Final report

Phosphorus Map of North Queensland Grazing Lands

Project code: B.GBP.0063

Prepared by: Stephen Leo, Niilo Gobius, Evan Thomas
Queensland Department of Environment, Science and Innovation
Queensland Department of Agriculture and Fisheries

Date published: 2024

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

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

Abstract

Livestock productivity in grazing lands is closely tied to the amount of plant available phosphorus (P) in the soil. To overcome P deficiency, graziers often resort to P supplementation and fertilisation. Soil maps are commonly used to inform P management or guide further sampling. However, soil maps of P availability are not available for the northern Queensland grazing regions, which contain a significant portion of the state's grazing land. The aim of this project was to fill that gap and update the existing map by producing a revised state-wide digital map of soil P. In this project, soil bicarbonate extractable P data was collated and supplemented with additional sampling in the northern grazing regions. Spatial covariates were intersected with the available soil P data and integrated into a quantile random forest model. The model performed well at predicting soil P with up to a 48% improvement in validation metrics compared to the previous soil P map. Final prediction maps including uncertainties were generated across Queensland at a 30 m resolution and will be available through LongPaddock by December 2024. The maps will provide essential information to graziers regarding P management across Queensland to support a more productive red meat industry.

Executive summary

Background

Phosphorus (P) is an essential nutrient required by cattle for almost every vital bodily function. Cattle consume the majority of their P intake through their diet of grazed pasture plants, which in turn absorb P from the soil. Many Australian soils are known to be inherently low in available P and can result in P deficiency in grazing cattle. Without sufficient P, overall cattle health can be significantly impacted and overall productivity reduced. Graziers often resort to P supplementation to overcome P deficiency and rely on soil maps to inform them of the P status of their lands. A digital map of soil P was recently produced to inform on P status but did not extend into the Gulf and Cape York regions of Queensland, where a significant portion of the state's grazing land exists and that are known to be P-deficient. The outcomes of this project will fill that gap and update the existing map, therefore providing graziers with access to high-resolution soil P information to inform P management strategies across all of Queensland. This will result in greater productivity and profitability across the red meat industry.

Objectives

The overall aim of this project was to update and extend the soil P map into the northern regions of Queensland. The specific objectives were to:

1. review the existing soil P datasets available for the entire extent of Queensland
2. review current availability of soil P data from the northern regions of Queensland
3. identify geographical gaps in the data set and collect/analyse new samples
4. update the previous digital soil mapping methodology with new machine learning models and spatial covariates
5. produce a predicted soil P (0-10 cm) map of Queensland at a 30 m resolution
6. incorporate the map into extension materials (e.g., upload to the LongPaddock website).

All objectives were successfully achieved, with the exception of point 6 which is subject to the final approval of this report.

Methodology

The methods undertaken to complete the soil P map of Queensland involved extracting all available soil bicarbonate extractable P data in Queensland from multiple sources. This was supplemented with additional soil sampling in under-represented areas in the northern grazing regions. Spatial covariates representing soil-forming factors were intersected with the available soil P data and integrated into a quantile random forest (QRF) model. The QRF model was trained on 80% of the dataset and the remaining 20% was used for validation.

Results/key findings

A total of 6796 unique site location samples were identified across Queensland as suitable for modelling following data cleaning. Of these, 228 samples were sourced in this project, either through soil sampling campaigns or archived data. The QRF model predicted soil P with a concordance correlation coefficient of 0.48, which is a 20% improvement compared to the previous soil P map. Final prediction maps including uncertainties (5^{th}, 50^{th} and 95^{th} percentiles) were generated across Queensland at a 30 m resolution and will be available through LongPaddock and QSpatial by December 2024.

Benefits to industry

The soil P map produced will provide general information to graziers regarding soil P status across their properties. However, it is important to note that soil mapping in Queensland is not carried out with sufficient precision for definitive usage at the paddock scale. The soil P map is indicative and should be complemented with site-specific soil tests to obtain accurate soil P levels. This will equip graziers with the knowledge necessary to refine P management for P-deficient soils, ultimately driving enhanced productivity and profitability in the red meat sector.

Future research and recommendations

Several future research directions were identified in this project including (i) the creation of a P buffering index map to complement the soil P map, which provides information on the soils capacity to adsorb and hold onto P, (ii) extension of the updated mapping methodology into the northern regions of Australia, including the Northern Territory and Western Australia which are known to be P-deficient, and (iii) areas of high uncertainty where future sampling campaigns could be targeted to improve model predictions and reduce uncertainty.

Table of contents

1. Background

Phosphorus (P) is an essential nutrient required by cattle for almost every vital bodily function. Phosphorus is required in the development of bones and teeth, energy metabolism (e.g., adenosine triphosphate), metabolism of fat, carbohydrates and protein, and cell structure (Dixon et al., 2020). The amount of P required by cattle highly depends on the season (wet or dry) and class of stock (reproductive status, growth and lactation). Cattle consume the majority of their P intake through their diet of grazed pasture plants, which in turn absorbs P from the soil. Many Australian soils are known to be inherently low in P due to extensive weathering, particularly in the northern regions (McCosker & Winks, 1994). Without sufficient P, a significant reduction in appetite of the animal can occur, which limits pasture and protein intake. This can lead to a number of health and welfare issues including poor growth, increased breeder mortality rates, reduced fertility and milk production, bone breakage, and therefore, significantly impacting productivity and profitability of beef enterprises.

To overcome P deficiency, graziers often develop a P management plan which involves supplementing P in the form of loose licks, liquid supplements, blocks or via water medication. The supplementation of P via these forms has demonstrated consistent results in improving pasture intake, liveweight gain and reproductive performance when cattle are P-deficient (Dixon et al., 2020; Schatz et al., 2023; Winks, 1990). One component of the P management plan is to undertake soil testing to inform P status. Various laboratory methods have been developed to assess soil P, some of which aim to quantify the total P content, while others estimate the plant available P. Total P tests measure both the soluble and insoluble P in the soil. Traditional soil surveys use total P to understand the nature of the soil (e.g., the degree of weathering), however, this test has limited use in indicating plant available P. In Queensland, plant available P is often measured through bicarbonate-extractable P (Colwell-P, method 9B2, Rayment & Lyons (2011)) and several traditional (polygon) soil maps have been developed (Ahern, 1994; McCosker & Winks, 1994). However, these existing maps of plant available P are generally of low resolution, limited extent and are not readily updatable with new data.

