil health according to both concepts for the current time frame (approx. 2010 to 2020), a pre-disturbance time frame (ie, entirely natural) and future time frames (under different climate, land management and fire scenarios).

1.2 Empirical approach

This approach will use existing soil-landscape units, with breakdown into smaller, more pedo-physiographically uniform units or "facets" for pilot projects in certain regions.

Existing data will be applied to all units, recording exact location of points, type of survey (eg, MER, soil-landscape mapping) and soil field and laboratory data.

Conclusions on baseline condition and trends will be derived from mainly qualitative or semi-quantitative analysis of the available data. The process will assist in clearly identifying new sites for further sampling in future programs (eg, revisiting of MER sites or poorly represented soil landscape units and facets)

Key steps

- Compile best soil-landscape mapping data over the entire subject NRC forest area
- Spatially identify known sample points, including recording of sampling quality, eg, full lab or only field data), identify key gaps in data
- Within pilot project areas, divide into facets, based on soil landscape data and using topographic modelling. Such spatial division could follow techniques developed in the 2007 Coastal Comprehensive Assessment (CCA) program (Yang *et al.* 2008), the 2018 ASRIS program using topographic wetness index (Gray and Young, unpublished) or recent assessment programs using Landform 7.
- Prepare dataset over each facet, including, where available; information on soils landscape, forest type, known management history, bushfire and hazard reduction history
- Undertake an assessment of soil health over each facet. This will involve identifying and clearly listing the key indicators for soil health. Apply the agreed assessment scheme, eg, those used in the 2008-09 MER program (Figure 1a & b). Apply in respect to:
  - condition relative to natural state: scheme to broadly involve identifying conditions in completely undisturbed equivalent soil facets, then assess difference in condition to other physically near-equivalent facets
  - soil-landscape constraints for forest productivity: to broadly assess soil and land constraints, including physical and chemical fertility indicators. It will not consider

climate directly (a major driver of forest productivity, but possibly as an influence in soil condition)

- Draw conclusions on temporal trends in individual facets where possible, eg,
  - where differences occur in physically similar units under different forest management,
  - where repeat soil sampling has occurred.
- Attempt to quantify changes and trends semi-quantitatively where possible, but some analysis may be qualitative only

%OC	Plains GSG Group			Slopes GSG Group			Tablelands GSG Group			Tablelands high altitude GSG Group			Coastal GSG Group			Organic soils GSG Group
	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	D
>16.0	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
12.0 to 16.0	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	4
8.0 to 12.0	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	3
6.0 to 8.0	5	5	5	5	5	5	5	5	5	4	5	5	4	5	5	2
5.0 to 6.0	5	5	5	5	5	5	5	5	5	3	4	5	3	4	5	2
4.0 to 5.0	5	5	5	5	5	5	5	5	5	3	3	4	3	3	4	2
3.0 to 4.0	5	5	5	5	5	5	5	5	5	3	3	3	2	3	3	1
2.6 to 3.0	5	5	5	5	5	5	4	5	5	2	3	3	2	2	3	1
2.0 to 2.6	5	5	5	4	5	5	3	4	5	2	2	3	2	2	3	1
1.6 to 2.0	4	5	5	3	4	5	3	3	4	2	2	2	1	2	2	1
1.2 to 1.6	3	4	5	3	3	4	2	3	3	2	2	2	1	2	2	1
1.0 to 1.2	3	3	4	2	3	3	2	2	3	2	2	2	1	2	2	1
0.8 to 1.0	2	2	3	2	2	2	1	2	2	1	2	2	1	1	2	1
0.6 to 0.8	2	2	2	1	2	2	1	1	2	1	1	2	1	1	2	1
<0.6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Figure 1a: Rating table from MER program: Allocation of soil carbon levels into soil condition classes

Test	Value	Surface soil type (Lawrie et al. 2002)							
		Loose sand	Fragile coarse textured	Fragile medium textured	Sodic surface	Coarse structured clay surface soils	Friable soil surfaces	Self-mulching clays	Highly organic surface
Drop shatter test (from VSA)	>1.5	5	5	5	4	4	5	5	5
	1.0 to 1.5	4	3	3	3	3	3	3	4
	0.5 to 1.0	3	3	2	2	2	3	3	3
	<0.5	3	3	2	1	1	3	3	3
Soil porosity (from VSA)	>1.5	5	5	5	4	4	5	5	5
	1.0 to 1.5	4	3	3	3	3	3	3	4
	0.5 to 1.0	3	3	2	2	2	3	3	3
	<0.5	3	3	2	1	1	3	3	3
Bulk density (0–10cm)	<1.2	5	5	5	4	4	5	5	n/a
	1.2 to 1.4	4	4	4	4	4	5	5	n/a
	1.4 to 1.6	3	3	3	3	3	4	4	n/a
	1.6 to 1.8	2	2	2	2	2	2	2	n/a
	>1.8	2	2	1	1	1	1	1	n/a
Soil carbon (carbon % 0 –10 cm)	>3.10	5	5	5	4	4	5	5	n/a
	1.72 to 3.10	4	4	4	3	3	4	4	n/a
	1.00 to 1.72	4	3	3	2	2	4	4	n/a
	0.70 to 1.00	3	3	2	2	2	3	3	n/a
	<0.70	3	2	2	2	2	3	3	n/a

Figure 1b: Rating table from MER program: Soil structure condition class based on soil test values and surface soil type

1.3 Pragmatic DSMM approach

This approach will employ digital soil modelling and mapping (DSMM) techniques that are designed to be "transparent" and relatively easily understood by many potential users of the results, but also robust and credible.

