final report

Project code: B.GBP.0043

Prepared by: Peter Zund, Jon Walton, Max De Antoni, Ben Harms and Evan Thomas
Queensland Department of Environment and Science

Date published: 18 February 2022

PUBLISHED BY
Meat and Livestock Australia Limited
PO Box 1961
NORTH SYDNEY NSW 2059

Soil bicarbonate-extractable P (Colwell-P) map of Queensland main grazing lands

Meat & Livestock Australia acknowledges the matching funds provided by the Australian Government to support the research and development detailed in this publication.

Abstract

Pasture agronomists in Queensland have highlighted that bicarbonate-extractable soil phosphorus (P) maps produced for Queensland are not of a suitable scale and needed to be updated to a finer resolution. Moreover, given the low availability of total soil P to plants, current total P maps provide limited information for agricultural purposes. Bicarbonate-extractable P is considered a more reliable test for plant available P. For these reasons, Queensland Department of Agriculture and Fisheries (DAF) staff have reiterated the need for improved soil P maps.

The aim of this project was to use Digital Soil Mapping (DSM) to produce a map of bicarbonate-extractable P levels in surface soils at a one-hectare pixel resolution of Queensland's grazing lands.

The bicarbonate-extractable P map was produced by fitting a Cubist model to data point observations of P that were intersected with environmental covariate data which represent factors of soil formation. This bicarbonate-extractable P map raster map represents the predicted natural (unfertilised) surface soil bicarbonate-extractable P of Queensland grazing land (total area: 1,074,410 km^2) south of the Gulf and Cape York Natural Resource Management Regions. The cell size of the raster map is 90 x 90 m^2, each with a predicted value of P ranging between 1 and 150 mg kg^{-1}.

As part of the DSM process, maps of P prediction uncertainty were also produced to assist in understanding map predictions. Uncertainty was calculated as the difference (range) between the 5^{th} and 95^{th} percentile median predictions, and relative uncertainty was calculated by dividing the range by the predicted median bicarbonate-extractable P value.

The P map is considered a reasonable prediction of P at the subcatchment to property scale. It is useful for distinguishing areas of high P from low P. While the level of uncertainty is higher in areas of more fertile soils (with higher P variability), across the majority of the map, where P levels are lower, predictions are more accurate.

The raster maps can be downloaded from the Queensland Spatial catalogue (QSpatial), with a pdf map available from publications.qld.gov.au.

Executive summary

Background

In recent years pasture agronomists from the Queensland Department of Agriculture and Fisheries (DAF) have highlighted that previous soil P maps produced for Queensland (Ahern *et al.* 1994; McCosker and Winks 1994) were not produced at a suitable scale and needed to be updated to a finer spatial resolution. Moreover, current 'total P' maps provide limited information regarding the availability of soil P to plants and therefore have limited value for agricultural purposes. Bicarbonate-extractable P is considered to be a more reliable test for plant available P. For these reasons, over recent years DAF staff have reiterated the importance of having improved soil P maps.

Objectives

The aim of this project was to produce a bicarbonate-extractable P map (and associated prediction uncertainty maps) of surface soil at a one-hectare pixel resolution (3 arc seconds pixel grid) of Queensland's grazing lands. Stage 3 comprises land south of the Gulf and Cape York Natural Resource Management Regions. This will be available for use in MLA and Queensland government publications and websites.

Another objective of the project was to establish a robust Digital Soil Mapping (DSM) methodology for producing bicarbonate-extractable P and prediction uncertainty maps, and a strategy for subsequent activities that further develop soil P mapping across defined areas for the Northern beef sector.

The project was delivered in stages, covering the following regions: the Fitzroy, Burnett, Southeast Queensland, Darling Downs, Balonne-Maranoa, Burdekin, Herbert, Wet Tropics, Atherton Tablelands, Mitchell grasslands of the Thompson and the Channel Country (total area: 1,074,410 km^2)

Methodology

Fundamentally, DSM uses statistical models derived from relationships between soil data and various spatial environmental covariates related to soil formation to predict soil properties across landscapes. Key advantages of DSM are that (i) its statistical models are quantitatively based, flexible and repeatable, (ii) it can be applied rapidly across large areas, and (iii) estimates of uncertainty can be routinely produced as a guide to the reliability of the DSM outputs.

The bicarbonate-extractable P and uncertainty maps were produced using the Cubist machine learning method/algorithm (Quinlan 1992; Kuhn *et al.* 2012) to build models (i.e. relationships) between soil site point data and the spatial layers of environmental covariates. Based on the rules developed in the models and raster maps of environmental covariates, P levels were predicted across the study area.

Models were validated using the k-fold cross validation method with 10 folds. The Cubist rules were mapped out at each k-fold iteration for the 5th (lower), 50th (median) and 95th (upper) percentile predictions. The final surfaces produced with this approach were determined using the median of the 10 k-folds, with the predicted bicarbonate-extractable P map being the median of the 50th percentile predictions. Prediction uncertainty was calculated as the difference (range) between the 5th and 95th percentile median predictions, and relative uncertainty was calculated by dividing the range by the predicted median bicarbonate-extractable P value.

Key findings

The final product is a raster map of predicted natural (unfertilised) surface soil bicarbonate-extractable P (and associated prediction uncertainty) of Queensland grazing land south of the Gulf and Cape York Natural Resource Management Regions. The modelled bicarbonate-extractable P concentrations were symbolised in six categories using a soil P rating system originally developed by Ahern et al. (Ahern *et al.* 1994). The size of the cells of the raster map is 90 x 90 m^2, each with a predicted value of P ranging between 1 and 150 mg kg^{-1}. The raster map can be downloaded from the Queensland Spatial catalogue (QSpatial), with a pdf map available from publications.qld.gov.au.

The P map is considered a reasonable prediction of P at the subcatchment to property scale. It is useful for distinguishing areas of high P from low P. While the level of uncertainty is higher in areas of more fertile soils (with higher P variability), across the majority of the map, where P levels are lower, predictions are more accurate.

Future research and recommendations

The outcomes of this project emphasise the importance of the availability of improved P maps for the red meat industry. Since the focus of this project was the development of a robust DSM methodology for the creation of bicarbonate-extractable P maps for Queensland's main grazing lands, future R & D efforts have been identified.

Table of contents

1 Background

1.1 The importance of soil P

Plant available soil P stocks are a key indicator of livestock productivity in grazing lands. P is an essential macronutrient for pasture legumes which improve nitrogen availability and hence pasture quality. Ensuring adequate P levels in plant biomass is fundamental for livestock, as it enables satisfactory growth and lactation rates.

Livestock P deficiency is a major limiting factor to productivity and profitability in the northern grazing industry. However, P supplementation for livestock and the application of P fertiliser to soil are major costs for graziers. Improving the ability of graziers to target supplementation and fertilisation can provide significant financial benefits to the beef industry and therefore its productivity and profitability. A key tool for improving P management is a high quality 'plant available phosphorus' map that is useful at the property (sub-catchment) scale.

1.2 The P cycle

P is present in soil in several forms, with complex interactions occurring between them. A basic understanding of the P cycle is essential to understand what this map represents. For plants, P is available when it is dissolved in the soil solution. The primary source of P in the soil solution is from the weathering of rock minerals (e.g. apatite), an extremely slow process. The availability of soil P however is limited by the capacity of the soil to chemically adsorb (fix) P, also known as the soil's P buffering capacity. The P buffering capacity is the ability of the soil solution to resist a change in its P concentration (e.g. removed by plant uptake or added in fertilisers). The P buffering capacity of a soil is linked to its concentration of iron and aluminium oxides since these compounds adsorb P, particularly in acid soils, and make it unavailable to plant roots. The plant-availability of soil P is also reduced by calcium carbonates, which are particularly effective at adsorbing P in alkaline soils. In tropical areas, acidic red soils may show a relatively high bicarbonate-extractable P but behave as if they are P deficient due to their high P buffering capacity. Plant and animal material also contains forms of organic P, which can be mineralised and return to the soil with the help of soil microbes. The remaining P not mineralised, adsorbed or taken up by plant and animals is said to be in the soil solution and available to plant roots.

1.3 Soil P tests

Many laboratory methods have been developed to measure soil P – some with the aim of measuring the total pool of soil P, and others to estimate plant available P.

The tests for 'total P' measure the overall reservoir of P in the soil, both soluble and insoluble, with results expressed as a % by dry weight (method 9A1, Rayment and Lyons 2011). Historically, this test has been used in soil surveys to help characterise the nature of the soil (e.g. the degree of weathering), but has only a limited relationship to the potential availability of P to plants. However, maps of total P have been prepared for Australia at a continental scale (e.g. Viscarra Rossel and Bui, 2016).

In Queensland, plant available soil P has commonly been measured using either acid-extractable P (BSES method) (method 9G2-Rayment and Lyons 2011) or bicarbonate-extractable P (Colwell-P) (method 9B2 - Rayment and Lyons 2011). Results are expressed in mg/kg of dry soil.