High-resolution, large coverage and easily updatable maps can be produced through the process of digital soil mapping (McBratney et al., 2003). In digital soil mapping, observed data is combined with spatial covariates that represent soil forming factors across the landscape and are integrated into statistical models to spatially predict soil attributes. Using digital soil mapping techniques, a soil P map (0-10 cm) across most of Queensland was recently published at a 90 m resolution (Zund et al., 2022). The map covered a large portion of Queensland and demonstrated significant interest by graziers. However, the map did not extend into the Gulf and Cape York regions, where a significant portion of the state's grazing lands exist and are known to be P-deficient (McCosker & Winks, 1994). A nation-wide map of soil P was also developed across six depths (0-5, 5-15, 15-30, 30-60, 60-100 and 100-200 cm) as part of the Soil and Landscape Grid of Australia (Zund, 2022). However, the topsoil layer (0-5 cm) did not align well with the P categories used to define whether the soil P is deficient or adequate as it typically considers the top 10 cm of soil.

To date, no reliable map of soil P is available for the northern grazing lands of Queensland. Therefore, the aim of this study was to produce a map across all of Queensland extending into the Gulf and Cape York regions using digital soil mapping techniques. The outcomes of this study will provide graziers with access to high-resolution soil P information to inform P management. This will result in greater productivity and profitability across the red meat industry.

2. Objectives

The objectives of this project were to:

1. review the existing soil P datasets available for the extent of Queensland to be mapped in this project
2. review current availability of soil P data from the northern regions of Queensland,
3. identify geographical gaps in the data set and collect/analyse new samples
4. update the previous digital soil mapping methodology with new machine learning models and spatial covariates
5. produce a soil P (0-10 cm) map of Queensland at a 30 m resolution
6. incorporate the map into extension materials (e.g., upload to the LongPaddock website).

All objectives of this project were successfully met, with the exception of objective 6 which is subject to the successful review of this final report and map.

3. Methodology

3.1 Site P data

3.1.1 Existing data

All available soil P data was extracted from the Queensland Soil and Land Information (SALI) database. All data extracted was analysed using the bicarbonate extraction method for soil P (commonly known as Colwell-P) - lab methods 9B1 or 9B2 as outlined in Rayment & Lyons (2011). The majority of soil P data extracted from SALI has been sourced from various land resource assessment projects undertaken over several decades, legacy data and more recent sites analysed in the previous P map project (Zund et al., 2022).

Following extraction, the data was cleaned to represent native P concentrations by removing (i) samples collected outside of the 0-10 cm layer in the soil profile, (ii) samples with concentrations greater than 150 mg kg^{-1}, which likely indicates previous fertilisation, (iii) cultivated or highly disturbed sites according to the recorded site description (McDonald et al., 2009, p. 128), and (iv) samples with concentrations greater than 7 mg kg^{-1} if within 75 m of a:

- plantation forest,
- modified pasture,
- cropping area,
- tree crop,
- feedlot or intensive animal farm,
- infrastructure,
- mine site,
- water storage system or channel.

Site exclusion was conducted using data from the most recent land use mapping of Queensland (Queensland Department of Environment and Science, 2019).

In cases where sites had multiple results, samples that were bulked and/or collected most recently were kept. If multiple samples still existed, the average per site was calculated.

3.1.2 Additional data

The Conditioned Latin Hypercube Sampling (cLHS) algorithm was used to determine additional sampling sites in the Southern Gulf, Northern Gulf and Cape York regions. The cLHS algorithm is widely used in digital soil mapping and has been proven to be an efficient soil sampling strategy (Yang et al., 2020). Briefly, cLHS is a stratified random sampling method, which samples ancillary data based on their multivariate distribution (Minasny & McBratney, 2006). The objective of the algorithm is to maximally stratify the distribution when selecting samples in order to best capture the variability in the dataset. The ancillary data included in the cLHS algorithm included spatial covariates identified with high importance in a preliminary model. This included radiometrics (thorium, potassium, uranium), elevation, weathering index, median slope averaged over 300 m and range in elevation averaged over 1000 m.

Prior to running the cLHS algorithm, the ancillary data was filtered using the most recent land use mapping of Queensland (Queensland Department of Environment and Science, 2019) to exclude areas outside the scope of the project or that are inaccessible, such as wetlands and rivers. The dataset was further filtered to remove areas more than 500 m from roads, which are likely to be inaccessible or too time-intensive to access. Once filtered, a total of 100 additional sites were allocated per region in the Southern Gulf, Northern Gulf and Cape York. The existing sites collected within each region were included in the cLHS process to avoid sampling areas already represented in the soil P dataset. If the selected sites by the cLHS algorithm were still not accessible, k-means clustering was used on the filtered dataset to determine similar areas based on the ancillary data within 10 km of the original sampling point (Fig. 1). K-means clustering is a method used to assign data into different groups that have relatively similar properties (in this case, ancillary data). The k-means clustering algorithm works by selecting random centroids in the dataset and then iteratively optimises the position of the randomly selected centroids by minimising the sum of squares from each point to the centroid (MacQueen, 1967).

Figure 1. Example site (red point) determined by the cLHS algorithm and potential other area (green polygon) to sample with similar spatial covariate data determined using k-means clustering.

Once a suitable site was located, the sampling process involved bulking 10 samples collected down to 10 cm spread over a 20 x 20 m area. A photo was captured at each site and additional information such as coordinates, location description, soil type and vegetation community were also recorded where possible.

3.2 Spatial modelling

3.2.1 Spatial covariates

Spatial covariate data was downloaded from the Terrestrial Ecosystem Research Network (TERN) datastore (Searle et al., 2022). A total of 156 covariates were downloaded from TERN at a 30 m resolution, which represent the soil forming factors of climate, organism (vegetation), relief and parent material (geology) (McBratney et al., 2003). Further details on the TERN spatial covariates used can be found at https://esoil.io/TERNLandscapes/Public/Pages/SLGA/GetData-COGSDataStore_30m_Covariates.html. An assessment of all the spatial covariates downloaded was initially conducted to remove covariates which displayed patterns unlikely to represent native soil P variability, such as seasonal Landsat-derived normalised difference vegetation index. In replacement, a few additional covariates representing pre-settler vegetation including pre-1750 major vegetation groups (Department of Climate Change, Energy, the Environment and Water, 2023a) and subgroups (Department of Climate Change, Energy, the Environment and Water, 2023b), and potential above-ground biomass (Roxburgh et al., 2019) were included. All covariates downloaded were clipped to the extent of Queensland and projected in EPSG 4326. The collected soil P data was subsequently intersected with the pre-processed spatial covariates to create the modelling dataset used in section 3.2.2. All processing was conducted in R (R Core Team, 2021).