Modelling techniques will include multiple linear regression (MLR) and random forest (RF) techniques. Bootstrapping processes involving up to 100 iterations of maps would be applied. These allow the production of digital maps, typically to 100 m resolution, presenting current levels of the various soil indicators: organic carbon, pH, macro nutrients, total P, EC etc. These would be similar to those presented in OEH (2018) and available on SEED. Assessment of soil erosion will be modelled based on techniques developed in Yang (2020).

Substitution modelling techniques will be employed. These involve the development of models then the substitution of key component layers to represent changed conditions. Similar approaches have been termed "space for time" modelling. The changes proposed for examination include:

- current climate with future climate layers (derived from NARCliM program), as demonstrated in Gray and Bishop (2019)
- current fractional vegetation cover layer with lower vegetation cover layer, as may be indicative of higher forest clearing levels, eg, 2001-2019 average reduced by 10%. A similar approach is being adopted by Gray *et al.* (in prep) for soil carbon sequestration with additional vegetation cover (Figure 2)
- a proposed forest disturbance index (FDI, ie, 1: no forest disturbance to 4: high forest disturbance), apply with a higher disturbance rating, eg, FDI 1 changed to FDI 2
- current bushfire regime to be substituted with a higher intensity, more frequent bushfire regime. Also include changes to hazard reduction regime

From resulting maps, derive boxplot results for different environmental conditions, using either three, four of five variables. A similar approach is presented in Figure 3 for potential carbon sequestration (Gray *et al.* in prep). Results could be based on the following division of classes:

(i)   soil type (eg, 5 fertility class),
(ii)   climate (eg, 4 rainfall or rain/temp classes),
(iii)   vegetation cover/forest type (eg, 5 classes, tall closed forest, woodland, etc),
(iv)   level of logging disturbance (low, mod, high),
(v)   fire history