Acid extractable P (P-BSES) is an indicator of various sources of P that the plant may access slowly over time and it has historically been used in high rainfall cropping areas such as sugarcane cropping systems. Bicarbonate-extractable P however, is considered to be the most useful single measure of soil P availability in pastures and is also used to determine the quantity of starter P to use in cropping systems. For this reason, the modelling and mapping activities conducted in this project used only data obtained with the bicarbonate-extractable P method.

A bicarbonate-extractable P test is ideally interpreted in conjunction with a P buffering index (PBI) test (method 9I2b - Rayment and Lyons 2011). PBI tests have been less commonly carried out on soil samples analysed for P in Queensland, therefore, this project relies on bicarbonate-extractable P results to estimate plant available P.

1.4 The need for a soil bicarbonate-extractable P map in Queensland

In recent years pasture agronomists from the Queensland Department of Agriculture and Fisheries (DAF) have highlighted that previous P maps produced for Queensland (Ahern *et al.* 1994; McCosker and Winks 1994) were not of a suitable scale and needed to be updated to a finer spatial resolution. Moreover, current 'total P' maps do not provide useful information regarding the availability of soil P to plants and therefore limited information for agricultural purposes. For these reasons, DAF staff have reiterated the importance of having improved soil P maps. Such information would be a key aid to target extension activities aimed at informing graziers regarding the costs and benefits associated with using P fertilisers and/or introducing legumes into Northern grazing systems.

The idea of an updated bicarbonate-extractable P map being produced was discussed at the 2016 Northern Beef Research Update Conference (NBRUC) in Rockhampton. After the conference, a prototype of what could be achieved using new DSM modelling and mapping techniques was produced for the Fitzroy Natural Resource Management (NRM) Region using existing bicarbonate-extractable P data. Even though the input data for the prototype map was not validated, it was found that there was sufficient P data in the study area to allow soil P to be modelled with reasonable accuracy.

After assessing the prototype mapDAF and DES expressed interest in seeing the mapping extended state-wide. DAF agreed to collate as much as possible the P data they held in research projects and Meat and Livestock Australia (MLA), provided funds for the collection of extra soil samples in areas which have insufficient data.

2 Objectives

2.1 Overall aim

The overall aim of this three-year project was to produce a soil bicarbonate-extractable P map (and associated prediction uncertainty maps) of surface soil at a one-hectare pixel resolution (3 arc seconds pixel grid) of Queensland's key grazing regions. Due to time and budget limitations, the production of a state-wide P buffering index (PBI) map was deemed out of the scope of this project. As mentioned elsewhere in this report, it would be highly desirable to produce a PBI spatial layer to be used in conjunction with the bicarbonate-extractable P map to more accurately interpret the available P status of soils across the state.

2.2 Contract objectives

The stated aim was achieved by means of the following objectives:

- Develop a robust methodology for producing a bicarbonate-extractable P and uncertainty maps for the Fitzroy, Burnett, Southeast Queensland, Darling Downs, Balonne-Maranoa, Burdekin, Herbert, Wet Tropics, Atherton Tablelands, Mitchell grasslands of the Thompson and the Channel Country (total area: 1,074,410 km^2).
- Provide a strategy for subsequent activities that further develop soil P mapping across defined areas for the Northern beef sector.

3 Methodology

3.1 Mapping approach

Modelling and mapping activities were conducted in three stages (listed below) to enable the collation of legacy data, and progressive collection and analysis of soil samples in conjunction with other projects.

- Stage 1 (2019): Fitzroy, Burnett, Southeast Queensland, Darling Downs and Balonne-Maranoa regions (517,044 km^2).
- Stage 2 (2020): Burdekin, Herbert, Wet Tropics and Atherton Tablelands regions (306,938 km^2).
- Stage 3 (2021): Mitchell grasslands of the Thompson and Channel Country regions (250,428 km^2).

The output of each stage was a map of predicted bicarbonate-extractable P for surface soil (0 – 0.1 m soil depth). The footprint of Stages 2 and 3 included areas covered in the previous stages. Importantly, the maps produced represent the inherent (unfertilised) or natural background extractable P concentration in soils and not the current P status in cropped or disturbed areas.

The Stage 3 boundary aligns with the northern and western boundaries of the Desert Channels, Burdekin and Wet Tropics Natural Resource Management (NRM) Regions (total area: 1,074,410 km^2).

3.2 Digital soil mapping (DSM) method

3.2.1 Cubist modelling

The bicarbonate-extractable P soil maps were produced using the Cubist machine learning method/algorithm (Quinlan 1992; Kuhn *et al.* 2012) to build models (i.e. relationships) between soil site point data and spatial layers of environmental covariates, which are landscape, soil and climate attributes that represent factors of soil formation. P levels were predicted across the study area based on the Cubist rules in these models and the raster maps of environmental covariates.

The Cubist models were validated using the k-fold cross validation method (Malone, Minasny, & McBratney, 2017). The k-fold approach allows distributions of the validation statistics to be derived as a means of assessing the stability and sensitivity of the models and parameters (Malone, Minasny, & McBratney, 2017). This validation method uses a data sub-setting process, rather than collecting an independent set of observations to test the models. Using the k-fold approach, the P site data was randomly split into 10 groups (folds) of equal size, with nine folds used for model calibration (internal validation) and one fold used for model testing (external validation). This process was repeated until each fold had been used once for external validation (i.e. 10 iterations). The Cubist rules were mapped out at each k-fold iteration for the 5th (lower), 50th (median) and 95th (upper) percentile predictions. The 90% confidence interval was the difference between the lower and upper percentiles. The final map produced for each of these predictions was determined using the median of the ten k-folds.

The final bicarbonate-extractable P map is the 50th percentile prediction (i.e. median). The reliability of the predicted P map refers to the likelihood of the modelled value being equivalent to an actual observation. To display uncertainty, two raster maps were produced. Prediction uncertainty was calculated as the difference (range) between the 5th and 95th percentile median predictions, and relative uncertainty was calculated by dividing the range by the predicted median bicarbonate-extractable P value.

3.2.2 Bicarbonate-extractable P data

Data from a total of 7,014 surface soil (0 – 0.1 m) samples were used to produce the Stage 3 map, equating to an average density of one site per 153 km^2 (Figure 1) within the 1,074,410 km^2 modelling area. All the samples used to calibrate and validate the Cubist model had been analysed for bicarbonate-extractable (Colwell) P (Rayment and Lyons 2011). The vast majority of the soil data used in this project were legacy data, sourced from various government land resource assessment projects undertaken over many decades and stored in the Queensland Soil and Land Information (SALI) database. 54 sites were selected from a dataset collected for a soil organic carbon project on the Darling Downs.

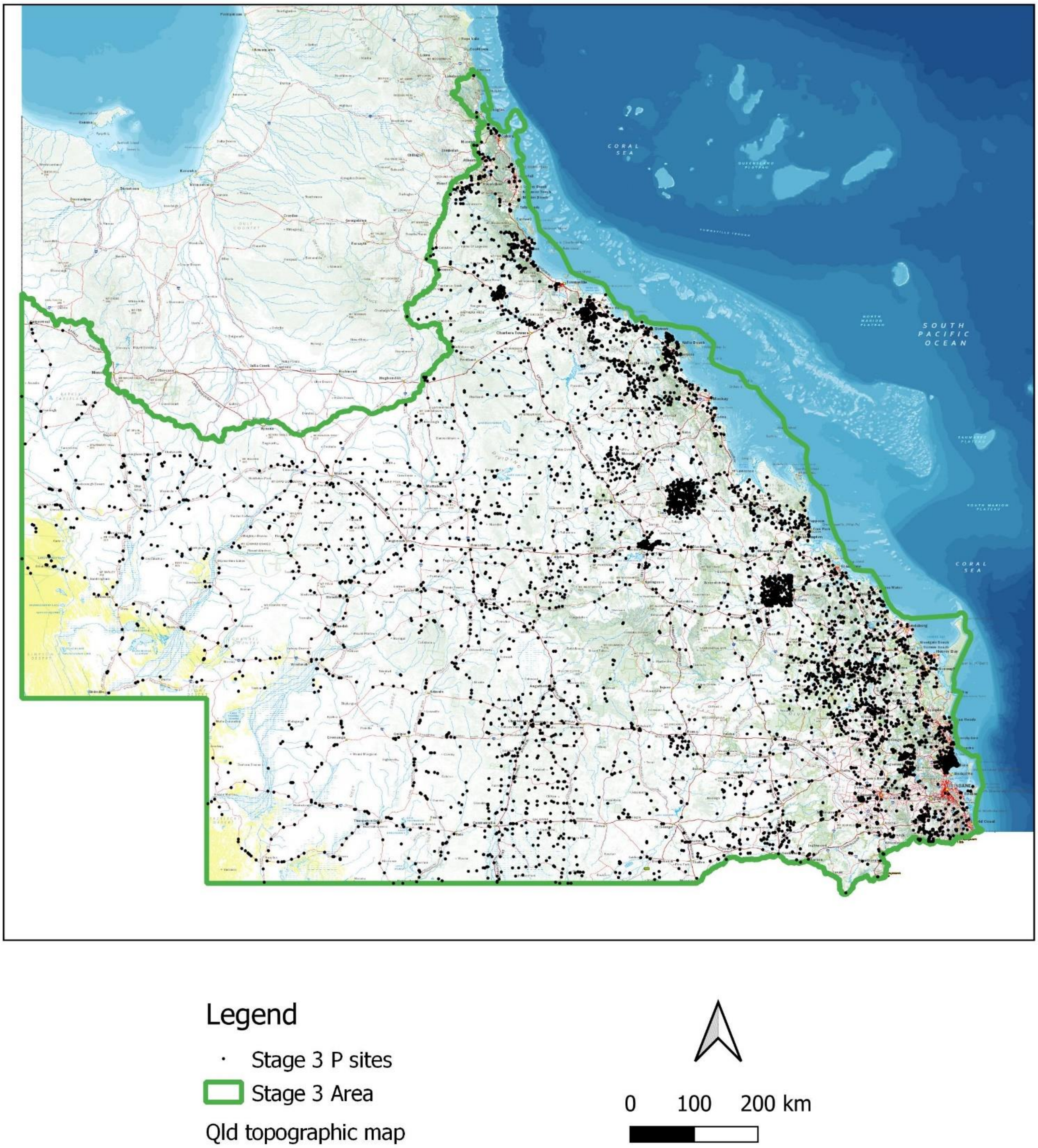

Figure 1 - Distribution of sample sites used for model calibration and validation.

