Applied Geochemistry 66 (2016) 275–288

Contents lists available at ScienceDirect

Applied Geochemistry

journal homepage: www.elsevier.com/locate/apgeochem

Predictive geochemical mapping using environmental correlation

John Wilford [a, *], Patrice de Caritat [a, b], Elisabeth Bui [c]

[a] Geoscience Australia, GPO Box 378, Canberra, ACT 2601, Australia
[b] Research School of Earth Sciences, Australian National University, Canberra, ACT 2601, Australia
[c] CSIRO Land and Water, GPO Box 1666, Canberra, ACT 2601, Australia

ARTICLE INFO

Article history:
Received 2 July 2015
Received in revised form
18 August 2015
Accepted 20 August 2015
Available online 4 September 2015

Keywords:
Geochemistry
Critical zone
Regolith
Landscape processes
Data mining
Regression trees
Yilgarn craton
Australia

ABSTRACT

The distribution of chemical elements at and near the Earth's surface, the so-called critical zone, is complex and reflects the geochemistry and mineralogy of the original substrate modified by environmental factors that include physical, chemical and biological processes over time.

Geochemical data typically is illustrated in the form of plan view maps or vertical cross-sections, where the composition of regolith, soil, bedrock or any other material is represented. These are primarily point observations that frequently are interpolated to produce rasters of element distributions. Here we propose the application of environmental or covariate regression modelling to predict and better understand the controls on major and trace element geochemistry within the regolith. Available environmental covariate datasets (raster or vector) representing factors influencing regolith or soil composition are intersected with the geochemical point data in a spatial statistical correlation model to develop a system of multiple linear correlations. The spatial resolution of the environmental covariates, which typically is much finer (e.g. ~90 m pixel) than that of geochemical surveys (e.g. 1 sample per 10-10,000 km^2), carries over to the predictions. Therefore the derived predictive models of element concentrations take the form of continuous geochemical landscape representations that are potentially much more informative than geostatistical interpolations.

Environmental correlation is applied to the Sir Samuel 1:250,000 scale map sheet in Western Australia to produce distribution models of individual elements describing the geochemical composition of the regolith and exposed bedrock. As an example we model the distribution of two elements – chromium and sodium. We show that the environmental correlation approach generates high resolution predictive maps that are statistically more accurate and effective than ordinary kriging and inverse distance weighting interpolation methods. Furthermore, insights can be gained into the landscape processes controlling element concentration, distribution and mobility from analysis of the covariates used in the model. This modelling approach can be extended to groups of elements (indices), element ratios, isotopes or mineralogy over a range of scales and in a variety of environments.

1. Introduction

The compilation of regional, national, continental and ultimately global multi-purpose geochemical atlases, based on the systematic sampling and analysis of soils or sediments, has been recognized as a strategic research priority since the pioneering work of John Webb in the UK in the 1960s (e.g. Darnley et al., 1995; Garrett et al., 2008; Plant et al., 2001; Thornton, 2012). It has demonstrated the pivotal role of geochemical surveys and geochemical data in the exploration for mineral resources, optimisation of land use, efficient production of crops and livestock, conservation of wildlife, management of pollution and development of sustainable environmental policies, at a range of spatial scales. Mapping of elemental baselines can be used to monitor land degradation and pollution in the context of natural geochemical variations and future environmental change.

Reimann (2005) reviewed and discussed various methods used for the production of geochemical maps. Initially and perhaps most faithfully to the original data and spatial distribution, geochemical data is represented as point source maps, using an appropriate (black and white or colour) symbology to indicate concentration (e.g. Caritat and Cooper, 2011). Interpolated or contour maps

* Corresponding author.
E-mail address: john.wilford@ga.gov.au (J. Wilford).

http://dx.doi.org/10.1016/j.apgeochem.2015.08.012
0883-2927/

(models) are often used as the final publication of geochemical atlases because they are easier to read to the non-specialist. The process of interpolation, which attempts to fit a function to the data to estimate values between data points, can follow different avenues (inverse distance weighting, kriging, spline, etc.) with different results (Lam, 1983). Interpolation methods based on kriging (e.g. Caritat and Grunsky, 2013; Reimann et al., 1998) and inverse distance weighting (e.g. Cohen et al., 2012; Smith et al., 2014) are among the most common and reliable methods (Li and Heap, 2011; Reimann, 2005). Recently, spatial regression trees and random forest methods, which are predictive modelling and mapping methods used in digital soil mapping (e.g. McBratney et al., 2003), have been applied to soil pollution analysis (Bou Kheir et al., 2014) and lithological mapping (Cracknell and Reading, 2014; Harris and Grunsky, 2015; Lacoste et al., 2011). Adhikari et al. (2013) compared the performance of ordinary kriging, stratified ordinary kriging, regression trees and rule-based regression kriging for mapping soil clay content over the Midwestern part of Denmark.

The composition of surface soil and sediment (regolith) is the result of complex processes occurring at the intersection of the geosphere, biosphere, atmosphere and hydrosphere (Scott and Pain, 2008; Taylor and Eggleton, 2001). Not only are bio-physico-chemical processes paramount, but time adds a further degree of complexity. Understanding the composition and evolution of the regolith is a fundamental scientific pursuit. The regolith ('everything between fresh rock and fresh air'; Eggleton et al., 2001) is a key component of the critical zone ('everything between unweathered bedrock and treetops'; Brantley et al., 2007) and supports plants and crops, sustains ecosystems, filters groundwater, stores carbon, supports infrastructure, and can conceal or host mineral deposits.

The geochemical composition of the regolith is inextricably linked to the makeup of its component solid particles (minerals, organic matter, …) and fluids (water with its dissolved species, complexes and colloids, and gases). The nature, abundance, composition, grain size, Eh, pH, pCO_2 and other characteristics of the regolith depend strongly on the geological substrate, climate, biota and landscape processes through time. Despite this complexity some profound understanding of soil, regolith and weathering has been achieved over several decades (from Jenny, 1941; to Drever, 2014), yet we are far from possessing a holistic, deterministic, 4D predictive capability of soil properties. Thus, every effort that can enhance this capability, at any scale, is important. This is why a method of quantitatively modelling the composition of surface regolith between observations that tests the dependency on fundamental variables can be more instructive than fitting a smoothing function with no regards to what controls the variability of conditions between the data points.

In this paper we demonstrate the application of decision tree-based modelling using points with measured elemental geochemistry coupled with environmental covariates at those points to predict and then extensively map the distribution of major and trace element geochemistry of the surface regolith of a 1:250,000 scale map sheet in Western Australia. Decision tree-based modelling is one of several environmental correlation approaches used in digital soil mapping that aim to quantitatively predict individual soil properties or classes with statistically determined uncertainty (Henderson et al., 2005; McBratney et al., 2003; McKenzie and Ryan, 1999). Here we compare this predictive modelling approach to more traditional methods of data interpolation including Ordinary Kriging (OK) or Inverse Distance Weighting (IDW) and discuss more broadly its merits for mapping and understanding the environmental controls on regolith geochemistry.

2. Study area – climate, geology, geomorphology and vegetation

The study area is the Sir Samuel 1:250,000 scale map sheet (from 27.0° to 28.0° S and 120.0°–121.5° E) in central Western Australia, which covers an area of approximately 16,000 km^2 on the Yilgarn Craton (Fig. 1 – inset). The climate is arid to semi-arid with hot summers and cool to mild winters (Liu et al., 2002). Average annual rainfall of 239 mm at Yeelirrie (27.28° S and 120.09° E) (http://www.bom.gov.au/climate/averages/tables/cw_012090.shtml) is greatly exceeded by average potential evaporation (>3200 mm). The heaviest rainfall events tend to occur as summer thunderstorms; otherwise rainfall is fairly evenly distributed through the year. Elevation varies from 425 to 609 m above sea level.

The geology of the study area (Fig. 1) comprises Archean greenstone belts consisting of metamorphosed mafic and ultramafic igneous rocks (mostly volcanic), sedimentary rocks, felsic volcanic and volcaniclastics rocks, minor cherts and banded iron formations (Bunting and Williams, 1979; Liu et al., 2002). Mafic and ultramafic lithologies include peridotite, komatiite dunite and high Mg basalt. These greenstone rocks are deformed and folded along north-northwest to north-northeast trending axes. Younger Archean granitoid rocks separating the greenstone belts include granodiorite, monzonite, quartz syenite, tonalite, gneiss and adamellite. Rock exposures are poor and generally occur over local bedrock rises. Most of the area is covered by Cenozoic regolith that masks the geological and topographic complexity of the basement with a gently undulating surface (Anand and Paine, 2002). The regolith includes deeply weathered (up to 100 m) in situ bedrock profiles and transported materials comprising colluvial, lacustrine and aeolian sediments (Anand and Paine, 2002). Cenozoic sediments are preserved in paleochannels, which are largely coincident with the present drainage system and are up to 125 m thick (Johnson, 2004). The upper part of the weathering profiles can be indurated with iron and silica to form ferruginous duricrust (also referred to as laterite) and silcrete. These indurated materials typically occur over local basement highs/ridges and form the cap rock of breakaways or erosional scarps (for regolith terms see Eggleton et al., 2001).

Overall the area has very low relief with rises and low hills defining local drainage divides. The greenstone belts typically are expressed as low hills with local prominent ridges formed by more resistant chert and banded iron formations. Granitoid rocks can form tor mantled rises and low domed hills (Bunting and Williams, 1979). Bedrock rises are often delineated by erosional scarps or breakaways that mark a boundary between more highly weathered materials above the scarp and less weathered materials immediately below. Down slope from these erosional features sheet-flood fans and floodplains form extensive low relief depositional plains. The floodplains tend to be saline with salt lakes associated with the larger drainage systems.

The vegetation type and distribution is strongly associated with the type of regolith, bedrock and landform. The area is largely dominated by acacia (mulga) and mixed eucalypt (mallee) scrub (Beard et al., 2013). Mulga and mallee plants are commonly found over mafic rocks and associated ferruginous soils. Granitic landscapes are associated with mulga and other *Acacia* sp. whilst extensive sand plains typically support spinifex grasses with scattered mulga and eucalypt species. Salt tolerant halophytes dominate the valley floors and playa lakes (Beard et al., 2013).

