Digital soil mapping of lithium in Australia

Wartini Ng[1*], Budiman Minasny[1], Alex McBratney[1], Patrice de Caritat[2], John Wilford[2]

[1] Sydney Institute of Agriculture, School of Life and Environmental Sciences, The University of Sydney, NSW 2015, Australia
[2] Geoscience Australia, Canberra, ACT 2601, Australia

5 *Correspondence to*: Wartini Ng (wartini.ng@sydney.edu.au)

Highlights

- Machine learning can be used to relate soil data and environmental covariates
- The first Australian digital soil map of lithium concentration is presented
- The prediction map can be used to delineate potential areas of anomalous Li
10
- Elevated soil Li observed and modelled in a number of States

Abstract

With a higher demand for lithium (Li), a better understanding of its concentration and spatial distribution is important to delineate potential anomalous areas. This study uses a digital soil mapping framework to combine data from recent

15 geochemical surveys and environmental covariates that affect soil formation to predict and map aqua regia-extractable Li content across the 7.6 million km^2 area of Australia. Catchment outlet sediment samples (i.e. soils formed on alluvial parent material) were collected by the National Geochemical Survey of Australia at 1315 sites, with both top (0 – 10 cm depth) and bottom (on average ~60 – 80 cm depth) catchment outlet sediments sampled. We developed 50 bootstrap models using a Cubist regression tree algorithm for each depth. The spatial prediction models were validated on an independent Northern Australia

20 Geochemical Survey dataset, showing a good prediction with a root mean square error of 3.32 mg kg^{-1} (which is 44.2 % of the inter-quartile range) for the top depth. The model for the bottom depth has yet to be validated. The variables of importance for the models indicated that the first three Landsat 30+ Barest Earth bands (red, green, blue) and gamma radiometric dose have a strong impact on the development of regression-based Li prediction. The bootstrapped models were then used to generate digital soil Li prediction maps for both depths, which could identify and delineate areas with anomalously high Li

25 concentrations in the regolith. The predicted maps show high Li concentration around existing mines and other potentially anomalous Li areas that have yet to be verified. The same mapping principles can potentially be applied to other elements. The Li geochemical data for calibration and validation are available at: (de Caritat and Cooper, 2011b; http://dx.doi.org/10.11636/Record.2011.020) and (Main et al., 2019; http://dx.doi.org/10.11636/Record.2019.002) respectively. The covariates data used for this study were sourced from the Terrestrial Ecosystem Research Network (TERN)

30 infrastructure, which is enabled by the Australian Government's National Collaborative Research Infrastructure Strategy

(NCRIS) https://esoil.io/TERNLandscapes/Public/Products/TERN/Covariates/Mosaics/90m/ (TERN, 2019). The final predictive map is available at: https://doi.org/10.5281/zenodo.7895482.

Keywords: digital soil mapping, lithium, machine learning, geochemical survey, critical minerals, anomalous, critical element

35 **1   Introduction**

Minerals have become essential commodities in modern human society. Many minerals are fundamental to technological and industrial advancement, particularly those utilised in renewable energy systems, electric vehicles, consumer electronics and telecommunications (Kabata-Pendias, 2010). These minerals can be considered critical, in the sense that they are of high importance and have a high risk of supply disruption. Methods for quantifying mineral criticality are discussed in detail in

40 Graedel et al. (2012).

Lithium (Li) is an important chemical element as the world transitions towards a lower-carbon economy. It has been listed as a critical element by various countries, including Australia, Canada, the European Union, Japan, the Republic of Korea and the United States of America (Mudd et al., 2018; David Huston, Geoscience Australia, pers. comm. March 2022). Australia is endowed with significant resources of many of the critical elements and the critical minerals hosting them, including Li.

45 Currently, Australia's ranking for economic resources of Li is second, but it ranks first for its production (Senior, 2022), with potential for additional discoveries. According to a recent survey (Senior, 2022), Australia produced 40 kilotons (kt) of Li (in terms of spodumene, $LiAlSi_2O_6$, concentrates; assuming 6% of Li_2O in spodumene concentrates) in 2020, or 49% of the global production; a significant increase from 21.3 kt of Li in 2017 (Champion, 2019).

The two primary sources for Li are brine stores and mineral deposits, where Li is hosted mainly in spodumene. A 2013

50 investigation by Geoscience Australia found that the potential of Li-rich salt lakes in Australia was relatively low in comparison to those, for instance, in the Americas (Jaireth et al., 2013; Mernagh et al., 2013; Mernagh et al., 2016). Most of the Li in Australia exists as mineral deposits (Champion, 2019). Despite Australia's current position as the world's leading supplier of Li, it has limited prospects for immediate expansion as the potential for similar deposits in Australia has not yet been fully investigated (Mudd et al., 2018). This study aims to contribute to filling this knowledge gap by providing the first digital map

55 of Li concentration in Australian soils.

Lithium values ranges from $<1 – 15$ mg kg^{-1} in ultramafic rocks, $5.5 – 17$ mg kg^{-1} in mafic rocks, whereas felsic rocks (granite, rhyolite and phonolite) contain higher Li concentrations, between $30 – 70$ mg kg^{-1} (de Vos et al., 2006). Lithium concentration in clay minerals ranges between $7 – 6000$ mg kg^{-1} (Starkey, 1982). With developments in technology, a process of extracting Li as Li-carbonate from certain minerals, other than spodumene, such as lepidolite ($KLi_2Al(Si_4O_{10})(F,OH)_2$) and petalite

60 ($LiAlSi_4O_{10}$), has been identified (Sitando and Crouse, 2012; Vieceli et al., 2018). Lower Li concentration is found in salt lake brines ($0.17 – 1.5$ mg kg^{-1}) (Grosjean et al., 2012). Extraction of Li from salt lake brine is in the form of Li-chloride, which needs to undergo an energy-intensive process to be converted to Li-carbonate from the Li-metal forms for use in batteries.

Lithium is found in trace amounts in all soil types, primarily in the clay fraction, with slightly lower concentrations in the organic soil fraction (Kabata-Pendias, 2010). Possible means by which Li is bound to clay have been reviewed elsewhere (Starkey, 1982). Across Europe, values of Li ranging from 0.28 – 271 mg kg^{-1} have been reported (Salminen et al., 2006), with smaller concentration ranges in agricultural soil (0.161 – 136 mg kg^{-1}) and grazing soil (0.1 – 153 mg kg^{-1}) (Reimann et al., 2014). Négrel et al. (2019) reported aqua regia soluble Li concentration of 11.3 mg kg^{-1} in European agricultural soil. In New Zealand, a study of Li concentration in soil reported a range between 0.08 – 92 mg kg^{-1} (Robinson et al., 2018). de Caritat and Reimann (2012) reported median Li concentrations (after aqua regia digestion) of 12 and 5.7 mg kg^{-1} in European agricultural topsoils and Australian surface sediments, respectively, both in the coarse (< 2 mm) fraction. Subsequently, Reimann and de Caritat (2017) published the first continental map (Supplementary Material; Fig.2SM) of Li in Australian soils, based on National Geochemical Survey of Australia (NGSA) data, showing that regions of high and low concentrations are found across all States. The amount of soil-available Li in has been found to be relatively low, about 3 – 5% of the total Li content in the surface layers both in south-eastern of USA (Anderson et al., 1988) and Siberia (Gopp et al., 2018), ranging from 0.24 – 0.68 mg kg^{-1}. A total Li concentration within a range of 5.27 – 400 mg kg^{-1} had been reported for catchment sediment samples in China (Liu et al., 2020) and within a range of <1 – 300 mg kg^{-1} in the USA topsoils (Smith et al., 2019).

Higher concentrations of Li are often found in the deeper layers of soil profiles (Merian and Clarkson, 1991). Typically, Li enters the soil column through the weathering of sedimentary minerals in the underlying saprolite and bedrock (Aral and Vecchio-Sadus, 2008). Because clay minerals predominantly drive the mineralisation and dissolution of Li, the clay mineral fraction will play a significant role in determining the Li concentration. The Li content of soil is controlled more by the soil formation conditions than by the composition of the parent materials (Kabata-Pendias, 2010). Similar observations are found in Négrel et al. (2019), where the aqua regia-extractable Li concentrations can be linked with known mineralisation process observed within Europe. This was also shown in the study by Luecke (1984) who explored the use of the enriched elements (Rb, Ba, Sr, Cu and Zn among others) information to aid predicting the distribution of Li pegmatites.

Mineral exploration aims to find ore deposits for mining purposes. Therefore, delineating target areas for mineral exploration through a series of mapping activities is a crucial initial stage leading to discovery (Carranza, 2011). Mineral prospectivity mapping (or modelling; MPM) is a method to quantify the probability of mineralisation in a selected area for mineral exploration purposes (Zuo, 2020). This prioritisation allows for the selection of smaller, higher-potential areas for detailed prospecting investment to minimise exploration costs, e.g. the number of drillholes.

Two common paradigms for creating MPM are knowledge-driven and data-driven models (Carranza, 2011). Knowledge-driven models do not require any data on mineral deposits, but rely on expert knowledge of spatial associations between mineral deposits and geological features, field experience and conceptual models to develop evidential maps that enables the discovery of mineral deposit (Carranza, 2008). Conversely, data-driven models utilise existing knowledge on the location of mineral occurrences, various survey datasets and spatial statistical methods to represent the likelihood of mineral occurrence within prospective areas (Carranza, 2008). Numerous data-driven models have been derived for the detection of anomalous mineral occurrences. Benedikt (2018) utilised Tellus regional stream sediment geochemistry to screen for anomalous metal

abundances within minerals in Southeast Ireland. Roshanravan et al. (2023) and Harris et al. (2023) also implemented data-driven machine learning model to develop predictive maps of gold prospects.

With the development of machine learning and technology (computer hardware, software and geographic information system (GIS) technology), there have been growing applications of MPM in the decades (Carranza, 2011; Porwal et al., 2015; Zuo, 2020). Several studies have demonstrated the use of remote sensing to explore various deposit types, such as gold (Au) deposits (Crósta et al., 2010), copper (Cu) deposits (Pour and Hashim, 2015) and iron (Fe) ores (Ducart et al., 2016). The application of remote sensing for Li deposits has also emerged. Gopp et al. (2018) explored the use of Normalised Difference Vegetation Index (NDVI) to develop a predicted map of the plant available content of Li in southwestern Siberia soil. Cardoso-Fernandes et al. (2018) and Cardoso-Fernandes et al. (2020) evaluated the potential use of Sentinel-2 in Li mapping in the Fregeneda-Almendra region across the Spain-Portugal border. Similarly, Köhler et al. (2021) further explored the use of combined geological data and Sentinel-2 data for Li potential mapping in Portugal. Antezana Lopez et al. (2023) used Sentinel-2, ASTER, JILIN GP, and PROBA CHRIS satellite data to study surface reflectance, as well as soil physicochemical properties to predict Li concentration in Bolivian salt flats.

In soil science, digital soil mapping (DSM) has been widely used to produce quantitative maps of soil attributes based on the known distributions of environmental covariates (i.e. rainfall, parent material, vegetation and landforms), that affect soil formation. The DSM framework is derived from the conceptual model developed by McBratney et al. (2003) in which a certain soil attribute results from the interaction of soil-forming factors. These factors are modified from Jenny (1941) and include soil (s), climate (c), organisms (o), relief (r), parent material (p), age/time (a) and spatial position (n), or *scorpan*. The factors are measured or approximated from various data types, including point observations, maps (polygons), survey data, and remote sensing data and derivatives thereof (e.g. gradients, buffer distances, etc.); these can be numerical or categorical data types.

In this study, we attempt to model Li distribution in the surface and subsurface soils of Australia by invoking the NGSA soil geochemistry dataset and various environmental covariates commonly used in DSM related to soil formation in Australia. In detail, the objectives of this study are thus to:

(i) evaluate the use of DSM framework to predict Li concentration in Australian soils, and

(ii) delineate anomalous areas potentially attractive for Li exploration and discuss their interpretations.