3.2.2 Modelling

A quantile random forest (QRF) was selected for modelling the spatial variability of soil P. A QRF model is a variation of the random forest model that also has the capacity to estimate uncertainty. Random forest is an ensemble-learning method that consists of many decision trees (Breiman, 2001). Each decision tree represents a subset of the training dataset, randomly sampled to create a predictive model. The predictive model from each tree is subsequently used to calculate a final prediction result for each quantile (in this case, the 5^{th}, 50^{th} and 95^{th}). To optimise the QRF algorithm, the number of covariates used at each split (referred to as mtry) was adjusted and the mtry with the lowest root mean square error (RMSE) was selected as optimal. The model employed 500 trees and underwent a 10 k-fold cross validation. All model fitting was completed using the '*ranger'* function (Wright & Ziegler, 2017) within the *'caret'* package (Kuhn, 2022) in R.

Prior to fitting the QRF model, the soil P data was log-transformed as it displayed a strong right-skewed (positive) distribution (Fig. A1). A variance inflation factor (VIF) analysis was also conducted to reduce multicollinearity and the potential of model overfitting due to the high number of spatial covariates. This was completed separately for each soil forming factor including climate, parent material and relief using a VIF threshold of 10. This excluded 27 of the 41 climate covariates, 2 of the 18 relief covariates and 1 of the 14 parent material covariates.

The extracted soil P dataset used in the QRF model was split into a training and testing dataset for model evaluation. 80% of the dataset was split using the cLHS algorithm for training the model, while the remaining 20% was used for model validation. Model performance was evaluated based on the coefficient of determination (R^2), bias, RMSE, normalised RMSE (NRMSE) according to the difference between the maximum and minimum observed values and Lin's concordance correlation coefficient (CCC) (Lin, 1989). The final prediction maps for the 5^{th}, 50^{th} and 95^{th} percentiles were mapped with the best performing model and the uncertainty was calculated as:

$$\text{Uncertainty} = \frac{95^{\text{th}}\text{percentile} - 5^{\text{th}}\text{percentile}}{\text{Max value }(95^{\text{th}}\text{percentile} - 5^{\text{th}}\text{percentile})}$$

4. Results

4.1 Soil P data

Following the data cleaning process, a total of 6796 unique site location samples were identified across Queensland as suitable for modelling (Fig. 2). Of these, 228 samples were sourced in this project, either through soil sampling campaigns or archived data. The soil P concentrations across all of Queensland ranged from 1 to 150 mg kg^{-1} (Table 1). The Fitzroy grazing land management (GLM) region had the highest number of sites at 1340, followed by the Burdekin at 1034. The lowest number of sites (57) were found in the Mary GLM region. All other GLM regions had between 111 (Border Rivers) and 507 (Inland Burnett) sites with soil P data. In terms of concentrations by GLM region, the Darling Downs had the highest median and standard deviation of 33 and 48 mg kg^{-1}, respectively. This large standard deviation observed is mostly attributed to high variation in parent materials (lithology) across the Darling Downs. Cape York, Desert Uplands and Northern Gulf had the lowest median soil P values of less than 5 mg kg^{-1}.

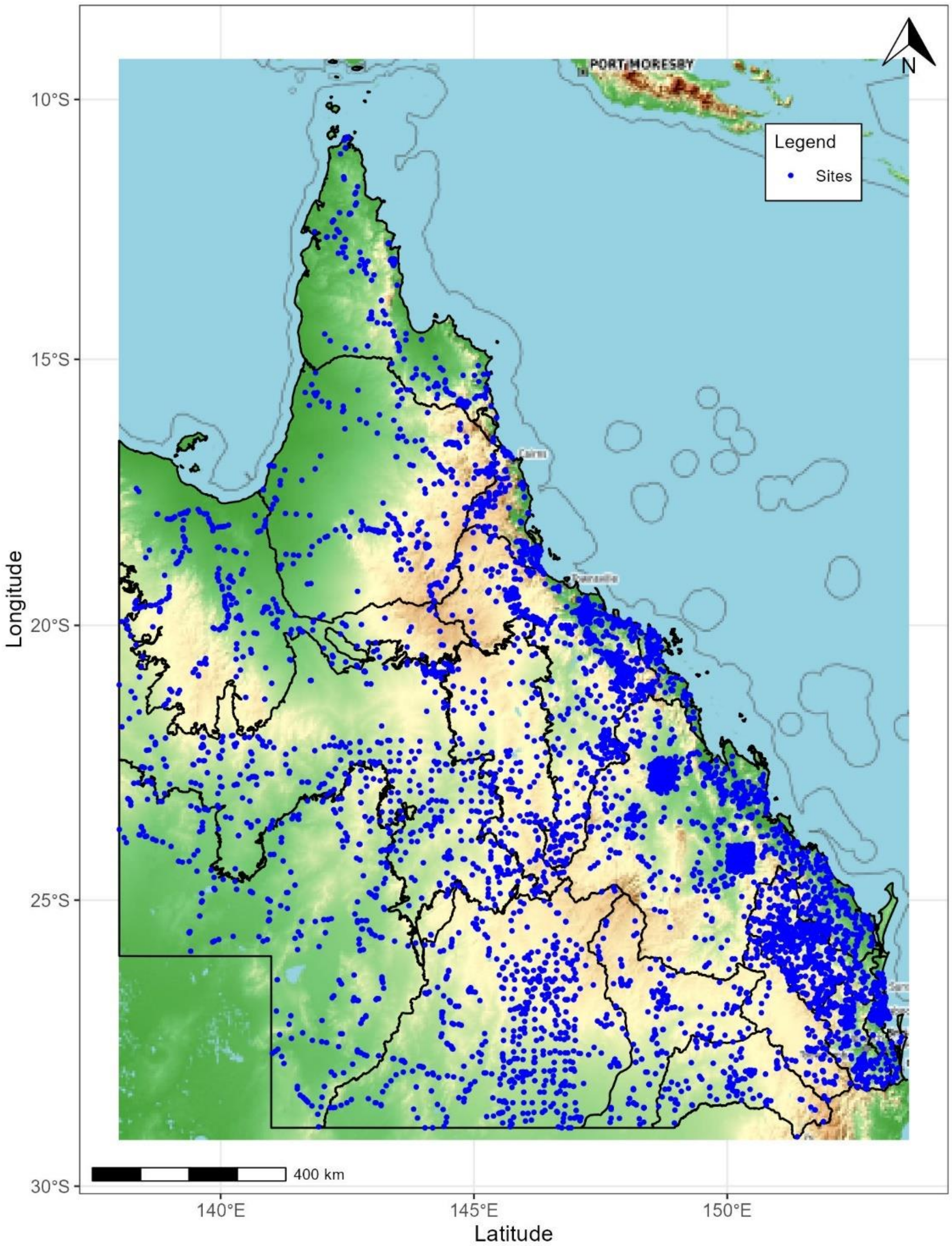

Figure 2. Soil sampling sites (blue dots) across Queensland used in the spatial modelling of soil P.