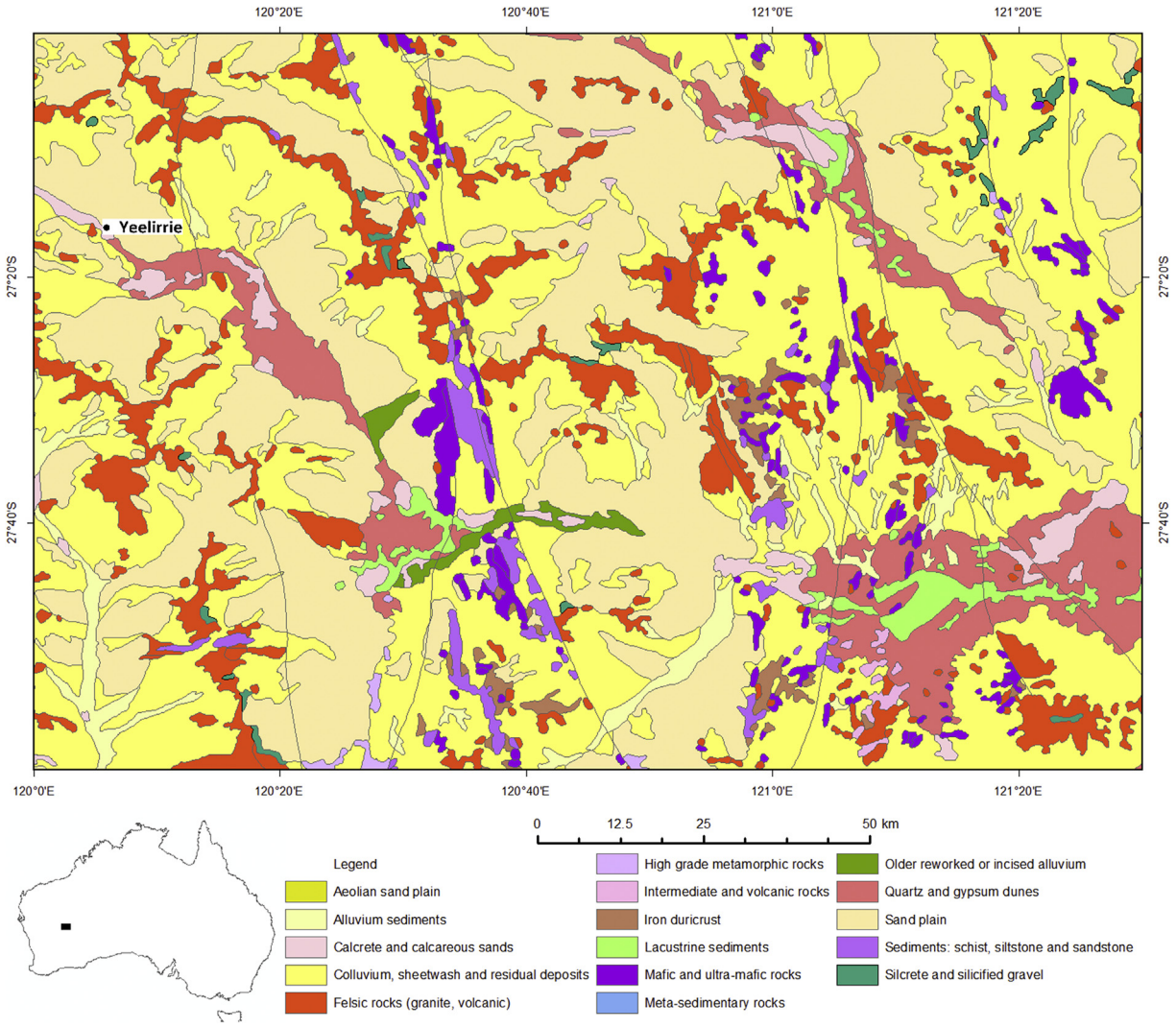

Fig. 1. Location of the Sir Samuel 1:250,000 scale map sheet (inset) and major geology and regolith units (Bunting and Williams, 1979).

3. Methods

3.1. Geochemical data and statistical analysis

The geochemical data analysed here is part of the Geological Survey of Western Australia's geochemistry program, which carried out ~4 km × 4 km grid sampling of surface regolith over several 1:250,000 scale map sheets between 1994 and 2001 (Morris et al., 1999). The program's aim was to provide baseline information on the distribution and composition of regolith for the mineral exploration industry. Over the Sir Samuel map sheet (Fig. 2), 1026 samples of sheetwash (n = 421), stream (397), and lake (52) sediments, as well as soil (156) were collected from just a few cm below the surface (to avoid incorporation of any windblown material, vegetation or other organic material) down to 10–40 cm depth (Kojan et al., 1996). The resulting average sample density is one site per ~16 km^2.

Each sample was sieved with the 0.45–2 mm fraction retained (to avoid wind deposited fines), milled and analysed for 47 elements as well as Loss on Ignition (LOI), pH and electrical conductivity. Analysis was carried out by commercial laboratories under strict quality assurance/quality control protocols including laboratory and in-house standards, duplicates and blanks (Kojan et al., 1996; Morris, 2000). Samples were analysed using a suite of analytical methods including multi-acid (HF/$HClO_4$/HNO_3) digestion followed by Inductively Coupled Plasma-Emission Spectrometry (ICP-ES) for Ca, Na, P, Cu, Ni, S, Sc, V and Zn, or Inductively Coupled Plasma-Mass Spectrometry (ICP-MS) for Ag, As, Ba, Be, Bi, Ce, Co, Ga, In, La, Li, Mo, Nb, Pb, Rb, Sb, Sn, Sr, Ta, Te, Th, U, W, Y and Zr; alkaline oxidative fusion with Na_2O_2 followed by ICP-ES for Si, Ti, Al, Fe, Mn, Mg, K and Cr; fire assay lead collection followed by ICP-MS for Au, Pd and Pt; alkaline fusion followed by Ion-Specific Electrode (ISE) measurement for F; aqua regia digestion followed by ICP-MS for Se; and finally gravimetry for LOI (Kojan et al., 1996). Data originally reported as oxides was transformed to elemental concentrations, with all concentrations expressed as parts per million (ppm; 1 ppm = 1 mg/kg).

We selected chromium (Cr) and sodium (Na) to illustrate the application of environmental correlation for predictive modelling of geochemical patterns. The distribution, geochemical associations and mobility of these elements differ markedly and therefore this choice provides complementary datasets for investigation. Only few results (8 for Cr, 1 for Na) were below the lower detection limit and those censored values were imputed using the robCompositions R package of Hron et al. (2010) before statistical analysis. Raw distributions of both Cr and Na (in ppm) were heavily skewed (Table 1, Fig. 3) and thus were logarithmically (Log(10))

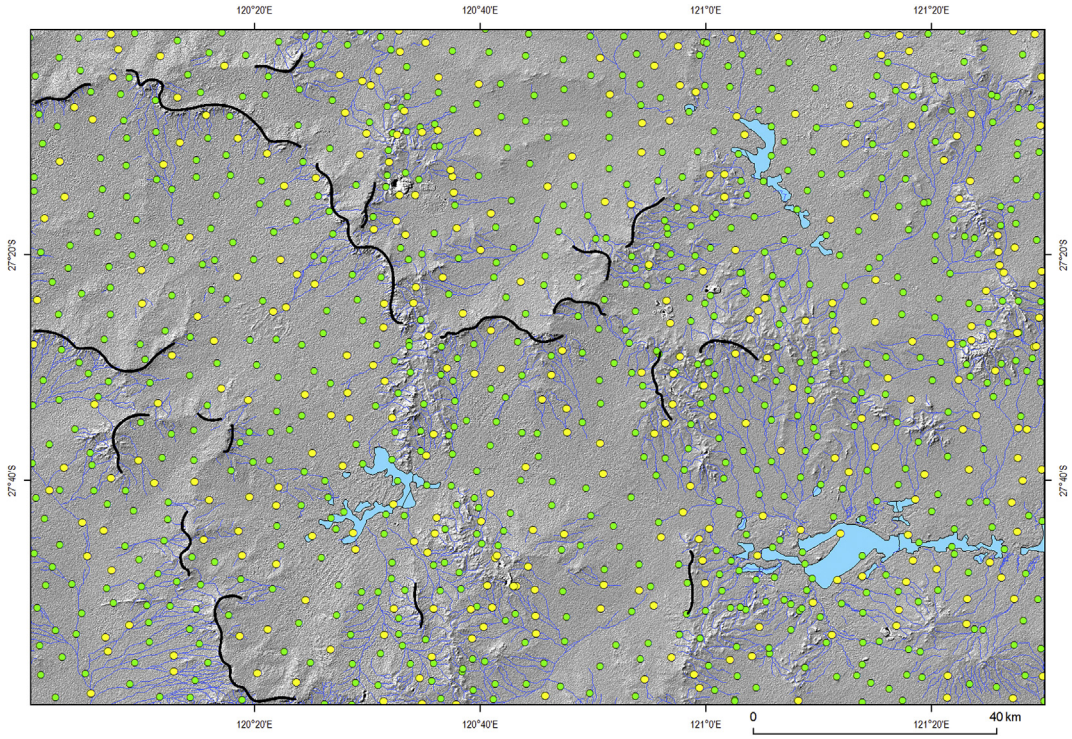

Fig. 2. Sample sites (718 training samples, green dots; 308 validation samples, yellow dots) overlain on a digital elevation model. Drainage and salt lakes are shown in blue and erosional scarps as thick black lines. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Summary statistics of the selected raw (Cr_ppm, Na_ppm) and Log(10) transformed (Cr_Log, Na_Log) regolith geochemistry data from the Sir Samuel study area. Values below the lower detection limit were imputed (see text).

	Cr_ppm	Cr_Log	Na_ppm	Na_log
Detection limit (DL)	20	1.30	74	1.87
N	1026	1026	1026	1026
N < DL	8	8	1	1
Minimum	14.98	1.18	54	1.74
Median	127	2.10	964	2.98
Mean	320	2.21	2998	3.06
Maximum	9714	3.99	49,040	4.69
Standard deviation	596	0.45	4744	0.61
Kurtosis	74.01	0.42	15.79	−0.74
Skewness	6.70	0.80	3.21	0.25

transformed. Fig. 3 and Table 1 show that considerable improvement in distribution was obtained by the transform.

