Spatial data mining with machine learning to reveal mineral exploration targets under cover in the Gawler Craton, South Australia

Nathaniel Butterworth*1* and Nicholas Barnett-Moore*2*

*1*Contact [nathaniel.butterworth@sydney.edu.au](mailto:nathaniel.butterworth@sydney.edu.au) Sydney Informatics Hub, University of Sydney, Australia,  *2TBA*

## **Abstract**

South Australia’s Gawler Craton hosts a variety of prominent natural resources, essential for both strategic and economic reasons. Amongst these is the distinguished Olympic Dam deposit, which currently represents the world’s fifth largest copper resource, third largest gold resource, and also the world’s largest uranium deposit. Despite this, the economic viability of pursuing new ore deposits remains precarious as changes in costs to exploration, extraction, and production fluctuate in accordance with evolving global demand. Such financial uncertainty is exacerbated by the present-day geological outlay of the Gawler Craton, mostly under thick regional cover, limiting the exposure of crucial basement outcrops required for sound exploration models. The continued increase in computational power as well as accessibility of high-quality publicly available geological and geophysical data has recently made feasible the development of original data-driven exploration methodologies. We seek to improve the efficiency of greenfield exploration for mineral commodities (focussing on gold and copper) across the Gawler Craton through the application of spatial data mining. Our study utilises 32 datasets from 12 publicly available data sources partitioned as 94 geophysical data ‘layers’ in the workflow. The values of each layer at locations of currently known ore deposits are used to build a model of the geophysical parameters which are most strongly associated with the known ore deposits. We then apply this workflow across the Gawler Craton to predict where new ore deposits are likely buried under cover. Regionally, our preferred models predict a concentration in high-prospectivity targets for copper along the Gawler’s eastern margin near Olympic Dam as well as to the north across Prominent and Cairn Hill, and high-prospectivity targets for gold across the central western Gawler Craton near the Challenger and Tarcoola sites as well as Prominent and Cairn Hill to the north. We propose these results reflect the previously observed dichotomy in important large-scale crustal controls on known major mineralisation events between the eastern and western domains of the Gawler Craton, subsequently, predicting targets along the eastern margin rich in copper with a notable absence in the west. We also observe several interesting prospectivity areas that are highly reproducible with each iteration of the model, independent of local training points. These clusters, located towards the north eastern extent of the Gawler region and its south western extent, suggest localised areas of further research and interest.

Keywords: *Data Mining, Mineral Exploration, Gawler Craton, Gold, Copper, Machine Learning*

## **Introduction**

Located in southern Australia, the Gawler Craton continues to attract widespread geological interest because of important ore deposit discoveries made within its world-class mineral systems (Fig. 1; Reid, 2019). Its premier allure, the prolific Olympic Dam deposit, currently represents the world’s fifth largest copper and third largest gold resource, whilst also boasting the world’s largest uranium deposit (Ehrig et al., 2012). The economic importance of the Gawler Craton has been further enhanced by other ore discoveries through time, including the iron-oxide copper-gold (IOCG) deposits at Prominent Hill (Heithersay, 2002), as well as Carrapateena (Fairclough, 2005); a member of the broader Olympic iron-oxide Copper-Gold Province (OP; Fig. 1: Hand et al., 2007), and the numerous gold discoveries across the central Gawler Gold Province (CGP; Fig. 1: Budd, 1998, 2004, Ferries and Schwarz, 2003, Budd, 2007). Yet, the economic viability of pursuing new deposits continues to alter as changes in costs to exploration, extraction, and production fluctuate in accordance with evolving global demand for commodities (Pohl, 2011). Across the Gawler, known limitations in available geological datasets compound these economic issues further, hindering advances in exploration. One example is the scarcity in crucial basement outcrops, most of which is under thick regional Proterozoic to Phanerozoic sedimentary cover (Daly et al., 1998). This represents a prominent limitation when developing a sound data-rich tectonic model of the space-time evolution of the Gawler Craton (Hand et al., 2007), impeding any real advancement in understanding Earth’s ancient geological setting required to produce such large-scale mineralisation events as Olympic Dam. This challenge promotes the employment of known geophysical techniques such as seismic and resistivity, which can image crustal structures deep into the Earth’s surface beneath thick sedimentary cover (Motta et al., 2019, Drummond et al., 2005, McLean and Betts, 2003). Yet, there is usually an increase in uncertainty in distinguishing ore-grade deposits from the surrounding unmineralised rock with depth when using geophysical methods alone (Eaton et al., 2003). This places pressure on the costs and associated risk of generating prospective targets deeper into Earth’s crust (Malehmir et al., 2012).

Considering these well-known limitations, a new exploration theme is emerging from within the field of computational geoscience. The increase in computational power and accessibility of publicly available geological and geophysical data has given rise to the development of original data-driven exploration methodologies with the purposes of creating a cost effective approach to uncovering new ore bodies (Merdith et al., 2015). Analytical techniques utilising spatiotemporal data mining provide a novel four-dimensional statistical approach to understanding the interplay of geodynamic parameters associated with the formation of a commodity of interest (Richards, 2013, Landgrebe et al., 2013, Cracknell and Reading, 2014). Here, we seek to improve the efficiency of greenfield exploration of gold and copper across the Gawler Craton by quantifying the geophysical signatures from the interplay of tectonic processes that enable their mineralisations through the application of spatial data mining. Our study utilises 32 datasets from 12 publicly available data sources partitioned as 94 geophysical data ‘layers’ in the workflow. The values of each layer at locations of currently known ore deposits are used to build a model of the geophysical parameters which are most strongly associated with the known ore deposits. We then apply this workflow across the Gawler Craton to predict where new ore deposits are likely buried under cover. Most commodities have competing ideas for the process that lead to their formation and the geophysical parameters which are associated with them (e.g. Gold; Sillitoe, 2008) This is likely due to the formation mechanisms (and the geophysical signatures they leave behind) being a highly complex mix of tectonic and geological processes (Tassara et al., 2017). The method presented here captures the complexity and range of all the explored parameters simultaneously and the interactions between them, to produce a simple and effective exploration targeting map for any commodity.

## **Background**

### **2.1 A Brief Geological History of the Gawler Craton**

We include a brief geological history of the Gawler region to ensure our models render interesting and insightful results within the context of the region’s exploration prospectivity. In the presence of a general geological knowledge of the area we can soundly assess our model results against public geological records and previous exploration attempts, ultimately providing future greenfield exploration targets.