The SALI database was filtered to exclude samples with P levels above a threshold value of 150 mg kg^{-1}, since such values were likely to be affected by fertilisation. Samples were also excluded from the modelling if they were from a SALI site that was known to having been cultivated at some stage or that was highly disturbed (McDonald *et al.* 2009 p. 128) according to the description in the site record. Samples were also excluded if their P concentration was >7 mg kg^{-1} *and* they were within a 75 m radius from a:

- cultivated field
- plantation forest
- improved or exotic pasture
- tree crop
- feedlot
- intensive animal farm
- infrastructure
- mine site
- water storage system or channel.

The exclusion of such sites was conducted using data from Australian Bureau of Agricultural and Resource Economics and Sciences (ABARES 2016) and the current Queensland Government land use map (Queensland Land Use Mapping Program 2019).

Note that the practical quantitation limit (PQL) for the Colwell P method is 2 mg $kg^{-}1$, however data extracted from the SALI database were in numeric value format, allowing results less than 2 mg kg^{-1} to be included.

Prior to modelling, the acquired data was log-transformed to resemble a normal distribution, which enables better performance of the Cubist machine learning algorithm. Validation statistics were calculated on both log-transformed and back-transformed data. As part of map production, the final map predictions were back-transformed to display P levels in mg kg^{-1}.

3.2.3 Environmental covariates

Environmental covariates are landscape, soil and climate factors prepared as raster spatial layers. In DSM, they are intended to represent factors of soil formation (SCORPAN factors), and they are used to interpolate the target soil variable based on models that include the covariates and the soil sample point data. The SCORPAN model (McBratney *et al.* 2003) represents soil (as either soil classes, Sc, or soil attributes, Sa) at a point in space and time as an empirical quantitative function of seven environmental covariates: soil (s), climate (c), organisms (o), relief (r), parent material (p), age (a), and spatial location (n):

$$S_{c,a} = f\,(s, c, o, r, p, a, n)$$

Covariates are typically derived from satellite, radar, geophysical and climate data. The majority of covariates used in this project were selected from a pool of covariates from the Terrestrial Ecosystems Research Network (TERN) collection. The selection was initially conducted using a Random Forest model that was fitted to 55 covariates and 3,058 surface soil sample data points. The 55 covariates were ordered by their importance to the Random Forest model and assessed using the correlation function in the R Statistics package to group them based on level of autocorrelation.

The first selection identified 39 covariates based on variable importance and variable correlation information. Further covariates were deleted because they had a coarse spatial resolution (e.g. 1-5 km^2 cell size) or had artefacts (patterns that do not appear to follow natural landscape patterns) that would have negatively impacted the quality of the final map prediction. Covariates which had patterns related to vegetation clearing were removed as soil P is typically more related to parent material than vegetation. The final selection resulted in a total of 27 covariates being used in the modelling phase of Stage 3 (Table 1). These are ranked in order of importance in the modelling.

To avoid the potential influence of land use upon prediction of P, a vegetation covariate was created for this project using Pre-clearing Broad Vegetation Groups of Queensland, derived from regional ecosystem mapping (Queensland Herbarium 2021a). The polygon mapping was converted to a raster dataset in GIS based on the 1:1 000 000 scale Broad Vegetation Groups, of which there were 99 classes.

Parent material is a key soil forming factor that can be represented by a range of covariates (e.g. geology mapping, gamma radiometrics). For pedological purposes, the most important aspect of parent material is lithology (i.e. rock type), in particular its mineralogy and chemical composition. Gray created a lithology covariate layer using DSM methods as part of a project which predicted six key soil properties across New South Wales (Gray *et al.* 2016). Twelve lithology classes were developed to group rock types primarily by silica content. Lithology was found to be the most important covariate (more important than gamma radiometrics) in predicting the soil properties.

A lithology covariate layer was created for this project to aid subsequent soil predictions. To better identify different landforms and potential soil properties, the 12 classes by Gray et al (2016) were subdivided into 19 subclasses, to distinguish igneous rocks, sedimentary rocks and alluvia (Table 2). Given that state-wide coverages of geology mapping and soil mapping had varying scales, the dataset *Vegetation management pre-clear regional ecosystem map - version 12.0* (Queensland Herbarium 2021b) was used as a base polygon layer which represented geology, soil and vegetation. In QGIS, the spatial join tool was used to assign key data to the regional ecosystem (RE) mapping polygons from the best available geology and soil mapping datasets, based on the dominant geology/soil type present within each RE polygon. In Microsoft Access, lithology classes were assigned to rock geology units based on the dominant lithology recorded for each unit. For alluvial units, which are often broad and can contain soils from varying parent material with varying soil properties, lithology classes were assigned based on comparison of the best mapping available soil mapping and RE mapping of alluvial/aeolian land zones (i.e. land zones 1 - 6) (Wilson and Taylor 2012). The polygon mapping was then converted to a raster dataset in GIS based on the lithology class assigned to each polygon.

Table 1 - Environmental covariates used in the Stage 3 modelling phase.

SCORPAN factor	Covariate	Units	Resolution	Reference/source	Importance rank
Organism	Pre-clearing broad vegetation groups	none	25 m	Pre-clearing broad vegetation groups (1:1 000 000) scale, Queensland Herbarium.	18.8
Climate	Average annual rainfall	mm	2.5 km	Bureau of Meteorology (2013). Climatological Gridded Data – Average annual rainfall between 1961 and 1990. Commonwealth of Australia.	11.9
Relief	3” SRTM Derived Digital Elevation Model	m	90 m	Geoscience Australia and CSIRO Land & Water (2010) 1” SRTM-Derived Digital Elevation Model (DSM, DEM, DEM-S and DEM-H) User Guide. Version 1.0. Geoscience Australia. ANZCW0703014182	9.1
Parent material	Weathering intensity index		90 m	Wilford, J., Roberts, D. (2019). Weathering Intensity Model of Australia. Geoscience Australia, Canberra. http://dx.doi.org/10.26186/5c6387a429914.	7.4
Parent material	Queensland Lithology classes		90 m	Walton, J., 2022. Lithology classes representing silica content assigned to pre-clearing regional ecosystem mapping polygons, geology mapping and best available soil polygon mapping.	6.7
Relief	Multi-resolution Valley Bottom Flatness	Index	90 m	Gallant J, Dowling T & Austin J (2012). Multi-resolution Valley Bottom Flatness (MrVBF, 3" resolution). v2. CSIRO. Data Collection. 10.4225/08/512EF27AC3888.	6.7
Climate	Average summer rainfall	mm	2.5 km	Bureau of Meteorology (2013). Climatological Gridded Data – Average summer rainfall between 1961 and 1990. Commonwealth of Australia.	6.5
Parent material	Radiometric filtered potassium	%	90 m	Radiometric grid of Australia (Radmap) v4 2019 - Potassium	6.5
Climate	Prescott Index modified by Linda Gregory (CSIRO)	Index		Prescott J A (1950) A climatic index for the leaching factor in soil formation. European Journal of Soil Science 1, 9-19.	6.5
Climate	Average daily minimum summer temperature	Co	2.5 km	Bureau of Meteorology (2013). Climatological Gridded Data – Average daily minimum summer temperature between 1961 and 1990. Commonwealth of Australia.	5.6
Parent material	Gravity		850 m	Australian National Gravity Database (2008) 0.5’ offshore – onshore gravity grid - July 2008. Geosciences Australia	4.5
Climate	Average daily maximum winter temperature	°C	2.5 km	Bureau of Meteorology (2013). Climatological Gridded Data – Average daily minimum summer temperature between 1961 and 1990. Commonwealth of Australia.	4.4
Relief	Rel_foc		90 m	Relief_foc (further information not available)	4.4