3.2. Predictive modelling approach

We apply the Cubist data mining package (Quinlan, 1992; www.rulequest.com) to establish predictive relationships between site geochemical concentrations and the environmental covariates (see next Section). The Cubist model structure consists of decision tree splits defined by conditional statements coupled with multiple linear regression models operating at the leaves. Continuous as well as categorical variables can be used in the decision tree to split the data into more homogeneous sub-regions. Values within these sub-regions are then predicted by the linear regression models. Many linear models are generated in any one prediction and where model predictions overlap an average prediction value is calculated. This modelling approach enables local linear correlations of the data to be made within sub-regions defined by the decision trees. Detailed explanation of the Cubist modelling approach is described by Henderson et al. (2005), Bui et al. (2009) and Kuhn and Johnson (2013).

The Cubist model structure consists of many rules. An example is shown below (where c1, c2 and c3 are constants; see Table 2).

Rule 1:

If (conditional statement; decision tree splits)
Relief < 50 m
Gamma-ray K > 2.1%
Lithology = granite
Then (linear model)
Property = c1 × Topographic wetness index + c2 × Slope + c3 × Gamma-ray K

3.3. Environmental covariates

In digital soil mapping environmental covariates are selected to broadly reflect the soil-forming factors as described by Jenny (1941):

$S = f(cl,o,r,p,t)$

where the soil in its current state (S) is a function f of the following soil-forming factors: climate (cl), biological organisms (o), relief (r), parent material (p) and time (t). These factors are also important in predicting surface geochemistry, particularly in highly weathered landscapes where pedogenic/weathering and geomorphic processes exert a strong control on element distribution and concentration.

The major environmental covariate themes used in the geochemical prediction are briefly summarised below and are individually listed in Table 2 and in the Supplementary Data.

3.3.1. Climate

Key climate attributes of rainfall and temperature do not show

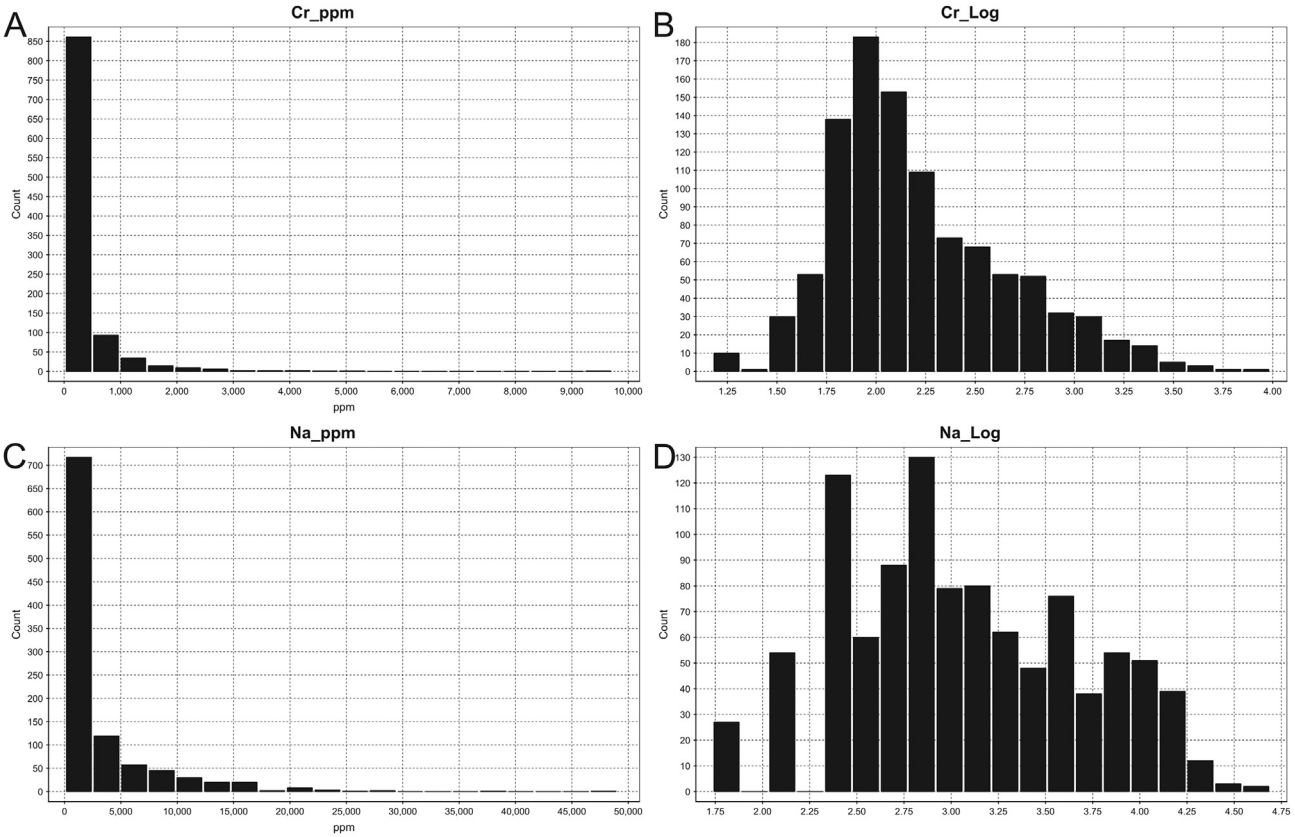

Fig. 3. Histograms of the selected raw (A – Cr_ppm, C – Na_ppm) and Log(10) transformed (B – Cr_Log, D – Na_Log) regolith geochemistry data from the Sir Samuel study area.

significant changes over the study area. We only included grids of average annual rainfall and the Prescott index, a leaching index that takes into account rainfall and evaporation (Gallant and Austin, 2012; Prescott, 1950). The leaching index is an estimate of excess water that can potentially percolate through the soil profile.

3.3.2. Soil and geology

Soil map units were derived from a digital version of the Atlas of Australian Soils at 1:2 million scale (Northcote et al., 1960–1968). Geology polygons were derived from the 1:1 million Surface Geology of Australia (Raymond, 2012). Geological units were classified into consolidated (bedrock) and unconsolidated materials (e.g. colluvium and alluvial sediments). Bedrock units were classified on their silica content (based on chemical compositions of common rock types; Gray and Murphy, 1999).

3.3.3. Terrain

Information on surface morphology was derived from an improved version of the 3 arc second (~90 m) Shuttle Radar Topographic Mission (SRTM) digital elevation model. The improved version has vegetation features removed and is smoothed to reduce noisy artefacts present in the original data (Gallant et al., 2011). The elevation model was used to generate a suite of terrain attributes which are listed in Table 2.

3.3.4. Satellite imagery

Advanced Spaceborne Thermal Emission and Reflectance Radiometer (ASTER) imagery provided information on surface mineralogy and vegetation. Moderate Resolution Imaging Spectroradiometer (MODIS) satellite imagery was processed to derive 12 coefficients based on Enhanced Vegetation Index (EVI) time-series data from 2000 to 2008 (Tan et al., 2009; Lymburner et al., 2011). These coefficients relate to a range of phenological parameters reflecting changes in greenness over that time period. These MODIS derived greenness coefficients are listed in Table 2 and a more detailed explanation of the coefficients is described by Tan et al. (2009). The coefficients were included to explore relationships between vegetation type and dynamics on the one hand and surface geochemistry on the other.

3.3.5. Geophysics

Available geophysical datasets included airborne magnetics, gamma-ray radiometrics and ground station gravity surveys. The magnetic intensity and the first vertical derivative of the magnetics were used as part of the covariate datasets. Airborne gamma-ray spectrometric data measures the surface concentration of three gamma-ray emitting radioelements ^{40}K, ^{208}Tl (used to estimate the abundance of Th = eTh) and ^{214}Bi (used to estimate the abundance of U = eU) (Wilford and Minty, 2007). The distribution of these elements varies in response to bedrock type and secondary weathering processes and soil formation (Wilford, 2012). Individual and ratio grids for K, eTh and eU were compiled from the Radiometric Map of Australia (Minty et al., 2009). Bouguer gravity anomaly data were extracted from the Australian National Gravity Database (Murray, 1997).

3.4. Model implementation, evaluation and comparison with geostatistical approach

The modelling approach involved randomly withholding a subset comprising 30% of the geochemical site data (N = 308) for external validation and using the remaining 70% (N = 718) as

Table 2
Environmental covariate predictive variables used in the Cr and Na Cubist models (see Supplementary Data for more detail).

Theme	No	Covariate	Description
Terrain	1	Elevation	Digital elevation model ~90 m resolution
	2	Regional elevation	Low band pass filtered elevation
	3	Multi-resolution valley bottom flatness	Defines low lying area, i.e. valley floors
	4	Aspect	Relief aspect
	5	Saga wetness index	Wetness index based on a modified catchment area
	6	Wetness index	Topographic wetness index
	7	Relief	Relief defined as the elevation range within a circular window of 300 m
	8	Local relief	Relief defined as the elevation range within a circular window of 100 m
	9	Regional relief	Relief defined as the elevation range within a circular window of 2000 m
	10	Slope	Degree of elevation fall of the topographic surface
	11	Local slope	Median slope within a 300 m radius
	12	Erosional-Depositional	Geology split on bedrock and transported units
Climate	13	Prescott index	The Prescott Index is a measure of water balance
Geology/geochemistry	14	Potassium (%)	Airborne gamma-ray derived from radioelemental potassium
	15	Thorium (ppm)	Airborne gamma-ray derived from radioelemental thallium-208
	16	Uranium (ppm)	Airborne gamma-ray derived from radioelemental bismuth-214
	17	Ratio Th/K	Ratio of gamma-ray measured thorium/potassium
	18	Gravity	Image based on ground based gravity measurements
	19	Magnetic intensity	Airborne magnetic survey
	20	Regional magnetic	Average magnetic intensity within a 1000 m circular window
	21	Weathering intensity	Index representing intensity of surface weathering
	22	Lithology type	Surface Geology of Australia 1:1 million scale
	23	Lithology silica	Lithology classified on silica content
	24	Soil units	Australian Soil Classification (ASC)
	25	Distance from outcrop	Euclidean distance from outcrop
	26	Gravity gradient	Measures rate of change in the earth's gravity field
	27	[a]ASTER regolith 1	ASTER satellite ratio bands 3/2
	28	ASTER regolith 2	ASTER satellite ratio bands 3/7
	29	ASTER regolith 3	ASTER satellite ratio bands 34/7
	30	ASTER Ferrous Fe	ASTER satellite ratio bands 5/4
Vegetation	31	[b]MODIS 1	Mean of [c]EVI values of the time period
	32	MODIS 2	Standard deviation of EVI values over the time period
	33	MODIS 3	Flatness − measure of low EVI values
	34	MODIS 4	Rate of rise (average rate of rise within the time period)
	35	MODIS 5	Rate of drop (average rate of drop within the time period)
	36	MODIS 6	Global minimum (lowest value within the time period)
	37	MODIS 7	Mean length of a cycle between peaks and troughs of the EVI
	38	MODIS 8	Global maximum (highest value within the time period)
	39	MODIS 9	Max ratio (global maximum/annual maximum)
	40	MODIS 10	Mean of the timing of the maximum
	41	MODIS 11	Standard deviation of the timing of the maximum
	42	MODIS 12	Annual minimum