The Gawler Craton is the oldest and largest geological province in South Australia. It records a complex geological history, spanning from the Archean to the Mesoproterozoic era. Late Archean rocks estimate a geographic outlay of the Craton, comprising the Sleaford Complex to the south and Mulgathing Complex in the central-western area of the Craton (Fig. 1: Hand et al., 2007). The oldest rocks in the region include the Mesoarchean Cooyerdoo Granite (~3150 Ma; Frazer et al., 2010, McAvaney et al., 2010). The evolution of the Gawler Craton can be described by a series of major tectonic events confined to two periods spanning in its ~1 billion-year geological history: The Late Archean and late Paleoproterozoic-early Mesoproterozoic (Hand et al., 2007). Basin development during the Late Archean occurred alongside significant magmatism (Hand et al., 2007). Then the ~2465–2410 Ma Sleafordian Orogeny occurred, resulting in high temperature metamorphism, and compressional deformation within the Mulgathing and Sleaford complexes (McFarlane, 2006). This event led to ~400 Myr of tectonic quiescence (Hand et al., 2007). Over the interval ~2000 to 1690 Ma, a series of rift basins developed within and on the currently preserved margins of the late Archean Gawler Craton (Hand et al., 2007). At c. 1850 Ma, the Donington Suite was emplaced along the eastern margin of the Gawler Craton (Reid and Hand, 2012), associated with a brief compressional orogenic phase known as the Cornian Orogeny that momentarily truncated these rifting events (Reid et al., 2008a). Volcano-sedimentary packages were then deposited at c. 1790 Ma, consisting of the Myola Volcanics and Broadview Schist in the northern Eyre Peninsula (Reid and Hand, 2012). These packages were succeeded by the Wallaroo Group volcanics and interlayered sediments (Cowley et al., 2003). Rifting and subsequent sedimentation continued across the Gawler Craton until its termination via a major orogenic event in the region’s south, the Kimban Orogeny, at c. 1730–1690 Ma. This orogeny is also recorded in the northern and western Gawler Craton, expressed as strongly deformed Paleoproterozoic metasedimentary sequences (Payne et al., 2008; Howard et al., 2011; and Reid, 2010), indicating the craton-wide effects of this orogeny (Fanning et al., 2007). During the waning stages of the Kimban Orogeny, c. 1690–1670 Ma, the Tunkillia Suite was emplaced, representing a late to post-orogenic magmatic event (Payne et al., 2010). The Kimban orogeny was then followed by a rejuvenation in rifting between 1680 to 1640 Ma, resulting in local magmatism and sedimentation (Hand et al., 2007). At ~1620–1615 Ma, the arc-related St. Peter Suite in the southern Gawler Craton is formed, inferring a nearby active plate margin (Hand et al., 2007). St Peter Suite magmatism was followed by the c. 1592 Ma Gawler Range Volcanics and c.1600–1570 Ma Hiltaba Suite (Reid and Hand, 2012). The Gawler Range Volcanics forms part of a felsic large igneous province estimated to occupy some 100,000 cubic kilometers (Wade et al., 2012). They are dominantly felsic, although minor basalts are also present (Allen et al., 2008), indicating widespread crustal melting was associated with mantle melting (Reid and Hand, 2012). From 1500 Ma the Gawler region experienced only localised tectonic events, remaining relatively tectonic quiescent until present time, buried beneath Phanerozoic cover, which allowed for the preservation of most economically viable mineralisation events pursued today.

### **2.2 Machine Learning and Data Mining for Geoscience Exploration**

Data mining and machine learning methods continue to be utilised and improved by the discipline of geoscience for understanding Earth systems and for greenfield exploration (e.g. Zuo and Carranza, 2011; Mountrakis et al. 2011; Rodriguez‐Galiano et al ., 2015; Merdith et al. 2015). The complex interplay of mixed parameter types makes the causal links between geology, tectonics, and subsequent formation of mineral deposits difficult to understand (Chiaradia et al., 2012). These multi-parameter mechanisms are well-suited to study through modern machine learning approaches, which can utilise all data simultaneously (Lesher et al., 2017). Various approaches have been used specifically for exploration targeting previously and met with mixed success (Yousefi et al., 2019; Farahbakhsh et al., 2019). Each method has its strengths and limitations, and depends on different assumptions around mechanisms for mineralisation and deposition. Depending on the data and the anticipated output, different models may be used toward appropriate aims. The recent review of Bergen et al. (2019) summarises much of the work in machine learning and its applications to geoscience exploration. The variety of methods and approaches for these unique problems make data-driven machine learning exploration the next frontier. Many models perform well by typical performance metrics; however, their field results are usually hidden beyond the realms of open science. To this end we aim to establish a robust workflow centered on a machine learning framework that empowers the domain-expert user to adjust inputs as they demand, whilst still being confined to a fundamental automated processing workflow. This ensures our application of machine learning is useful safely at our desks, but importantly maintains a high accuracy in the field.

## **Methods**

We use Python’s Scikit-Learn (Pedregosa et al., 2011) and MATLAB’s *Classification Learner* (Mathworks, 2018a) tool to benchmark around 20 different Machine Learning classifiers, manually tune hyper-parameters, identify the robustness of features, clean data and identify missing values, and decide on the most robust and successful classifier to use in the final automated model presented here. Subsequently, our preferred algorithm is a *Random Forest Classifier* (Breiman, 2001). This is implemented into our model written in the Python language (full models, accompanying code, and exploration targeting maps are available at <https://github.com/natbutter/gawler-exploration/>).

### **3.1 Technical Description**

Our approach starts with the dataset *South Australian all mines and mineral deposits* (2020), which contains all known locations of mineral commodities across the region. We subset this into its individual commodities, analysing and generating prospectivity maps on an individual-commodity basis. In this example we subset out Gold (Au; Fig 2a) and Copper (Cu; Fig. 2b). We then merge 32 different geophysical datasets (e.g. Fig. 2), made publicly available by the South Australian Government (SA Government, 2020), to identify the value for every geophysical parameter from the datasets at each location of known Au and Cu deposits. Finally, we use this co-registered data matrix to build a machine learning classifier and make predictions for exploration targets across the Gawler region.

In the example of Cu, the locations dataset yields 1569 known deposits, as plotted in Figure 2b. From the 32 geophysical datasets we extract up to 94 associated geophysical parameter values at each of the 1569 sites and store these co-registered values in a matrix.

Our method requires an additional set of points that equitably represent locations which are not related to the commodity under investigation, which we label “non-deposits”. Ideally these points should resemble “background” geophysical values that are uncorrelated to predictive feature values (Carranza, 2008). To this end, we pick the non-deposits by randomly generating location points throughout the spatial extent of the known deposits. A balanced set of true-positive examples (deposits) and true-negative examples (non-deposits) are ensured by picking random non-deposits equal to the number of known deposits. Finally, the non-deposits are also co-registered with their corresponding values for each of the 94 geophysical parameters. The merged data matrix of deposits and non-deposits is then used to build a predictive model detailing the geophysical parameters (known as model features) that have the strongest association with Cu.