Table 1. Number of sites (*n*) and the minimum, mean, median, maximum and standard deviation (SD) of soil P values (mg kg^{-1}) within each Grazing Land Management (GLM) region across Queensland.

GLM region	*n*	Minimum	Mean	Median	Maximum	SD
Border Rivers	111	1	25	16	150	28.6
Burdekin	1034	1	15.3	8	142	19.1
Cape York	197	1	5.9	3	105	13.3
Channel Country	263	2	14.3	11	68	11.4
Coastal Burnett	260	1	11.9	7	140	17.2
Darling Downs	125	2	54.8	33	150	48
Desert Uplands	154	1	7.2	4	93	10.4
Fitzroy	1340	1	21.9	12	146	25.1

Inland Burnett	507	2	25.9	13	150	31
Mackay Whitsunday	178	1	15.2	9.5	100	17.3
Maranoa Balonne	177	1	22.8	14	150	26.5
Mary	57	2	15.6	9	83	15.7
Mitchell Grass Downs	425	2	15.5	12	91	11.6
Moreton	367	1	31.9	18	137	31.2
Mulga	477	1	13.1	8	129	13.6
Northern Gulf	258	1	7.6	4	63	9.7
South East	447	1	15.3	7	139	22
Southern Gulf	223	2	14	8.5	86	14.1
Wet Tropics	192	1	16.3	6	135	26

4.2 Model fit

The QRF model provided an overall moderate performance against the validation dataset with an R^2 of 0.4, bias of -4.28 mg kg^{-1}, RMSE of 18.4 mg kg^{-1}, NRMSE of 12.4% and CCC of 0.48 (Table 2). The performance of the statistical model in predicting soil P is consistent with or superior to that reported in other studies (Gray, 2023; Kaya et al., 2022; Shahbazi et al., 2019; Zund, 2022; Zund et al., 2022). Compared to the previous Queensland soil P map (Zund et al., 2022), an improvement in all model metrics were observed. The R^2, bias and CCC improved by 48%, 21% and 20%, respectively, while the RMSE reduced by 17%. The relatively high RMSE observed in the current study is mostly attributed to the large range in soil P values modelled in the dataset (1-150 mg kg^{-1}). However, the NRMSE provides a more realistic metric on the model accuracy as the RMSE has been normalised according to the difference between the maximum and minimum observed soil P values. The bias indicates that the QRF model tends to underestimate predictions compared to the observed soil P values, which is also likely attributed to the large range but also the highly-skewed distribution of soil P values (Fig. A1).

Table 2. The QRF model performance metrics (R^2, bias, RMSE, NRMSE and CCC) against the validation dataset.

Metric	Value
R^2	0.4
Bias	-4.28 mg kg^{-1}
RMSE	18.4 mg kg^{-1}
NRMSE	12.4%
CCC	0.48

4.3 Covariate importance

The covariate importance analysis indicated that parent material factors were the most important covariates in the QRF model (Table 3). Weathering index was identified as the most important covariate, which represents the degree primary minerals have been altered to secondary components via weathering processes. Weathering index was followed by equinox rainfall seasonality (ratio of spring to autumn rainfall) and several radiometric covariates representing parent material. The significance of these spatial covariates in the QRF model aligns with the P cycle, reflecting how P is gradually released from parent materials through weathering processes and is further influenced by the variances in P content inherent to different lithologies. Pre-settler major vegetation subgroups was also identified as an important covariate, representing native vegetation across the state, which can be a surrogate for soil and parent material variability. A few relief factors were also identified as important covariates in the QRF model, which are likely to represent the movement of P throughout the landscape via erosional processes.

Table 3. Covariate importance ranking of the spatial covariates used the QRF model and their respective soil forming factors.

Soil forming factor	Covariate information	Units	Importance
Parent Material	Weathering intensity	-	100
Climate	Ratio of Spring (Sep-Nov) to Autumn (Mar-May) cumulative precipitation	ratio	84.3
Parent Material	Radiometrics (uranium potassium ratio)	ratio	82.9
Parent Material	Radiometrics (thorium potassium ratio)	ratio	62
Parent Material	Radiometrics (potassium)	%	61.7
Organism	Pre-settler major vegetation subgroups	-	60.6
Soil	Estimated smectite in clay minerals (0-20 cm)	%	57.4
Climate	Maximum monthly minimum temperature	℃	49.9
Soil	Estimated illite in clay minerals (0-20 cm)	%	49.1
Parent Material	Estimated geology silica content	%	48.6
Climate	Minimum differences in atmospheric water deficit seasonality between successive months	mm	45.6
Organism	Potential biomass (dry matter)	tonnes ha^{-1}	39.2
Parent Material	Minimum geology age	-	30.2
Soil	Estimated kaolinite in clay minerals (0-20 cm)	%	24.3
Climate	Maximum differences in precipitation between successive months	mm	23.4
Climate	Annual thunder days	-	20.6
Parent Material	Radiometrics (thorium)	ppm	20.3
Relief	Digital elevation model	m	19.7
Climate	Maximum monthly mean diurnal temperature range	℃	18.6
Climate	Ratio of Summer (Dec-Feb) to Winter (Jun-Aug) cumulative precipitation	ratio	18.3
Parent Material	Total magnetic intensity (reduced to pole)	-	15.9
Climate	Minimum monthly precipitation	mm	15.8
Relief	Median slope over 300 m	%	15.4
Climate	Minimum difference in temperature between successive months	℃	10.1
Relief	Elevation range over 1000 m	m	9.3
Parent Material	Isostatic residual gravity	-	8.6
Climate	Maximum difference of shortwave solar radiation between successive months	MJ m^2 day^{-1}	5.2

Soil forming factor	Covariate information	Units	Importance
Climate	Maximum difference of potential evaporation between successive months	mm	3.5
Relief	Prescott index	-	1.9
Climate	Annual potential evaporation	mm	0

4.4 Final prediction maps

The final prediction map of soil P across all of Queensland is displayed in Fig. 3, aligning with the seven P categories in Table 4. The resolution of the map is projected at 30 m, which is a large improvement compared to the previous soil P map at 90 m (Zund et al., 2022). The continuous layer ranges from 1 to 150 mg kg^{-1}, which will be available on QSpatial and integrated into LongPaddock upon review and acceptance of this report. The highest reported soil P values were observed in the Darling Downs region, while the lowest soil P values were mostly observed in the Northern Gulf and Cape York regions. 30% of Queensland was predicted to have deficient to acutely deficient (<6 mg kg^{-1}) concentrations of soil P (Table 4). The largest portion of the state was predicted to have moderate soil P concentrations (24%). Only 10% and 3% of the state were predicted to have high (16-25 mg kg^{-1}) and very high (>25 mg kg^{-1}) concentrations of soil P, respectively.