[a] ASTER = Advanced Spaceborne Thermal Emission and Reflectance Radiometer.
[b] MODIS = Moderate Resolution Imaging Spectroradiometer
[c] EVI = Enhanced Vegetation Index.

training data (Fig. 2). The Cubist model was then run 30 times using bootstrap with sample replacement to produce a unique rule set for each run. Sample replacement insured that approximately 30% of the sites were randomly left out of the analysis for each model since any given site could be selected more than once. This procedure improved the stability and uncertainty of the model and allowed calculation of model uncertainty. The Cubist rule sets generated were then evaluated and applied spatially calculating a median predicted value (i.e. the final model). The difference between the 95th and 5th quantiles (Q95-Q5) of the 30 model runs was used as a measure of statistical dispersion. This dispersion was normalised to the median value (Q95-Q5/Q50) as an estimate of the model uncertainty (Arrouays et al., 2014). Higher uncertainties reflect a greater spread of the predicted values (relative to the median).

In contrast a geostatistical approach to geochemical mapping relies on the construction of variograms to establish the predictability of values from place to place (Clark and Harper, 2000). The constructed semi-variogram models the relationship between a

Table 3
Parameters of the Ordinary Kriging Cr and Na interpolations with 718 sites.

Cr
Model type: exponential
Nugget variance: 0.0293
Sill: 0.1674
Range or lag distance: 45,584.64 m
Na
Model type: exponential
Nugget variance: 0.0427
Sill: 0.3348
Range or lag distance: 18,905.88 m

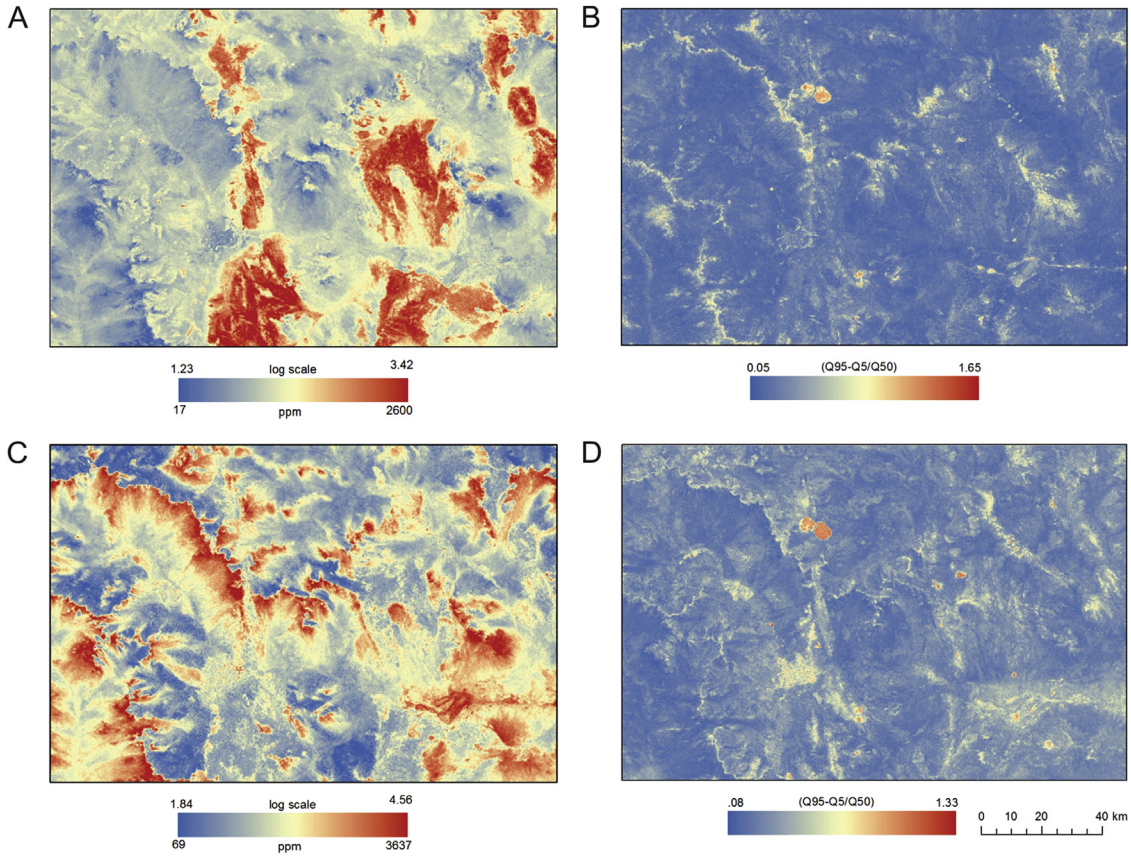

Fig. 4. A – Cubist prediction map for Cr; B – Cubist prediction uncertainty map (Q95-Q5/Q50) for Cr; C – Cubist prediction map for Na; D – Cubist prediction uncertainty map (Q95-Q5/Q50) for Na.

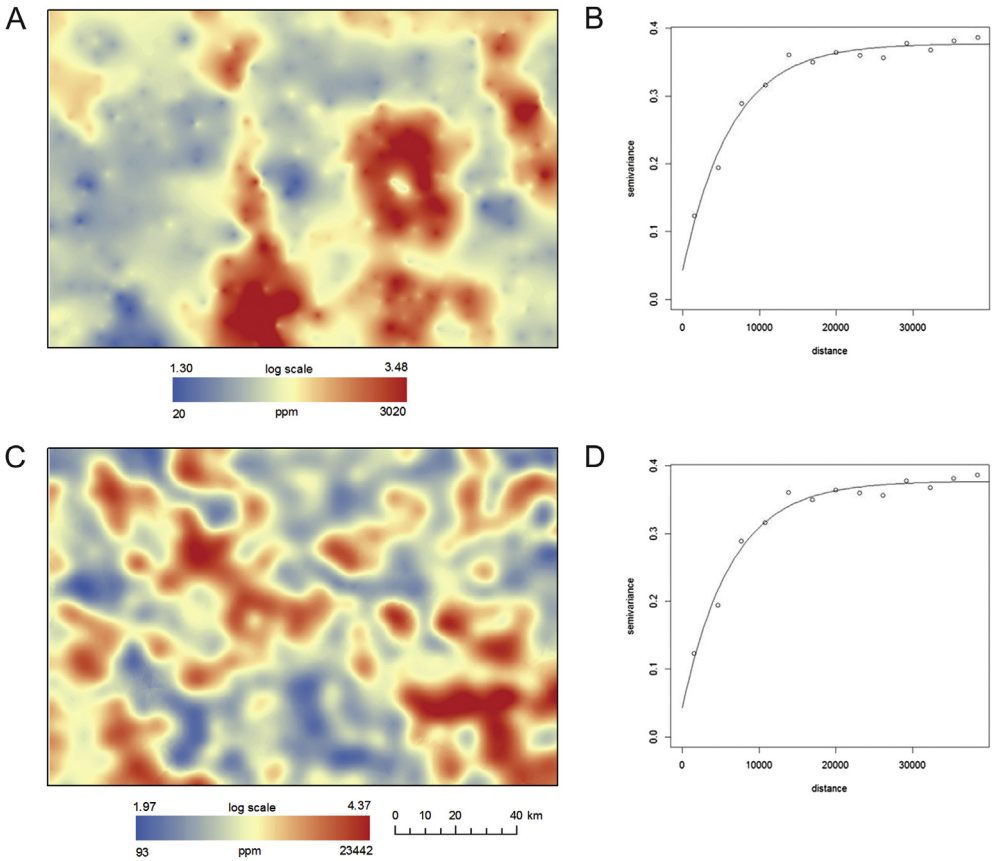

Fig. 5. A – Ordinary Kriging map for Cr; B – Corresponding variogram (distance in m); C – Ordinary Kriging map for Na; D– Corresponding variogram (distance in m).

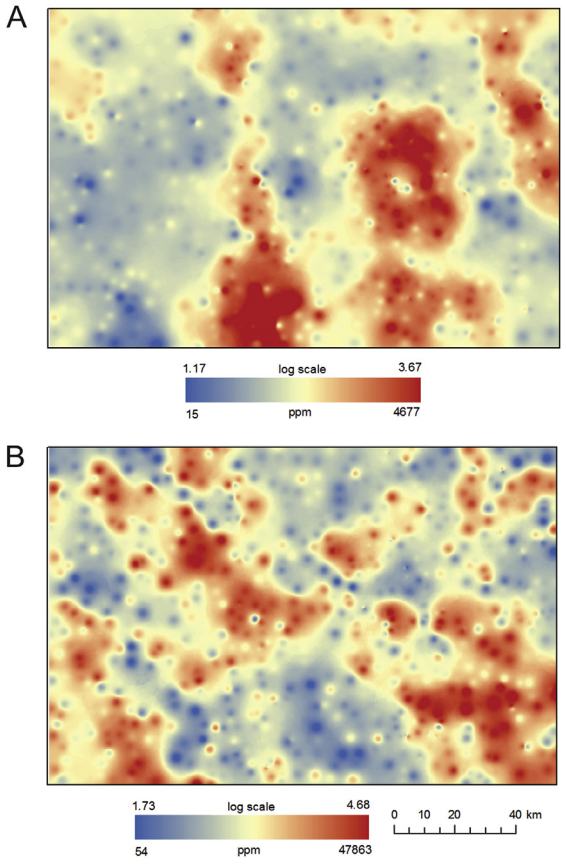

Fig. 6. A – Inverse Distance Weighting map for Cr; B – Inverse Distance Weighting for map Na.

value at one location and the value at another as a function of the distance and direction between them, accounting for spatial autocorrelation. Ordinary Kriging (OK) with an exponential semivariogram model was used to predict the value of Cr and Na concentrations at unsampled points. We also compared the covariate model with the IDW interpolation method. Both these spatial interpolation methods are based on the assumption of spatial autocorrelation (samples closer together are more likely to be similar than those further apart). Inverse distance weighted interpolation is a local deterministic method where the weighting factor diminishes with increasing distance. Both methods were implemented in ARCMap 10.2 – Geospatial Analyst (www.esri.com). Kriging parameters are shown in Table 3; parameters used for IDW included 'power2' interpolation with a searching neighbourhood of 15 (at least 10), major and minor semiaxis of 46,749 m.