The predictive model is built using a *Random Forest Classifier* (Breiman, 2001) as implemented in Python’s Scikit-learn package (Pedregosa et al., 2011), a *supervised* learning technique which utilises a forest of “decision trees” to classify the deposits. Each node in a tree represents a value of a randomly selected geophysical feature and its known binary classification of being a “deposit” or a “non-deposit”. Each branch of a tree leads to another geophysical feature and another classification. Many combinations of features are sampled in each decision tree throughout the forest, eventually leading the algorithm to build a model of geophysical feature values that can predict new classifications.

Using all 1,569 Cu deposits to train the model without any distinction between Cu grade or economic significance tightens the variance in the classification and prediction (Fig. 3a), however, the liberal coverage of non-deposits can over-train the model in regions where there are no known Cu deposits (i.e. having too many false-negative training points). This over-training can restrict greenfield targets from being generated. To counteract this over-training effect, we concurrently restrict our model to be trained with only the 99 known “significant” Cu deposit locations (specified in the attribute table of the SA Mines and Minerals (2020) dataset, indicating the deposit is of economic quality) and an equivalent 99 random non-deposit locations (Fig. 3b)

Finally, we generate a regularly spaced grid of 0.01-degree resolution throughout the extent of the Gawler Target Region and co-register each point with the 94 geophysical parameters. We then apply our trained models on these features, resulting in a prediction of how likely a Cu deposit will be present or not at each location over the grid (Fig 3). Prospectivity in these maps is the predicted classification probabilities of the input geophysical features, computed as the mean predicted class probabilities of the forest of decision trees in the model. We repeat the methodology for the 1,515 Au deposits (Fig. 3c), and again restrict the Au deposits to the 93 labelled as “significant” deposits to build a second model (Fig. 3d).

### **3.2 Datasets and pre-processing**

Our workflow incorporates up to 94 geophysical parameter features to train and test the machine learning model. These layers are derived from 32 different datasets from 12 unique data sources, as listed in Table 1. One additional layer is artificially included throughout the model as a benchmark for uncorrelated geophysical parameters (i.e. random performance). Additional geophysical layers can be easily incorporated into the workflow, with exact details depending on the source data type. Currently, we use four data types (1) raster grids in netCDF format, (2) shapefile polygons, (3) shapefile polylines, and (4) xyz text data.

|  |  |
| --- | --- |
| **Data Source** | **Dataset/Geophysical Feature** |
|  | *RASTERS* |
| Cudahy et al., 2012. ASTER Mineral Maps for SA from ExploreSA. | 1 AlOH group composition  2 AlOH group content  3 FeOH group content  4 Ferric oxide content  5 Ferrous iron content in MgOH  6 Ferrous iron index  7 MgOH group composition  8 MgOH group content  9 Green vegetation  10 Kaolin group index  11 Opaque index .  12 Quartz index  13 Regolith B3/B7  14 Regolith B4/B7  15 Silica index |
| SA GRAVITY, 2016 | 16 Gravitational Bouguer Anomaly |
| SA Radiometrics, 2016 | 17 Dose Rate  18 K concentration  19 Th concentration  20 U concentration |
| SA TMI, 2016 | 21 Total Magnetic Intensity (TMI)  22 TMI Reduction to Pole (RTP)  23 TMI RTP Low Pass filtered second vertical derivative |
| Hutchinson et al., 2008. DEM. | 24 Digital Elevation Model |
| Cowley et al., 2018. Depth to Crystalline Basement. | 25 Basement Elevation  26 Depth to Basement |
|  | *VECTOR POLYGONS* |
| Archaean-Early Mesoproterozoic Geology, 2020 | 27 Archaean-Early Mesoproterozoic Basement Geology Map Unit |
| 7M Geology, 2020 | 28 Regional Surface Geology Map Unit |
|  | *VECTOR POLYLINES* |
| Archaean-Early Mesoproterozoic faults, 2020 | 29 Distance to Archaean-Early Mesoproterozoic faults |
| Neoproterozoic-Ordovician faults, 2020 | 30 Distance to Neoproterozoic-Ordovician faults |
| Gairdner Dolerite, 2020 | 31 Distance to Gairdner Dolerite |
|  | *REGULARLY GRIDDED DATA* |
| AusLAMP, 2020 | 32-94 Resistivity for 63 layers from -25m to -300km. |

Table 1. List of datasets used to train and test the model. Full references included at end of report. Visual examples of some of the datasets are plotted in Figure 1 and Figure 2.

The Random Forest classification regime has the advantage of simultaneously utilising all geophysical features by being flexible to handle multiple data-types including categorical data (e.g. “Archaean-Early Mesoproterozoic geology”), the point/line data (e.g. “Gairdner Dolerite”), regularly gridded numerical data (e.g. “Gravitational Bouguer Anomaly”), and 3-dimensional numerical gridded data (e.g. “AusLAMP 3D resistivity model”). However, as all these datasets are delivered in various formats with inconsistent model resolutions it is best to pre-process each dataset individually to be readily fed into the model. For this we use *gdalwarp* and *gdal\_translate* (GDAL, 2020) to convert all raster data sets into netCDF formatted data and resample the resolution to 0.01 degrees and crop the datasets (if required) to the SA area of interest.

The model will automatically learn the important features for correct classification. Thus, we do not need to know which features should be included in the analysis. Further to this, it does not substantially impede model performance by including redundant features in the workflow (see the sensitivity analysis along with Table 3). As a result, we also use this opportunity to learn the types of geophysical features that are most relevant to the predictive signatures of a commodity (Fig. 4) which indicate specific geophysical data that would be most beneficial to target in the future.

## **Results**

Our results include two sets of model runs (i.e. Model I and Model II) for each commodity of interest (i.e. Cu and Au). The difference in model runs is constrained by the inclusion/exclusion of point locations (see Fig. 2a, 2b and Methods section), with the selection methods employed for each commodity being:

Model I. All known deposits throughout SA (all deposits: Fig. 3a, 3c)

Model II. Deposits labelled as “significant” only (significant deposits: Fig 3b, 3d).

These two models work towards highlighting the adaptability of our workflow and validating the robustness of prospective exploration regions in the context of the model input parameters. There are infinite permutations and iterations for any machine learning method where we may tune parameters and data inputs for optimal model performance. Thus, to constrain our algorithm, we define “optimal model performance” as the geologist-expert inferred *chosen* input deposit locations and chosen geophysical data layers used to predict similar exploration targets in unexplored locations. Our two deposit-selection models provide a framework for generating targets in under explored areas and to inform future data collection.

In its current state Model I has educational utility when determining which geophysical parameters are the most important features for classifying Cu and Au mineralisation occurrences (Fig. 3a, 3c) without regard for grade or quality. We use Model I primarily to benchmark our classification accuracy and overall model validity by highlighting regions of currently known deposit provinces (Fig. 3). Regions of high likelihood prospectivity (should and do) overlap with currently known deposit locations, and the reverse is true of regions of low likelihood prospectivity, overlapping with the randomly generated non-deposit data points.