2 Materials and methods

2.1 Li measurement

This study used two soil datasets, referred to as the calibration and validation datasets. The calibration dataset was used to build the spatial prediction model and the validation dataset was used to test the prediction quality of the calibrated model. The calibration dataset data were generated as part of the NGSA project (www.ga.gov.au/ngsa), a collaborative project between Geoscience Australia and the States/NT between 2007 – 2011, which aimed to document the soil geochemical concentration

levels and patterns across Australia. Details on the project, analysis, sampling methods and the measurement of other parameters can be found in de Caritat and Cooper (2011b), de Caritat and Cooper (2015) and (de Caritat, 2022).

130 The NGSA collected samples at 1315 sites (including field duplicates) at or near the outlet of large catchments with a total area coverage of 6.17 million km^2 and an average sampling density of 1 site for every 5200 km^2 (de Caritat and Cooper, 2011b). The target sampling medium was floodplain sediments away from river channels, though in various places in Australia, aeolian modification of floodplain sediments can be important; thus, the medium was called 'catchment outlet sediment' rather than floodplain sediment. These geomorphological entities are typically vegetated and biologically active (plants, worms, ants,

135 etc.), thereby making the collected materials true soils (e.g., SSSA, 2022), albeit soils all developed on transported alluvium parent material. Due to limitations in access, samples from some parts of South Australia and Western Australia could not be obtained.

Samples were collected from two depths, namely 'top outlet sediment' (TOS) from 0 – 10 cm depth, and 'bottom outlet sediment' (BOS) from, on average, ~60 – 80 cm depth. All of the samples were air-dried, homogenised and dry sieved to <2

140 mm and <75 µm prior to various analyses for 60+ elements (see de Caritat et al. (2009) and de Caritat et al. (2010), for a full description of the NGSA sample preparation and analytical methods, respectively).

In this contribution, we use Li concentration after aqua regia digestion, as the NGSA did not report total Li. A 0.50 ± 0.02 g aliquot of sample (<2 mm) was digested in aqua regia (1.8 mL of HCl + 0.6 mL of HNO_3) at 90 ± 3°C for 2 hours to leach acid-soluble components. Once the sample had cooled to room temperature, 17.5 mL of diluent was added, and the sample

145 was inverted 10 times to homogenise the content. The sample was further diluted 50 times prior to analysis, using inductively coupled plasma mass-spectrometry (ICP-MS) in a commercial laboratory (de Caritat et al., 2010). For the remainder of the paper, any reference to Li concentrations is understood to mean aqua regia-extractable Li unless otherwise noted. Any Li measurements that fell below the detection limit (0.1 mg kg^{-1}) were replaced with half the detection limit (0.05 mg kg^{-1}). A detailed quality assessment of the NGSA data is given in de Caritat and Cooper (2011a), where a relative analytical precision

150 (repeat analysis of TILL-1 Certified Reference Materials (CRM)) of 12% and a relative overall precision (based on field duplicates) of 39% were reported. The distribution of sampling sites and Li concentration levels for both TOS and BOS are shown in Figure 1.

As an independent validation dataset, we used the geochemical dataset from the Northern Australia Geochemical Survey (NAGS) project (Main et al., 2019). This dataset contains 773 observations located in the Tennant Creek – Mt Isa region in

155 the Northern Territory and Queensland, with an approximate sampling density of one sample every 500 km^2 and collected in 2017. The distribution of these samples is also shown in Figure 1. These samples were collected, prepared and analysed following the NGSA protocols (de Caritat and Cooper, 2011b), albeit at a higher sampling density. However, only TOS samples were collected in NAGS. Furthermore, these NAGS samples were collected at a different time and analysed in a different laboratory compared to the NGSA dataset. To address the analytical variation that could potentially arise, a levelling

160 method was applied using the TILL-1 CRM standards (Main and Champion, 2022). First, the subset of the NGSA dataset that covers the spatial area of the NAGS dataset was extracted. Then a Kolmogorov-Smirnov test was used to verify if the samples

from the two datasets (subset of the NGSA and NAGS) were similar. A correction factor to relate the two datasets based on the TILL-1 CRM standards was then calculated and applied as a multiplier to the NAGS dataset to level its data to the NGSA dataset.

165 [Figure 1]

2.2   Environmental covariates

A total of 19 environmental covariates (Table 1) characterising the factors of climate, parent material, soil, and topography, which contribute to soil formation, were considered in this study.

The first factor is climate. Water (humidity) and temperature affect the rate of mineral weathering and thus soil formation.
170 Hence, we included precipitation, evaporation and temperature data (Harwood, 2019), along with the topographic wetness index (TWI) data (Gallant and Austin, 2012b), informing on the relative wetness within a landscape. In short, the TWI was derived from the partial contributing area product, which was computed from a Hydrologically enforced Digital Elevation Model, and from the percent slope product, which was computed from the Smoothed Digital Elevation Model (DEM-S; Gallant and Austin, 2012b).

175 The second factor is parent material (i.e. degree of weathering and mineralogical composition), including gamma-ray radiometric and total magnetic intensity. Gamma-ray radiometric surveys provide estimates for the concentrations of gamma-ray-emitting radioelements potassium (K), uranium (U) and thorium (Th) at/near the soil surface. The gamma-ray radiometric data were measured from airborne surveys throughout most of Australia (Poudjom Djomani et al. (2019). In this study, we used a complete gamma-ray survey grid where gaps in the airborne coverage were filled in using covariate machine learning
180 (Wilford and Kroll, 2020). Gamma-ray radiometric data have been found to be a useful covariate in identifying surface processes such as sediment transport and weathering (Wilford, 2012; Wilford et al., 1997) and detecting radioactive mineral deposits and occurrences (Alhumimidi et al., 2021; Dickson et al., 1996; Dickson and Scott, 1997; Wilford et al., 2009). Total magnetic intensity (TMI), which measures variations in the Earth's magnetic field intensity caused by the contrasting content of various rock-forming minerals in the crust (Poudjom Djomani et al., 2019), could also potentially identify geological
185 features and processes.

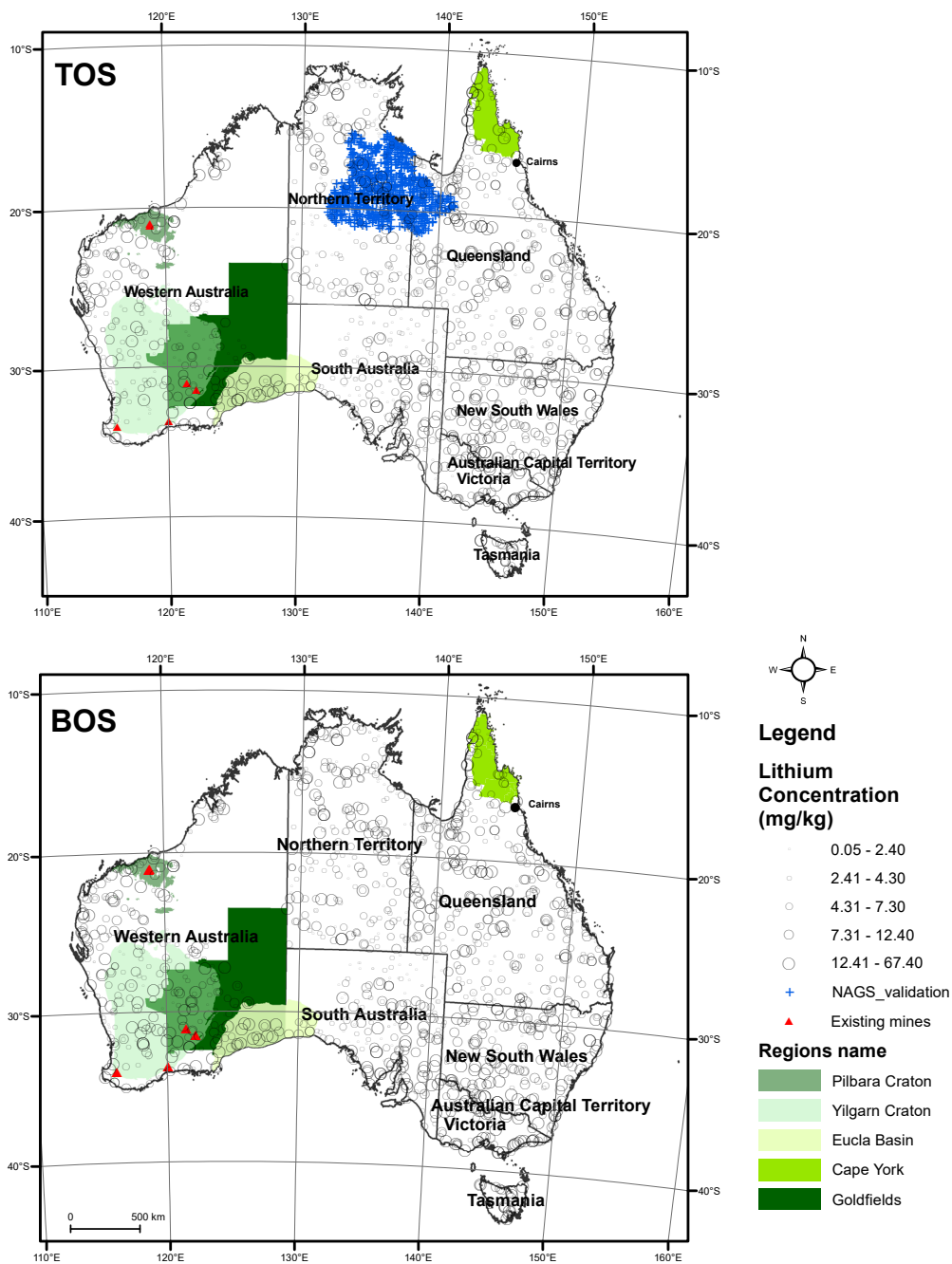



Figure 1. Distribution of sampling sites from the National Geochemical Survey of Australia (NGSA, black circles) for both depths: top outlet sediment (TOS) 0 – 10 cm, top; and bottom outlet sediment (BOS) ~60 – 80 cm, bottom. Distribution of sampling sites from the Northern Australia Geochemical Survey (NAGS, blue plus signs) for TOS only, top. All data refer to the coarse fractions (< 2mm). Aqua regia-soluble Li concentrations (mg kg⁻¹) are categorised in five quantile classes. Regions discussed in the text are highlighted in various shades of green. Projection: Australian Albers Equal Area (EPSG:3577). Data sources: de Caritat and Cooper (2011b), Hughes (2020), and Main et al. (2019).

190

The third factor is the soil itself, particularly the relevant physical soil properties. As previous studies, e.g. by Kabata-Pendias (1995) and Robinson et al. (2018), highlighted the high correlation between Li and clay content of soil, soil texture was used as a covariate. The soil texture spatial information (sand and clay contents) was derived from Malone and Searle (2021), which contained updated information on soil texture across Australia derived using a digital soil mapping approach. The sand and clay fractions were developed by integrating field morphological (n = 180,498) and laboratory measurements of soil texture fractions (n = 17,367) from the Soil and Landscape Grid of Australia (SLGA). The SLGA is based on a comprehensive compilation of soil attributes across Australia, including the NGSA dataset. These sand and clay content maps (Malone and Searle, 2021) used were for specific depth intervals (0 – 5 cm, 5 – 15 cm, 15 – 30 cm, 30 – 60 cm, 60 – 100 cm, and 100 – 200 cm). They were converted to the depths corresponding to the NGSA Li measurement (0 – 10 cm and ~60 – 80 cm) using the mass-preserving spline function, described in Bishop et al. (1999) and modified by Malone et al. (2009). Soil reflectance in the visible, near-infrared (NIR), and short-wave-infrared (SWIR) spectra captured by remote sensing images provides information on soil composition. However, the unprocessed images consist of a mixture of soil, bedrock, vegetation and clouds. By removing the influence of seasonal vegetation, Roberts et al. (2019) were able to document the 'barest' state of soil, so critical in mapping the physical characteristics of soil and rock. This was done by combining Landsat 5, 7, and 8 observations of the past 30 years to remove the contamination by seasonal vegetation, cloud cover, shadows, detector saturation and pixel saturation. The model used to develop the Barest Earth product was validated using the NGSA spectral archive (Lau et al., 2016).