Key steps

- produce digital soil maps of values of key indicators properties under current conditions
- derive preliminary estimate of condition (forest productivity) of each indicator
- combine indicators into a single condition rating for forest productivity
- apply substitution process with DSMM to determine values of indicators in past and future under different environmental/land management conditions
- by comparing with natural (pre-European) conditions, derive estimate of condition relative to natural for each indicator. Combine these into a single condition rating
- use results to determine potential quantitative estimates of change, use plots such as box plots to demonstrate the change and trends (eg, similar to Figure 3).
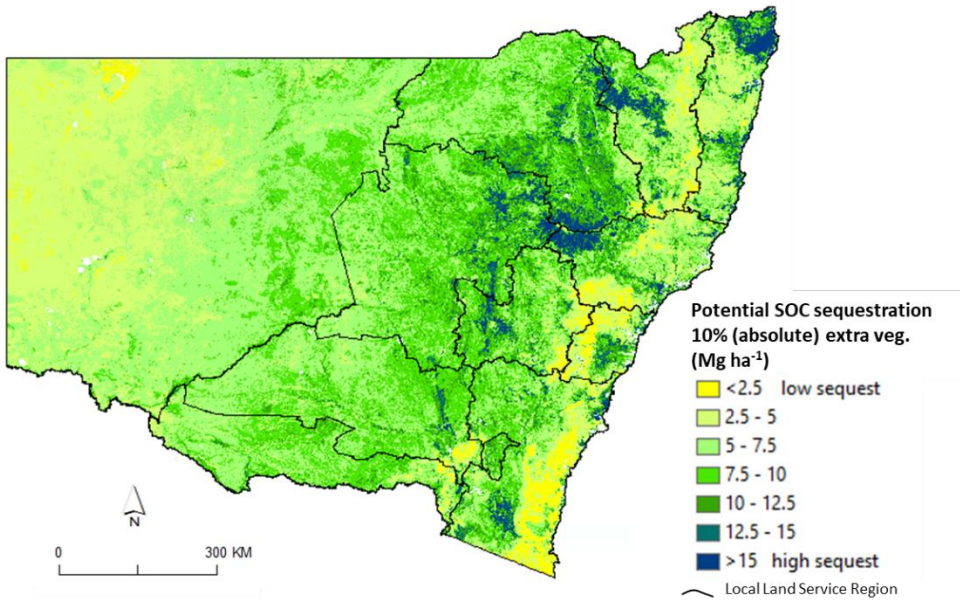
Figure 2: State-wide potential SOC sequestration with 10% absolute extra vegetation cover (Gray *et al.* in prep)
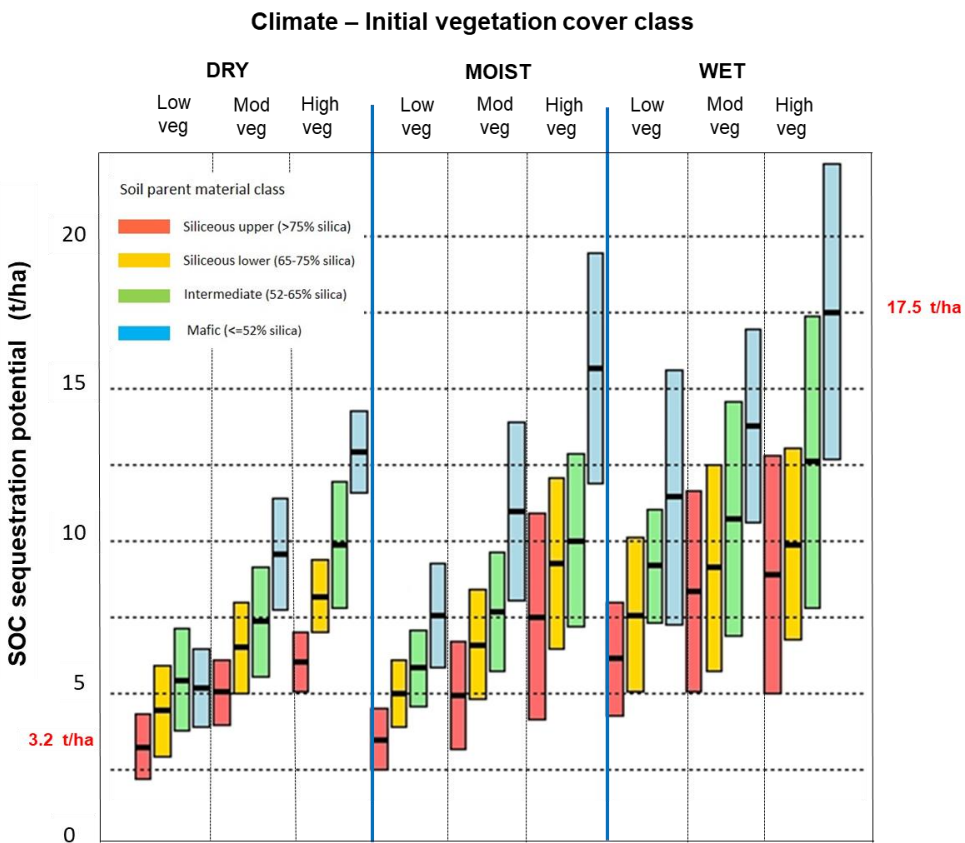

Figure 3: SOC sequestration potential (with 10% absolute vegetation cover increase) by enviro class (0-30 cm, Mg ha^{-1}, mean and upper & lower 90% confidence intervals) (Gray *et al.* (in prep)

1.4 Data cube approach

The focus of this part of the project will be to build a model to predict soil health and stability in space and time. Soil organic carbon has been described as the cement that binds soil particles together and stores the reservoir of soil biology that contribute to soil health. Due to its importance and expected abundance of observations the main focus for this approach will be soil organic carbon. The target will be predictions of the top 30 cm of the soil profile due to this being the layer used in the national carbon account system (NCAS) to simulate changes in SOC (Richards and Evans, 2004) and being most impacted by disturbances such as fire and erosion (Heath et al, 2009).

Pedo-transfer functions related to soil erodibility (the K factor in the Revised Universal Soil Loss Equation) and bulk density (compaction) (Tranter et al., 2007) which use carbon and particle size fractions as inputs will be used to assess erosion and compaction in a more direct manner. Similarly to soil carbon, particle size fraction observations will be more abundant in historical databases than measures of soil stability such as bulk density and strength. The use of PTFs allows us to leverage off the more abundant particle size fraction and carbon to predict trajectories in space and time of soil stability as represented by compaction and erodibility.