Table 4. Area of Queensland (%) covered by each of the seven P categories.

P Category	P Range (mg kg^{-1})	Area of Queensland (%)
Acutely deficient	≤4 mg kg^{-1}	18
Deficient	4 – ≤6 mg kg^{-1}	12
Marginal	6 – ≤8 mg kg^{-1}	17
Low	8 – ≤10 mg kg^{-1}	16
Moderate	10 – ≤16 mg kg^{-1}	24
High	16 – ≤25 mg kg^{-1}	10
Very high	>25 mg kg^{-1}	3

There were a few artefacts (visual anomalies) identified in the map due to the spatial covariates. One of these was caused from weathering index covariate, which can be seen starting in Jericho and travelling south to Bayrick. This artefact becomes less noticeable the further you travel east. There are also a few small areas across the state on K’gari, Stradbroke Island and Cape Gloucester with missing data due to gaps in the spatial covariates.

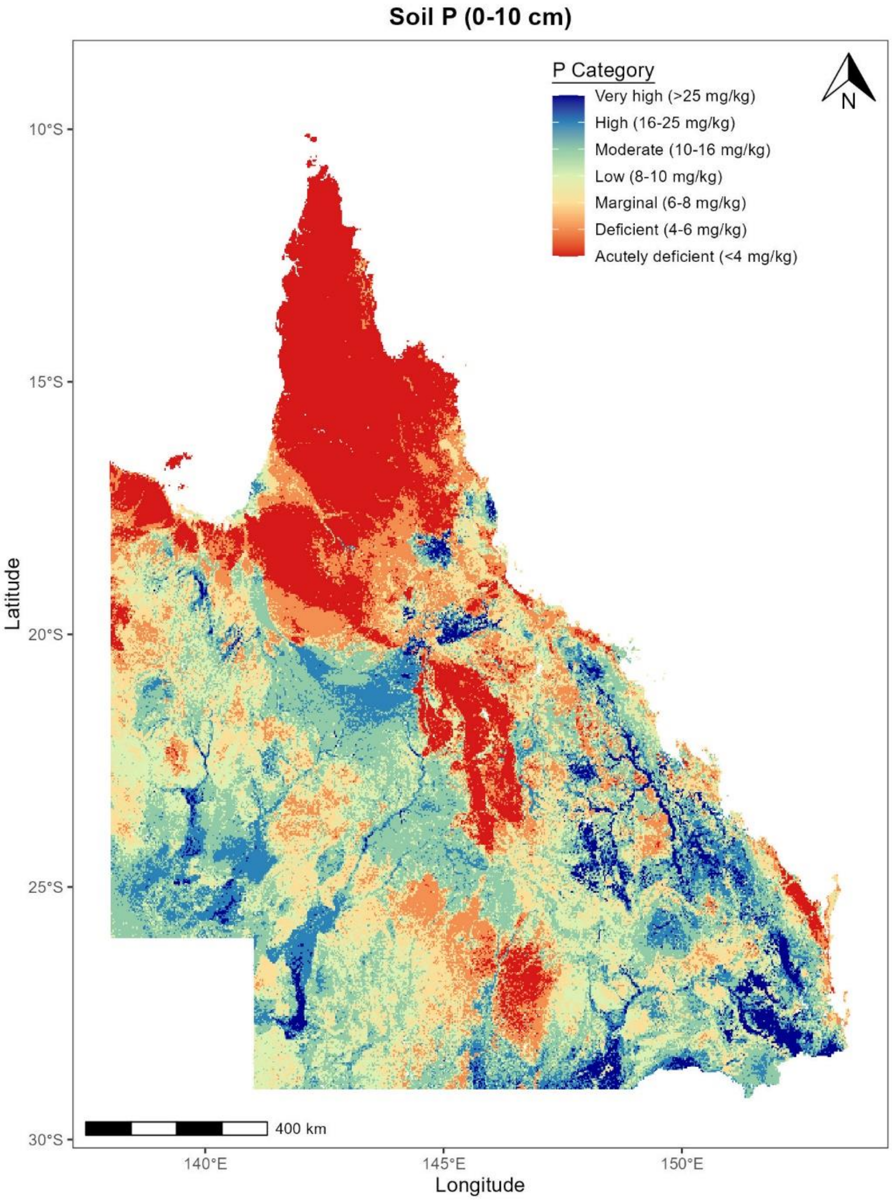

Figure 3. Soil P predictions in the 0-10 cm layer across Queensland categorised into seven P classes.

The uncertainty predictions followed a similar pattern to the soil P predictions, with high uncertainty associated with high concentrations of predicted soil P (Fig. 4). Areas within the Darling Downs, around Innisfail and Central Queensland all displayed high levels of uncertainty. The higher uncertainty is likely to be generated from a large range in observed data within that region or in some cases, from limited data. Low uncertainty levels were mostly associated with areas within the Northern Gulf and Cape York regions, where low soil P values were relatively consistent.