Finally, topography is represented by elevation and slope. These factors also play an important role, as they affect how water is added to and/or lost from soil. The elevation was derived from the DEM-S which was obtained from the 1 arc-second resolution Shuttle Radar Topography Mission (SRTM) data acquired by NASA in February 2000 (Gallant, 2011). The slope covariate was also calculated from DEM-S using the finite difference method (Wilson and Gallant, 2000). The different spacing in the E-W and N-S directions due to the geographic projection of the data was accounted for by using the actual spacing in metres of the grid points calculated from the latitude.

All covariates were reprojected to EPSG: 3577 (GDA94 datum; Australian Albers equal area projection) and resampled to a common spatial resolution of 3 km prior to any analysis. All the environmental covariates used are shown in Table 1.

[Table 1]

The correlation matrix of the Li concentrations to all the other element concentrations and environmental covariates was generated using Pearson's correlation method. Strong correlation was defined as >0.5, moderate was defined as 0.35 to 0.5, and weak correlation was defined as <0.35. Note that this classification was generated to facilitate interpretation of this dataset only and is not implied to be a general rule.

Table 1. Environmental covariates used for digital soil mapping of Li.

Covariate	Description	Source	Original resolution
PTA	Annual precipitation (mm)	Harwood (2019)	90 m
EPA	Annual potential evaporation (mm)	Harwood (2019)	90 m
TRA	Annual temperature range (°C)	Harwood (2019)	90 m
Dose	Radiometrics: filtered dose (nGy/h)	Wilford and Kroll (2020)	0.001 degree
K	Radiometrics: filtered K element concentrations (%)	Wilford and Kroll (2020)	0.001 degree
Th	Radiometrics: filtered Th element concentrations (ppm)	Wilford and Kroll (2020)	0.001 degree
Th/K	Radiometrics: derived Th to K ratio (ppm/%)	Wilford and Kroll (2020)	0.001 degree
TMI	Total magnetic intensity (nT/m)	Poudjom Djomani et al. (2019)	90m
Sand	Sand content (%)	Malone and Searle (2021)	90 m
Clay	Clay content (%)	Malone and Searle (2021)	90 m
Landsat band 1*	Blue (450 – 510 nm)	Wilford and Roberts (2019)	25 m
Landsat band 2*	Green (530 – 590 nm)	Wilford and Roberts (2019)	25 m
Landsat band 3*	Red (640 – 670 nm)	Wilford and Roberts (2019)	25 m
Landsat band 4*	Near infrared NIR (850 – 880 nm)	Wilford and Roberts (2019)	25 m
Landsat band 5*	Shortwave infrared SWIR1 (1570 – 1650 nm)	Wilford and Roberts (2019)	25 m
Landsat band 6*	Shortwave infrared SWIR2 (2110 – 2290 nm)	Wilford and Roberts (2019)	25 m
Elevation	3 Second DEM - Shuttle Radar Topography Mission (m asl)	Gallant (2011)	1 arc-second
Slope	Elevation gradient (%)	Gallant and Austin (2012a)	90 m
TWI	Topographic wetness index (dimensionless)	Gallant and Austin (2012b)	30 m

*All Landsat bands referred here are from the Landsat 30 + Barest Earth products

2.3   Modelling

Here, we used the machine learning model Cubist to relate soil observations to the environmental covariates. Cubist is a tree-based regression algorithm based on the M5 theory (Quinlan, 1993). This algorithm creates partitions of data with similar spectral characteristics and creates one or more rules for each partition. If the partition rules are satisfied, then the linear regression of that partition is used to create the prediction (Eq. 1). Each rule can be defined as:

  If [condition is true], then [regression], else [apply next rule]   (Eq. 1)

The Cubist model has two tuning parameters: *committees* (number of sequential models included in the ensemble) and *neighbours* (number of training instances that are used to adjust the model-based prediction). A comprehensive combination of *committees* (5, 10, 20, 30, 40, 50) and *neighbours* (0, 1, 5, 9) was tested to tune the Cubist model. To obtain the best estimates of optimum parameters, a 10-fold cross-validation approach was utilised. Based on the optimum parameters, 50 bootstrap models ('sampling with replacement') were trained. The flowchart of the process is shown in Figure 2.

[Figure 2]

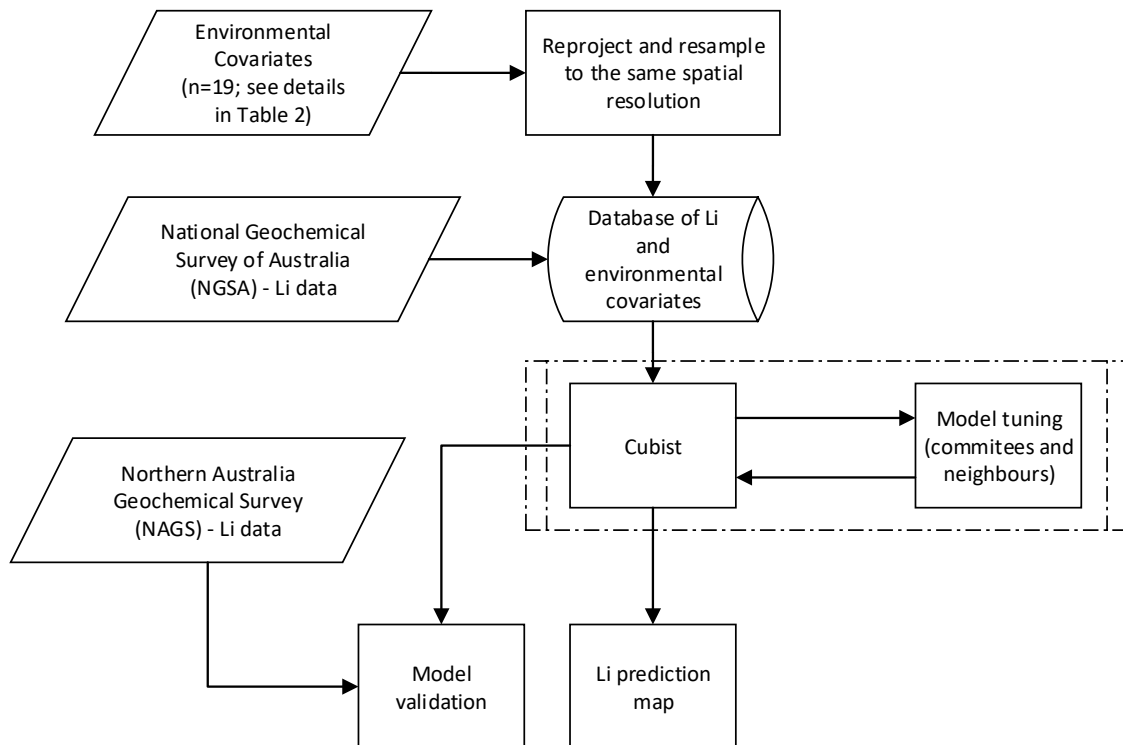



Figure 2. Flowchart of Cubist model training to generate Li prediction map along with model validation.

The performance of the prediction models was then evaluated using both an internal evaluation and the external, independent validation dataset. An internal evaluation of the model was conducted using "out of bag" samples, which were not used during the development of the bootstrap models. The NAGS dataset was used to evaluate the performance on the independent dataset (top depth only). The following metrics, briefly explained below, were used: adjusted coefficient of determination (R^2_{adj}), Lin's concordance correlation coefficient (LCCC) (Lin, 1989), root mean square error (RMSE), bias, and ratio of performance to interquartile distance (RPIQ). R^2_{adj} is a measure of the linear association between observed and predicted values; LCCC measures the agreement between the observed and predicted values in relation to the 1:1 line while accounting for the magnitude of the differences; RMSE is a measure of the differences between the observed and predicted values; bias is the measure of the difference between the mean of the observed and the mean of the predicted values; and RPIQ is a measure of performance that takes into account the distribution of the values, and can be calculated as a fraction of the interquartile range of the observed values (Q_3–Q_1) and the RMSE (RPIQ = (Q_3–Q_1)/RMSE) (Bellon-Maurel et al., 2010).

Variable importance analysis was also conducted to evaluate the contributions of each covariate to the Li prediction. The relative variable importance is measured as the percentage of times the environmental covariate is either used as *conditions* for a rule or as *predictors (usages)* within the linear regression model when certain conditions are met. These bootstrap models were then used to generate output maps with the same extent and resolution. The final map output was derived based on the mean prediction of the bootstrap models; similarly, the standard deviation map was obtained based on the standard deviation of the prediction from the bootstrap models.

2.4 Data processing and statistical computing

All the data analytics, modelling, and mapping procedures in this study were conducted in the R statistical open-source software (R Core Team, 2021). Besides the base R functionality, the R packages used in this study included "Cubist" (Kuhn and Quinlan, 2021) for fitting cubist models; "caret" (Kuhn, 2022) for tuning the hyperparameter of the Cubist model; and "raster" (Hijmans, 2021) for handling raster layers and generating soil map predictions. All soil maps were produced in ArcMap version 10.8 (ESRI 2019) using the Albers equal area projection (EPSG:3577).

3 Results and discussion

3.1 Descriptive analysis

The distribution of 1315 *aqua regia-soluble* Li concentration values (NGSA dataset, de Caritat and Cooper, 2011b) was positively skewed (Figure 3) with concentrations ranging from 0.05 – 67.4 and 0.05 – 56 mg kg^{-1}, for TOS and BOS respectively. Only limited observations above 20 mg kg^{-1} of Li concentrations were found in this study for both TOS (n = 76) and BOS (n = 95). The mean concentration of TOS (7.6 mg kg^{-1}) was slightly lower than that of BOS (8.8 mg kg^{-1}). These concentrations were lower than those observed for the mean *aqua regia-soluble* Li concentrations in European soil at 11.3 mg kg^{-1} (Négrel et al. 2019), and those found in upper continental crust (both in loess and shales) at 35 mg kg^{-1} *total* Li (Teng et

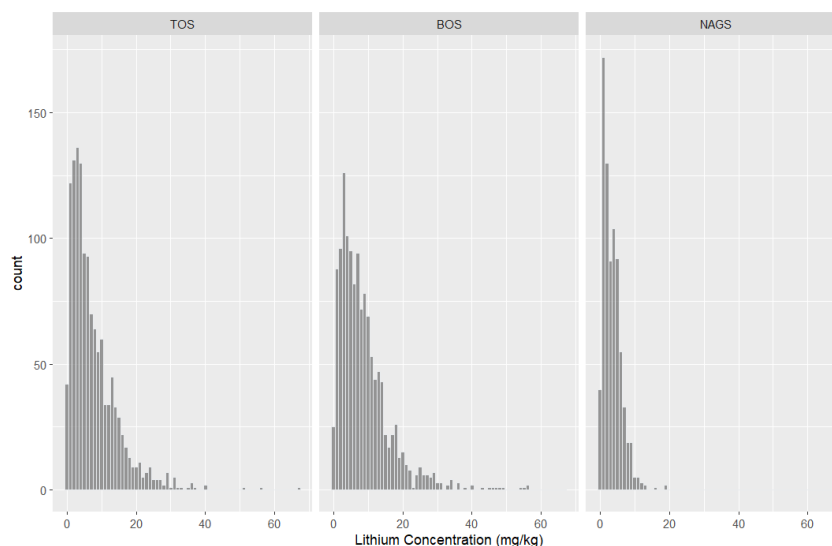
al., 2004). A soil geochemical survey in the USA shows soil *total* Li concentration with a range of <1-300 mg kg^{-1} (median 20 mg kg^{-1}) for soils from 0-5 cm and a range of <1-280 mg kg^{-1} (median 24 mg kg^{-1}) for soil samples from the C horizon (Smith et al., 2019). Similarly, *total* Li concentration of up to 400 mg kg^{-1} has been reported in China (Liu et al., 2020). These latter Li concentrations, measured using four acid extraction, were considerably higher than the aqua regia extraction data from the NGSA dataset.