Measures of model performance based on the 30% out-of-sample subset included the coefficient of determination (R^2), root mean square error (RMSE) and goodness-of-prediction (G) estimate. The R^2 values were calculated from the Pearson correlation coefficient (R) and indicate how well predictions and observations match. The closer R^2 is to unity the better.

The RMSE gives an indication of the accuracy of the model or how well the model is able to predict observed values, according to:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(z(x_i) - \hat{z}(x_i)\right)^2}$$

where $z(x_i)$ is the observed value at location i, $\hat{z}(x_i)$ the predicted value at location i, and n is the sample size. The smaller the RMSE the better.

The G estimate assesses how good the prediction or interpolation is compared with just using the sample mean (Karydas et al., 2009), according to:

$$G = \left(1 - \frac{\sum_{i=1}^{n}\left[z(x_i) - \hat{z}(x_i)\right]^2}{\sum_{i=1}^{n}\left[z(x_i) - \bar{z}(x_i)\right]^2}\right) \times 100\%$$

where all symbols as per above, and $\bar{z}(x_i)$ is the sample mean. If predictions are more reliable than the use of the sample mean, G values are positive and range up to 100% (which is a perfect model).

4. Results and discussion

4.1. Maps/models and uncertainties

The Cr and Na maps generated using the Cubist, OK and IDW methods are shown in Figs. 4–6. All maps were evaluated with the same 30% out-of-sample sites. Each method was compared in terms of the R^2, RMSE and G estimate (Table 4). Both interpolation methods (OK and IDW) were similar in terms of their accuracy and mapping effectiveness. The Cr interpolation maps had significantly higher R^2 and G with corresponding lower RMSE compared to Na (Table 4). The Cubist maps for Cr and Na are shown in Fig. 4A,C and model uncertainty maps in Fig. 4B,D. The Cubist models for Cr and Na were more robust in terms of R^2, RMSE and G compared with the interpolation methods, particularly for the Na predictions (Table 4).

Comparison of the range of concentration values (Fig. 3, Table 1) to the predictive model outputs (Fig. 4A,C) suggests that Cubist tends to under predict the target variable. Both Cubist uncertainty maps (Fig. 4B,D) show the largest uncertainty at two circular anomalies in the northwest quadrant of the study area. This is the Mt Keith open pit nickel mine and its tailings, an environment that was not directly targeted by the regolith sampling. As a result Cubist struggles to fit consistent rules and predicted values in this unsampled environment, yielding the largest variation/noise among the 30 model runs generated here. Outside of the Mt Keith mine site, the Cubist uncertainty is generally much lower.

4.2. Element distributions and insight into processes from the environmental predictors

4.2.1. Predictive maps

The Cubist models presented here are based on statistical correlations between the Cr and Na values and the covariate datasets. The resolution of the prediction grids therefore reflects that of the covariates, approximately 90 m. The accuracy of the model will depend on the strength of the correlation between the element and the covariates. In contrast, OK and IDW grids of element concentrations are only interpolations between points ignoring changes in any other property between them. The accuracy of the interpolation reflects the degree of correlations between points and how well these points represent the area. Geochemical variations between points will largely reflect the scale at which landscape processes operate. For example, sample density and landscape processes scale are likely to explain the differences in prediction accuracy between Cr and Na.

The distribution of Cr is spatially clustered and genetically related to the mafic and ultramafic rocks forming the north-northeast and north-northwest trending greenstone belts (Fig. 1). Chromium is relatively immobile in the weathering environment and elevated Cr tends to be largely restricted to these lithologies (Figs. 1 and 7A). The sampling density of 718

Table 4
Model performance based on 30% out-of-sample validation (see maps on Figs. 4–6).

	Cubist Cr	OK Cr	IDW Cr	Cubist Na	OK Na	IDW Na
R^2	0.70	0.66	0.65	0.67	0.32	0.30
RMSE	0.25	0.27	0.28	0.34	0.49	0.50
G	70%	66%	65%	65%	31%	29%

Table 5
Model performance based on 90% out-of-sample validation (sensitivity analysis).

	Cubist Cr	OK Cr	IDW Cr	Cubist Na	OK Na	IDW Na
R^2	0.58	0.35	0.32	0.47	0.07	0.09
RMSE	0.29	0.36	0.37	0.44	0.59	0.58
G	58%	34%	30%	45%	6%	6%

sites (= 70% training) over the map sheet (or 1 site per ~22 km^2) broadly captures the Cr distribution adequately (OK Cr $R^2 = 0.66$, IDW Cr $R^2 = 0.65$). Sodium on the other hand is less clustered with local highs corresponding to both primary and secondary processes (Fig. 7B). The same sampling density poorly captures the Na variability because the landscape process scale that controls the distribution of Na is much finer than the sampling density. In contrast the Cubist models provide considerable more spatial detail than the OK and IDW maps and are less dependent on the number of site observations to explain or predict the distribution of the elements. The predictive power of correlative modelling approach is particularly evident when reducing the density of the site observations. Even when only 10% of the sites are used for training and the remaining 90% for validation (Table 5) the correlative modelling technique is significantly better than the interpolation methods. Furthermore, the Cubist approach can be used to predict in areas without site measurements as long as the rules and associated covariate relationships hold true for the new area (e.g. areas where the environmental controls on element distribution are similar).

4.2.2. Process understanding

Explicitly expressed rules generated by the Cubist data mining approach allow exploration of the factors controlling the distribution of Cr and Na. Analysis of how the environmental covariates are used in the model prediction can improve our understanding of the often complex inter-relationships between environmental/

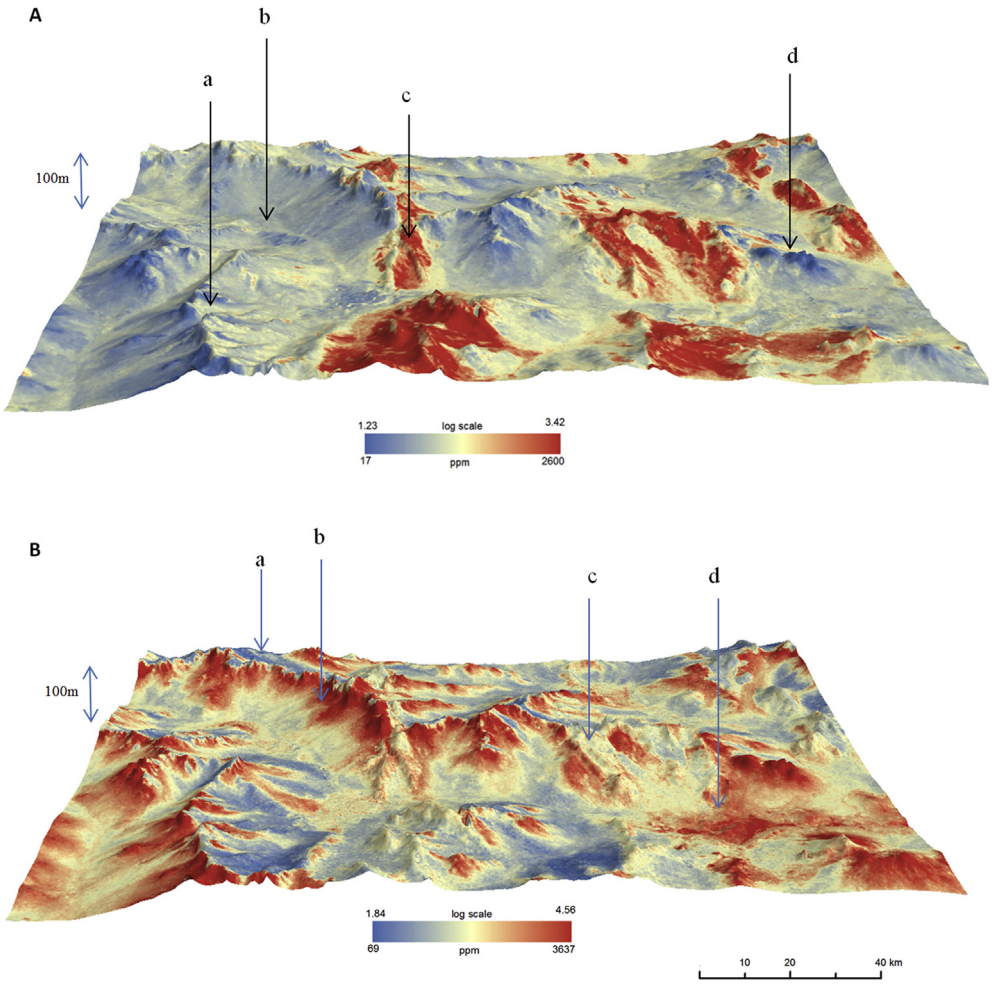

Fig. 7. A – 3D drape of Cubist prediction model for Cr over digital elevation model (a – moderate Cr values developed on very highly weathered granite; b – low Cr associated with sheet flood fans and valley floors; c – elevated Cr associated with mafic rocks; d – low Cr corresponding to exposed granitic rocks); B – 3D drape of Cubist prediction model for Na over digital elevation model (a – low Na corresponding to highly leached regolith profiles; b – elevated Na associated with exposed granitic rocks with Na-plagioclase; c – moderate to low Na over mafic lithologies; d – high Na associated with evaporitic salts and lacustrine sediments). All models looking north, east-west extent ~150 km, vertical exaggeration ~20.