SCORPAN factor	Covariate	Units	Resolution	Reference/source	Importance rank
Climate	Annual rainfall variability	Index	25 km	Bureau of Meteorology (2013). Climatological Gridded Data – Annual rainfall variability between 1900 and 2003. Commonwealth of Australia.	3.8
Climate	Average annual thunderstorm days	days	25 km	Bureau of Meteorology (2013). Climatological Gridded Data – Average annual thunderstorm days between 1990 and 1999. Commonwealth of Australia.	3.7
Parent material	Total Magnetic Intensity RTP			Total Magnetic Intensity (TMI) grid of Magnetic Map of Australia. 3rd Edition, 1999 survey - RTP	2.4
Relief	Slope	%	90 m	Gallant J & Austin J (2012). Slope (3" resolution) derived from 1" SRTM DEM-S. v3. CSIRO. Data Collection. 10.4225/08/50A9DF115250E.	2.4
Parent material	Distance to outcrop			Distance from the nearest rock outcrop	2.4
Parent material	Radiometric filtered thorium	ppm	90 m	Radiometric grid of Australia (Radmap) v4 2019 - Thorium	2.3
Parent material	Radiometric filtered uranium	ppm	90 m	Radiometric grid of Australia (Radmap) v4 2019 - Uranium	2.2
Relief	Geomorphons		90 m	Classification of 3s DEM into Geomorphons (patterns of landforms) by Malone. CSIRO.	2.2
Relief	Topographic Wetness Index	Index	90 m	Gallant J & Austin J (2012). Topographic Wetness Index (3" resolution) derived from 1" SRTM DEM-H. v1. CSIRO. Data Collection. 10.4225/08/50A9DF3968422.	1.6
Relief	Slope relief class		90 m	Gallant J, & Austin J (2012). Slope Relief (3" resolution) derived from 1" SRTM DEM-S. v1. CSIRO. Data Collection. 10.4225/08/50A9D1D1BA1C1	1.6
Parent material	Total Magnetic Intensity TMI			Total Magnetic Intensity (TMI) grid of Magnetic Map of Australia. 3rd Edition, 1999 survey - TMI	1.5
Relief	Multi-resolution Ridge Top Flatness	Index	90 m	Gallant, John; Dowling, Trevor; Austin, Jenet (2013): Multi-resolution Ridge Top Flatness (MrRTF). v2. CSIRO. Data Collection. https://doi.org/10.4225/08/56EA312A5E63B	1.3
Parent material	Total Magnetic Intensity 1VD			Total Magnetic Intensity (TMI) grid of Magnetic Map of Australia. 3rd Edition, 1999 survey - 1VD	1.1
Relief	Topographic Position Index	Index	90 m	Topographic Position Index derived from 1" SRTM DEM-S	1

Table 2 - Queensland lithology covariate classes

Class ID	Lithology class	Silica (SiO_2) % Range	Lithology Examples	Common soils (Australian Soil Classification)
1	Extremely siliceous	>85	Quartz sands (beach, riverine or aeolian), chert, pure quartzite, jasper, quartz reefs, silicified rocks	Arenosols, Arenic Rudosols; Podosols
22	Siliceous upper (sedimentary)	77-85	Quartz sandstone, quartz siltstone, unqualified quartzite	Arenosols, Kandosols; sandy Tenosols
23	Siliceous upper (alluvial)	77-85	Alluvial sands	Arenosols, Kandosols; sandy Tenosols
31	Siliceous mid (igneous)	70-77	Granite, rhyolite and siliceous tuff	Kandosols; dystrophic (low fertility) Chromosols, Kurosols & Sodosols
32	Siliceous mid (sedimentary)	70-77	Arkose sandstone, most unqualified sandstone	Kandosols; dystrophic (low fertility) Chromosols, Kurosols & Sodosols
41	Siliceous lower (igneous)	65-70	Adamellite, granodiorite, dacite, monzogranite, siliceous/intermediate tuff	Mesotrophic (moderate fertility) Chromosols, Kurosols & Sodosols
42	Siliceous lower (sedimentary)	65-70	Most greywacke & lithic sandstone, unqualified siltstone	Mesotrophic (moderate fertility) Chromosols, Kurosols & Sodosols
51	Intermediate upper (igneous)	60-65	Syenite, trachyte	Dermosols; eutrophic (high fertility) Chromosols, Kurosols & Sodosols
52	Intermediate upper (sedimentary)	60-65	Most argillaceous rocks (mudstone, claystone, shale, slate, phyllite and schist)	Dermosols; eutrophic (high fertility) Chromosols, Kurosols & Sodosols
53	Intermediate upper (alluvial)	60-65	Alluvial loams and non-cracking clays	Dermosols; eutrophic (high fertility) Chromosols, Kurosols & Sodosols
61	Intermediate lower (igneous)	52-60	Monzonite, trachy-andesite, diorite, andesite, intermediate tuff	Grey & brown Vertosols
62	Intermediate lower (sedimentary)	52-60	Cracking clays on sedimentary rocks	Grey & brown Vertosols
63	Intermediate lower (alluvial)	52-60	Alluvial cracking clays (not black)	Grey & brown Vertosols
71	Mafic (igneous)	45-52	Gabbro, dolerite, basalt, mafic tuff, amphibolite	Black Vertosols
72	Mafic (alluvial)	45-52	Alluvial black cracking clays	Black Vertosols
8	Ultra-mafic	≤45	Serpentinite, dunite, peridotite, tremolite-chlorite-talc schists	Vertosols (high heavy metals)
9	Calcareous	Variable	Limestone, dolomite, calcareous shale, calcareous sands	Dermosols, Calcarosols (pedogenic)
101	Sesquioxide (mafic)	Variable	Deeply weathered mafic igneous rocks	Ferrosols, Red Kandosols, Red Dermosols
102	Sesquioxide (igneous, non-mafic)	Variable	Deeply weathered acid to intermediate (lower) igneous rocks	Red Kandosols, Red Dermosols, Ferrosols,
103	Sesquioxide (sedimentary)	Variable	Laterite, bauxite, ferruginous sandstone, ironstone	Red Kandosols, Red Dermosols, Ferrosols
11	Organic	Variable	Peat, coal, humified vegetative matter	Organosols
12	Evaporite	Variable	Gypsum, halite, anhydrite	Hydrosols

4 Results

4.1 P input data

The final training dataset had a soil P range between 0.1 and 149.0 mg kg^{-1}, with the distribution of P data skewed toward the lower values of the population (Figure 2a). The bicarbonate-extractable P concentration was below 10 mg kg^{-1} (low to very low (Ahern *et al.* 1994)) in 47% of samples and was below 25 mg kg^{-1} in 77% of samples, with only 2% of samples recording P levels above 100 mg kg^{-1}. A log transformation was applied to the dataset to reduce the skewness and approximated the data population to a normal distribution (Figure 2b) in preparation for Cubist modelling. The skewness of the P concentrations in the samples demonstrated the generally low levels of soil P in Queensland soils.

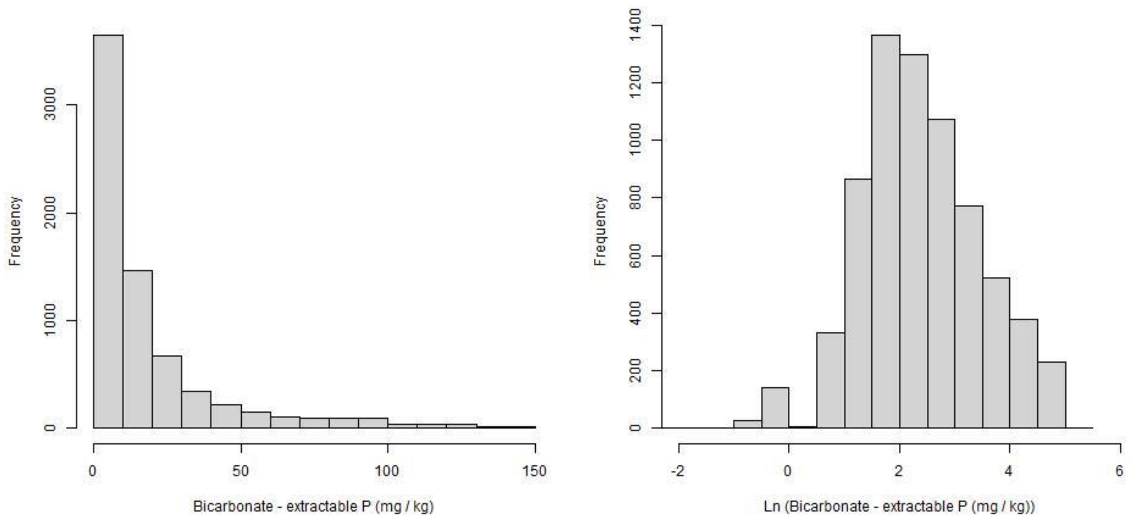

Figure 2 - (a) Distribution of training data bicarbonate-extractable P levels and (b) distribution after transformation using a natural logarithm.

Apart from the Darling Downs and Moreton regions, the soil P content did not vary greatly across the grazing land management regions represented in the study area, as displayed by the boxplots in

Figure 3. The boxes represent the 50th percentile range in data, with the median value shown by the dark bar in the middle of the box. The Darling Downs and Moreton regions had a high proportion of sites spread across a large range of values. The skewed nature of the distribution is shown by the wide range in outlier values (circles) of high P levels. Note that only part of the Southern Gulf region is included in Stage 3 of this project.