[Figure 3]

Figure 3. Histograms of Li concentrations for both NGSA depths: top outlet sediment (TOS) 0 – 10 cm, left; and bottom outlet sediment (BOS) ~60 – 80 cm, middle, and NAGS. Data source: de Caritat and Cooper (2011b) and Main et al. (2019).

Based on the data collected by the NGSA project, the highest concentrations of Li for both TOS and BOS were found in northernmost Queensland (Cape York Peninsula), as shown in Figure 1 and Table 2. Other regions that have elevated concentrations of Li were located in the Goldfields-Esperance region (Table 2) in Western Australia, which has been recognised as one of the most resource-rich areas on the planet (Champion, 2019), and the region around the Victoria-New South Wales border (Figure 1). Some of the findings correlate well with the existing Li mine sites in Australia (red triangles in Figure 1). The largest deposit of Li found in Australia is the Greenbushes deposit, south of Perth. Other regions include Mount Marion and Earl Grey in the Yilgarn Craton, and Pilgangoora in the Pilbara Craton (Champion, 2019; see Table 1). In July 2019, Strategic Metals Australia (SMA) found a new Li exploration target near Cairns, in the Georgetown province of north Queensland (Gluyas, 2019). However, this discovery has not been updated in the data collected by Geoscience Australia because considerable work such as drilling, modelling, resource calculation and feasibility studies are needed to bring the discovery to the feasibility stage.

[Table 2]

Table 2. Aqua regia-extractable lithium concentrations across various regions of Australia.

Region (n = number of samples)	Range (mg kg⁻¹)	Median (mg kg⁻¹)
Pilbara Craton (n = 12)	1.2 – 15.7	6.80
Yilgarn Craton (n = 101)	0.05 – 32.7	3.50
Eucla Basin (n = 29)	1.6 – 22.6	12.40
Cape York (n = 20)	0.3 – 67.4	3.95
Goldfields (n = 78)	0.1 – 32.7	5.80

3.1.1 Correlation between Li with other measured properties

Despite other studies reporting strong correlations between Li and Mg (Kashin, 2019; Robinson et al., 2018), and between Li and other elements elsewhere, including Al, B, Fe, K, Mn and Zn, the NGSA data only show strong correlations (as defined above) between Li and Al (Pearson's correlation coefficient $r = 0.74$), Ga ($r = 0.69$), Cs ($r = 0.68$), and Rb ($r = 0.66$) for TOS, and slightly lower correlations for BOS: Al ($r = 0.69$), Ga ($r = 0.64$), Cs ($r = 0.62$), Rb ($r = 0.61$). Correlations between Li and K and Mg were only moderate for both TOS ($r = 0.48$ and 0.43) and BOS ($r = 0.46$ and 0.33). de Vos et al. (2006) also observed good correlations ($r > 0.4$) between total Li and Al, Ga and Rb within the floodplain sediment samples. Similarly, Cardoso-Fernandes et al. (2022) found strong correlation between total Li and Sn, B, Rb, Cs and F in stream sediment samples using geochemical pathfinder analysis.

The Li concentration in soil was (strongly) negatively correlated with measured sand content from the NGSA dataset ($r = -0.55$), and (moderately) positively correlated with clay content ($r = 0.44$). This is consistent with the findings of Kabata-Pendias (2010) and Robinson et al. (2018), who noted the tendency of clay minerals to concentrate Li. It has been suggested that Li may be located internally within clay minerals, mainly kaolinite, illite, smectites including hectorite, palygorskite and sepiolites, in ditrigonal cavities via isomorphous substitution, rather than on exchange sites (Anderson et al., 1988; Starkey, 1982) as a result of subsolidus cation exchange reactions with residual pegmatitic fluids (London and Burt, 1982).

3.1.2 Correlation with environmental covariates

Overall, the correlation between Li concentration and the environmental covariates was weak (Figure 4). The correlation with sand and clay content derived from digital soil maps was lower in comparison to the measured (NGSA) values discussed above, with $r = -0.28$ and 0.25, respectively, for TOS; and $r = -0.23$ and 0.22, respectively, for BOS.

For TOS, the Landsat bands 3 (Red), 5 (SWIR1) and 6 (SWIR2) had similar (weak) negative correlations with Li content ($r = -0.15$ to -0.17). For gamma-ray radiometric data, both total dose and K content had weak correlations with Li ($r = 0.10$ to 0.14). These positive correlations are expected as the associations of Li deposits and felsic rocks (high in both total dose and K) due to the observed incompatibility in mineral structures (Benson et al., 2017). Precipitation had a weak positive correlation ($r =$

0.12), while both temperature and elevation had weak negative correlations (r = -0.12) with Li content. TWI and slope had negligible correlation with Li content (r = -0.02 to 0.05).

335 For BOS, similar observations on the correlations between Li content and environmental covariates were found where temperature and Landsat bands 3, 5 and 6 had (weak) negative correlations (r = -0.11 to -0.16) with Li, while radiometric K (r = 0.10) and dose (r = 0.09) had (weak) positive correlation with Li. Similarly, TWI and slope showed negligible correlation with Li (r = -0.01 to 0.05).

[Figure 4]

340

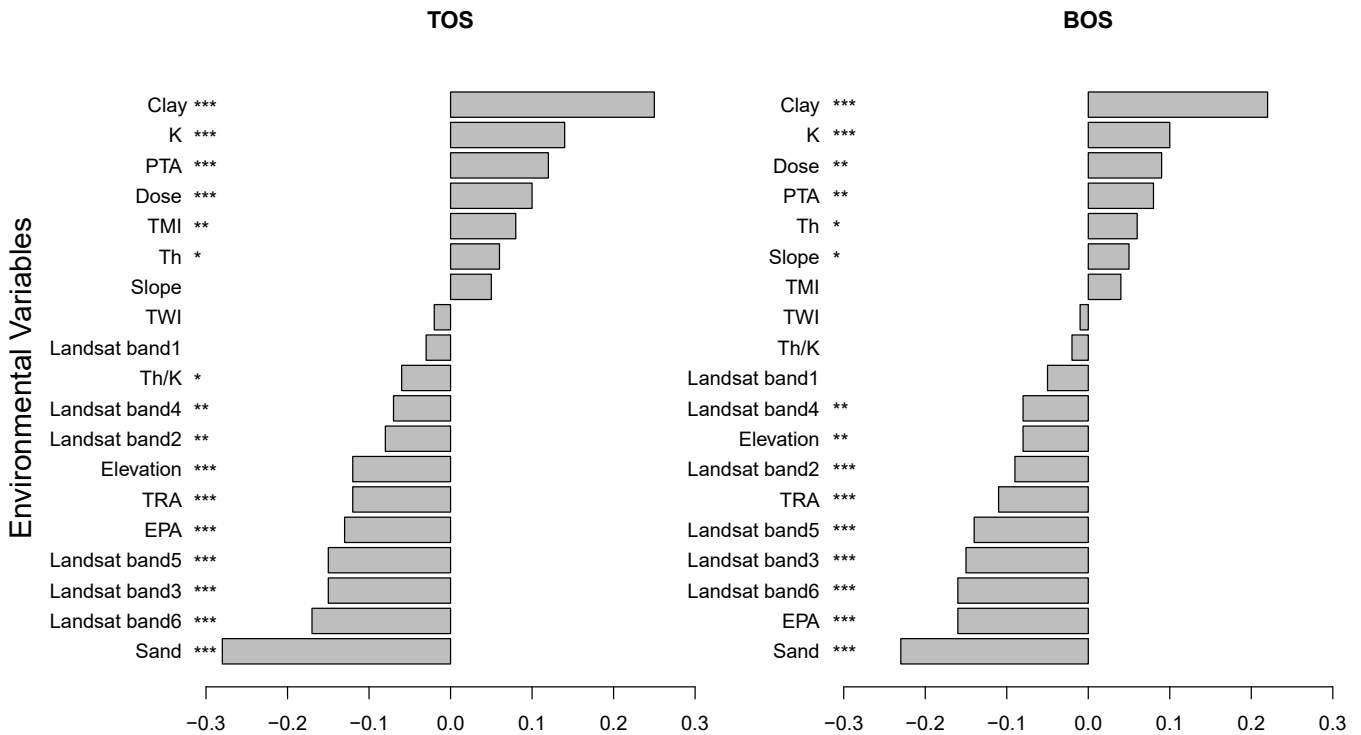



Figure 4. Pearson's correlation coefficient (r) between Li content and the environmental covariates (*scorpan*) for both NGSA depths: top outlet sediment (TOS) 0 – 10 cm, left; and bottom outlet sediment (BOS) ~60 – 80 cm, right. Data sources: de Caritat and Cooper, 345 2011b; Gallant, 2012a; Harwood, 2019; Wilford, 2019; Wilford and Kroll, 2020; Malone and Searle, 2021. See Table 1 for abbreviations. *** Correlation is significant at the 0.001 level; ** Correlation is significant at the 0.05 level; * Correlation is significant at the 0.01 level.

3.2 Model evaluation

The final Cubist model was tuned with 20 committees and 9 neighbours, which resulted in the lowest RMSE compared to the other combinations of hyperparameters, indicating an optimised Cubist model.

3.2.1 Internal evaluation

Validation statistics based on internal evaluation using the out-of-bag data for the Li predictions are presented in Table 3. There was a slightly lower accuracy on the prediction for BOS (R^2_{adj} = 0.12; LCCC = 0.29; RMSE = 7.28 mg kg^{-1}) compared to TOS (R^2_{adj} = 0.20; LCCC = 0.36; RMSE = 6.29 mg kg^{-1}). This is expected as most of the environmental covariates reflected soil surface conditions. To the best of our knowledge, the machine learning models developed in most mineral exploration studies were assessed based on classification accuracy (i.e. presence or absence of specific minerals in sample), instead of regression accuracy (Jooshaki et al., 2021). In addition, remote sensing studies on mapping Li minerals are rarely validated (e.g. Cardoso-Fernandes et al. (2019)). Hence, no comparison can be made with other studies.

[Table 3]

Table 3. Internal model evaluation and validation results for the prediction of Li concentrations using Cubist model for both NGSA depths: top outlet sediment (TOS) 0 – 10 cm; and bottom outlet sediment (BOS) ~60 – 80 cm. External independent validation is based on comparing predictions to the NAGS dataset Li concentrations.

Depth	R^2_{adj}	LCCC	RMSE	bias	RPIQ
NGSA - TOS (0 – 10 cm)	0.20	0.36	6.29	-0.80	1.20
NGSA - BOS (~60 – 80 cm)	0.12	0.29	7.28	-0.76	1.14
Independent Validation: NAGS - TOS (0 – 10 cm)	0.36	0.45	3.32	2.18	1.03

3.2.2 Independent validation dataset

The predictive model performance was also externally evaluated using an independent dataset (NAGS, TOS only) that was not part of the calibration dataset. To address the analytical variation that could potentially arise from the use of predictive model from the NGSA dataset for the NAGS dataset, a levelling method was implemented. A subset of NGSA dataset within the extent of the NAGS dataset was extracted (range = 0.05 – 28.7 mg kg^{-1}; median = 4.15 mg kg^{-1}) and compared to the NAGS dataset (range = 0.1 – 19.5 mg kg^{-1}; median = 3 mg kg^{-1}) using a two-sample Kolmogorov-Smirnov test (D = 0.24, p < 0.01). Because the samples were not deemed to have similar distribution at a 1% significance, a correction factor was calculated to level the NAGS dataset to the NGSA dataset using TILL-1 CRM standards. Upon levelling, the two datasets were deemed to have similar distribution with the two-sample Kolmogorov-Smirnov test (D = 0.18, p = 0.012).

We reported the performance of model validation the same way the model evaluation was conducted (Table 3; Figure 5). The model validation resulted in higher accuracy (R^2 = 0.36; LCCC = 0.45). The RMSE was also slightly lower (RMSE = 3.32 mg

375 kg^{-1}) than those observed in the TOS model evaluation (RMSE = 6.29 mg kg^{-1}), most likely due to lower observation values within the NAGS validation dataset.