As a proxy for quantifying model accuracy we perform a 10-fold cross-validation classification on each of the models (Table 2). This score, as with any machine learning model metric, must be considered in the context of what the model is trying to achieve in practice, regardless it provides a simple test for model validity. Whilst Model I performs exceptionally at correctly classifying the training points, it suffers from over-training, and because of this it can be considered generally poor for exploration targeting. We recognise Model I is highly constrained by excessive coverage of “non-deposit” locations and excessive over-counting of known deposit locations for both Cu (Fig. 3a) and Au (Fig. 3c). We believe "over-training" occurs in Model I (Fig. 3a, 3c) because there are too many random non-deposit points within the bounds of the training region, teaching the model to classify the majority of the target region as a “non-deposit” area. Additionally, the datasets for both Cu and Au have multiple “deposit” points in spatially coexistent areas (e.g. the abundance of Cu deposits between around 139°E and 31°S visible in Figure 2b). These “multiple counts” for one geophyscially homogeneous region will bias the model toward classifying only these types of regions as being a “deposit”. Finally, with the intention of producing exploration targets, it is a poor training choice for Model I to equally weight all deposits simply as mineral occurrences and not account for grade or economic quality of the deposits.

To account for these shortcomings, Model II restricts the deposits used for training to only those labelled “significant”. With this tightening of “deposit” classification the model has fewer examples to train on and a more geologically informed agenda with an arguably higher quality training set. To our aim of identifying exploration targets the classification score increases (Table 2), however, as expected, with a slight increase in the model variance. This is expected with this data-driven approach, as the training set is now relatively disparate over a large area.

|  |  |  |
| --- | --- | --- |
| **Commodity/Model** | **Score** | **Standard Deviation** |
| Cu Model I | 0.86 | 0.07 |
| Cu Model II | 0.89 | 0.08 |
| Au Model I | 0.85 | 0.11 |
| Au Model II | 0.80 | 0.16 |

Table 2. The 10-fold cross-validation accuracy score for the example Cu and Au datasets and both deposit selections Models. This score provides a reasonable idea for the accuracy of what the model has been trained to classify. A score of 1 implies it is successfully classifying all training points as “deposits” or “non-deposit”. A score of 0.5 implies random performance.

Encouragingly, Modell II successfully identifies regions where there are known Cu deposits even without any “significant” deposits in those locations to train on (Fig. 3b). Specifically, regions in the northern part of the Gawler, on the central eastern boundary, and in the south eastern region all contain no “significant” points for training (Fig. 3b) but have other known Cu deposits (Fig. 3a). This suggests optimal model performance and encourages additional greenfield exploration in regions Model II has identified in the western Gawler, and revisiting known occurrences in the north.

The performance of both Models for Au deposits follows a similar pattern to Cu. Model I is good at classifying the already known deposits, but again, is over-trained by the abundance of non-deposits, clustering, and multiple counting of known Au deposits, restricting the predictive map to generally already known gold-fields (Fig 3c). Interestingly, we note the lower scores and higher variance of both Models for Au compared to Cu, discussed further in *Model Flexibility and Sensitivity*.

As for Cu, when we restrict the choice of deposits in Model II for Au the predictive map becomes more useful for exploration (Fig. 3d). In this map, several regions predict high probability of prospectivity with no known “significant” training points. These prospective locations are validated with known less significant Au deposits being present (Fig. 3c). We note the tight predictions of Au within the IOCG province and around Olympic Dam, specifically. We detail this further in the *Discussion on Model Flexibility and Sensitivity* section.

All models identify the geophysical features that are “important” to predicting the outcome of the classification (Fig. 4). The model disavows any geological expertise and relies purely on the data to drive the classification. We can compare these “important” features with our preconceived geological domain expertise as a check against the realistic response of the model. Features highlighted by the model as “important” can also aid in furthering our understanding of mineral systems and formation processes, and guide future geophysical data collection requirements.

Numerous iterations could be made to choose input data with additional restrictions on specific deposit types, ore sizes and grades, known equivalent formation mechanisms, additional geophysical parameters, etc. Our workflow is highly adaptable, automatic, and reproducible, so can seamlessly incorporate new data as it becomes available and include any domain-expertise decisions a human-operator may provide. Nevertheless, we find Model II for Cu and Au provides a fair balance of human-input-decisions and quantifiably performs robustly for exploration to identify potential sites to drill.

1. **Discussion**

### **5.1 Copper**

To verify our results we overlay our preferred models against major Gawler basement geological domains and interrogate our prospectivity targets in the context of the proposed Olympic Dam IOCG province and the Central Gawler Gold Province (Fig. 5) as well as previously published prospectivity efforts (Fig. 6; Wise, 2019, Wise and Katona, 2015). Generally, we find good agreement between current proposed metallogenic provinces and our models as they better constrain targeting within these regions (previously broad brushes guided by assumed rock sequence boundaries). Cu model results highlight two localised ‘clusters’ of high-prospectivity and several smaller medium-prospectivity clusters (Fig. 5a). The first concentrated Cu cluster lies along the eastern margin of the Gawler region within the core of the Olympic Dam IOCG province, extending from just north of Olympic Dam southward past Carrapateena (Fig. 5a). This Cu cluster overlays most of the localised Hiltaba Suite, Donington Suite, borders the eastern edge of the Gawler Range Volcanics to the west and the Wallaroo Group to the east. This prospectivity cluster nicely reproduces major Cu mining sites (Fig. 5a) and suggests reasonable agreement with previous efforts that have investigated the unique geological setting of the Olympic Dam region and its potential for large-scale i.e. Mt, Cu recovery. On a regional scale, Hand et al. (2007) determined variations in crustal composition and key differences in Hiltaba Suite petrogeneses between the two major postulated provinces. In the Olympic Dam IOCG province Hiltaba-aged granites are isotopically more evolved, richer in Uranium and Thorium, and oxidized compared to similar aged granites in the central Gawler gold province (Hand et al., 2007). The central Gawler gold province Au deposits differ from IOCG deposits of the Olympic Cu-Au province in their lack of significant iron oxide, Cu, and U enrichments (Skirrow et al., 2002). In addition, discrepancies exist in modern-day heat flow between the IOCG province (90 ± 10 mW m–2) and the central Gawler Gold Province (54 ± 5 mW m–2), suggesting important lithospheric compositional differences between the two metallogenic provinces (Hand et al., 2007). Here, we propose our Cu model captures these observed regional-scale geological differences as our results do not predict any viable high-prospective Cu targets over the same Hiltaba Suite outcropping within the bounds of the Central Gawler Gold Province to the west of the IOCG province (Fig. 5a). This also suggests important geological controls on the formation of Cu, and the role of potential geological variations on Cu prospectivity between the provinces. The second high-prospectivity Cu cluster is located north of Olympic Dam near Prominent Hill. This overlays the approximate lithological boundary between the Mulgathing Complex and metasedimentary and igneous rocks of the northern Gawler Craton (Fig. 5a). Cu-Au mineralisation at Prominent Hill is hosted by a Mesoproterozoic sequence of bimodal volcanic rocks (Gawler Range Volcanic equivalents) and volcaniclastic and sedimentary rocks (Belperio et al., 2007). Across model iterations and fine tuning of the workflow we observe an abrupt boundary that is constantly reproduced, separating the Olympic Dam Cu cluster and the Prominent Hill Cu cluster. This suggests a distinct set of geological controls responsible for the setting of Prominent Hill and Cairn Hill mineralisation events, separate from Olympic Dam.