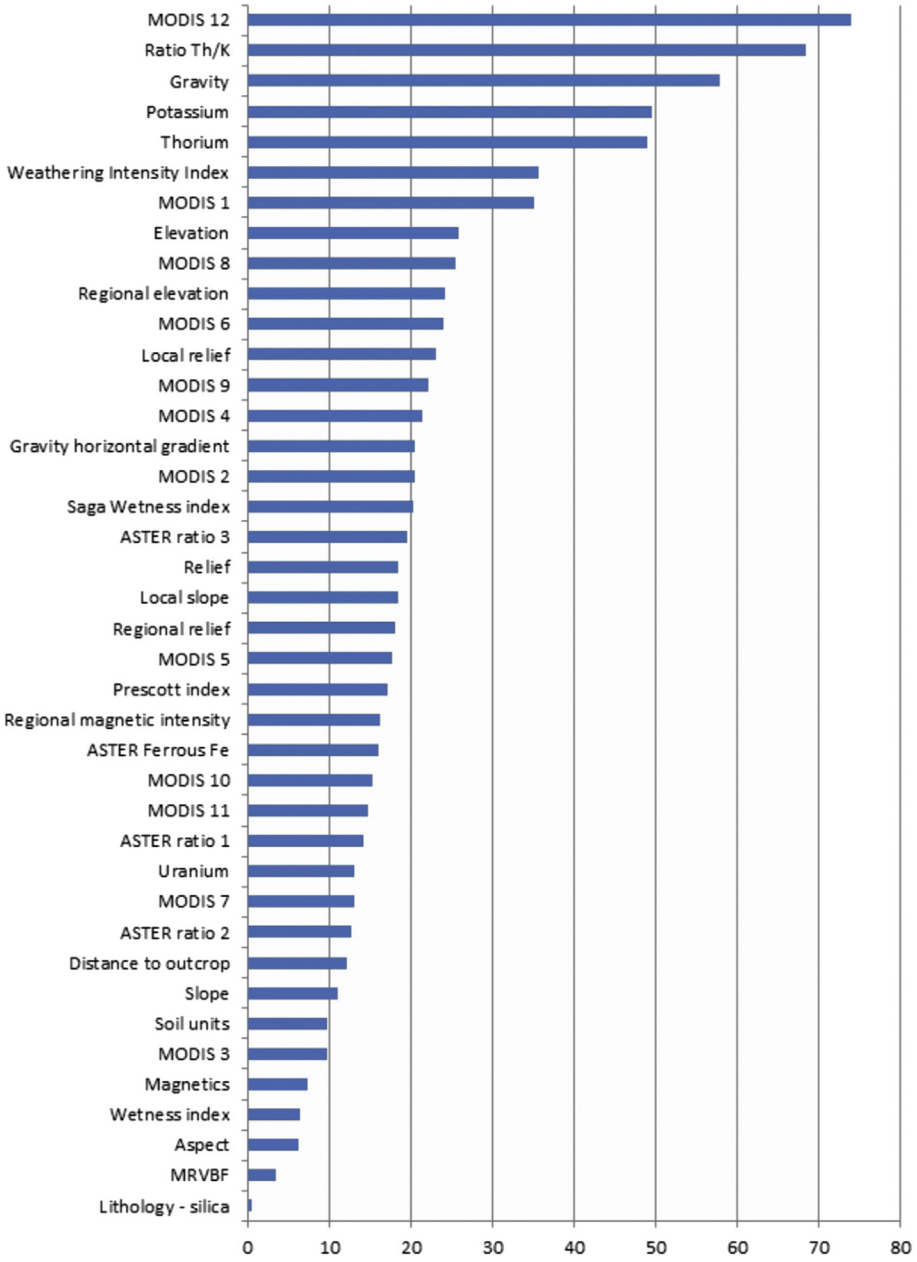

Fig. 8. Percentage covariate usage in both the condition and model components of the Cr prediction.

landscape processes and element distribution. Furthermore, the rules provide a framework to test our conceptual models of landscape geochemistry and in the case of Cr can be used to target potential areas of mineralisation.

Chromium in the Yilgarn Craton is mainly found as the oxide mineral chromite, which is largely associated with mafic lithologies and the ferruginous weathered materials derived therefrom (Scott, 1990). The Cr prediction model is mostly based on MODIS greenness coefficients, gamma-ray radiometrics, particularly the Th/K ratio, gravity, weathering intensity and to a lesser extent terrain attributes and ASTER (Fig. 8). The emergence of MODIS 12 greenness coefficient – low annual greenness – as the most important predictor of spatial Cr pattern is interesting and requires further investigation. Its use may have a plant physiological explanation because Cr is toxic to growth (Shanker et al., 2005). Alternatively a geobotanical relationship may exist where a particular vegetation type is growing preferentially on highly ferruginous soils developed over mafic rocks. Soil biological processes may be involved too since Cr is readily absorbed by organic matter (Bartlett and Kimble, 1976; Scott and Pain, 2008) and can be mobilised by microorganisms (McFarlane et al., 1994). Mafic rocks and residual weathering profiles developed on them tend to be high in Th and low in K. Hence, high Th/K ratio values are a characteristic feature of mafic/ultramafic rocks. These lithologies are also denser and consequently have higher gravity response compared to adjacent granitic rocks and sediments.

From a mineral exploration perspective the Cr model is delineating elevated Cr values within the exposed mafic greenstone belts but also high Cr values in areas of shallow cover where mafic rocks are concealed. This is evident when comparing the Cr model

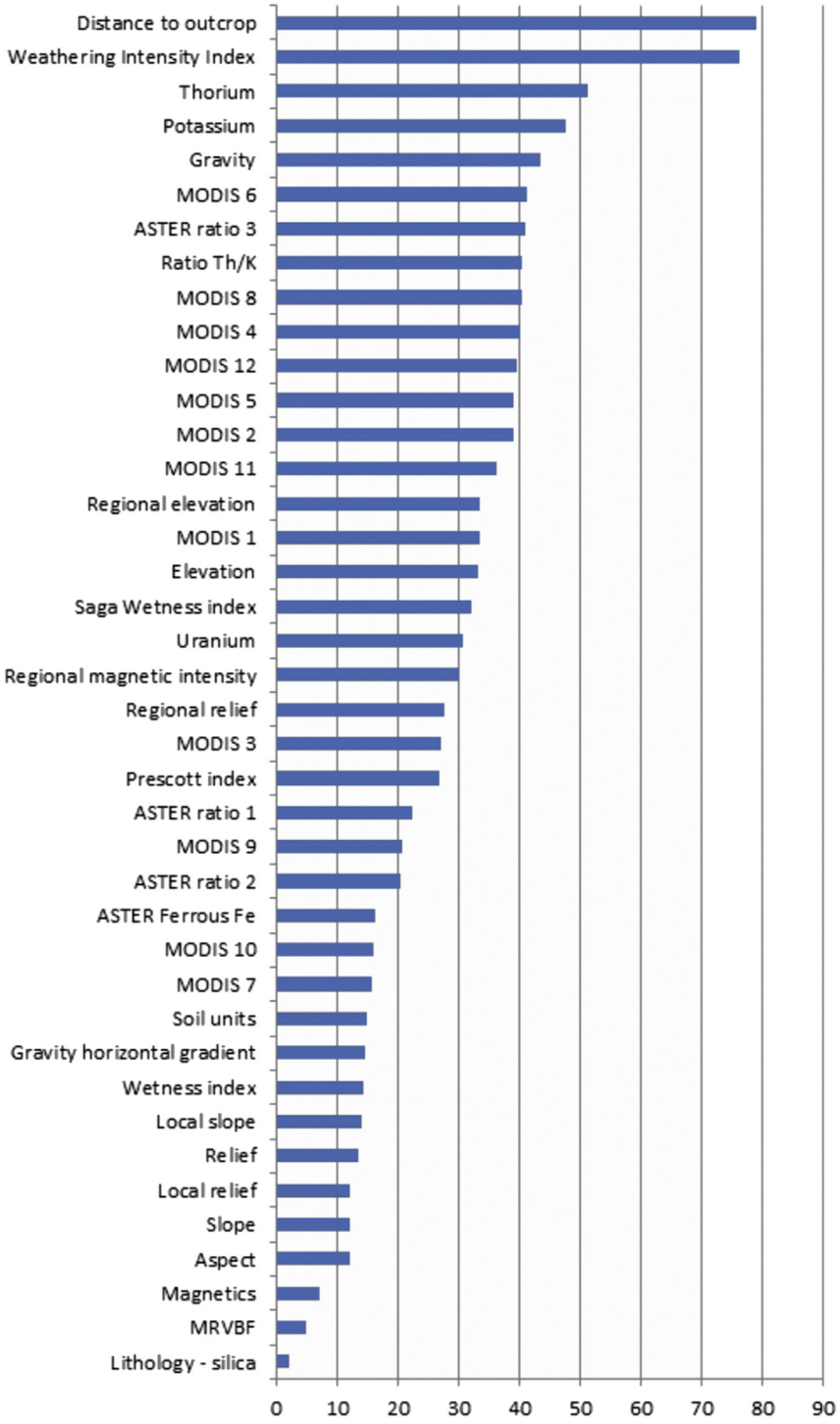

Fig. 9. Percentage covariate usage in both the condition and model components of the Na prediction.

(Fig. 4A) with the surface geology map (Fig. 1). Mafic rocks are generally poorly exposed however high modelled Cr values link these scattered outcrops through relationships between residual accumulation of Cr in the regolith and the environmental covariates (particularly magnetic intensity and MODIS). The result of this approach is thus more insightful than predictive lithological mapping (e.g. Harris and Grunsky, 2015; Martelet et al., 2006; Schetselaar et al., 2000).

The Na model prediction is largely based on distance to outcrop, weathering intensity, Th, K, gravity and to a lesser extend MODIS greenness coefficients, ASTER and terrain attributes (Fig. 9). Sodium concentrations are associated with plagioclase in exposed felsic rocks including granite, rhyolite and syenite. Sheet flood fans shedding sediments down slope from these felsic lithologies are also high in Na (Fig. 7B). Sodium is highly mobile during weathering and pedogenesis and is associated with evaporitic minerals (e.g.