[Figure 5]

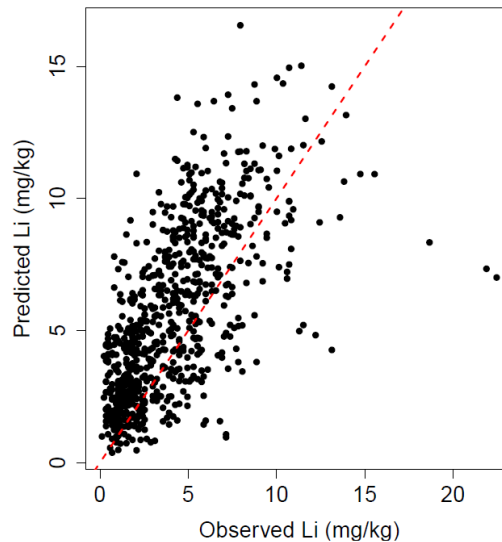



Figure 5. Goodness-of-fit plot showing observed vs predicted Li concentrations based on the independent validation dataset (NAGS,
380 **TOS only). Red dashed line is the 1:1 line.**

3.3 Variable importance analysis

From the Cubist model, we can infer the relative importance of the covariates by calculating the percentage of times a covariate is being used in the model. The variables used by Cubist model can be further split in terms of "importance as conditions within rule" and "frequency of usage as predictors in models".

385 For Li prediction in TOS, the variables clay, PTA, TRA, and EPA are of higher importance in the conditions than other variables (Figure 6). This implies that the model separates out prediction values based on climate covariates along with clay content. However, within the regression models, the top five variables most frequently used in the regression were the Landsat band 2, band 6, band 1, band 3 and gamma radiometric total dose. The first three Landsat bands (red, green, and blue) and band 6 (SWIR2) have been commonly used to predict soil properties and delineate geological boundaries, and differentiate
390 vegetation zones (Khorram et al., 2012), while the gamma radiometric dose discriminated the various soil types and their mineral makeup. The next set of covariates were annual precipitation and clay and sand contents, indicating they have lower importance as predictors. As indicated in the correlation analysis, slope was not significant.

For the BOS model, TRA variable had the highest importance in the conditions of the model (Figure 6) for Li predictions, separating high and low values. EPA, clay content and PTA also affect model conditions. Overall, parameters that were more
395 frequently used as predictors in the BOS model were similar to those for TOS, i.e. the top-five are gamma radiometric dose, and Landsat bands 2, 1, 6, and 3. In the BOS model, however, there was a higher importance of the clay content (sixth most

used) compared to the TOS model (ninth). The usage of slope covariate as predictor is similarly low (last) for both TOS and BOS.

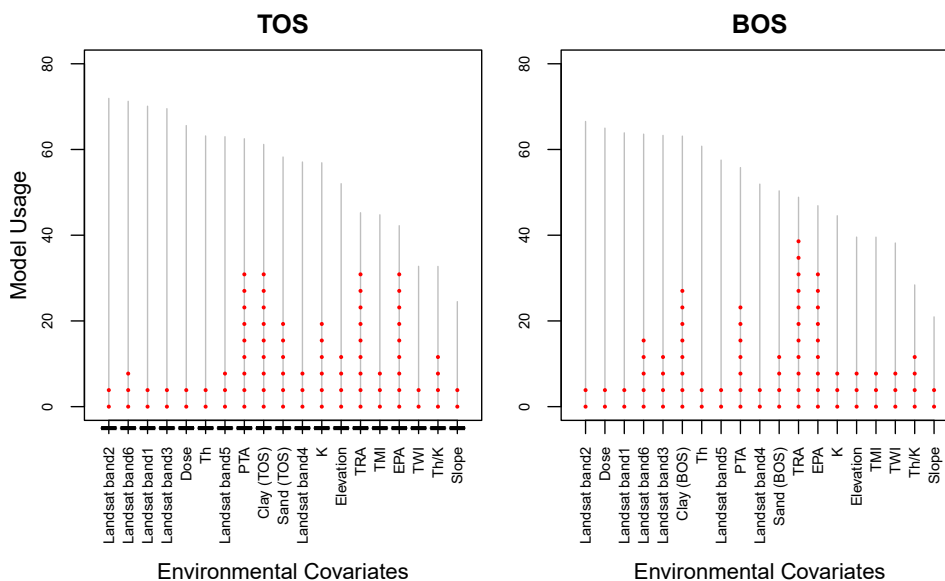
[Figure 6]

400

Figure 6. Variable importance of covariates in terms of importance as conditions (red dotted lines) and frequency of usage as predictors (grey lines) by the Cubist algorithm for both NGSA depths: top outlet sediment (TOS) 0 – 10 cm, left; and bottom outlet sediment (BOS) ~60 – 80 cm, right. Covariates are sorted in order of decreasing frequency of usage.

3.4   Li prediction maps

405   The Cubist model led to the generation of spatial predictions of aqua regia-soluble Li concentration in alluvium-derived soils across Australia at two depths (Figure 7). So far, there are only five known Li mines in Australia (mostly in Western Australia), all of which are located within areas that were predicted to have a higher background concentrations of soil Li, especially for the BOS model (>~8 mg kg^{-1}) (Figure 8).

[Figure 7] [Figure 8]

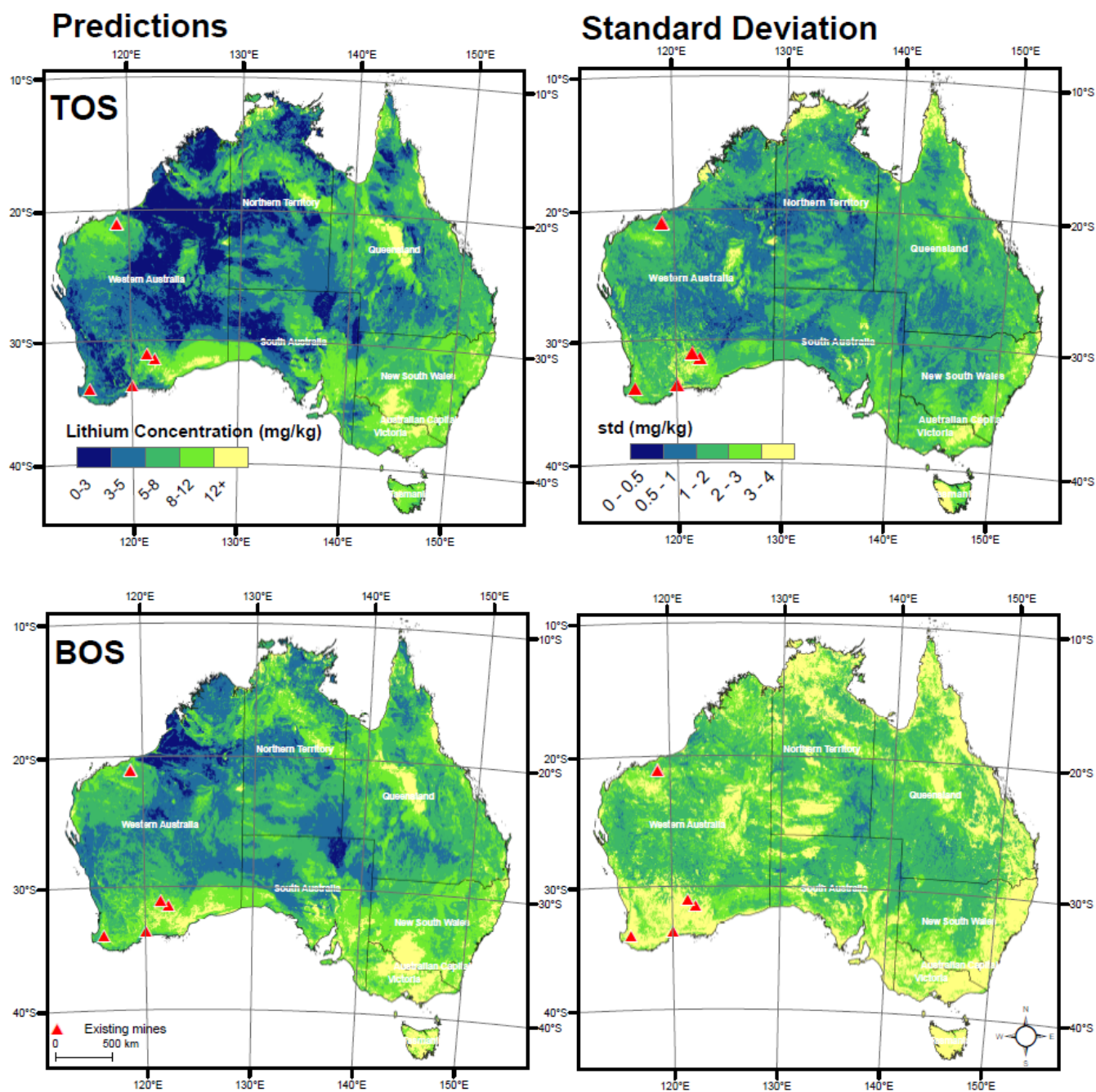



Figure 7. Spatial distributions of predicted aqua regia-soluble Li concentrations (mg kg^{-1}) in coarse fraction (<2 mm) alluvial soils across Australia, left, and standard deviations (mg kg^{-1}), right, for both National Geochemical Survey of Australia (NGSA) depths: top outlet sediment (TOS) 0 – 10 cm, top; and bottom outlet sediment (BOS) ~60 – 80 cm, bottom.

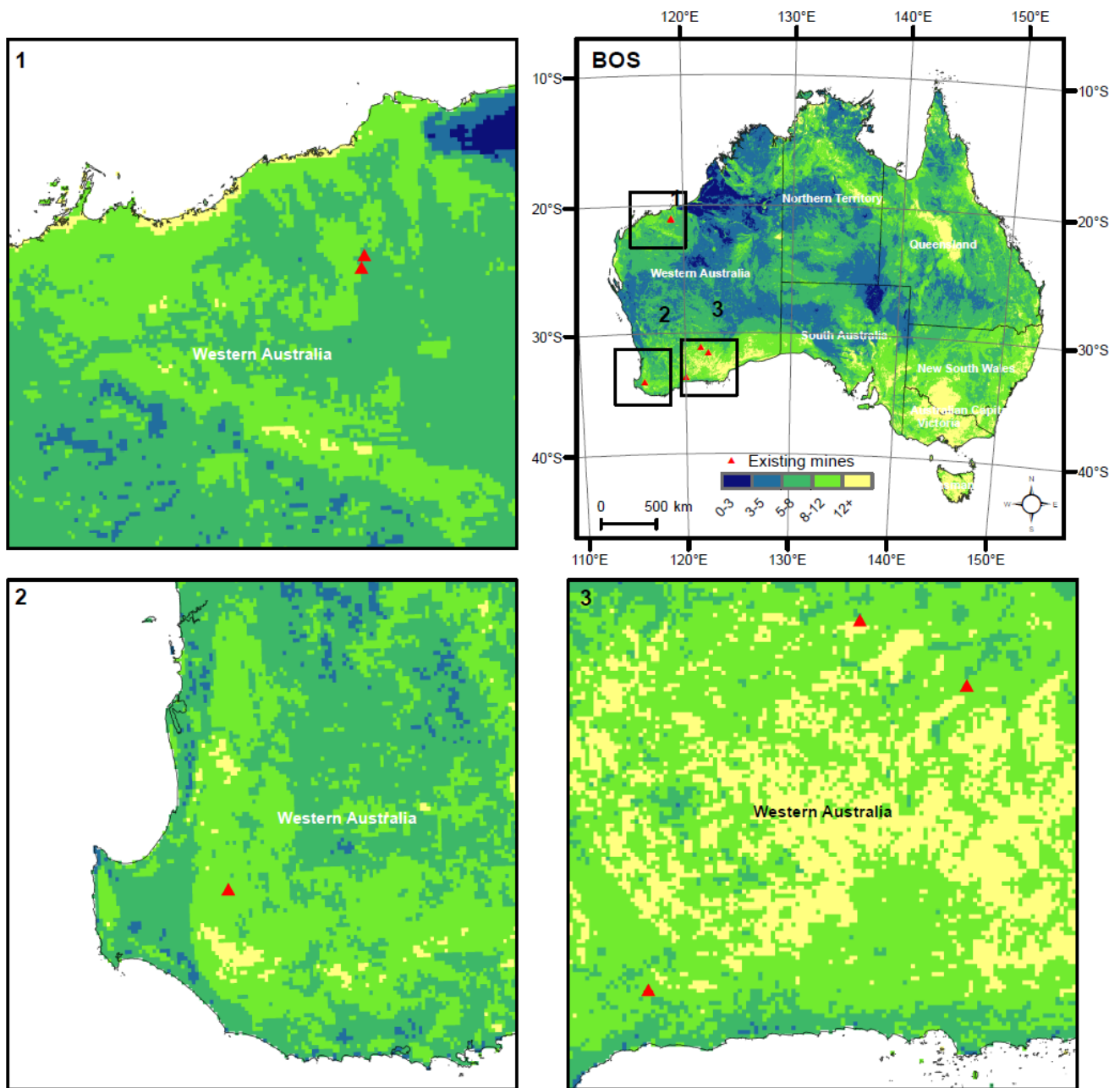



Figure 8. Distribution of Li mines on the digital soil map of Li in Australia for bottom outlet sediment (BOS) ~60 – 80 cm depth.