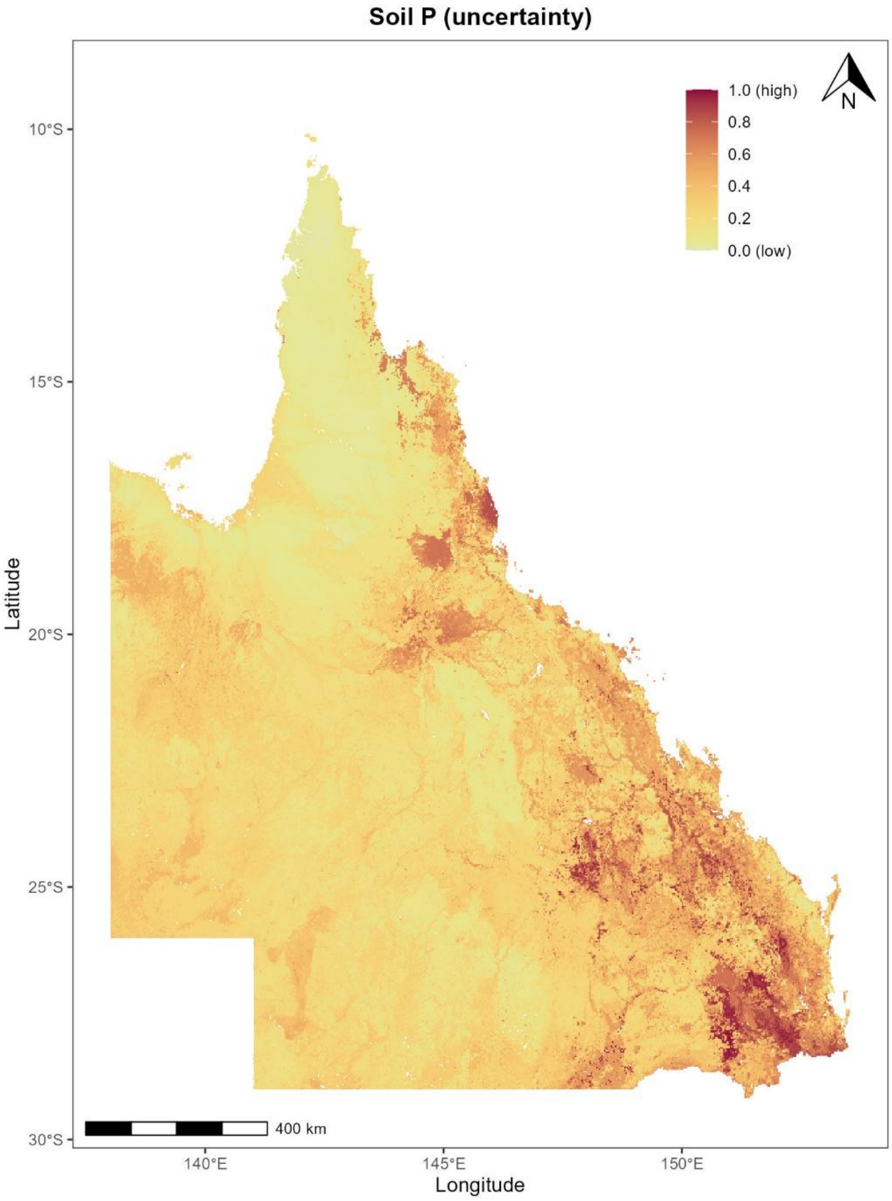

Figure 4. Uncertainty (uncertainty range/max value of uncertainty range) of soil P predictions across Queensland.

4.4.1 2022 vs 2024 soil P map

There are some significant differences in the soil P predictions between that developed in this project and the previous Queensland soil P map (Zund et al., 2022), as shown in Fig. 5. The differences are likely attributed to the increase in training data and a different statistical model used. A larger training dataset can impact how the model is developed by building different relationships between the observed soil P data and the spatial covariates. This is particularly true given that the northern regions had extremely low soil P values, which was not in the previous map. Furthermore, a quantile random forest (QRF) model was used in this study compared to a cubist model in the previous map (Zund et al., 2022). A QRF model builds multiple independent decision trees using a random subset of the data. In contrast, a cubist model is a rule-based decision tree model which builds linear regression models at the terminal nodes of each tree (Quinlan, 1992). A QRF model was selected for this study as the QRF model demonstrated greater prediction accuracy compared to cubist, gradient boosting machines and multiple linear regression models in previous milestones but also in other studies (e.g. Gomes et al., 2019; Zeraatpisheh et al., 2019).

The main differences between the two maps are concentrated in the areas of high uncertainty (Fig. 5). 16% of the total area across the new and old soil P map was calculated to have zero change in soil P values, while 60% of the total area had an absolute difference between 0 and 2 mg kg^{-1} (

Table 5.). 86% of the total area was found to have less than a 5 mg kg^{-1} absolute difference between the new and old soil P map.