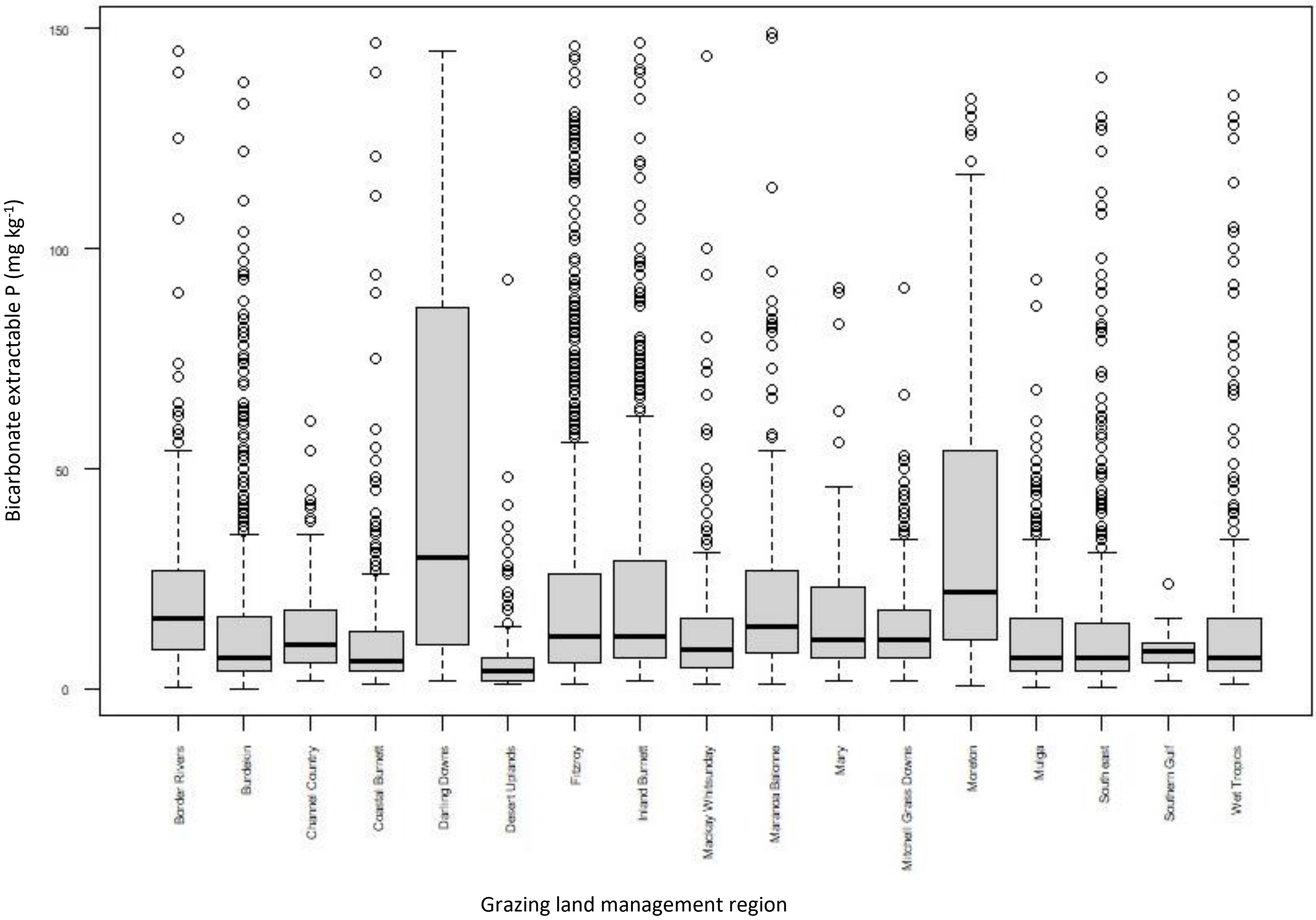

Figure 3 - Boxplot of bicarbonate Extractable P training data values (mg kg^{-1}) by Grazing Land Management Region.

4.2 Bicarbonate-extractable P map

The final product of this project was a raster map of predicted natural (unfertilised) surface soil bicarbonate-extractable P of Queensland grazing lands south of the Gulf and Cape York Natural Resource Management Regions (Figure 4). The size of the cells of the raster map was 90 x 90 m^2, each with a predicted value of P ranging between 0.1 and 148 mg kg^{-1}. The modelled bicarbonate-extractable P concentrations are displayed below using the P rating categories originally developed by Ahern et al. 1994. The raster map can be downloaded from the Queensland Spatial catalogue (QSpatial).

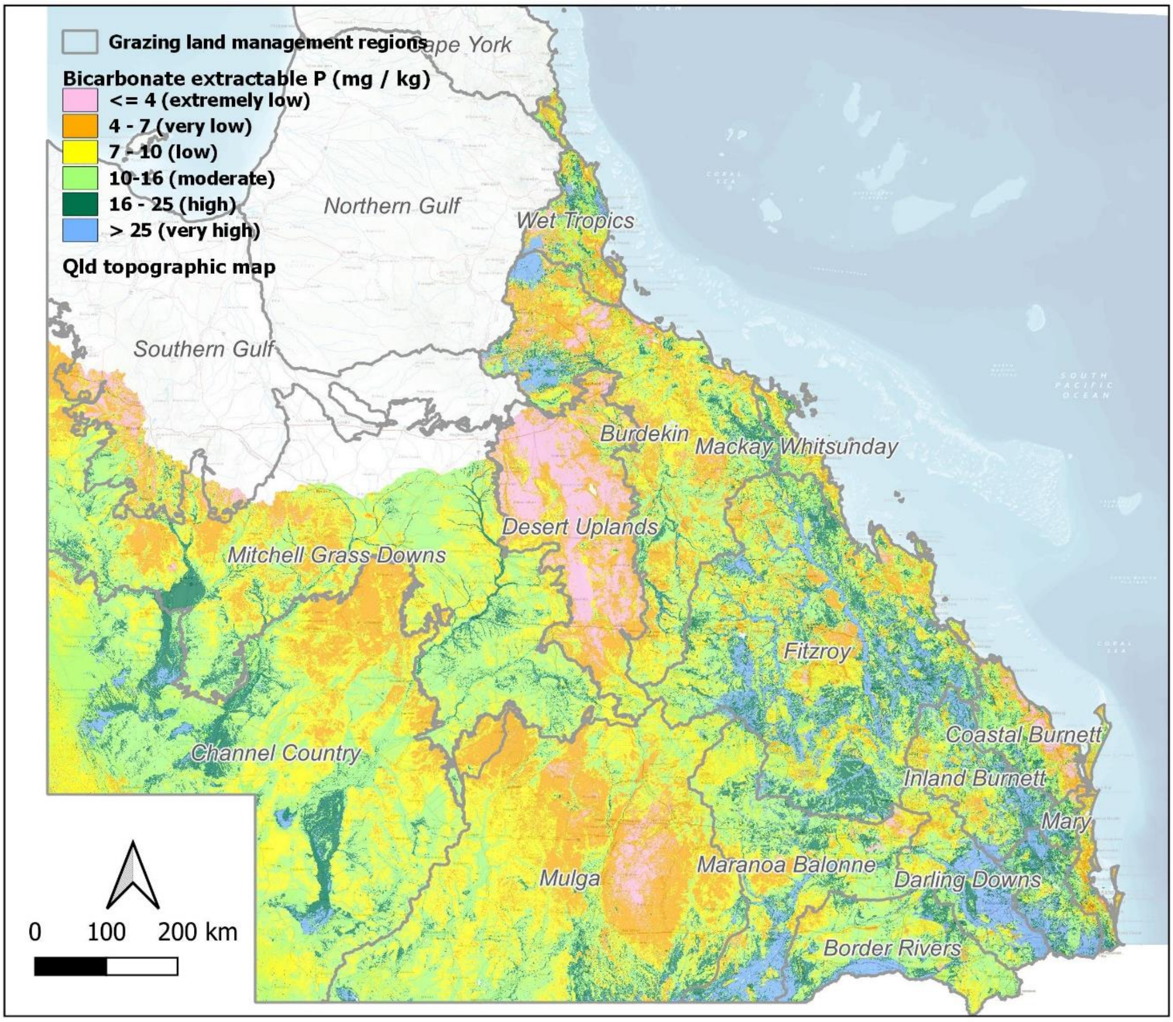

Figure 4 - Predicted bicarbonate-extractable P (0 – 0.1 m) displayed using the Ahern et al. (1994) soil P rating system for pastures.

An extract of the bicarbonate-extractable P map over the Middlemount - Windeyers Hill area (northeast of Rockhampton) is shown below. The mapping matches 1:100 000 scale soil mapping (Burgess 2003) boundaries (black linework) project well. It also picks up trends in the broader scale land systems mapping (Story *et al.* 1967) (purple linework) where sites density is much lower, with alluvium having high P, and duricrusted Tertiary sediments being low.