In Australia, the largest producer of spodumene is the Greenbushes Li operation, located approximately 250 km south-southeast of Perth. In the most recent public report, the company reported combined measured and indicated resources of 118.4 million tons (Mt) of ore at 2.4% Li_2O containing proved and probable reserves of 61.5 Mt grading 2.8% Li_2O (Champion, 2019). Other locations explored for Li include Mount Cattlin and Mount Marion in the Goldfields-Esperance region, and

Pilgangoora of East Pilbara. In a recent review (Champion, 2019), these projects' reports estimated Li resources ranging from 11.8 to 71.3 Mt at 1.01 to 1.37% Li_2O. The predicted soil Li concentrations at the known Li mine sites range from 4.5 to 7.3 mg kg^{-1} for TOS and from 7.1 to 12.6 mg kg^{-1} for BOS. The highest TOS and BOS concentrations of Li proximal to a known mine site are for the Mount Marion deposit in Western Australia.

425 Although most Li exploration to date has been conducted in Western Australia, our map indicates that other regions in Australia are potentially anomalous in Li (Figure 7). These areas are located for instance within the central west region of Queensland and visually correspond to areas of widespread black cracking (smectite-rich) soils, or vertosols (Isbell, 2021). Elevated concentration of Li was also observed over parts of the Eucla Basin, which has a widespread distribution of Fe-oxide rich regolith with carbonate accumulations (Johnson, 2015; Wilford et al., 2015). The sources of carbonate include weathered

430 Proterozoic and Palaeozoic carbonate bedrock, vast marine sediments that extend across the low-lying and offshore areas associated with Cenozoic sedimentary basins and abundant widespread pedogenic carbonates (Johnson, 2015). This is in line with de Vos et al. (2006) observations, where higher Li concentrations of up to 56 mg kg^{-1} were identified in calcareous soil (high carbonate accumulation) in comparison to those of organic soil (1.3 mg kg^{-1}). The Fe in the Fe-oxides and oxyhydroxides that help retaining Li may be released from oxidation of primary minerals during weathering (Kabata-Pendias, 2010). The

435 ultimate origin of Li within these clay-, iron- and carbonate-rich soils remains to be established in the case of Australia. Other regions of potential interest occurring on different soil types are located in southern New South Wales and parts of Victoria. We further explored the correlation of Li concentration against soil orders (Searle, 2021). Figure 9 shows the range of Li concentration across various soil types identified at the sampling locations. The Li concentration tended to be slightly higher on Vertosols, Calcarasols, as well as Dermosols. These observations indicate Li accumulated in a more uniform soil profile

440 with less differentiation between top and subsoils. In addition, clay soils (Vertosols) and soils with high $CaCO_3$ (Calcarosols) appeared to have larger Li concentrations. These observations supported the anomalous Li predictions in various parts of Australia mentioned earlier.

[Figure 9]

The highest predicted values on the Li digital soil maps are 28 mg kg^{-1} and 22 mg kg^{-1} in TOS and BOS, respectively. Although

445 higher Li concentration was expected to be observed in the deeper layer, the model used in this study was not able to support such predictions yet. This is most likely because the covariates used within the model represent observations from TOS instead of BOS. The variance of covariates within BOS was not obtained, and hence yielding lower accuracy predictions.

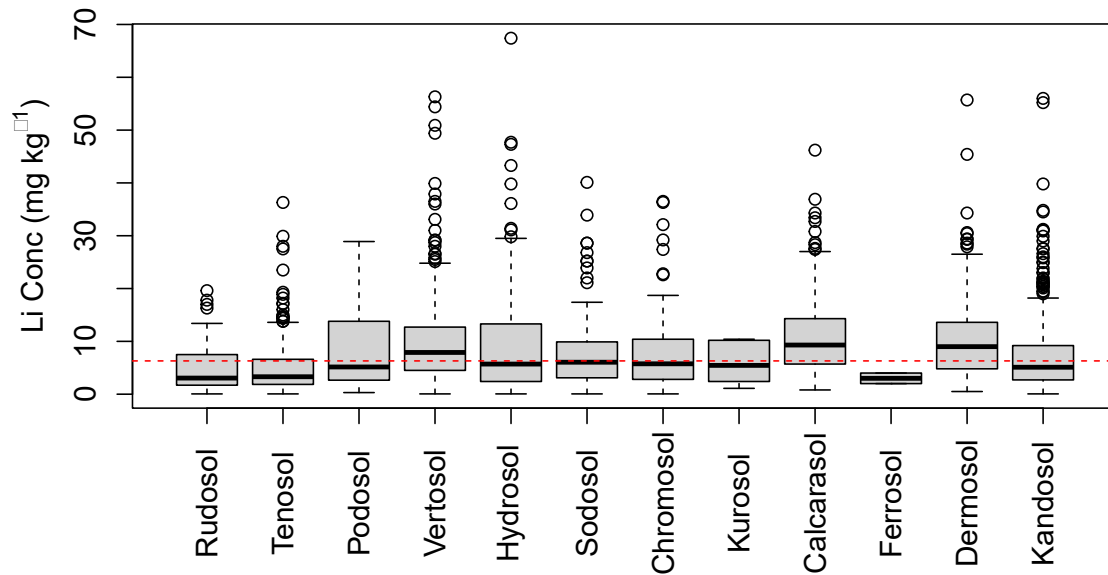



Figure 9. Boxplots of Li concentration in both TOS and BOS across various soil orders based on the Australian Soil Classification
(ASC) system. The boxes indicate the interquartile interval, the bold black lines in the middle of the boxes represent the median,
while the values outside 1.5-times the interquartile interval are indicated by circles. Red dashed line represented the median values
of Li across both TOS and BOS depths.

3.5 Study limitations

While we have successfully modelled soil Li distribution in Australia and validated it using an independent sample dataset,
we recognise that there are limitations to this study's approach. (1) The NGSA data used apply to catchment outlet sediment
representing the local accumulation of mainly detrital minerals. Therefore, strictly speaking, the predictions developed herein
apply only to similar alluvial soils. (2) The NGSA data were measured using an aqua regia digestion that only extracts a portion
of the total Li found in soil. The results could potentially be improved if total Li was measured. Most of the observations
collected had relatively low concentration; having more representative samples at higher concentrations might improve the
prediction accuracy. (3) Despite the large amount and spread of data, the NGSA does not cover the whole of Australia. Notably,
there is a data gap in parts of Western Australia and South Australia. However, no more extensive geochemical dataset than
the NGSA exists in Australia. (4) The environmental covariates used in the study were selected based on our understanding of
relevant soil-forming processes. (5) There is also limited information on how the covariates vary with depth, except for the
soil texture (sand and clay content) data. The inclusion of more environmental covariates related to depth and soil mineralogical
information may improve the predictive capability of these machine learning models. Note that quantitative mineralogical data
are currently being acquired on the NGSA samples, both as X-ray diffraction data on whole sediment samples and clay
fractions (de Caritat and Troitzsch, 2021) and as automated mineralogy using energy dispersive spectrometry on heavy mineral
fractions (de Caritat et al., 2022a; de Caritat et al., 2022b; de Caritat et al., 2022c).

The final product was only validated in one area within Australia (Tennant Creek – Mt Isa region in the Northern Territory and Queensland). Despite our predictions of elevated soil Li in parts of Queensland, New South Wales and Victoria, ground-truthing is required to confirm them and further work is necessary to determine the origin of the contained Li.

4   Conclusions

Spatial prediction models have been increasingly utilised to help minimise risk and thus cost of mineral exploration. In this study, digital soil mapping of Li concentrations at two different depths (TOS: 0 – 10 cm, BOS: ~60 – 80 cm) based on the Cubist model was carried out across Australia using the National Geochemical Survey of Australia dataset and publicly available environmental covariates. Geology and mineralogy are of high importance in predicting soil Li anomalies, as demonstrated by the reliance of the model on the Landsat and gamma-ray radiometric covariates. Despite most mineral exploration for Li being conducted in Western Australia, other regions (such as Queensland, New South Wales and Victoria) have elevated predicted Li concentrations and could become potential areas of interest. The model accuracy tested on the independent Northern Australia Geochemical Survey (TOS only) was reasonable compared to the calibration model performance. Overall, the model performance was on the low side and inclusion of the results into a prospectivity framework needs to consider the model uncertainties. This approach provides an estimate of the environmental background concentration of Li, which is reflecting a range of processes including source rock geochemistry from which the sediments were derived, weathering (including pedogenesis), and geomorphic processes. The work provides a framework to better understand the processes controlling Li concentration at the surface (as revealed through the covariate relationships) and the modelling effectively delineates regions with locally higher Li background. Despite the low prediction accuracy, this paper demonstrates a step forward in the development of machine learning in generating predictive geochemical maps. It also highlights the importance of establishment of national geochemical survey databases enabling the exploration of various elements and minerals nationally and globally, and not limited to Li. Future work should include obtaining other relevant environmental covariates and new mineralogy data, which could further improve model performance, ground-truthing of anomalous regions, and investigation of ultimate Li sources. As more survey data are collected, the use of more complex models can also be explored, including the use of Li concentrations in bedrock materials.

CRediT authorship contribution statement

WN: conceptualisation, data curation, analysis, writing – original draft, review and editing. **BM**: conceptualisation, methodology, writing –review and editing. **AM**: conceptualisation, methodology, writing –review and editing. **PdeC**: conceptualisation, data provision, data curation, writing –review and editing. **JW**: conceptualisation, data curation, writing – review and editing.

Declaration of competing interest

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

Data availability

505   The Li geochemical data for calibration and validation are available at: http://dx.doi.org/10.11636/Record.2011.020 (de Caritat and Cooper, 2011b) and http://dx.doi.org/10.11636/Record.2019.002 (Main et al., 2019) respectively. The covariates data used for this study was sourced from Terrestrial Ecosystem Research Network (TERN) infrastructure, which is enabled by the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS) https://esoil.io/TERNLandscapes/Public/Products/TERN/Covariates/Mosaics/90m/ (TERN, 2019). The final predictive map
510   is available at: https://doi.org/10.5281/zenodo.7895482.

Acknowledgments

BM is supported by the ARC Discovery project Forecasting soil conditions (DP200102542). AMcB is supported by the ARC Laureate Fellowship 'A calculable approach to securing Australia's soil' (FL210100054). The National Geochemical Survey
515   of Australia (NGSA) project (http://www.ga.gov.au/ngsa) was funded by the Australian Government's Onshore Energy Security Program (OESP 2007-2011). The Northern Australian Geochemical Survey (NAGS) project was funded by the Australian Government's Exploring for the Future (EFTF 2020-2024) initiative. We acknowledge the traditional custodians of the lands on which these samples were collected and thank all landowners for granting access to the sampling sites. We are grateful to the Geoscience Australia laboratory staff for their assistance with sample preparation. We thank Geoscience
520   Australia reviewers for their detailed and constructive critique of our work. PdeC and JW publish with the permission from the Chief Executive Officer, Geoscience Australia.