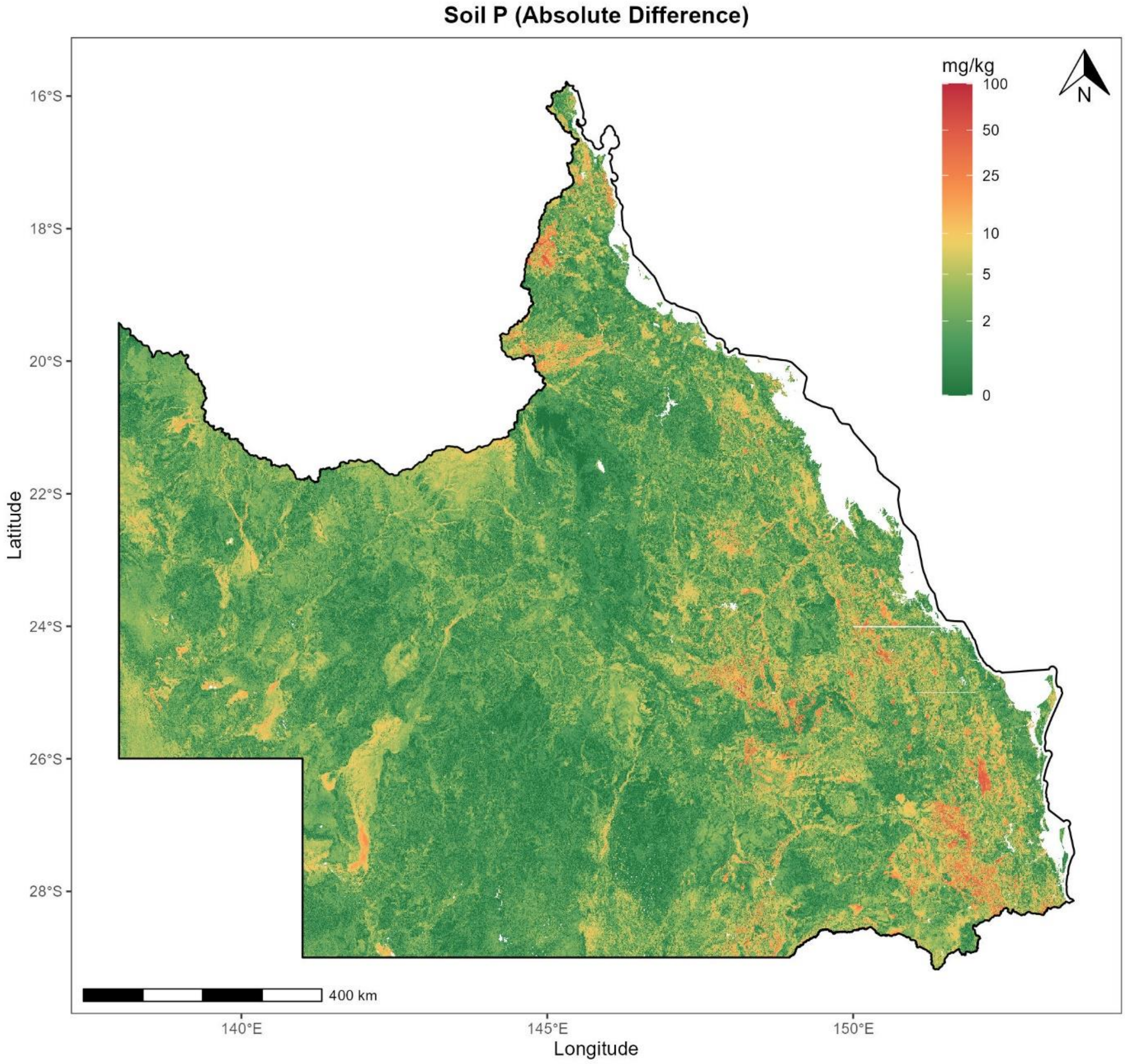

Figure 5. Absolute differences in soil P (mg kg^{-1}) predictions between the previous map (Zund et al., 2022) and the soil P map developed in this study (Fig. 3).

Table 5. Percentage of mapping area (%) categorised by the absolute differences in soil P (mg kg^{-1}) predictions between the previous soil P map (Zund et al., 2022) and the soil P map developed in this study (Fig. 3).

P difference (mg kg^{-1})	Map area (%)
0	16
≤1	42
≤2	60
≤3	72
≤4	80
≤5	86

5. Conclusion

5.1 Key findings

The key findings of this project are the development of a soil P map (plus uncertainties) for north Queensland, and the production of a revised soil P map for all of Queensland at a 30 m resolution. The maps provide an accurate estimate of soil P, particularly in areas of low uncertainty. The project also developed a digital soil mapping framework which is easily repeatable and updatable as new data is provided.

5.2 Benefits to industry

The soil P map developed provides graziers and stakeholders with a prediction of soil bicarbonate extractable P across the entire state of Queensland. The soil P predictions can be used to identify P-deficient soils to inform P management via supplementation. The map also provides levels of uncertainty, which assists graziers in determining whether additional soil sampling is required, particularly given no soil mapping across Queensland is conducted at a sufficient intensity to be used at the paddock level.

The map developed could further aid the red meat industry in:

- identifying areas with sufficient P suitable for the introduction of legume pastures such as Leucaena, Stylos and Desmanthus
- identifying areas with high uncertainty to guide future soil sampling campaigns
- raising awareness of the soil P cycle and P deficiencies in grazing systems and how this can impact overall productivity
- identifying areas where soil P depletion might be occurring and how different grazing management strategies can be implemented to prevent this.

6. Future research and recommendations

The outcomes of this project identified several key future research directions including:

- The development of a P buffering index map to complement the soil P map and provide an accurate estimate of plant available P as the P buffering index indicates the soil’s capacity to adsorb and release P,
- The extension of the updated mapping methodology into the northern regions of Australia including the Northern Territory and Western Australia which are known to be P-deficient, and
- The location of areas with high uncertainty where future sampling campaigns could target to improve model predictions and reduce uncertainty.

Further recommendations include organising workshop activities to promote the use of the soil P map and how it can be integrated into P management plans for the red meat industry.

7. References

Ahern, C. R. (1994). *The soil fertility of central and north-east Queensland grazing lands, ISBN 0724259201*. Queensland Department of Primary Industries, Q194065.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324/METRICS

Department of Climate Change, Energy, the Environment and Water. (2023a). *Australia—Pre-1750 Major Vegetation Groups—NVIS Version 6.0 (Albers 100m analysis product)* [Dataset]. https://fed.dcceew.gov.au/maps/992a30c936c6445c9f07a60c1036d519/about

Department of Climate Change, Energy, the Environment and Water. (2023b). *Australia—Pre-1750 Major Vegetation Subgroups—NVIS Version 6.0 (Albers 100m analysis product)* [Dataset]. https://fed.dcceew.gov.au/maps/cde3109f9105411c88901e69e7f0ccbe/about?_gl=1*1yvz410*_ga*MTk0OTIxNzY3NS4xNjgwNTA2ODEw*_ga_1M2TBC9WWS*MTY5MzQ2Mzk3NC4xOS4xLjE2OTM0NjQ0MzIuMC4wLjA.

Dixon, R. M., Anderson, S. T., Kidd, L. J., & Fletcher, M. T. (2020). Management of phosphorus nutrition of beef cattle grazing seasonally dry rangelands: A review. *Animal Production Science*, *60*(7), 863. https://doi.org/10.1071/AN19344

Gomes, L. C., Faria, R. M., De Souza, E., Veloso, G. V., Schaefer, C. E. G. R., & Filho, E. I. F. (2019). Modelling and mapping soil organic carbon stocks in Brazil. *Geoderma*, *340*, 337–350. https://doi.org/10.1016/j.geoderma.2019.01.007

Gray, J. (2023). *Digital soil mapping of key soil properties over New South Wales, version 2.0*. NSW Department of Planning and Environment.

Kaya, F., Keshavarzi, A., Francaviglia, R., Kaplan, G., Başayiğit, L., & Dedeoğlu, M. (2022). Assessing machine learning-based prediction under different agricultural practices for digital mapping of soil organic carbon and available phosphorus. *Agriculture*, *12*(7), 1062. https://doi.org/10.3390/agriculture12071062

Kuhn, M. (2022). *caret: Classification and Regression Training* (Version R package version 6.0-93) [Computer software]. https://CRAN.R-project.org/package=caret

Lin, L. I.-K. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, *45*(1), 255. https://doi.org/10.2307/2532051

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, 281–297.

McBratney, A. B., Mendonça Santos, M. L., & Minasny, B. (2003). On digital soil mapping. *Geoderma*, *117*(1–2), 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4

McCosker, T., & Winks, L. (1994). *Phosphorus nutrition of beef cattle in northern Australia, ISBN 072425417X*. Information series (Queensland. Department of Primary Industries) QI94012.

Minasny, B., & McBratney, A. B. (2006). A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Computers and Geosciences*, *32*(9), 1378–1388. https://doi.org/10.1016/j.cageo.2005.12.009

Quinlan, J. R. (1992). Learning with Continuous Classes. *Proceedings of Australian Joint Conference on Artificial Intelligence*, 343–348.