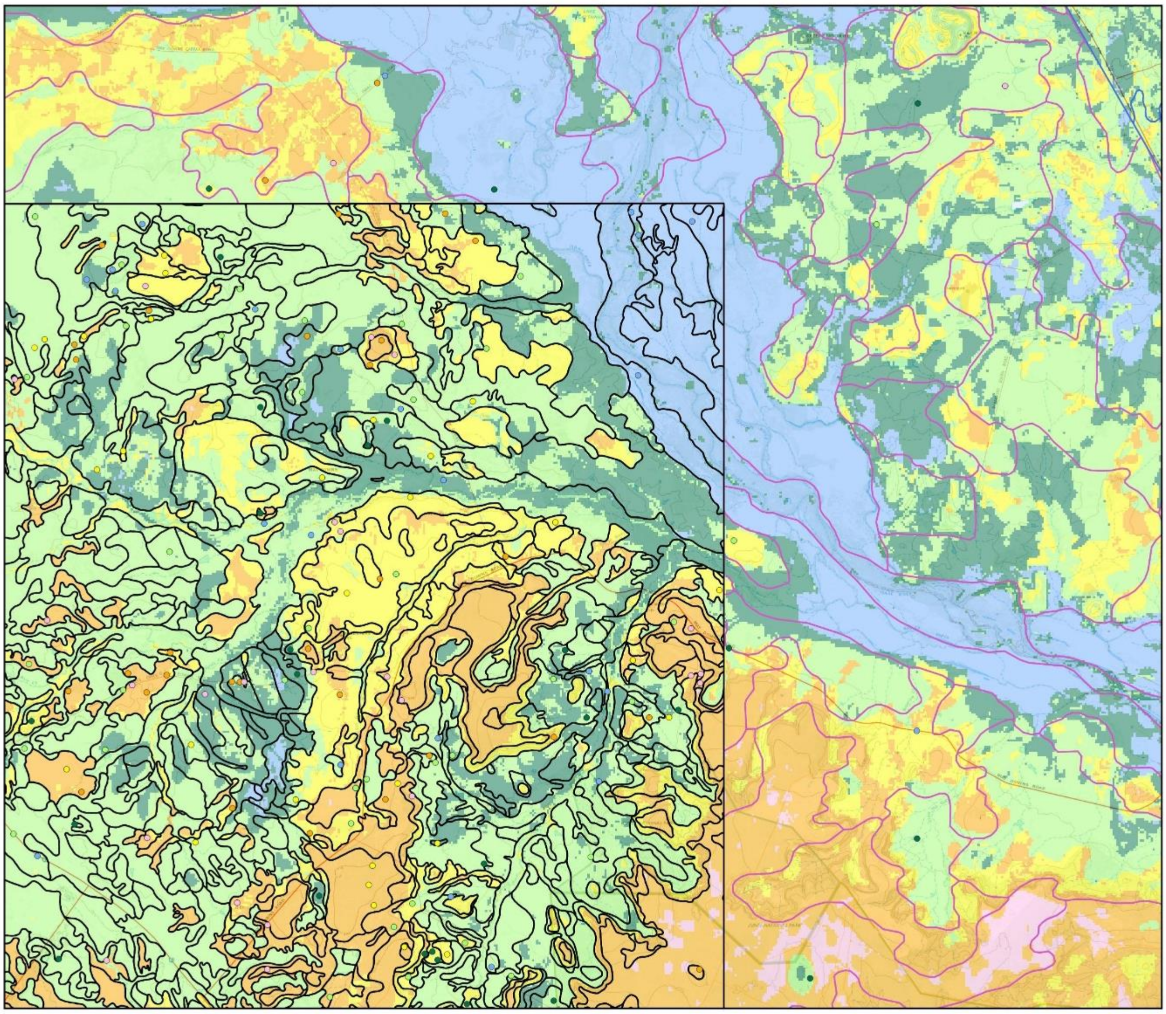

Figure 5 - Predicted bicarbonate extractable P map extract, Windeyers Hill area.

4.3 Model performance

The validation statistics of the Cubist modelling are shown in Table 3. The average magnitude of errors between the observed and the predicted values (root *mean* square error (RMSE) = 22 mg kg^{-1}) was relatively large, representing approximately 16% of the of predicted bicarbonate-extractable P range. This RMSE was greater than most of the critical P concentration intervals (<4, 4 – 7, 7-10, 10 – 16, 16 – 25 & >25 mg kg^{-1}) developed by Ahern (Ahern *et al.* 1994). Based on the mean bias result, the model tended to underpredict P (Bias = -5 mg kg^{-1}). The relatively high RMSE is likely to be the result of the skewed distribution of the observed bicarbonate-extractable P data (Figure 2 (a)) and relatively small number of predictions with very large errors (probably underprediction of some very high P in some samples). 22% of the Colwell P concentrations are above 25 mg kg^{-1}, with only 2% of values above 100 mg kg^{-1}. A comparison of observed versus predicted Colwell P concentrations across all the training data sites supports this reasoning, with a general trend of large underprediction when P concentrations are high.

The root *median* square error (RMedSE) result however, provides a more positive picture on the accuracy of the map. The RMedSE result is less affected by the skewed distribution of the bicarbonate-extractable P data (i.e. small number of very high values) and this alternative measure of the predictions provides further insight in relation to the withheld validation data used in cross validation. The RMedSE result of 5.2 mg kg^{-1} indicates that in 50% of cases, errors with magnitude of less than 5.2 mg kg^{-1} are expected. Similarly, the median bias (0.1 mg kg^{-1}) is closer to zero as it is less affected by extreme values.

The concordance (0.4) between the training data and the predictions was considered reasonable for a model based mostly on legacy data. Importantly, for 88% of the validation data, the 90% prediction interval contained the true soil P concentration (close to the ideal percentage of 90%), indicating that the model is providing a fair and realistic assessment of prediction uncertainty.

Table 3 – Model validation statistics calculated using the k-fold (Cubist) method.

Statistic	**Validation: Prediction performance**
Root mean standard error (RMSE) [a]	22.3 mg kg^{-1}
Root median standard error (RMedSE)[b]	5.2 mg kg^{-1}
Coefficient of correlation (r^2)[c]	0.27
Bias [d]	-5.4 mg kg^{-1}
Bias (median) [e]	0.1 mg kg^{-1}
Concordance [f]	0.4
Percent within upper and lower limits [g]	88

[a] RMSE is a measure of accuracy, with a result of 0 indicating a perfect fit between prediction and observation. It is calculated using a mean value.
[b] RMedSE is a measure of accuracy, with a result of 0 indicating a perfect fit between prediction and observation. It is calculated using a median value, which is less affected by skewed datasets.
[c] r^2 provides a measure of how well observed outcomes are replicated by the model. This measure is based on the proportion of total variation of outcomes explained by the model.
[d] Bias quantifies the difference between a model's estimation of a parameter and a parameters true value. An unbiased model will converge to the correct value as data availability increases.
[e] Bias (median) quantifies the difference between a model's estimation of a parameter and a parameters true value. An unbiased model will converge to the correct value as data availability increases.
[f] Concordance is a single statistic that both evaluates the accuracy and precision of the relationship. It is often referred to as the goodness of fit.
[g] Associated with each predicted value is a 90% prediction interval, an interval which has a 90% chance of capturing the true soil P concentration. Ideally, 90% of the 90% prediction intervals should capture the true soil P concentration.

The validation method used (k-fold cross validation) and reliance on legacy data may have introduced potential sources of bias to the modelling process.

According to Malone et al. (2017), validation of trained models with data subsetting mechanisms could result in some bias in the validation statistics.

The distribution of samples is skewed due to legacy data being dominated by soil surveys of cropped regions and along roads which are more accessible. This location bias in the training and validation data is also likely to have resulted in sub-optimal coverage of the range of values in the 'covariate space'. More fertile areas suitable for irrigated cropping are often sampled for land suitability assessment projects, with a large amount of legacy data being present in these areas, resulting in bias toward soils with high P concentrations rather than low P. This bias was moderated however, by deliberately excluding sites from cultivated or highly disturbed areas to avoid samples with unnaturally high P levels due to anthropogenic activities (e.g., fertilisation).

It is important to note that an approach that could have enabled substantially better performance of the model would have entailed the collection of an entirely new or supplementary dataset, with sites spread throughout the modelling area based on a statistically based sampling strategy linked to the covariates used. Given the timeline and the budget allocated to this project however, this approach was not possible.

4.4 Prediction uncertainty

DSM maps are predicted soil attribute surfaces, and as such they need to be interpreted with care. Prediction maps are never perfect, always deviating somewhat from reality. The reliability of the predicted P map refers to the likely level of agreement between the modelled values and actual observations; uncertainty maps provide a measure for assessing and communicating this reliability.

High uncertainty values in an area can be due to a number of reasons, related to the modelling method used, training data and covariates. To assist with the interpretation of the bicarbonate-extractable P map (Figure 4), two uncertainty maps were produced to illustrate the variations in P concentrations predicted during the Cubist modelling. These raster maps can be downloaded from the Queensland Spatial catalogue (QSpatial).