5 References

Alhumimidi, M. S., Aboud, E., Alqahtani, F., Al-Battahien, A., Saud, R., Alqahtani, H. H., Aljuhani, N., Alyousif, M. M., and Alyousef, K. A.: Gamma-ray spectrometric survey for mineral exploration at Baljurashi area, Saudi Arabia, Journal of Radiation Research and Applied Sciences, 14, 82-90, 10.1080/16878507.2020.1856600, 2021.

Anderson, M. A., Bertsch, P. M., and Miller, W. P.: The distribution of lithium in selected soils and surface waters of the southeastern USA, Appl Geochem, 3, 205-212, 1988.

Antezana Lopez, F. P., Zhou, G., Paye Vargas, L., Jing, G., Oscori Marca, M. E., Villalobos Quispe, M., Antonio Ticona, E., Mollericona Tonconi, N. M., and Orozco Apaza, E.: Lithium quantification based on random forest with multi-source geoinformation in Coipasa salt flats, Bolivia, Int J Appl Earth Obs, 117, 103184, 10.1016/j.jag.2023.103184, 2023.

Aral, H. and Vecchio-Sadus, A.: Toxicity of lithium to humans and the environment--a literature review, Ecotoxicol Environ Saf, 70, 349-356, 10.1016/j.ecoenv.2008.02.026, 2008.

Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J. M., and McBratney, A.: Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy, Trac-Trend Anal Chem, 29, 1073-1081, 10.1016/j.trac.2010.05.006, 2010.

Benedikt, S.: Using Tellus stream sediment geochemistry to fingerprint regional geology and mineralisation systems in Southeast Ireland, Irish Journal of Earth Sciences, 36, 45-61, 10.3318/ijes.2018.36.45, 2018.

Benson, T. R., Coble, M. A., Rytuba, J. J., and Mahood, G. A.: Lithium enrichment in intracontinental rhyolite magmas leads to Li deposits in caldera basins, Nat Commun, 8, 270, 10.1038/s41467-017-00234-y, 2017.

Bishop, T. F. A., McBratney, A. B., and Laslett, G. M.: Modelling soil attribute depth functions with equal-area quadratic smoothing splines, Geoderma, 91, 27-45, Doi 10.1016/S0016-7061(99)00003-8, 1999.

Cardoso-Fernandes, J., Lima, A., and Teodoro, A.: Potential of Sentinel-2 data in the detection of Lithium (Li)-bearing pegmatites: a study case, SPIE Remote Sensing, SPIE Remote Sensing, Berlin, Germany, https://doi.org/10.1117/12.2326285, 2018.

Cardoso-Fernandes, J., Teodoro, A. C., and Lima, A.: Remote sensing data in lithium (Li) exploration: A new approach for the detection of Li-bearing pegmatites, Int J Appl Earth Obs, 76, 10-25, 10.1016/j.jag.2018.11.001, 2019.

Cardoso-Fernandes, J., Teodoro, A. C., Lima, A., and Roda-Robles, E.: Semi-Automatization of Support Vector Machines to Map Lithium (Li) Bearing Pegmatites, Remote Sens-Basel, 12, 2319, 10.3390/rs12142319, 2020.

Cardoso-Fernandes, J., Lima, J., Lima, A., Roda-Robles, E., Kohler, M., Schaefer, S., Barth, A., Knobloch, A., Goncalves, M. A., Goncalves, F., and Teodoro, A. C.: Stream sediment analysis for Lithium (Li) exploration in the Douro region (Portugal): A comparative study of the spatial interpolation and catchment basin approaches, Journal of Geochemical Exploration, 236, 106978, 10.1016/j.gexplo.2022.106978, 2022.

Carranza, E. J. M.: Handbook of Exploration and Environmental Geochemistry, Elsevier Science B.V., 189-247 pp., https://doi.org/10.1016/S1874-2734(09)70011-7, 2008.

Carranza, E. J. M.: Geocomputation of mineral exploration targets, Computers & Geosciences, 37, 1907-1916, 10.1016/j.cageo.2011.11.009, 2011.

Champion, D.: Australian Resource Reviews: Lithium 2018, 2019.

Crósta, A. P., De Souza Filho, C. R., Azevedo, F., and Brodie, C.: Targeting key alteration minerals in epithermal deposits in Patagonia, Argentina, using ASTER imagery and principal component analysis, Int J Remote Sens, 24, 4233-4240, 10.1080/0143116031000152291, 2010.

de Caritat, P., Cooper, M., Lech, M., McPherson, A., and Thun, C.: National Geochemical Survey of Australia: Sample Preparation Manual. Record 2009/08, 2009.

de Caritat, P., Cooper, M., Pappas, W., Thun, C., and Webber, E.: National Geochemical Survey of Australia: Analytical Methods Manual. Record 2010/15, 2010.

de Caritat, P. and Cooper, M.: National Geochemical Survey of Australia: Data Quality Assessment. Record 2011/021, 2011a.

de Caritat, P. and Cooper, M.: National Geochemical Survey of Australia: The Geochemical Atlas of Australia. Record 2011/020, Geoscience Australia [dataset], http://dx.doi.org/10.11636/Record.2011.020, 2011b.

de Caritat, P. and Reimann, C.: Comparing results from two continental geochemical surveys to world soil composition and deriving Predicted Empirical Global Soil (PEGS2) reference values, Earth and Planetary Science Letters, 319-320, 269-276, 10.1016/j.epsl.2011.12.033, 2012.

de Caritat, P. and Cooper, M.: A continental-scale geochemical atlas for resource exploration and environmental management: the National Geochemical Survey of Australia, Geochemistry: Exploration, Environment, Analysis, 16, 3-13, 10.1144/geochem2014-322, 2015.

de Caritat, P. and Troitzsch, U.: Towards a Regolith Mineralogy Map of the Australian Continent - A Feasibility Study in the Darling-Curnamona-Delamerian Region. RECORD 2021/35, http://dx.doi.org/10.11636/Record.2021.035, 2021.

de Caritat, P.: The National Geochemical Survey of Australia: review and impact, 22, geochem2022-2032, doi:10.1144/geochem2022-032, 2022.

de Caritat, P., Bastrakov, E., Walker, A. T., and McInnes, B. I. A.: The Heavy Mineral Map of Australia Project. Data Release 2. The Barkly Isa Georgetown Region, http://dx.doi.org/10.11636/Record.2022.043, 2022a.

de Caritat, P., McInnes, B. I. A., Walker, A. T., Bastrakov, E., Rowins, S. M., and Prent, A. M.: The Heavy Mineral Map of Australia: Vision and Pilot Project, Minerals, 12, 10.3390/min12080961, 2022b.

de Caritat, P., Walker, A. T., Bastrakov, E., and McInnes, B. I. A.: The Heavy Mineral Map of Australia Project -- Data Release 1: The Darling-Curnamona-Delamerian Region, http://dx.doi.org/10.11636/Record.2022.031, 2022c.

de Vos, W., Tarvainen, T., Salminen, R., Reeder, S., Vivo, B. D., Demetriades, A., Pirc, S., Batista, M. J., Marsina, K., Ottesen, R. T., O'Connor, P., Bidovec, M., Lima, A., Siewers, U., Smith, B., Taylor, H., Shaw, R., Salpeteur, I., Gregorauskienė, V., Halamić, J., Slaninka, I., Lax, K., Gravesen, P., Birke, M., Breward, N., Ander, E. L., Jordan, G., Ďuriš, M., Klein, P., Locutura, J., Bel-lan, A., Pasieczna, A., Lis, J., Mazreku, A., Gilucis, A., Heitzmann, P., Klaver, G. T., and Petersell, V.: Geochemical atlas of Europe. Part 2, Interpretation of geochemical maps, additional tables, figures, maps, and related publications, Geological Survey of Finland, Espoo, Finland, 225-228, http://weppi.gtk.fi/publ/foregsatlas/text/Li.pdf, 2006.

Dickson, B. L., Fraser, S. J., and KinseyHenderson, A.: Interpreting aerial gamma-ray surveys utilising geomorphological and weathering models, Journal of Geochemical Exploration, 57, 75-88, Doi 10.1016/S0375-6742(96)00017-9, 1996.

Dickson, B. L. and Scott, K. M.: Interpretation of aerial gamma-ray surveys - adding the geochemical factors, AGSO Journal of Australian Geology and Geophysics, 17, 187-200, 1997.

Ducart, D. F., Silva, A. M., Toledo, C. L. B., and de Assis, L. M.: Mapping iron oxides with Landsat-8/OLI and EO-1/Hyperion imagery from the Serra Norte iron deposits in the Carajas Mineral Province, Brazil, Braz J Geol, 46, 331-349, 10.1590/2317-4889201620160023, 2016.

Gallant, J. and Austin, J.: Slope derived from 1" SRTM DEM-S. v4. CSIRO. Data Collection, 2012a.

Gallant, J. and Austin, J.: Topographic Wetness Index derived from 1" SRTM DEM-H. v2. CSIRO. Data Collection, 2012b.

Gallant, J., Wilson, N., Dowling, T., Read, A., Inskeep, C: SRTM-derived 1 Second Digital Elevation Models Version 1.0. Record 1, 2011.

Gluyas, A.: Explorer makes significant Lithium discovery in North Queensland: https://www.australianmining.com.au/news/explorer-makes-significant-lithium-discovery-in-north-queensland/, last access: 14 March 2022, 2019.

Gopp, N. V., Savenkov, O. A., Nechaeva, T. V., and Smirnova, N. V.: The Use of NDVI in Digital Mapping of the Content of Available Lithium in the Arable Horizon of Soils in Southwestern Siberia, Izvestiya Atmospheric and Oceanic Physics, 54, 1152-1157, 10.1134/S0001433818090165, 2018.

Graedel, T. E., Barr, R., Chandler, C., Chase, T., Choi, J., Christoffersen, L., Friedlander, E., Henly, C., Jun, C., Nassar, N. T., Schechner, D., Warren, S., Yang, M. Y., and Zhu, C.: Methodology of metal criticality determination, Environ Sci Technol, 46, 1063-1070, 10.1021/es203534z, 2012.

Grosjean, C., Miranda, P. H., Perrin, M., and Poggi, P.: Assessment of world lithium resources and consequences of their geographic distribution on the expected development of the electric vehicle industry, Renewable & Sustainable Energy Reviews, 16, 1735-1744, 10.1016/j.rser.2011.11.023, 2012.

Harris, J. R., Ayer, J., Naghizadeh, M., Smith, R., Snyder, D., Behnia, P., Parsa, M., Sherlock, R., and Trivedi, M.: A study of faults in the Superior province of Ontario and Quebec using the random forest machine learning algorithm: Spatial relationship to gold mines, Ore Geology Reviews, 157, 10.1016/j.oregeorev.2023.105403, 2023.

Harwood, T.: 9s climatology for continental Australia 1976-2005: BIOCLIM variable suite. v1. CSIRO. Data Collection, 2019.

Hijmans, R. J.: raster: Geographic Data Analysis and Modeling. R package version 3.5-2 [code], 2021.

Hughes, A.: Australian Operating Mines Map 2019, Geoscience Australia [dataset], 2020.

Isbell, R.: Australian Soil Classification, CSIRO Publishing, Melbourne, Victoria, 192 pp., 10.1071/9781486314782, 2021.

Jaireth, S., Bastrakov, E. N., Wilford, J., English, P., Magee, J., Clarke, J., Caritat, P. d., Mernagh, T. P., McPherson, A., and Thomas, M.: Map of Salt Lake Systems Prospective for Lithium Deposits, 2013.

Jenny, H.: Factors of Soil Formation: A System of Quantitative Pedology, McGraw-Hill, New York, NY, 1941.