We also observe several important medium-prospectivity Cu clusters across the region (Fig. 5a). The importance of these clusters is founded on the premise that they represent true greenfield exploration targets outside of the comfort of “significant” locations to constrain them e.g. Olympic Dam and Prominent Hill clusters (Fig. 5a). Secondly, these medium-prospectivity clusters regularly reproduce themselves with each iteration of the model, independent of training points and parameter controls imposed. This suggests very robust findings within the workflow presented here. The first lies along the south eastern extent of the Gawler region, on the peripheral eastern edge of the Central Gawler Gold Province (Fig. 5a). We note no significant records of the proposed Hiltaba Suite geological controls within this part of the central Gawler Gold Province, and hypothesize there may be less metallogenic variation between the two provinces at this latitude and longitude (Fig. 11a; Hand et al., 2007). We base this hypothesis on the close spatial proximity of the two provinces along the southern border of the Gawler region. The second medium-prospectivity Cu cluster is interpreted to the far north east of the Gawler region (Fig. 5a). This little Cu cluster is most fascinating as it appears in every iteration of the model, whether known training deposits are utilised or not. This little Cu cluster also predicts Au in a similarly consistent fashion (discussed below). We propose this site as one of worthwhile interest for further research and exploration. The final medium-prospectivity Cu cluster is potentially the weakest, located sparsely across the north western extent of the Gawler region over meta-sedimentary and igneous rocks (Fig. 5a).

Through the application of simple GIS-derived workflows for the Olympic Dam IOCG province (Wise, 2019) and across Prominent Hill/Cairn Hill (Wise and Katona, 2015) previous efforts have also estimated prospectivity maps that suggest local areas of potential economic IOCG interest (Fig. 6). We acknowledge that any comparisons made between our models and those of Wise (2019) are qualitative and used to derive noteworthy similarities in predictions as well as differences. We understand our workflow is calibrated to the individual commodities i.e. Cu and Au, which comprise components of an IOCG deposit, whereas Wise (2019) and Wise and Katona (2015) are seeking IOCG targets more generally. We are most interested in the spatial comparison between the “green to red” relative potential areas of Wise (2019) and our white outline, which represents the 0.5 contour (Fig. 6; polygons from 0.5 to 1.0 classified as “deposits”, estimated to equal the “green to red” values of Wise) of our preferred Cu model.

Across Prominent Hill/Cairn Hill, we observe excellent first order agreement between our Cu model and Wise and Katona (2015; Fig. 6a). Our Cu model follows a similar spatial outline demarking the south-western and southern extent of potential prospectivity for this region similar to Wise and Katona (2015). Across the eastern margin of the local region our Cu model incorporates a significant block of prospectivity, which Wise and Katona (2015) do not. Northward any comparison is marred by the domain restrictions of Wise and Katona (2015). Finally, westward, the ‘peppered’ prospectivity predictions of our Cu model are interesting in the context of a broadly similar pattern of Wise and Katona (2015), which also model ‘blotches’ of high relative potential i.e. red, across this area (Fig. 6a). In general, we suggest good first order agreement between models, with our Cu model confirming further research and targeted exploration across the Prominent Hill/Cairn Hill region is recommended. Heading southward, across the Olympic Dam IOCG province, we immediately recognise a very tight agreement between both model’s estimation of a northern and north-western boundary, demarcating both model’s cessation in relative prospectivity potential of the Olympic Dam IOCG province (Fig. 6c). Along the north eastern margin, both models estimate a similar boundary of the province’s high-prospectivity potential (Fig. 6c). Comparably, though, along the south eastern margin we note that Wise (2019) does not convincingly capture the economic prospectivity of the Carrapateena site or significant sites within its immediate vicinity (Fig. 6c). Domain restrictions of Wise (2019) impede a valid comparison along the south western margin, although both models cover the IOCG region up until this distinct boundary condition (Fig. 6c).

In summary, these comparisons provide further support that our model predictions of Cu are reasonable in the context of previous efforts, lending support to the validity of our workflow and model predictions for Cu.

### **5.2 Gold**

In the instance of Au, we observe a greater number of high-prospectivity clusters stretched across the Gawler region (Fig. 5b) when compared against our more concentrated Cu clusters (Fig. 5a). In our interpretations of modelled Au clusters we must consider the unique differences in geological settings for major Au mineralisation in the Gawler. Firstly, the Challenger deposit located in the central western region, has been defined as a new class of deposit, termed a “migmatised gold deposit” (Tomkins and Mavrogenes, 2002). Previous efforts describe its unique differences from other high-grade metamorphic deposits in that during metamorphism, the gold was mobilized to such an extent that nearly all of it is now hosted by migmatitic leucosomes (Tomkins and Mavrogenes, 2002). Hosted by the Archean Christie Gneiss, a part of the basal unit of the Mulgathing Complex, the region experienced granulite facies metamorphism at ~ 2440 Ma (Daly et al., 1998). Comparatively, mineralisation in the Tarcoola Goldfield, located in the central Gawler region, is hosted dominantly by quartz veins cross-cutting both Palaeoproterozoic sedimentary rocks of the Tarcoola Formation and the Paxton Granite (Budd and Fraser 2004). Age estimations indicate that Au mineralisation at Tarcoola is temporally related to the 1595–1575 Ma Gawler Range – Hiltaba Suite geological units (Budd and Fraser, 2004). Similarly, the iron-oxide copper gold deposits found in the Olympic Cu–Au province (Skirrow et al. 2002) in the eastern part of the Gawler Craton are also of Hiltaba age (e.g. Olympic Dam, as discussed above). However, the central Gawler Gold Province gold deposits differ from IOCG deposits of the Olympic Cu-Au province in their lack of significant iron oxide, Cu, and U enrichments (Skirrow et al., 2002), echoed by a lack of Cu clusters predicted by our own Cu prospectivity maps across the Central Gawler Gold Province (Fig. 5).The correlation between broadscale Au mineralisation across the Gawler and the large-scale igneous body of Gawler Range – Hiltaba Suite Volcanics suggests a large deep crustal thermal heat source that facilitated mineralisation during this brief window of time (Hand et al., 2007, Budd and Fraser, 2004).