The *prediction uncertainty range map* (Figure 6) was produced by calculating the difference between the 95th (upper) and 5th (lower) percentile prediction rasters. This map shows a very similar spatial pattern to the map of the predicted P since, when predictions are calculated on a log-transformed variable, the back-transformed prediction intervals are generally larger for larger predicted values.

The *prediction relative uncertainty map* (Figure 7) provides a much more useful interpretation. This was created by dividing the prediction uncertainty range (as described above) by the predicted (median) bicarbonate-extractable P concentration. The *prediction relative uncertainty map* showed different spatial patterns compared with the *prediction uncertainty range map*. The relative uncertainty tended to be higher in the eastern half of the state, with highest uncertainty indicated in areas such as the Wet Tropics and Coastal Burnett.

In general, uncertainty was higher in regions where soil sample data was scarce, or where there was a large range of P concentrations in the sample data, in areas where soils have high natural fertility (e.g. Darling Downs and Moreton GLM regions). For example, at sites predicted to have extremely low P (P < 4 mg k^{-1}, (Ahern *et al.* 1994)) 90% of soil samples had P concentrations ranging from 0 to 6 mg kg^{-1}, whereas at sites predicted to have very high P (P > 25 mg k^{-1}) 90% of samples had P levels between 15 and 149 mg kg^{-1}. Greater variability in observed P levels translated to higher uncertainty levels in the modelled outputs. Therefore, high uncertainty levels in fertile areas do not necessarily mean that the modelled P levels were unreliable. Examples of this dynamic were alluvial plains, basalt uplands and areas along major rivers such as the Fitzroy basin, the Toowoomba Range and the Atherton Tablelands, where high uncertainty in the prediction was likely due to a large range of P concentrations in the soil samples. High values could also be caused by specific covariates with extreme values.

Areas where the lack of input data are likely to have led to elevated prediction uncertainty values include the Great Dividing Range, an area east of The Lynd Junction, the Great Basalt Wall region, the Coastal Ranges from North Queensland to the D'Aguilar Range in Southeast Queensland, the Moore Park-Hummock area, Mt. Barney, the McPherson Range and Mt. Tamborine.

It should be noted that the maps represent predictions and uncertainties on point-support data (i.e. samples from individual soil profile observations), however short-range variation in soil P can be comparatively high and may not be explainable by the covariates, which were of 90 m pixel size. Furthermore, it should be expected that averages of these predictions over larger areas would have smaller errors (if compared with data from multiple profiles averaged over the same area) and smaller prediction uncertainties. However, testing the nature of this effect would require careful consideration of appropriate validation data and was beyond the scope of the current study.

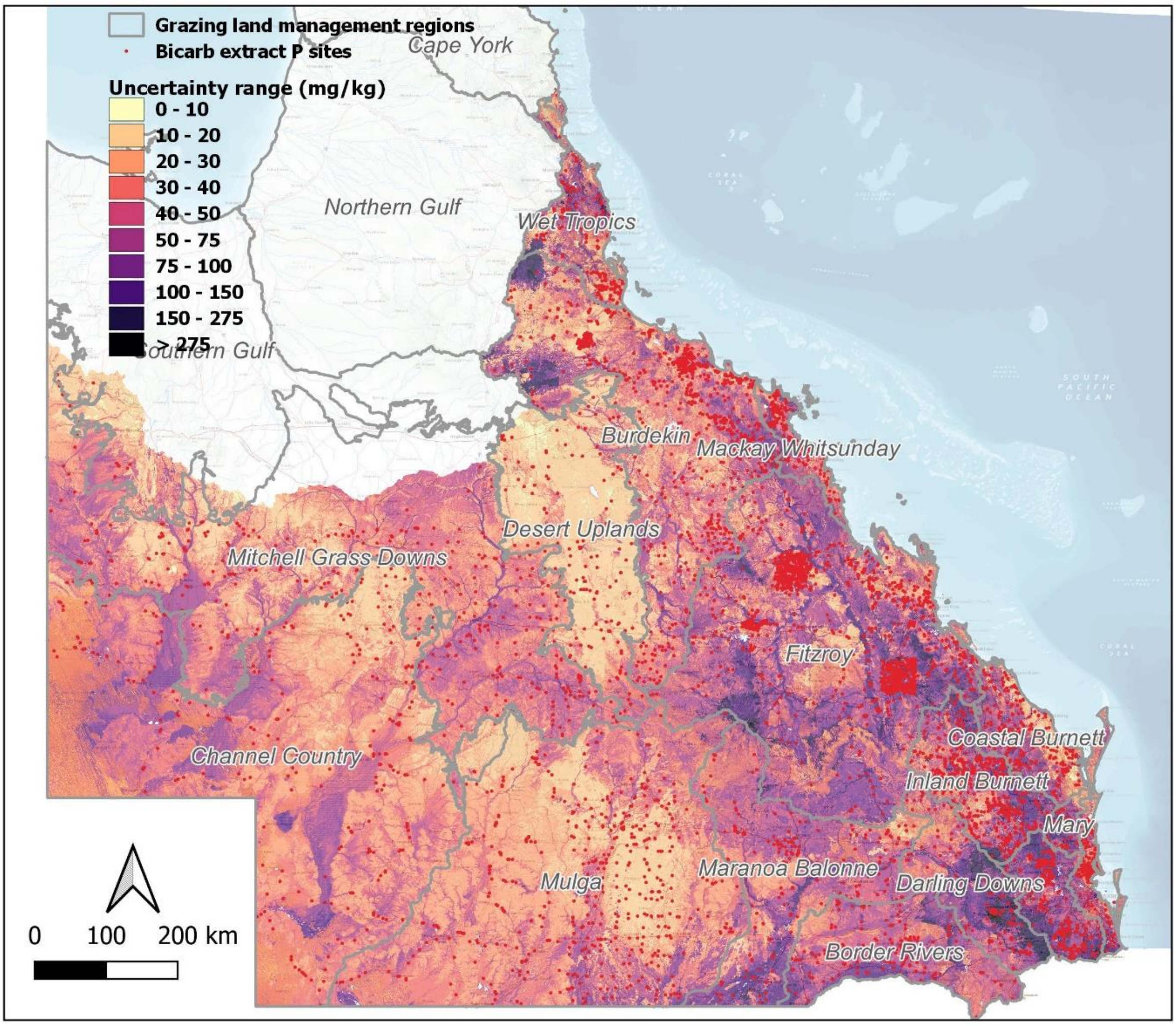

Figure 6 – Bicarbonate-extractable P prediction uncertainty range.

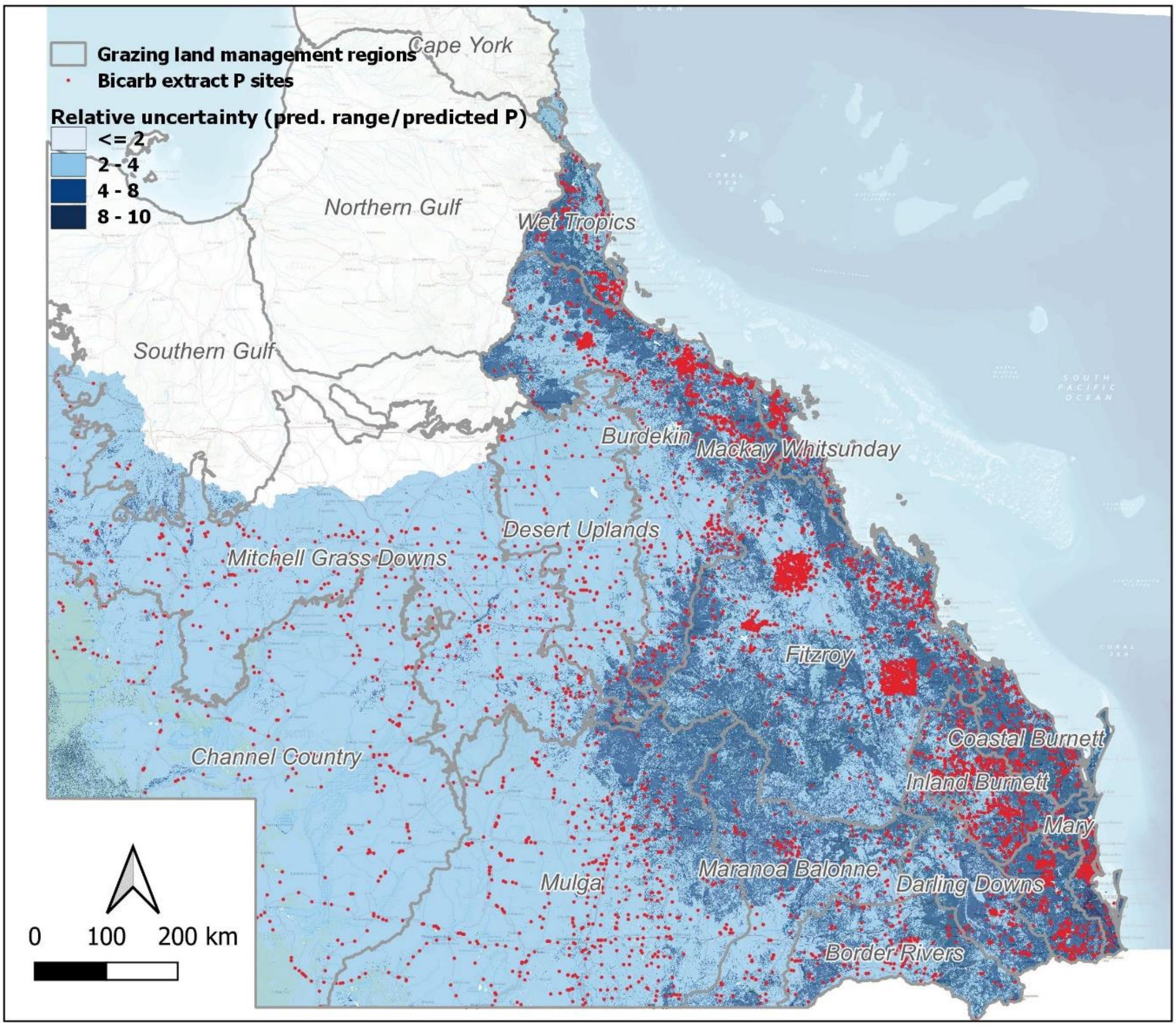

Figure 7 - Predicted bicarbonate-extractable P relative uncertainty.