A data cube approach is illustrated in Figure 4. This involves developing a data cube of potential predictors that vary in space, time and space and time. This approach has only been possible in the past decade with free and open access to continental scale datasets for terrain (Grundy et al. 2015) and remote sensing platforms (e.g. Landsat, MODIS, Sentinel) providing weekly images across the world (Lewis, Oliver et al. 2017). Using soil organic carbon as an example each measurement would have a spatial location and date of sampling. From this we could estimate the values for a range of covariates such as weather, terrain attributes and remote sensing for the same location and time when the sample was taken.
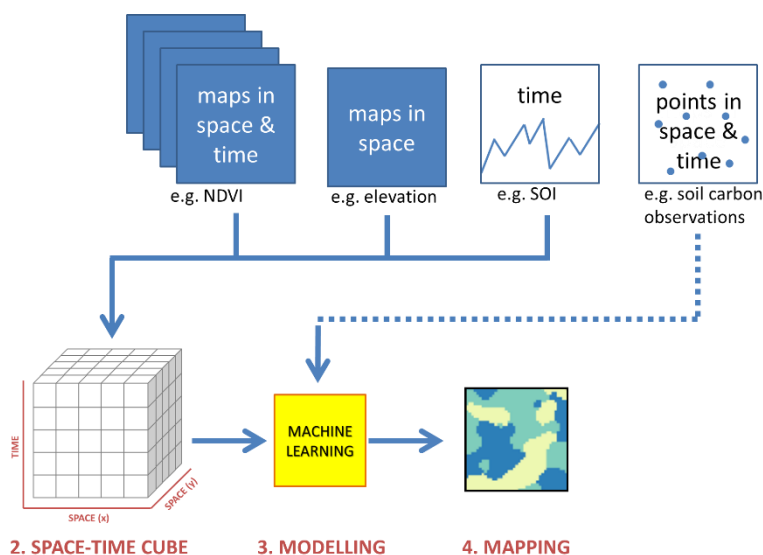

Figure 4. Conceptual framework for predicting soil health and stability

In the case of covariates that vary in time we would use a representation of current and past conditions. For example, mean monthly rainfall in the 3 years prior to sampling. This is because soil organic carbon is a function of current and past conditions. A data scientist with no domain knowledge in soil could naively put the whole time series of rainfall in the model. Feature extraction as

represented by this simple example of taking the mean value over the last 3 years has been shown to create more interpretable models. A machine learning method is then used to build a model to predict soil organic carbon based on the data cube of predictor variables. Members of the team have successfully applied this approach to soil moisture (Wimalathunge and Bishop 2019) and grain yield (Filippi, Jones et al. 2019).

We would apply the same approach to model soil organic carbon, bulk density and soil erodibility subject to data availability. Essential for this approach to work is sufficient data in space and time for organic carbon.

The approach will be implemented with Random Forest and Cubist models

The advantage of the data-driven approach is that disturbances such as road networks, fires and logging can be represented in the data cube. For example, distance to nearest road could be used as a predictor. In the case of fire/logging we could also add in a covariate related to time since the last disturbance event and a severity measure based on a change in a band ratio such as normalised difference vegetation index (NDVI) before and after the event.

Once a model is built we can use it to:

- understand key drivers of soil health and stability across jurisdictions and also for localised regions using interpretive machine learning approaches which would tell us for a particular location what is causing an indicator to increase or decrease (Biecek 2018)

- map in space and time for the whole record where we have soil observations and values for the temporal predictors in the data cube. However, care has to be taken when extrapolating to outside the bounds of where we have soil observations in the data cube attribute space

To address the issue of extrapolation and applicability of the models we will implement the approach described by Meyer and Pebesma (2020). Their method delineates the area of applicability (AOA) of a model based on the dissimilarity index (DI). This will show where in the data cube space (space-time) that we can make reliable predictions given our soil observations.

The output of our data-driven approach will be maps through time of indicators at a spatial resolution of 30 m on a monthly time step for the period 1984-2020. The time period coincides with the record of Landsat imagery on Google Earth Engine that is available as surface reflectance. Satellite imagery will be the most restrictive part of the data cube in terms of access to historical data and therefore be limit on how far back we can hindcast. The 30 m resolution matches the Landsat resolution and the monthly time step is based on the rate of change in the key indicator for this study of soil carbon.

The maps will have an associated 95% prediction interval allowing for assessment of the significant of the trend on a pixel-basis by seeing if the prediction intervals overlap between time periods.

As described earlier we can use interpretive machine learning methods to identify the dominant drivers of the change based on the variables from the data cube that are selected in the model and the magnitude of their effect on the predictions. This can be explored at different spatial scales from across all jurisdictions, to individual forest estates down to the pixel level.

Update 2 - Review and collation of existing datasets related to soil health & stability, identify key gaps, including aerial gaps

2.1 Soil point datasets

- NSW Soil and Land Information System (SALIS); comprises soil observations collected over several decades; includes detailed descriptions of site and soil profiles, including observed land degradation. Coverage over NSW RFA forested areas includes a total of approximately 6470 profile points with field data, of which approximately 1630 contain at least some laboratory data on key physical and chemical soil properties. (https://www.environment.nsw.gov.au/topics/land-and-soil/information/salis)

- 2008-09 NSW Monitoring Evaluation and Reporting (MER) program datasets: comprises extensive field and laboratory soil data over forest and woodland areas; over 170 points across NSW with approximately 50 points in RFA regions of all tenures (including unprotected freehold) and 30 points in specific tenures the subject of the current proposal. Laboratory data includes total carbon, carbon fractions (particulate, humus and resistant forms of carbon), bulk density and other properties