This complicated mixture of geological settings for Au mineralisation alludes to the difficulty in any attempt to successfully predict potential greenfield Au exploration targets. However, here, we argue our Au model produces both expected sites as well as some interesting potential new locations within geological proximity to known locations. The first observation we make is the stark contrast in high-prospectivity Au clusters between the Challenger, Tarcoola locations, in the west and the site-specific clusters to the east inside the Olympic Dam province (Fig. 5b). Our Au model reproduces all known “significant” Au sites within the local Olympic Dam region in a very precise manner. Our workflow’s calibration of geological and geophysical layers suggests a tight constraint on the specific formational controls of Au and its spatial distribution in this localised area. Further north, Prominent Hill and Cairn Hill predict the strongest high-prospectivity Au clusters (Fig. 5b). We note the northern extent of the Olympic Dam IOCG province terminates approximately here. This demarkes the IOCG Au mineralisation at Prominent Hill, hosted by a Mesoproterozoic sequence of bimodal volcanic rocks (Gawler Range Volcanic equivalents) and volcaniclastic and sedimentary rocks (Belperio et al., 2007). Migrating westward, the Challenger region predicts the largest cluster of any of our prospectivity maps presented here. This Au cluster appears to merge itself into the northern extent of the Tarcoola sites and Central Gawler Gold Province through a series of lower medium-prospectivity clusters (Fig. 5b). We note the high-prospective zone is focused within the vicinity of the known “significant” Au sites across the western extent of the Gawler region. In the context of medium-prospectivity Au clusters our model predicts an interesting cluster towards the north eastern extent of the region. Similar to the localised Cu clusters discussed above, this Au cluster is highly reproducible with or without control points in every iteration of the model, making it highly reproducible and favorable in a workflow such as this. Similarly, we also note an interesting highly reproducible medium-prospectivity Au cluster towards the south western edge of the Gawler region with potential for greenfield exploration (Fig. 5b).

Again, we compare our preferred Au model against the previous prospectivity efforts of Wise and Katona (2015; Fig. 6b) and Wise (2019; Fig. 6d) as in Cu, above. Across the Prominent Hill/Cairn Hill area our Au model generally agrees with that of Wise and Katona (2015). Our Au model covers the entire westward region of the Wise and Katona (2015) model, and matches most closely along the south western boundary (Fig. 6b). We also attempt to capture the small semi-isolated anomaly south west of Prominent Hill as identified and discussed by Wise and Katona (2015; Fig. 6b). However, our predictions estimate this to be a member of the larger Au cluster along the northern boundaries of the Central Gawler Gold province (Fig. 6b). Along the south eastern boundary our Au model is more generous in spatial area covered, whereas we observe a tighter match along the north eastern boundary (Fig. 6b). Again, the northern boundary comparison is impeded by the model conditions of Wise and Katona (2015). Southward, across the Olympic Dam IOCG province we are reminded of our prefered Au model’s precise predictions of Au prospectivity as opposed to the more general predictions of Wise (2019). We highlight the close spatial proximity in strong relative potential of Wise (2019; red blotches, Fig. 6d) and our tight Au prospectivity clusters.

Overall, we believe this comparison provides further support for our workflow and valuable new insights into the previously modelled resource potential of the Gawler’s eastern margin.

### 5.3 Prospectivity Summary

Regionally, our preferred models predict a trend of high-prospectivity Au mineralisation across the central western regions of the Gawler, and a comparable concentration of Cu clusters focused along the region’s eastern margin (Fig. 5-6). We believe this reflects our workflow’s capacity to extract a set of criteria that independently predicts the strong compositional differences in underlying regional geology (expressed by the model indicators of important features in Fig. 4), responsible for the dichotomy in mineralisation styles between the two highly economical metallogenic provinces of the Gawler Craton (Hand et al., 2007). We propose the specific localised regions of high prospectivity (i.e. red Fig. 5) are most suitable for prominent exploration targets. These targets are typically in the immediate vicinity of known “significant” deposits used to calibrate the model. We also recommend the highly reproducible medium-prospectivity cluster for both Au and Cu in the far north eastern area of the Gawler region as a potential greenfield target.

## **Model Flexibility and Sensitivity**

The resolution of the input layers is a quantifiable limitation on the output target map resolution. Realistically deposits are spread over an area and are likely to leave geophysical signatures over the entire deposit *zone*. Our target outputs are provided in 0.01-degree high-resolution grids, but visible from these are blocky-steps between adjacent pixel groups. This is due to the underlying resistivity model layers (AusLAMP, 2020) having a relative coarse resolution of ~55 km (or 0.5 degrees) compared to the ~1 km (0.01-degree) resolution of most other model layers. Regardless, we find the major prospectivity regions remain consistent and resolution improves (Appendix 1) when all resistivity layers are removed from the model. In spite of this, we still argue for the inclusion of this particular data set as previous efforts have linked resistivity anomalies to significant ore bodies (Lindsay et al., 2018, Heinson et al., 2018), and as the AusLAMP project continues to develop, we expect resistivity will be further constrained. Our preprocessing of the input grids overrides some of these artefacts as minimising the discrepancies between input layer resolutions generates a more balanced target dataset which is not artificially altered by any one input layer. Also, by smoothing the resolution of the grids, the point value of an individual deposit will be more representative of the definite spatial distribution of a deposit zone.

Our relative naive approach for data input selection (for Model I) often results in multiple points being counted in nearby locations, potentially leading to over-training of true-positive results. Realistically, nearby commodity points are probably related to the same deposit, indicating the same plausible tectonic mechanisms created similar geophysical parameter expressions. Related to this, the map location of a deposit is not a true representation of most commodity deposits, but rather a commodity would be deposited across a certain spatial extent continuing below the subsurface (e.g. for Cu; Singer et al., 2008). Choosing an appropriate subset of true-positive “deposits” and corresponding true-negative “non-deposits” is thus paramount to building a useful predictive model. We find by distinguishing the “significant” deposits in Model II to be a fair compromise. This approach means we perform training on more desirable deposit types, without model overfitting. There are additional inferences that could be made to building the training set, however the Model II approach provides a reasonable balance given the varying formation processes and grade of deposits in any collection of known deposits, this provides the necessary complexity and overlap of deposit types and geological settings to be useful to provide enormous value to exploration targeting.