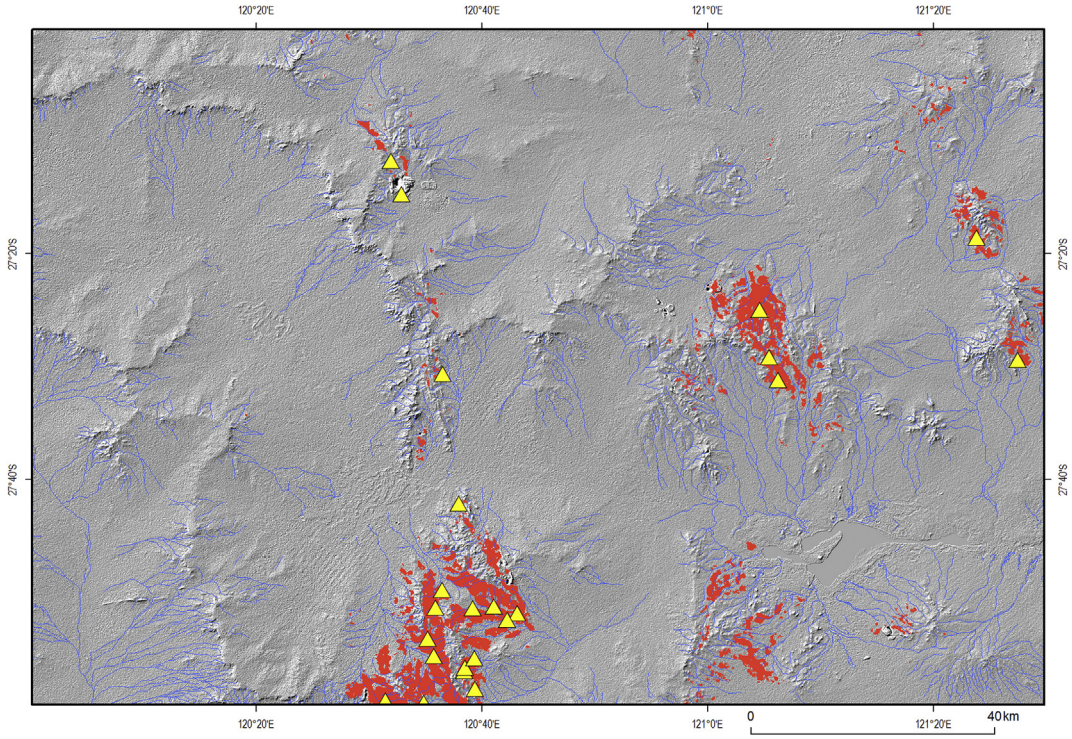

Fig. 10. Cubist prediction map for Cr with anomalous values (based on Tukey outliers; see text) highlighted (red pixels), overlain on a hill shaded digital elevation model with drainage (blue lines). The actual samples identified as outliers are shown as yellow triangles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

precipitation of NaCl) and discharge sites along valley floors and salt lakes. It is therefore not surprising that the distance-to-outcrop covariate, which effectively separates bedrock highs from valley floor deposits, and Saga wetness index are important predictors in the Na model. Radiometric K dose delineates felsic bedrock with high Na-plagioclase whilst high weathering intensity highlights areas where intense leaching has removed Na from the upper part of the weathering profile. These highly leached, Na depleted zones occur above local erosional scarps over granitic bedrock (Fig. 7B).

Regolith chemistry is a key attribute of the critical zone, the layer between unweathered bedrock and the tree tops where the geosphere, hydrosphere, biosphere, and atmosphere interface (Jordan et al., 2001). Regolith chemistry will reflect the natural composition of the bedrock and changes in response to chemical, physical and biological processes over time (Brantley and Lebedeva, 2011). The application of environmental correlation as described in this paper provides a mechanism to understand the processes operating within the critical zone and the interaction between the multiple factors controlling the nature of regolith geochemistry over extensive areas. It allows us to go beyond the traditional approach that relies on univariate functions of soil formation factors (toposequences, climosequences, lithosequences, biosequences, or chronosequences).

Correlative modelling is likely to be most efficient when various parts of the landscape have been sampled to capture the full spectrum of environmental conditions (terrain, climate, geology, vegetation, etc.). Therefore strictly adhering to a single, well constrained geochemical sampling medium is not a prerequisite for applying correlative modelling. Thus sampling schemes specifically targeting a range of conditions, such as Latin hypercube (McBratney et al., 2003), may be particularly well suited to this type of decision tree modelling. This also means that legacy geochemical datasets with multi-media geochemical results are potentially amenable to Cubist-type correlative environmental modelling, with ensuing interpretative power.

4.2.3. Application to mineral exploration

The ability to model the distribution of a chemical element in relation to a number of controlling variables offers the potential to map the distribution of pixels above the local baseline ('background'), which could highlight areas worthy of further investigation. To illustrate this, we identified and mapped the 'Tukey outliers' (Tukey, 1977) within the Cr population. Such outliers are not arbitrarily chosen, but their presence (or absence) is controlled by the internal structure of the dataset at hand (specifically, how tight the box in a boxplot is compared to the data's full range). Tukey outliers are defined as those samples for which the Log10 (Cr) concentrations exceed the upper fence; the upper fence is defined as 1.5 times the hinge spread (HS, or Q75-Q25) added to the 75th percentile (Log10 (Cr) > Q75 + 1.5 × HS). In the present case study, the upper fence is at Log10 (Cr) = 3.33, resulting in 21 outliers. Subsequently, we highlighted every pixel in the Cr Cubist model with a value exceeding 3.33–0.25 (the RMSE), or 3.08 (Fig. 10). The subtraction operation is justified by the concern that we did not want to miss any potential anomaly because of the uncertainty in the Cubist model. The anomalous regions highlighted in Fig. 10 are much more extensive than suggested by the anomalous samples alone and may warrant further exploration investment (e.g. for nickel, chromium, or platinum group elements).

5. Conclusions

Generating predictive element compositional maps from geochemical survey data using environmental correlative modelling has significant advantages over traditional geostatistical

interpolation methods such as kriging and inverse distance weighting. Application of environmental correlation in our study area produced quantitative geochemical landscape models with higher accuracy and higher resolution than the interpolation methods. The approach is particularly advantageous where the survey sampling density is not commensurate with the landscape process scale. This integrated modelling incorporates extensive high-resolution spatial data (e.g. from satellite and airborne sources) to value-add to geochemical survey data. The approach has the potential to model the full range of major and trace element compositions, principal components of these elements, various indices or ratios devised for exploration or environmental assessment purposes, and even isotopes. Furthermore, this covariate modelling approach provides a knowledge-discovery platform to globally improve our understanding of the complex web of physical, chemical and biological interactions occurring within the critical zone that control geochemical landscapes, be it for environmental, earth or resource science applications.

Acknowledgements

We wish to thank David Champion and David Huston from Geoscience Australia, as well as the journal referees, for their ideas, comments and corrections in reviewing the paper. We also wish to thank the Geological Survey of Western Australia and specifically Paul Morris for making the geochemistry dataset available. We appreciate the input from Ross Searle from CSIRO who helped with the implementation of the Cubist algorithm. John Wilford and Patrice de Caritat publish with permission from the Chief Executive Officer, Geoscience Australia.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.apgeochem.2015.08.012.