4.5 Artefacts in the predictions

Artefacts are patterns of prediction that do not appear to follow natural landscape patterns. The magnitude of the artefacts identified in Stage 1 and 2 was substantially reduced in Stage 3. This was likely due to some different covariates being used in the latter stage. One large rectangular pattern was still visible in the Mulga Region north of Augathella, due to an artefact in the radiometrics covariates. However, actual differences in the predicted values were only 2-5 mg kg^{-1} across the artefact boundaries. Other features are present northeast of Eidsvold, and between Inglewood and Tara, however from review of the covariates it is not clear what has caused these.

Some artefacts may have been due to the data partitioning phase that was conducted in Cubist before performing the regressions needed to predict the P value. During the data partitioning phase, the model uses the sample and relative covariate data to develop rules that define subsets of the data with similar characteristics (i.e. modelled by the same regression model). This can introduce artificial boundaries between rules. It was not possible to remove these artefacts due to the intrinsic structure of the Cubist model, but we note that the impact of rule boundaries was reduced by applying the Cubist modelling procedure based on the 10 subsets of the k-fold validation procedure.

However, it is important to note that these artefacts are easily identifiable and that the general distribution of predicted P was deemed sensible based on expert knowledge. The only exceptions were the high P levels predicted on K'Gari (Fraser Island) and other areas with large sand masses.

5 Conclusions

5.1 Key findings

The activities of this project led to the development of the following outputs:

- The first bicarbonate-extractable P and uncertainty maps of Queensland main grazing lands, produced at a 90m pixel resolution (3 arc seconds pixel grid). The maps cover a total area of 1,074,410 km^2 and include the following regions: Fitzroy, Burnett, Southeast Queensland, Darling Downs and Balonne-Maranoa, Burdekin, Herbert, Wet Tropics and Atherton Tablelands, Mitchell grasslands of the Thompson and Channel Country.
- A robust DSM methodology based on fitting a Cubist model (Quinlan 1992; Kuhn *et al.* 2012) to the point P data that were intersected with environmental covariate data.

The P map is considered a reasonable prediction of P at the subcatchment to property scale. It is useful for distinguishing areas of high P from low P. While the level of uncertainty is higher in areas of more fertile soils (with higher P variability), across the majority of the map, where P levels are lower, predictions are more accurate.

5.2 Benefits to industry

The soil bicarbonate-extractable P map produced by this project provides stakeholders of Queensland's red meat industry with a substantially better map of predicted plant available P compared with previous P maps.

The knowledge gained from mapping P throughout Queensland main grazing lands will assist MLA to target more accurately their investments promoting research and extension activities on soil P. For example, the maps produced in this project could be used to:

- Plan P supplementation across different landscapes to improve the productivity and profitability of grazing operations through enhanced P management.
- Prioritise and concentrate soil testing efforts in areas with high uncertainty.
- Prioritise the introduction of Stylos and Leucaena or other leguminous pasture species in areas with sufficient natural P to maintain growth.
- Identify areas where inadvertent mining of soil native P might be occurring.
- Identify the areas most suitable for hosting field experiments on soil P dynamics.
- Raise awareness of the P cycle and soil P deficiencies in Queensland grazing systems.

Additionally, the data displayed on the maps will also be beneficial to State and Commonwealth Departments to:

- Enhance catchment modelling capabilities to improve water quality simulations for the protection of the Great Barrier Reef.
- Identify areas with high native P levels that might contribute to eutrophication and water quality issues in waterways and dams.
- Improve the State of the Environment reporting.
- Target soil health monitoring activities.

6 Future research and recommendations

The outcomes of this project outlined the importance of the availability of improved P maps for the red meat industry. Since the focus of this project was the development of a robust DSM methodology for the development of a soil bicarbonate-extractable P for Queensland main grazing lands, future R&D effort should focus on the following activities:

- Employ the methodology established during this project to develop a bicarbonate-extractable P and uncertainty map for the whole Northern grazing region of Australia, including Far North Queensland, the Northern Territory and the Northern regions of Western Australia.
- Develop a P buffer index map for the whole Northern grazing region of Australia to be used in conjunction with the bicarbonate-extractable P map. Since the bicarbonate-extractable P method does not consider a soils P buffering capacity, if used alone it does not provide a precise estimate of actual plant available P. The availability of bicarbonate-extractable P and P buffer index data would enable the estimation of more accurate plant available P levels for the whole Northern grazing region of Australia.
- Perform a spatial analysis to assess how the previous P maps by Ahern et al. (1994), McCosker and Winks (1994) and(Ahern *et al.* 1994; McCosker and Winks 1994; Viscarra Rossel and Bui 2016) compare with the maps produced in this project.
- Improve the predicting power of the model used in this project by: i) collecting more samples in the areas with sparse data for model calibration and validation or where the uncertainty analysis indicated high variability in P concentrations; and ii) exploring the use of other covariates.
- Employ different machine learning algorithms (e.g., random forest or convolutional neural network) to eliminate the artefacts produced in this project with Cubist.

7 Acknowledgments

The authors of this report would like to express their gratitude to Tom Orton for the support provided on statistical matters and David Lawrence for Access to bicarbonate-extractable (Colwell) P data from the Darling Downs.

8 References

ABARES (2016) The Australian Land Use and Management Classification Version 8. (Canberra)

Ahern CR, Shields PG, Enderlin NG, Baker DE (1994) Soil Fertility of central and north-east Queensland grazing lands. Information Series Q104065. (Brisbane, Australia.)

Burgess JW (2003) Land Resource Assessment of the Windeyers Hill Area, Isaac–Connors and Mackenzie River Catchments, Central Queensland. (Brisbane)

Gray JM, Bishop TFA, Wilford JR (2016) Lithology and soil relationships for soil modelling and mapping. *Catena* **147**, 429–440. doi:10.1016/j.catena.2016.07.045.

Kuhn M, Weston S, Keefer C, Coulter N (2012) Cubist Models For Regression.

McBratney AB, Mendonça Santos ML, Minasny B (2003) 'On digital soil mapping.' doi:10.1016/S0016-7061(03)00223-4.

McCosker TH, Winks L (1994) 'Phosphorus nutrition of beef cattle in northern Australia.' (Queensland Department of Primary Industries, Meat Research Corporation (Australia): Brisbane)

McDonald RC, Isbell RF, Speight JG (2009) Land Surface. 'Aust. soil L. Surv. F. Handb.' (Ed NCST) pp. 127–145. (CSIRO Publishing: Melbourne)

Queensland Herbarium (2021a) Pre-clearing broad vegetation groups - Queensland. Version 5.1. https://qldspatial.information.qld.gov.au/catalogue/custom/viewMetadataDetails.page?uuid=%7B34FA2F14-B647-40B2-88EF-F796A9BE4CBA%7D.

Queensland Herbarium (2021b) Vegetation management pre-clear regional ecosystem map - version 12.0. https://qldspatial.information.qld.gov.au/catalogue/custom/detail.page?fid=%7BCB642237-0FB9-4F15-A27C-087987077FE8%7D.

Queensland Land Use Mapping Program (2019) Queensland Land Use.

Quinlan JR (1992) Learning with continuous classes. *Machine Learning*. doi:10.1.1.34.885.

Rayment G, Lyons DJ (2011) 'Soil chemical methods : Australasia (Australian soil and land survey handbook series: v.3).' (CSIRO Publishing: Collingwood, Victoria.)

Story R, Galloway RW, Gunn RH, Fitzpatrick AE (1967) Land of the Isaac-Comet area, Queensland. (Melbourne)

Terrestrial Ecosystems Research Network AusSoilsDSM.

Viscarra Rossel RA, Bui EN (2016) A new detailed map of total phosphorus stocks in Australian soil. *Science of the Total Environment* **542**, 1040–1049. doi:10.1016/j.scitotenv.2015.09.119.

Wilson P, Taylor P (2012) 'Land Zones of Queensland.'