There are various ways to produce a set of “non-deposits” (e.g. Qi et al., 2005, Abedi et al., 2012, Zuo and Carranza, 2011), each with tradeoffs relating to over-training/over-fitting, classification accuracy, but most importantly for our case, the predictive power of the model. We build upon the approach of Butterworth et al. (2016) to randomly select an equal amount of known-deposit-points across the region of interest and sample the underlying geophysical parameters at those random locations. By randomly generating the “non-deposit” locations we run the risk of selecting areas that could be undiscovered economic commodities. This would unhelpfully train the model to learn incorrect parameters associated with what we label “non-deposits” (i.e. a false-negative classification). Conversely, being too restrictive on the proximity selection of the non-deposit training set can lead to overfitting and reduce the predictive usefulness of the model (Carranza and Laborte 2015). Thus, we find with the size of the datasets of Cu and Au with Model II for deposit selection this is likely not a highly relevant issue. Additionally, as we regenerate the non-deposit datasets for each run through of the model (increasing the chance of generating appropriate true-negative classifications without overfitting due to deposit proximity choice) we have found all our major results consistent.

In an ideal case, if the definitive boundaries of the training deposits are known, and the input-layers resolution are increased, our approach can be further scaled-down to local exploration, making the workflow highly suited for multi-scale targeting. After the data has been co-registered (which only needs to be done once) the model takes less than 1 minute to train and generate a prospectivity map for any single commodity, operating on a standard laptop, so rapid prototyping and iterations of feature and training-point selection can be made as required. The efficiency of the co-registration of model layers has been improved by producing a parallel workflow (utilising the *Dask (2016)* Python framework).

If the model is trained on a mix of deposits with geologically distinct formation pathways, and subsequently displays broader differences in their geophysical signatures, there will be greater variance in classifying unknown locations (as is the case for Au in the Gawler region for our chosen training deposits for Model II). Functionally, this means in areas which have several similar example deposits to learn from (e.g. Prominent Hill and Tarcoola regions) the model will present with more false-positive results (i.e. greater distribution of high-prospectivity results). Conversely, location types that have few example deposits to learn from (e.g. Olympic Dam) will have more false-negative results (i.e. less coverage of high-prospectivity classification). This can be somewhat quantified by the standard deviation (representing the model variance) presented in Table 2.

Alternatively with a selection of deposits from a commodity that is mostly derived from a more consistent formation mechanism, like Cu (at least within the Gawler), we would expect the geophysical parameters at the known deposits to also be consistent, and thus a tighter and more predictable variance is expected, also distinguished in Table 2.

To further understand important model features we present three geophysics parameter histograms in Figure 7. These figures present the values of three different geophysical features, with each parameter sampled using 4 different relevant data subsets. The first label, SA, filled in blue, shows the geophysical parameters sampled across the entire spatial range of the dataset within South Australia (Fig. 7). The second, orange label, Gawler target zone, shows the geophysical parameters’ values sampled across just the spatial extent of the Gawler region. If the Gawler target zone and SA histograms are the same, we expect the Gawler region to be fairly similar to the overall signature of the SA-wide geophysical feature. It can be informative if these two subsets are distinct from each other, as it highlights the Gawler to be geophysically distinct overall, compared to SA. As the Gawler takes up a large portion of SA, we generally expect these two histograms to be comparable for most features (Fig. 7). The third, green, label shows the value counts for the geophysical feature underlying each of the randomly selected Non-Deposit data. Within our sampling criteria, if we have successfully represented the “background” geophysics we expect this to resemble the SA and Gawler target region histograms. Again, if the Gawler histogram is significantly different from the Non-Deposit data, it could be warranted to revisit the non-deposit data generation procedure, as we essentially want the non-deposits to fairly represent “background” geology/geophysics within our target region. Finally, the last histogram with the red label, is the feature value counts of the Deposit dataset. If the feature is important in predicting the classification in the model, we expect this histogram to be distinct from the other histograms. Figure 7a shows the Basement Elevation, which we have determined to be a relatively important feature (for the Cu Model I data), and the distinct “Deposit” histogram can be seen. Compared with Figure 7b, showing the Bouguer Gravity Anomaly feature that is determined to also be a semi-important feature (for the Cu Model I data). Finally, Figure 7c shows the histogram of the ASTER MgOH composition Index, which was determined to be a non-important feature for classification (for the Cu Model I data).

We further constrain the selection methods for our workflow by conducting a sensitivity analysis (Appendix 3), on Model II data for Cu, of the model’s geophysical parameter features.

## **Conclusion**

Regionally, our preferred models predict significant high-prospectivity targets for Cu along the Gawler’s eastern margin in the Olympic Dam IOCG province as well north of this at Prominent and Cairn Hill, and high-prospectivity targets for Au across the central western region near the Challenger and Tarcoola sites as well as Prominent and Cairn Hill to the north. These localised areas are all within proximity to known major deposit locations, suggesting robust areas of further research and targeted exploration. We propose these results, calibrated independently of previous geological observations, reflect the observed dichotomy in important large-scale lithospheric crustal controls on known major mineralisation events between the two major provinces. This is best represented through the prediction of deposits rich in Cu concentrated along the eastern margin of the Gawler and a notable absence in the central west. We also observe several smaller, prospectivity areas that are highly reproducible across each iteration of the model. These localised clusters are located towards the north eastern extent of the Gawler region and its south western extent, representing regions with greenfield exploration potential and further research.

Overall, our workflow is highly scalable and adaptable, but also robust in its predictive strength. This method provides realistic predictive power for highly non-linear multivariate data in a very flexible format. It is important to understand what the model has been trained to predict. In the case of Model II, the model was trained using only the significant deposits, optimising it to search for the included combination of geophysical feature layers most strongly associated with what an expert has distinguished as “significant”. Deviations from this input data are easily applied and happily recognised by the model outcomes. Additionally, if the formation mechanisms of the commodity of interest vary in their deposition processes, having different geophysical expressions, we highlight the need to account for this in model training and interpretation, as we have shown for Cu compared with Au.

Please see Supplementary Section for the application of this workflow on several other major commodities (Ag, Co, DIA, Fe, Mn, Ni, U, Zn) across the Gawler region of interest.

## 

## **Acknowledgments**

The authors acknowledge the computing facilities provided by the Sydney Informatics Hub, a Core Research Facility of the University of Sydney.

## **References**

Abedi, M., Torabi, S. A., Norouzi, G. H., Hamzeh, M., & Elyasi, G. R. (2012). PROMETHEE II: A knowledge-driven method for copper exploration. Computers & Geosciences, 46, 255-263.

Allen, S.R., McPhie, J., Ferris, G. and Cadd, A.G., 2008, Evolution and architecture of a large felsic igneous province in western Laurentia: The 1.6 Ga Gawler Range Volcanics, South Australia: Journal of Volcanology and Geothermal Research, v. 172, pp. 132–147.