- Soil datasets from a decade long University of Sydney bushfire research program led by Dr Bell which include sampling before and after bushfires and prescribed burns

2.2 Soil spatial datasets

- Polygonal soil map layers: these provide detailed descriptions and data on soil and landscape character. Extensive coverage over NSW forest areas are available through the Soil -landscape mapping series, typically at 1:100 000 scale but some at 1:250 000 scale. Several areas are covered by Soil and Land Resource Series mapping, typically at 1:250 000 or coarser (https://www.environment.nsw.gov.au/topics/land-and-soil/information/soil-maps)

- Digital soil maps: raster layers prepared through quantitative modelling techniques; includes a suite of physical and chemical soil properties across the State with resolution of 100 m and multiple standard depth intervals down to 2 m (https://datasets.seed.nsw.gov.au/dataset/digital-soil-maps-for-key-soil-properties-over-nsw)

- Modelled hillslope erosion layers, with monthly coverage back to January 2000 (https://datasets.seed.nsw.gov.au/dataset/modelled-hillslope-erosion-over-new-south-wales

- Soil change with climate change to approximately 2070 - from NARCliM program; includes high resolution maps of projected hillslope erosion and change in organic carbon, pH and macro-nutrients (https://climatechange.environment.nsw.gov.au/Impacts-of-climate-change/Soil)

2.3 Other key environmental datasets

- Fire extent and severity maps, from DPIE https://datasets.seed.nsw.gov.au/dataset/google-earth-engine-burnt-area-map-geebam
- High resolution DEMs (LiDAR 1-10 m)
- High resolution images (Sentinel, Landsat, ADS40/80). Landsat and Sentinel will be accessed from Google Earth Engine.
- Land use data: polygonal layers including 2007, 2017 releases, from DPIE: https://datasets.seed.nsw.gov.au/dataset/nsw-landuse-2013
- MODIS fractional vegetation cover data, from CSIRO http://www.auscover.org.au/datasets/fractional-cover-modis/

- Climate and weather data, derived from BoM http://www.bom.gov.au/climate/data/index.shtml?bookmark=201
- Climate projection data from NARCliM program, from DPIE (https://climatechange.environment.nsw.gov.au/

2.4 Example datasets for data cube

Table 1 gives some examples of the types of covariates that will be used in the data cube approach. These are in addition to those described in sections 2.1-2.4.

Table 1 Covariates

	Covariate	Source	Resolution	Characteristics
Spatial	DEM, slope	Geoscience Australia	30 m	Topographically controlled effects
	Land use	MODIS	500 m, annual	Land management
	Topographic Wetness Index (TWI)	ASRIS	90 m	Topographic control on hydrological processes
	Clay % (0-30,30-100 cm)	SLGA	90 m	Water holding capacity
Spatial & temporal	Evapotranspiration(ET)	MODIS	500 m, 8 day total	Plant water use
	Enhanced Vegetation Index (EVI)	Landsat/Sentinel	10 – 30 m	Seasonal vegetation greenness
	Precipitation(P)	BOM	5 km, daily	Relates to soil water content
	Temperature (min, max & average)	SILO	5 km, daily	Temperature difference effects which relate to ET
	Solar radiation	SILO	5 km, daily	Relates to evaporation