Johnson, A. K.: Regolith and associated mineral systems of the Eucla Basin, South Australia, Department of Geology ad Geophysics, Adelaide University, 2015.

Jooshaki, M., Nad, A., and Michaux, S.: A Systematic Review on the Application of Machine Learning in Exploiting Mineralogical Data in Mining and Mineral Industry, Minerals, 11, 816, 10.3390/min11080816, 2021.

Kabata-Pendias, A.: Biogeochemistry of Lithium, Proc. Int. Symp. Lithium in the Trophic Chain SoilPlant-Animal-Man, Warsaw, September 13–14,

Kabata-Pendias, A.: Trace elements in soils and plants (4th ed.), CRC press, 2010.

Kashin, V. K.: Lithium in Soils and Plants of Western Transbaikalia, Eurasian Soil Science, 52, 359-369, 10.1134/S1064229319040094, 2019.

Khorram, S., Koch, F. H., van der Wiele, C. F., and Nelson, S. A. C.: Remote Sensing, Springer, New York, 2012.

Köhler, M., Hanelli, D., Schaefer, S., Barth, A., Knobloch, A., Hielscher, P., Cardoso-Fernandes, J., Lima, A., and Teodoro, A. C.: Lithium Potential Mapping Using Artificial Neural Networks: A Case Study from Central Portugal, 11, 1046, 2021.

Kuhn, M. and Quinlan, R.: Cubist: Rule- And Instance-Based Regression Modeling. R package version 0.3.0 [code], 2021.

Kuhn, M.: Classification and Regression Training_. R package version 6.0-93 [code], 2022.

Lau, I., Bateman, R., Beattie, E., de Caritat, P., Thomas, M., Ong, C., Laukamp, C., Caccetta, M., Wang, R., and Cudahy, T.: National Geochemical Survey of Australia reflectance spectroscopy measurements. v4. CSIRO. Data Collection., https://doi.org/10.25919/5cdba18939c29, 2016.

Lin, L. I.: A concordance correlation coefficient to evaluate reproducibility, Biometrics, 45, 255-268, 10.2307/2532051, 1989.

Liu, H., Wang, X., Zhang, B., Wang, W., Han, Z., Chi, Q., Zhou, J., Nie, L., Xu, S., Yao, W., Liu, D., Liu, Q., and Liu, J.: Concentration and distribution of lithium in catchment sediments of China: Conclusions from the China Geochemical Baselines project, Journal of Geochemical Exploration, 215, 106540, https://doi.org/10.1016/j.gexplo.2020.106540, 2020.

London, D. and Burt, D. M.: Chemical-Models for Lithium Aluminosilicate Stabilities in Pegmatites and Granites, Am Mineral, 67, 494-509, 1982.

Luecke, W.: Soil Geochemistry above a Lithium Pegmatite Dyke at Aclare, Southeast Ireland, Irish Journal of Earth Sciences, 6, 205-211, 1984.

Main, P. T., Bastrakov, E. N., Wygralak, A. S., and Khan, M.: Northern Australia Geochemical Survey: Data Release 2 – Total (coarse fraction), Aqua Regia (coarse and fine fraction), and Fire Assay (coarse and fine fraction) element contents, Geoscience Australia, Canberra. [dataset], http://dx.doi.org/10.11636/Record.2019.002, 2019.

Main, P. T. and Champion, D. C.: Levelling of multi-generational and spatially isolated geochemical surveys, Journal of Geochemical Exploration, 240, 10.1016/j.gexplo.2022.107028, 2022.

Malone, B. and Searle, R.: Updating the Australian digital soil texture mapping (Part 2, Soil Research, 59, 435-451, 10.1071/sr20284, 2021.

Malone, B. P., McBratney, A. B., Minasny, B., and Laslett, G. M.: Mapping continuous depth functions of soil carbon storage and available water capacity, Geoderma, 154, 138-152, 10.1016/j.geoderma.2009.10.007, 2009.

McBratney, A. B., Santos, M. L. M., and Minasny, B.: On digital soil mapping, Geoderma, 117, 3-52, 10.1016/S0016-7061(03)00223-4, 2003.

Merian, E. and Clarkson, T. W.: Metals and their compounds in the environment, VCH, Weinheim, 1991.

Mernagh, T. P., Bastrakov, E. N., Clarke, J. D. A., de Caritat, P., English, P. M., Howard, F. J. F., Jaireth, S., Magee, J. W., McPherson, A. A., Roach, I. C., Schroder, I. F., Thomas, M., and Wilford, J. R.: A review of Australian salt lakes and assessment of their potential for strategic resources, 2013.

Mernagh, T. P., Bastrakov, E. N., Jaireth, S., de Caritat, P., English, P. M., and Clarke, J. D. A.: A review of Australian salt lakes and associated mineral systems, Australian Journal of Earth Sciences, 63, 131-157, 10.1080/08120099.2016.1149517, 2016.

670   Mudd, G. M., Werner, T. T., Weng, Z.-H., Yellishetty, M., Yuan, Y., McAlpine, S. R. B., Skirrow, R., and Czarnota, K.: Critical Minerals in Australia: A Review of Opportunities and Research Needs. Record 2018/51, 2018.

Négrel, P., Ladenberger, A., Reimann, C., Birke, M., Demetriades, A., and Sadeghi, M.: GEMAS: Geochemical background and mineral potential of emerging tech-critical elements in Europe revealed from low-sampling density geochemical mapping, Appl Geochem, 111, 104425, 10.1016/j.apgeochem.2019.104425, 2019.

675   Porwal, A., Das, R. D., Chaudhary, B., Gonzalez-Alvarez, I., and Kreuzer, O.: Fuzzy inference systems for prospectivity modeling of mineral systems and a case-study for prospectivity mapping of surficial Uranium in Yeelirrie Area, Western Australia, Ore Geology Reviews, 71, 839-852, https://doi.org/10.1016/j.oregeorev.2014.10.016, 2015.

Poudjom Djomani, Y., Minty, B. R. S., Hutchens, M., and Lane, R. J. L.: Total Magnetic Intensity (TMI) Grid of Australia 2019 - seventh edition - 80 m cell size, 2019.

680   Pour, A. B. and Hashim, M.: Hydrothermal alteration mapping from Landsat-8 data, Sar Cheshmeh copper mining district, south-eastern Islamic Republic of Iran, Journal of Taibah University for Science, 9, 155-166, 10.1016/j.jtusci.2014.11.008, 2015.

Quinlan, J. R.: C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers Inc., San Mateo, California, 1993.

Reimann, C., Birke, M., Demetriades, A., Filzmoser, P., and O'Connor, P.: Chemistry of Europe's Agricultural Soils - Part A:
685   methodology and interpretation of the GEMAS dataset, Schweizerbarth, Stuttgart, 2014.

Reimann, C. and de Caritat, P.: Establishing geochemical background variation and threshold values for 59 elements in Australian surface soil, Sci Total Environ, 578, 633-648, 10.1016/j.scitotenv.2016.11.010, 2017.

Roberts, D., Wilford, J., and Ghattas, O.: Exposed soil and mineral map of the Australian continent revealing the land at its barest, Nat Commun, 10, 5297, 10.1038/s41467-019-13276-1, 2019.

690   Robinson, B. H., Yalamanchali, R., Reiser, R., and Dickinson, N. M.: Lithium as an emerging environmental contaminant: Mobility in the soil-plant system, Chemosphere, 197, 1-6, 10.1016/j.chemosphere.2018.01.012, 2018.

Roshanravan, B., Kreuzer, O. P., Buckingham, A., Keykhay-Hosseinpoor, M., and Keys, E.: Mineral potential modelling of orogenic gold systems in the granites-tanami Orogen, Northern Territory, Australia: A multi-technique approach, Ore Geology Reviews, 152, 105224, 10.1016/j.oregeorev.2022.105224, 2023.

695   Salminen, R., Batista, M. J., Bidovec, M., Demetriades, A., Vivo, B. D., Vos, W. D., Ďuriš, M., Gilucis, A., Gregorauskienë, V., Halamiã, J., Heitzmann, P., Lima, A., Jordan, G., Klaver, G., Klein, P., Lis, J. z., Locutura, J., Marsina, K., Mazreku, A., O'Connor, P., Olsson, S. Å., Ottesen, R. T., Petersell, V., Plant, J. A., Reeder, S., Salpeteur, I., Sandström, H., Siewers, U., Steenfelt, A., and Tarvainen, T.: Geochemical Atlas of Europe. Part 1 - Background information, methodology and maps, Geological Survey of Finland, 2006.

700   Searle, R.: Australian Soil Classification Map. Version 1.0.0. Terrestrial Ecosystem Research Network [dataset], https://doi.org/10.25901/edyr-wg85, 2021.

Senior, A., Britt, A.F., Summerfield, D., Hughes, A., Hitchman, A., Cross, A., Sexton, M., Pheeney, J., Teh, M., Hill, J., Cooper, M.: Australia's Identified Mineral Resources 2021, Geoscience Australia, Canberra, http://dx.doi.org/10.11636/1327-1466.2021, 2022.

705   Sitando, O. and Crouse, P. L.: Processing of a Zimbabwean petalite to obtain lithium carbonate, International Journal of Mineral Processing, 102, 45-50, 10.1016/j.minpro.2011.09.014, 2012.

Smith, D. B., Solano, F., Woodruff, L. G., Cannon, W. F., and Ellefsen, K. J.: Geochemical and mineralogical maps, with interpretation, for soils of the conterminous United States, Reston, VA, Report 2017-5118, 10.3133/sir20175118, 2019.

SSSA: Soil Science Society of America Glossary. Available at: https://www.soils.org/publications/soils-glossary, last access:
710   07 September 2022, 2022.

Starkey, H. C.: The Role of Clays in Fixing Lithium, Report 1278F, https://doi.org/10.3133/b1278F, 1982.

Teng, F. Z., McDonough, W. F., Rudnick, R. L., Dalpe, C., Tomascak, P. B., Chappell, B. W., and Gao, S.: Lithium isotopic composition and concentration of the upper continental crust, Geochim Cosmochim Ac, 68, 4167-4178, 10.1016/j.gca.2004.03.031, 2004.

715   TERN: TERN Landscape Covariates 90m, Terrestrial Ecosystem Research Network [dataset], https://esoil.io/TERNLandscapes/Public/Products/TERN/Covariates/Mosaics/90m/, 2019.

Vieceli, N., Nogueira, C. A., Pereira, M. F. C., Durão, F. O., Guimarães, C., and Margarido, F.: Recovery of lithium carbonate by acid digestion and hydrometallurgical processing from mechanically activated lepidolite, Hydrometallurgy, 175, 1-10, 10.1016/j.hydromet.2017.10.022, 2018.

720   Wilford, J., Worrall, L., and Minty, B.: Radiometric map of Australia provides new insights into uranium prospectivity, Ausgeo News, 95, 1-4, 2009.

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

Wilford, J., de Caritat, P., and Bui, E.: Modelling the abundance of soil calcium carbonate across Australia using geochemical

725   survey data and environmental predictors, Geoderma, 259, 81-92, 10.1016/j.geoderma.2015.05.003, 2015.

Wilford, J. and Roberts, D.: Landsat 30+ Barest Earth, 2019.

Wilford, J. R., Bierwirth, P. N., and Craig, M. A.: Application of airborne gamma-ray spectrometry in soil/regolith mapping and applied geomorphology, AGSO Journal of Australian Geology and Geophysics, 17, 201-216, 1997.

Wilford, J. R. and Kroll, A.: Complete Radiometric Grid of Australia (Radmap) v4 2019 with modelled infill, 2020.

730   Wilson, J. and Gallant, J.: Primary topographic attributes, in: Terrain Analysis: Principles and Applications, edited by: Wilson, J. P., and Gallant, J. C., John Wiley & Sons, 51-85, 2000.

Zuo, R. G.: Geodata Science-Based Mineral Prospectivity Mapping: A Review, Natural Resources Research, 29, 3415-3424, 10.1007/s11053-020-09700-9, 2020.