R Core Team. (2021). *R: A language and environment for statistical computing* [Computer software]. https://www.R-project.org/

Rayment, G. E., & Lyons, D. J. (2011). Soil Chemical Methods—Australasia. In *Soil Chemical Methods—Australasia*. CSIRO Publishing. https://doi.org/10.1071/9780643101364

Roxburgh, S. H., Karunaratne, S. B., Paul, K. I., Lucas, R. M., Armston, J. D., & Sun, J. (2019). A revised above-ground maximum biomass layer for the Australian continent. *Forest Ecology and Management*, *432*, 264–275. https://doi.org/10.1016/j.foreco.2018.09.011

Schatz, T. J., McCosker, K. D., & Heeb, C. (2023). Phosphorus supplementation improves the growth and reproductive performance of female Brahman cattle grazing phosphorus-deficient pastures in the Victoria River District, Northern Territory, Australia. *Animal Production Science*, *63*(6), 544–559. https://doi.org/10.1071/AN21565

Searle, R., Malone, B., Wilford, J., Austin, J., Ware, C., Webb, M., Roman Dobarco, M., & Van Niel, T. (2022). *TERN Digital Soil Mapping Raster Covariate Stacks* [Computer software]. CSIRO. https://doi.org/10.25919/jr32-yq58

Shahbazi, F., Hughes, P., McBratney, A. B., Minasny, B., & Malone, B. P. (2019). Evaluating the spatial and vertical distribution of agriculturally important nutrients—Nitrogen, phosphorous and boron—In North West Iran. *CATENA*, *173*, 71–82. https://doi.org/10.1016/j.catena.2018.10.005

Winks, L. (1990). Phosphorus and beef production in northern Australia. 2. Responses to phosphorus by ruminants: A review. *Tropical Grasslands*, *24*, 140–158.

Wright, M. N., & Ziegler, A. (2017). Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, *77*(1). https://doi.org/10.18637/jss.v077.i01

Yang, L., Li, X., Shi, J., Shen, F., Qi, F., Gao, B., Chen, Z., Zhu, A. X., & Zhou, C. (2020). Evaluation of conditioned Latin hypercube sampling for soil mapping based on a machine learning method. *Geoderma*, *369*, 114337. https://doi.org/10.1016/j.geoderma.2020.114337

Zeraatpisheh, M., Ayoubi, S., Jafari, A., Tajik, S., & Finke, P. (2019). Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. *Geoderma*, *338*, 445–452. https://doi.org/10.1016/j.geoderma.2018.09.006

Zund, P. (2022). *Soil and Landscape Grid National Soil Attribute Maps—Available Phosphorus (3" resolution)—Release 1* (Version 1) [Dataset]. CSIRO. Data Collection. https://doi.org/10.25919/6qzh-b979

Zund, P., Walton, J., De Antoni, M., Harms, B., & Thomas, E. (2022). *Soil bicarbonate-extractable P (Colwell-P) map of Queensland main grazing lands*.

8. Appendix

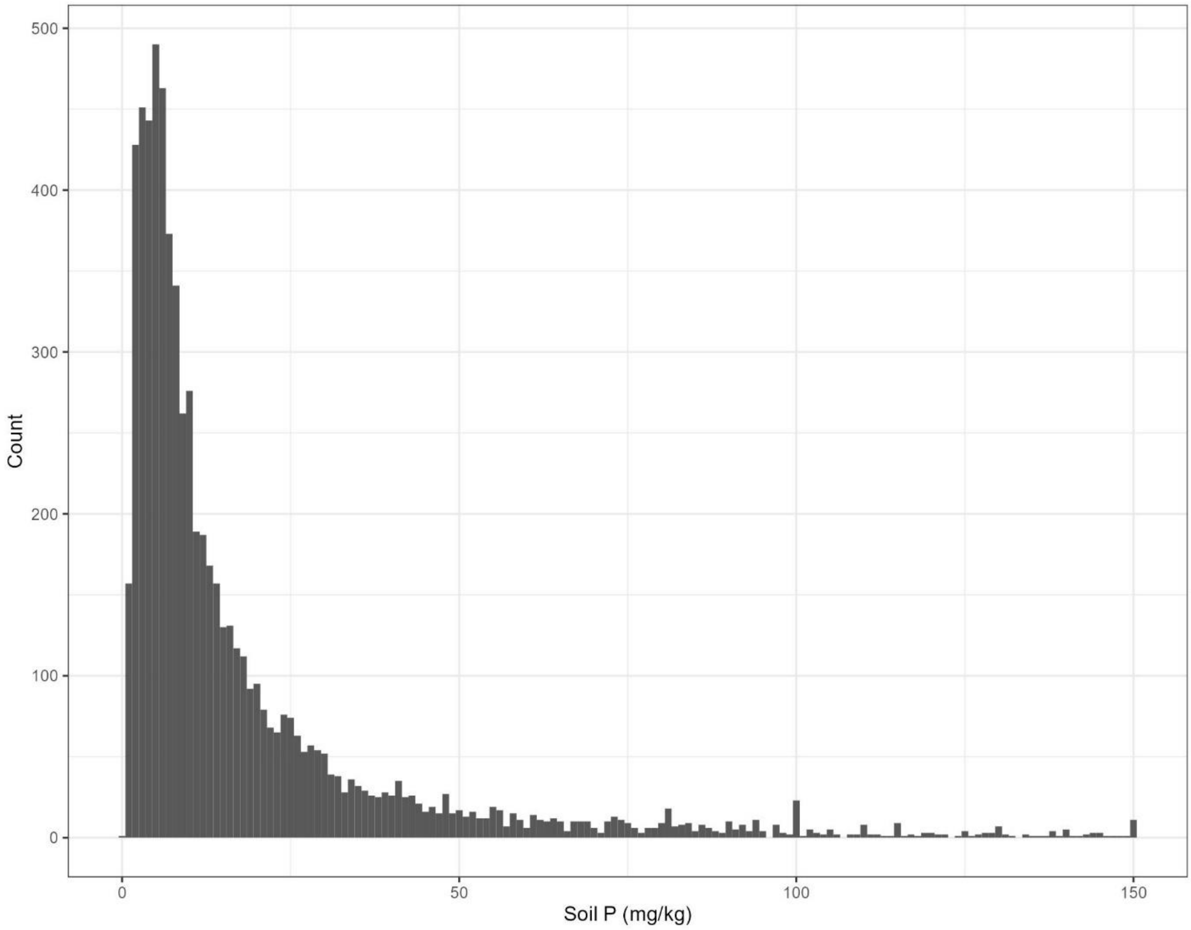

Figure A 1. Distribution of soil P values (mg kg^{-1}) across Queensland in the modelling dataset.