2.5 Aerial gaps in soil observations

Figures 5-7 present data for locations where surface organic carbon has been measured and are based on the combination of the MER and SALIS dataset held by DPI-E. Figures 6 and 7 present their spatial distribution for NSW and the RFA regions for different time slices. Figure 8 presents the same data but as number of observations per RFA region with different time slices.
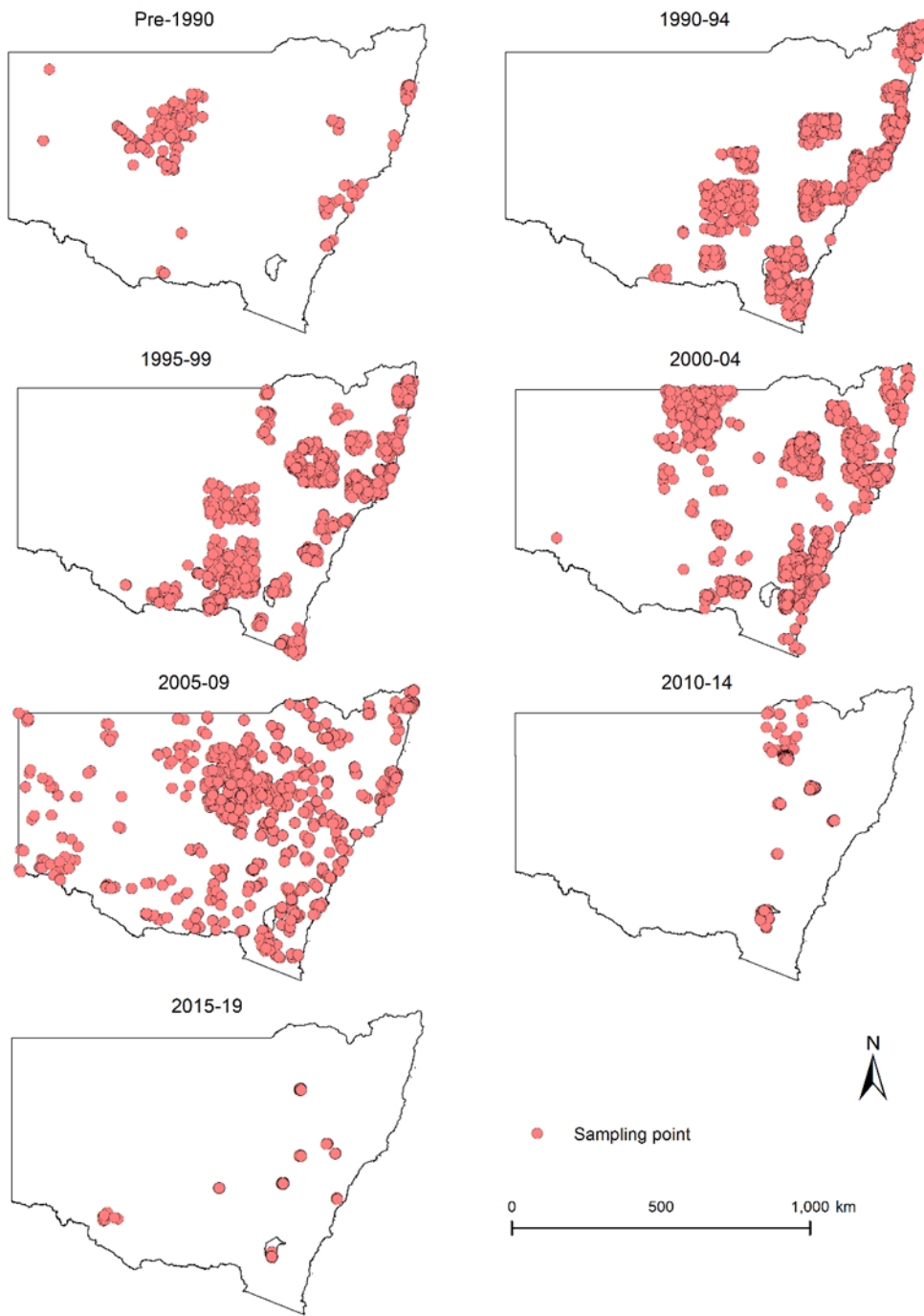
Figure 5. Spatial distribution of surface organic carbon measurements across NSW
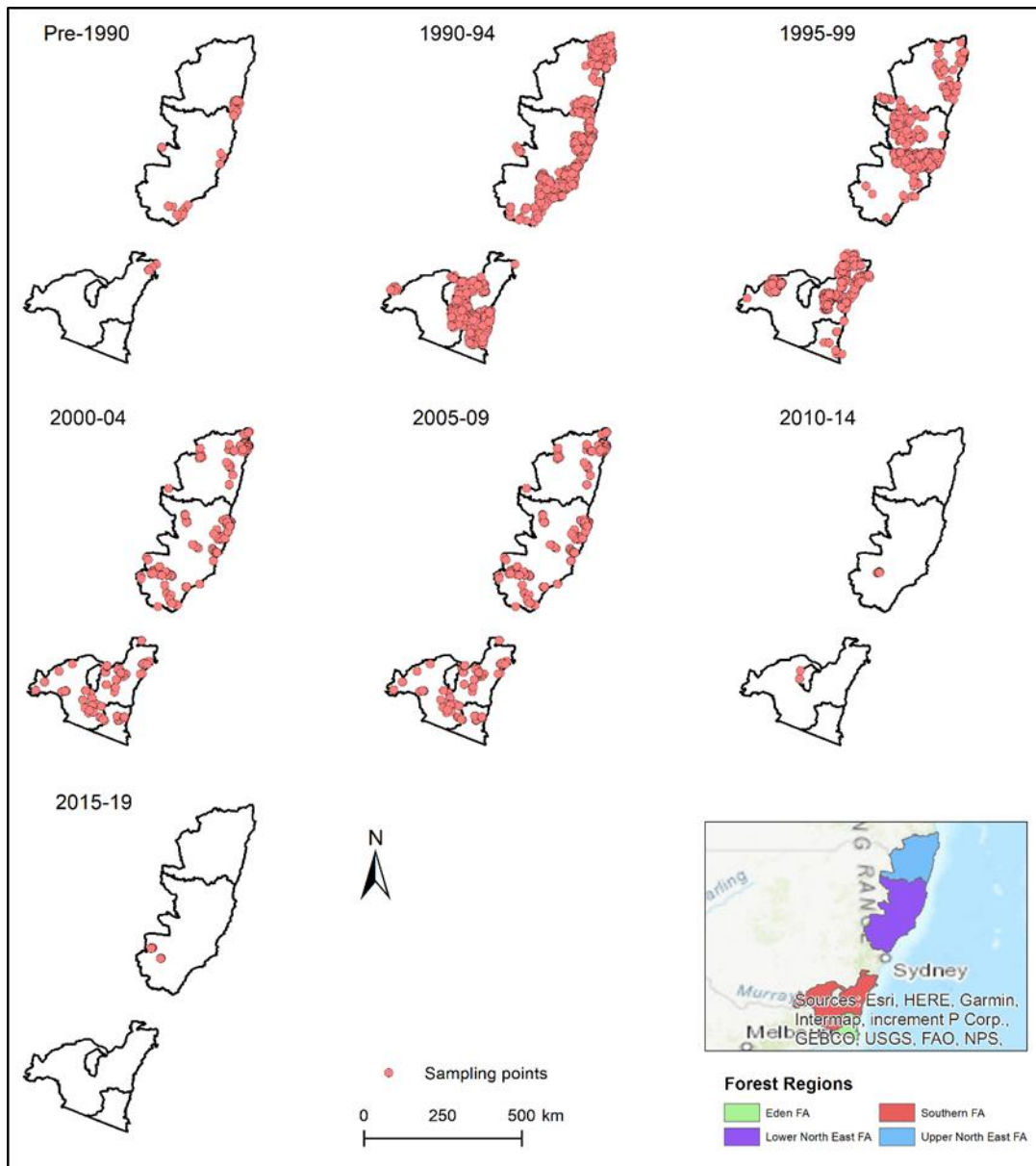
Figure 6. Spatial distribution of surface organic carbon measurements across RFA regions
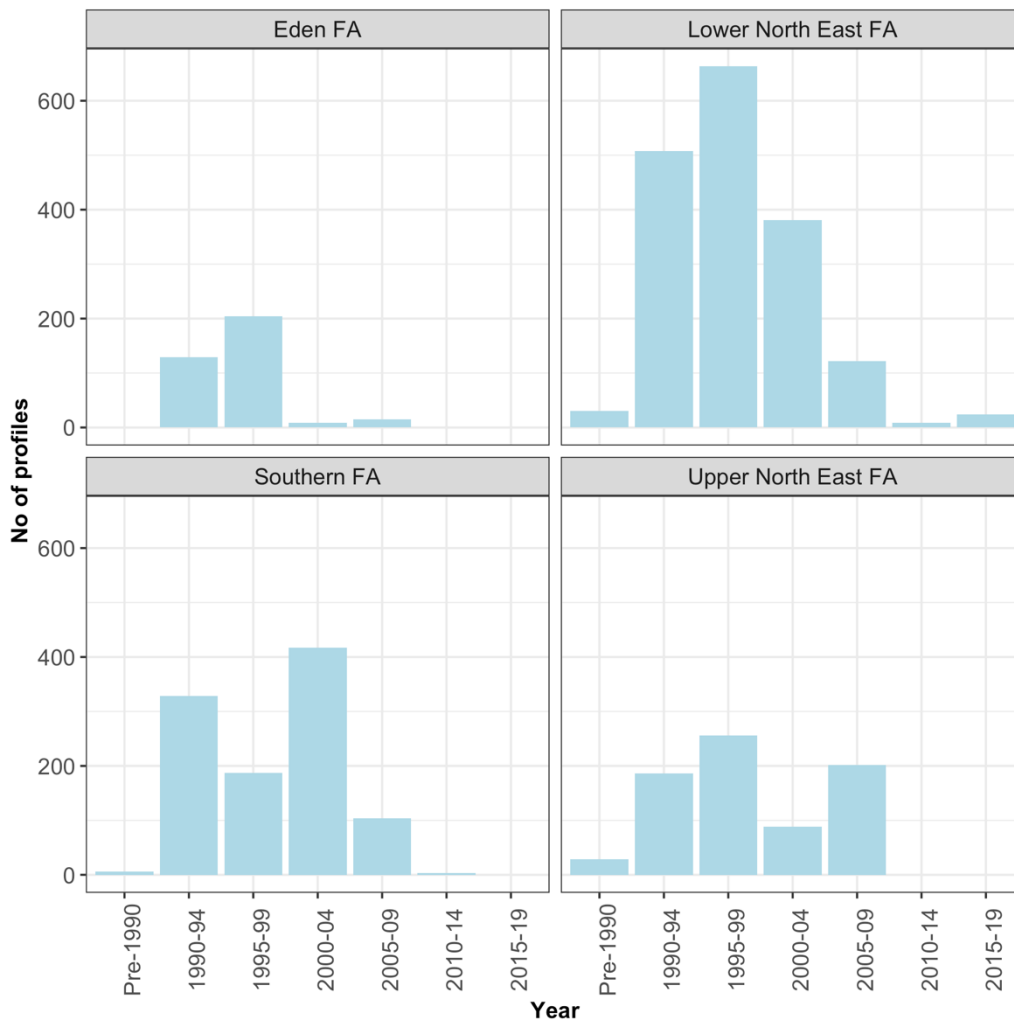
Figure 7. Number of surface organic carbon measurements across RFA regions