References

Adhikari, K., Bou Kheir, R., Greve, M.B., Greve, M.H., 2013. Comparing kriging and regression approaches for mapping soil clay content in a diverse Danish landscape. Soil Sci. 178 (9), 505–517.
Anand, R.R., Paine, M., 2002. Regolith geology of the Yilgarn Craton, Western Australia: implications for exploration. Aust. J. Earth Sci. 49, 3–162.
Arrouays, D., McKenzie, N., Hempel, J., de Forges, A.R., McBratney, A.B., 2014. Global Soil Map: Basis of the Global Spatial Soil Information System. Taylor and Francis.
Bartlett, R.J., Kimble, J.M., 1976. Behaviour of chromium in soils: I, trivalent forms. J. Environ. Qual. 5, 379–386.
Beard, J.S., Beeston, G.R., Harvey, J.M., Hopkins, A.J.M., Shepherd, D.P., 2013. The Vegetation of Western Australia at the 1:3,000,000 Scale. Explanatory Memoir, second ed.9. Conservation Science, WA, pp. 1–152.
Bou Kheir, R., Shomar, B., Greve, M.B., Greve, M.H., 2014. On the quantitative relationships between environmental parameters and heavy metals pollution in Mediterranean soils using GIS regression-trees: the case of Lebanon. J. Geochem. Explor. 147, 250–259.
Brantley, S.L., White, T.S., Ragnarsdottir, K.V., 2007. The critical zone. Elements 3, 307–338.
Brantley, S.L., Lebedeva, M., 2011. Learning to read the chemistry of regolith to understand the critical Zone. Annu. Rev. Earth Planet. Sci. 39, 387–416.
Bunting, J.A., Williams, S.J., 1979. Sir Samuel, W.A. West. Aust. Geol. Surv., 1:250 000 Geol. Ser. Explan. Notes 40.
Bui, E.N., Henderson, B., Viergever, K., 2009. Using knowledge discovery with data mining from the Australian soil resource information system database to inform soil carbon mapping in Australia. Glob. Biogeochem. Cycles 23. GB4033. http://dx.doi.org/10.1029/2009GB003506.
Caritat, P. de, Cooper, M., 2011. National geochemical survey of Australia: the geochemical atlas of Australia. Geosci. Aust. Rec. 2011/20 (2 volumes), 557. Available at: http://www.ga.gov.au/metadata-gateway/metadata/record/gcat_71973.
Caritat, P. de, Grunsky, E.C., 2013. Defining element associations and inferring geological processes from total element concentrations in Australian catchment outlet sediments: multivariate analysis of continental-scale geochemical data. Appl. Geochem. 33, 104–126.
Clark, I., Harper, W.V., 2000. Practical Geostatistics. Ecosse North American LLC, Columbus, OH, USA, p. 416.
Cohen, D.R., Rutherford, N.F., Morisseau, E., Zissimos, A.M., 2012. Geochemical patterns in the soils of Cyprus. Sci. Total Environ. 420, 250–262.
Cracknell, M.J., Reading, A.M., 2014. Geological mapping using remote sensing data: a comparison of five machine learning algorithms and response to variations in the spatial distribution of graining data and use of explicit spatial information. Comput. Geosci. 63, 22–33.
Darnley, A.G., Björklund, A., Bølviken, B., Gustavsson, N., Koval, P.V., Plant, J.A., Steenfelt, A., Tauchid, M., Xie, X., Garrett, R.G., Hall, G.E.M., 1995. A Global Geochemical Database for Environmental and Resource Management. Recommendations for International Geochemical Mapping, Final Report of IGCP Project 259. UNESCO Publishing, p. 122.
Drever, J.I., 2014. Surface and groundwater, weathering and soils. In: Holland, H., Turekian, K. (Eds.), Treatise on Geochemistry, , second ed.7. Elsevier, p. 562 http://dx.doi.org/10.1016/B978-0-08-095975-7.09986-1.
Eggleton, R.A., Anand, R.R., Butt, C.R.M., Chen, X.Y., Craig, M.A., Caritat, P. de, Field, J.B., Gibson, D.L., Greene, R., Hill, S.M., Jones, M., Lintern, M.J., McQueen, K.G., Pain, C., Pillans, B., Robertson, I.D.M., Smith, K., Taylor, G., 2001. The Regolith Glossary – Surficial Geology, Soils and Landscapes. Cooperative Research Centre for Landscape Evolution and Mineral Exploration, ISBN: 0-7315-3343-7144.
Gallant, J., Austin, J., 2012. Prescott Index Derived from 1″ SRTM DEM-s. V2, , CSIRO. Data Collection. http://dx.doi.org/10.4225/08/53EB2D0EAE377. Available at: https://data.csiro.au/dap/landingpage?pid=csiro%3A9636.
Gallant, J.C., Wilson, N., Dowling, T., Read, A., Inskeep, C., 2011. SRTM-derived 3 Second Digital Elevation Models Version 1.0. Available at: http://www.ga.gov.au/metadata-gateway/metadata/record/gcat_72760.
Garrett, R.G., Reimann, C., Smith, D.B., Xie, X., 2008. From geochemical prospecting to international geochemical mapping: a historical overview. Geochem. Explor. Environ. Anal. 8, 205–217.
Gray, J.M., Murphy, B.W., 1999. Parent Material and Soils. A Guide to the Influence of Parent Material on Soil Distribution in Eastern Australia, Technical Report No. 45. New South Wales Department of Land and Water Conservation, Sydney.
Harris, J.R., Grunsky, E.C., 2015. Predictive lithological mapping of Canada's North using random Forest classification applied to geophysical and geochemical data. Comput. Geosci. 80, 9–25.
Henderson, B.L., Bui, E.N., Moran, C.J., Simon, D.A.P., 2005. Australia-wide predictions of soil properties using decision trees. Geoderma 124, 383–398.
Hron, K., Templ, M., Filzmoser, P., 2010. Imputation of missing values for compositional data using classical and robust methods. Comput. Stat. Data Anal. 54, 3095–3107.
Jenny, H., 1941. Factors of Soil Formation – a System of Quantitative Pedology. McGraw Hill Book Co, New York.
Johnson, S.L., 2004. Hydrogeology of the Sir Samuel 1:250 000 Sheet, Department of Environment, Hydrogeological Map Explanatory Notes Series, Report HM, vol. 11, p. 26. Western Australia.
Jordan, T., Ashley, G.M., Barton, M.D., Burges, S.J., Farley, K.A., Freeman, K.H., Jeanloz, R., Marshall, C.R., Orcutt, J.A., Richter, F.M., Royden, L.H., Scholz, C.H., Tyler, N., Wilding, L.P., 2001. Basic Research Opportunities in Earth Science. National Academy Press, Washington, D.C.
Karydas, C.G., Gitas, I.Z., Koutsogiannaki, E., Lydakis-Simantiris, N., Silleos, G.N., 2009. Evaluation of spatial interpolation techniques for mapping agricultural topsoil properties in Crete. EARSeL eProceedings 8 (1), 26–39.
Kojan, C.J., Faulkner, J.A., Sanders, A.J., 1996. Geochemical Mapping of the Sir Samuel 1:250 000 Sheet, Explanatory Notes. Geol. Surv. of West. Aust. 69.
Kuhn, M., Johnson, K., 2013. Applied Predictive Modelling. Springer, New York, USA.
Lacoste, M., Lemercier, B., Walter, C., 2011. Regional mapping of soil parent material by machine learning based on point data. Geomorphology 133, 90–99.
Lam, N.S.N., 1983. Spatial interpolation methods: a review. Am. Cartogr. 10, 129–149.
Li, J., Heap, A.D., 2011. A review of comparative studies of spatial interpolation methods in environmental sciences: performance and impact factors. Ecol. Inf. 6, 228–241.
Liu, S.F., Champion, D.C., Cassidy, K.F., 2002. Geology of the Sir Samuel 1:250 000 Sheet Area, Western Australia: Geo. Aust. Rec. 2002/14, 57.
Lymburner, L., Tan, P., Mueller, N., Thackway, R., Thankappan, M., Islam, A., Lewis, A., Randall, R., Senarath, U., 2011. The National Dynamic Land Cover Dataset. Geo. Aust. Rec. 2011/31, 95. Available at: https://www.ga.gov.au/products/servlet/controller?event=GEOCAT_DETAILS&catno=71069.
Martelet, G., Truffert, C., Tourliere, B., Ledru, P., Perrin, J., 2006. Classifying airborne radiometry data with agglomerative hierarchical clustering: a tool for geological mapping in context of rainforest (French Guiana). Int. J. Appl. Earth Obs. Geoinformation 8 (3), 208–223.
McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52.
McFarlane, M.J., Bowden, D.J., Giusti, L., 1994. The behaviour of chromium in weathering profiles associated with the African surface in parts of Malawi. In: Robinson, D.A., Williams, R.B.G. (Eds.), Rock Weathering and Landform Evolution. John Wiley, Chichester, pp. 321–338.
McKenzie, N.J., Ryan, P.J., 1999. Spatial prediction of soil properties using environmental correlation. Geoderma 89, 67–94.
Minty, B., Franklin, R., Milligan, P., Richardson, L.M., Wilford, J., 2009. New radiometric map of Australia. Explor. Geophys. 40, 325–333.
Morris, P.A., 2000. Composition of Geological Survey of Western Australia Geochemical Reference Materials (GRMs). West. Aust. Geol. Surv. Rec. 2000/11.

Morris, P.A., Sanders, A.J., Pirajno, F., Faulkner, J.A., Coker, J., 1999. Regional-scale regolith geochemistry: identification of metalloid anomalies and the extent of bedrock in the Archean and Proterozoic of Western Australia. In: Taylor, G., Pain, C. (Eds.), New Approaches to an Old Continent. Proceedings of Regolith '98. Cooperative Research Centre for Landscape Evolution & Mineral Exploration, Perth, pp. 101–108.

Murray, A.S., 1997. The Australian national gravity database. AGSO J. Aust. Geol. Geophys. 17, 145–155.

Northcote, K.H., Beckmann, G.G., Bettenay, E., Churchward, H.M., Van Dijk, D.C., Dimmock, G.M., Hubble, G.D., Isbell, R.F., McArthur, W.M., Murtha, G.G., Nicolls, K.D., Paton, T.R., Thompson, C.H., Webb, A.A., Wright, M.J., 1960-1968. Atlas of Australian Soils, Sheets 1 to 10, with Explanatory Data. Commonwealth Scientific and Industrial Research Organisation and Melbourne University Press, Melbourne, Australia.

Plant, J., Smith, D., Smith, B., Williams, L., 2001. Environmental geochemistry at the global scale. Appl. Geochem. 16, 1291–1308.

Prescott, J.A., 1950. A climatic index for the leaching factor in soil formation. J. Soil Sci. 1, 9–19. http://dx.doi.org/10.1111/j.1365-2389.1950.tb00714.x.

Quinlan, J.R., November 1992. Learning with continuous classes. In: Adams, A., Sterling, L. (Eds.), AI 92: Proceedings of the 5th Australian Joint Conference on Artificial Intelligence, Hobart Tasmania 16-18. World Scientific, Singapore, pp. 343–348, 1992.

Raymond, O.L., 2012. Surface geology of Australia data package 2012 Edition. Digit. Dataset, Geosci. Aust. Available at: https://www.ga.gov.au/products/servlet/controller?event=GEOCAT_DETAILS&catno=74855.

Reimann, C., 2005. Geochemical mapping: technique or art? Geochem. Explor. Environ. Anal. 5, 359–370.

Reimann, C., Äyräs, M., Chekushin, V., Bogatyrev, I., Boyd, R., Caritat, P. de, Dutter, R., Finne, T.E., Halleraker, J.H., Jæger, Ø., Kashulina, G., Lehto, O., Niskavaara, H., Pavlov, V., Räisänen, M.L., Strand, T., Volden, T., 1998. Environmental Geochemical Atlas of the Central Barents Region. NGU-GTK-CKE Special Publication, Geological Survey of Norway, Trondheim, Norway. ISBN:82-7385-176-1745.

Scott, K., Pain, C., 2008. Regolith Science. CSIRO Publishing. ISBN:978-1-4020-8859-9472.

Scott, K.M., 1998. Electron microprobe studies of minerals from weathered profiles, Parkinson Pit and environs, Mt Magnet, WA. CSIRO Division of Exploration Geoscience Restricted Report 147R (reissued as Open File Report 25, 1990). Cooperative Research Centre for Landscape Evolution & Mineral Exploration, Perth.

Shanker, A.K., Cervantes, C., Loza-Tavera, H., Avudainayagam, S., 2005. Chromium toxicity in plants. Environ. Int. 31 (5), 739–753.

Schetselaar, E.M., Chung, C.J.F., Kim, K.E., 2000. Integration of Landsat TM, gamma-ray, magnetic, and field data to discriminate lithological units in vegetated granite-gneiss terrain. Remote Sens. Environ. 71 (1), 89–105.

Smith, D.B., Cannon, W.F., Woodruff, L.G., Solano, F., Ellefsen, K.J., 2014. Geochemical and mineralogical maps for soils of the conterminous United States, U.S. Geological Survey Open-file Report, 2014-1082, p. 386. http://dx.doi.org/10.3133/ofr20141082.

Tan, P., Lymburner, L., Thankappan, M., Lewis, A., 2009. Mapping broadacre cropping practices using MODIS time series: harnessing the data explosion. In: International Workshop on Impact of Climate Change on Agriculture, Ahmedabad, India.

Taylor, G., Eggleton, R.A., 2001. Regolith Geology and Geomorphology. John Wiley & Sons. ISBN:978-0-471-97454-3384.

Thornton, I., 2012. Environmental geochemistry: 40 years research at Imperial College, London, UK. Appl. Geochem 27, 939–953.

Tukey, J.W., 1977. Exploratory Data Analysis. Addison-Wesley, Reading, USA.

Wilford, J., 2012. A weathering intensity index for the Australian continent using airborne gamma-ray spectrometry and digital terrain analysis. Geoderma 183–184, 124–142.

Wilford, J., Minty, B., 2007. The use of airborne gamma-ray imagery for mapping soils and understanding landscape processes. In: Lagacherie, P., McBratney, A.B., Voltz, M. (Eds.), Digital Soil Mapping an Introductory Perspective. Elsevier, Developments in Soil Science, p. 31.