In broad terms the number and distribution of soil carbon measurements for each time slice is promising in terms of applying the data cube approach. An obvious issue is the lack of data for 2010 onwards but if the pre-2010 vegetation and weather variation represents the 2010+ period then extrapolation in time may not be such an issue. The area of applicability approach described earlier will be used to assess the level of extrapolation.

References

Amacher MC, O'Neil KP and Perry CH (2007). *Soil vital signs: A new Soil Quality Index (SQI) for assessing forest soil health*. Research Paper RMRS-RP-65WWW. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station

Chapman GA, Gray JM, Murphy BW, Atkinson G., Leys JF, Muller R, Peasley B, Wilson BR, Bowman G, McInnes-Clarke SK, Tulau MJ, Morand DT and Yang X (2011). *Assessing the condition of soils in NSW, Monitoring, evaluation and reporting program*, Technical report series, Office of Environment and Heritage, NSW Department of Premier and Cabinet, Sydney. http://www.environment.nsw.gov.au/ resources/soc/20110718SoilsTRS.pdf

Filippi P, et al. (2019). An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning. *Precision Agriculture,* 20: 1015-1029.

Gray JM, Bishop TFA (2019). Mapping change in key soil properties due to climate change over south-eastern Australia. *Soil Research*, 57, 467-481. https://doi.org/10.1071/SR18139_CO

Gray JM and Young MA, (unpublished). Soil-landscape data for NSW in ASRIS format, as submitted to CSIRO, March 2018 for the Australian Soil Resource Information System

Gray JM, Wang B, Water C, Cowie A, Orgill, S (in prep) Modelling of potential soil carbon sequestration from increasing vegetation cover across New South Wales, Australia. (in preparation, a project under the NSW Primary Industry Productivity and Improvement Program for Net Zero Emissions Program)

Grundy MJ, et al. (2015). Soil and Landscape Grid of Australia. *Soil Research*, 53(8): 835-844.

Heath, J, et al. (2015). Wildfire effects on soil carbon and water repellency under eucalyptus forest in Eastern Australia. *Soil Research*, 53: 13-23.

Lawrence GB, Fernandez IJ, Hazlett PW, Bailey SW, Ross DS, Villars TR, Quintana A, Ouimet R, McHale MR, Johnson CE, Briggs RD, Colter RA, Siemion J, Bartlett OL and Vargas O (2016). Methods of soil resampling to monitor changes in the chemical concentrations of forest soils, *Journal of Visualized Experiments* 117, e54815, doi:10.3791/54815.

Lewis A, et al. (2017). The Australian Geoscience Data Cube — Foundations and lessons learned. *Remote Sensing of Environment*, 202, 276-292.

Meyer H and Pebesma E, (2020). Predicting into unknown space? Estimating the area of applicability of spatial prediction models. https://arxiv.org/abs/2005.07939.

OEH (2014). *Soil condition and land management in NSW: final results from the 2008–09 monitoring evaluation and reporting program*. Technical Report. NSW Office of Environment and Heritage, Sydney. Available at: http://www.environment.nsw.gov.au/Research-and-publications/Publications-search/Soil-condition-and-land-management-in-New-South-Wales

OEH (2018). *Digital soil mapping of key soil properties over NSW, version 1.2.* Technical Report, NSW Office of Environment and Heritage, Sydney (prepared by J. Gray). Report and maps accessible on SEED: https://datasets.seed.nsw.gov.au/dataset/digital-soil-maps-for-key-soil-properties-over-nsw or via eSPADE: http://espade.environment.nsw.gov.au

Richards GP and Evans D, (2004) Development of a carbon accounting model (FullCAM Vers. 1.0) for the Australian continent, *Australian Forestry*, 67:4, 277-283.

Soil and Landscape Assessment Team (2020). *Soil health and stability monitoring in forests, Background and review of potential indicators,* Department of Planning, Industry and Environment, Sydney (advanced draft as submitted to NRC, 21.9.2020).

Tranter G, Minasny B, McBratney A, Murphy B, Mckenzie NJ, Grundy M, Brough D, (2007). Building and testing conceptual and empirical models for predicting soil bulk density. Soil Use and Management, 23(4), 437–443.

Wimalathunge, N and Bishop, TFA (2019). A space-time observation system for soil moisture in agricultural landscapes. *Geoderma*, 314, 1-13.

Yang X, Gray JM, Chapman GA, Young MA, (2008). Soil landscape constraint mapping for coastal land use planning using geographic information system. *Journal of Coastal Conservation,* 11, 143-151.

Yang X (2020). State and trends of hillslope erosion across New South Wales, Australia, *Catena* 186 (104361), 1-9.