Belperio, A., Flint, R. and Freeman, H., 2007. Prominent Hill: A hematite-dominated, iron oxide copper-gold system. *Economic Geology*, *102*(8), pp.1499-1510.

Bergen, K. J., Johnson, P.A., Van De Hoop, M., Beroze, G.C., 2019, Machine learning for data-driven discovery in solid Earth geoscience, Science Vol. 363, Issue 6433, DOI:10.1126/science.aau0323

Breiman, “Random Forests”, Machine Learning, 45(1), 5-32, 2001.

Budd, A., Wyborn, L. and Bastrakova, I., 1998. Exploration significance of the Hiltaba Suite, South Australia. *AGSO Research Newsletter*, *29*, pp.1-4.

Budd, A.R.; Fraser, G.L. Geological relationships and 40Ar/39Ar age constraints on gold mineralisation at Tarcoola, central Gawler gold province, South Australia. Aust. J. Earth Sci. 2004, 51, 685–700.

Budd, A.R.; Skirrow, R.G. The nature and origin of gold deposits of the Tarcoola goldfield and implications for the central Gawler gold province, South Australia. Econ. Geol. 2007, 102, 1541–1563.

Carranza, E. J. M. (2008). Geochemical anomaly and mineral prospectivity mapping in GIS. Handbook of Exploration and Environmental Geochemistry, Vol. 11. Amsterdam: Elsevier, 351 pp

Carranza, E. J. M., & Laborte, A. G. (2015). Data-Driven Predictive Modeling of Mineral Prospectivity Using Random Forests: A Case Study in Catanduanes Island (Philippines). Natural Resources Research, 25(1), 35–50. doi:10.1007/s11053-015-9268-x

Chiaradia, M., Ulianov, A., Kouzmanov, K., & Beate, B. (2012). Why large porphyry Cu deposits like high Sr/Y magmas?. Scientific Reports, 2, 685.

Cowley, W.M., Conor, C.H.H., and Zang, W., 2003, New and revised Proterozoic stratigraphic units on northern Yorke Peninsula: Minerals and Energy South Australia Journal, v. 29, p. 46–58.

Cracknell, M.J. and Reading, A.M., 2014. Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. Computers & Geosciences, 63, pp.22-33.

Daly, S.J.; Fanning, C.M.; Fairclough, M.C. Tectonic evolution and exploration potential of the Gawler Craton, South Australia. AGSO J. Aust. Geol. Geophys. 1998, 17, 145–168.

Dask Development Team (2016). Dask: Library for dynamic task scheduling, https://dask.org

Drummond, B.; Lyons, P.; Goleby, B.; Jones, L. Constraining models of the tectonic setting of the giant Olympic Dam iron oxide-copper-gold deposit, South Australia, using deep seismic reflection data. Tectonophysics 2006, 420, 91–103.

Ehrig, K.; McPhie, J.; Kamenetsky, V. Geology and mineralogical zonation of the Olympic Dam iron oxide Cu-U-Au-Ag deposit, South Australia. In Geology and Genesis of Major Copper Deposits and Districts of the World: A Tribute to Richard H. Sillitoe; Hedenquist, J.W., Harris, M., Camus, F., Eds.; Society of Economic Geologists Special Publication 16: Littleton, CO, USA, 2012; pp. 237–267.

Eaton, D.W., Milkereit, B. and Salisbury, M.H. eds., 2003. Hardrock seismic exploration. Society of Exploration Geophysicists.

Fairclough, M., 2005, Geological and metallogenic setting of the Carrapateena FeO-Cu-Au prospect—a PACE success story: Minerals and Energy South Australia Journal, v. 38, p. 4–7.

Fanning, C.M., Reid, A.J. and Teale, G., 2007, A geochronological framework for the Gawler Craton, South Australia: Geological Survey of South Australia, Bulletin 55.

Farahbakhsh, E., Chandra, R., Eslamkish, T., & Müller, R. D. (2019). Modeling geochemical anomalies of stream sediment data through a weighted drainage catchment basin method for detecting porphyry Cu-Au mineralization. Journal of Geochemical Exploration, 204, 12–32. doi:10.1016/j.gexplo.2019.05.003

Ferris, G.M., and Schwarz, M.P., 2003, Proterozoic gold province of the central Gawler craton: Minerals and Energy South Australia Journal, v. 30, p. 4–12.

Fraser, G., McAvaney, S., Neumann, N., Szpunar, M., & Reid, A. J. (2010). Discovery of early Mesoarchean crust in the eastern Gawler Craton, South Australia. Precambrian Research, 179(1–4), 1–21. https://doi.org/10.1016/j.precamres.2010.02.008

GDAL/OGR contributors (2020). GDAL/OGR Geospatial Data Abstraction software Library. Open Source Geospatial Foundation. URL <https://gdal.org>

Groves, D.I.; Bierlein, F.P.; Meinert, L.D.; Hitzman, M.W. Iron oxide copper-gold (IOCG) deposits through Earth history: Implications for origin, lithospheric setting, and distinction from other epigenetic iron oxide deposits. Econ. Geol. 2010, 105, 641–654.

Hand, M.; Reid, A.; Jagodzinski, E. Tectonic framework and evolution of the Gawler Craton, South Australia. Econ. Geol. 2007, 102, 1377–1395.

Heinson, G., Didana, Y., Soeffky, P. et al. The crustal geophysical signature of a world-class magmatic mineral system. Sci Rep 8, 10608 (2018).<https://doi.org/10.1038/s41598-018-29016-2>

Heithersay, P., 2002, Prominent Hill discovery—URN 1: The best copper gold intersection in 25 years: Minerals and Energy South Australia Journal, v. 24, p. 4–5.

Heithersay, P., 2002, Prominent Hill discovery—URN 1: The best copper-gold intersection in 25 years: Minerals.

Howard, K.E., Hand, M., Barovich, K.M., Payne, J.L. and Belousova, E.A., 2011, U-Pb, Lu-Hf and Sm-Nd isotopic constraints on provenance and depositional timing of metasedimentary rocks in the western Gawler Craton: Implications for Proterozoic reconstruction models: Precambrian Research, v. 184, pp. 43–62.

Landgrebe, T. C. W., Merdith, A., Dutkiewicz, A., & Müller, R. D. (2013). Relationships between palaeogeography and opal occurrence in Australia: A data-mining approach. Computers & Geosciences, 56, 76-82.

Lesher, M., Hannington, M., Galley, A., Ansdell, K., Astic, T., Banerjee, N., Beauchamp, S., Beaudoin, G., Bertelli, M., Bérubé, C., Beyer, S., Blacklock, N., Byrne, K., Cheng, L.-Z., Chouinard, R., Chouteau, M., Clark, J., D'Angelo, M., Darijani, M., Devine, M., Dupuis, C., El Goumi, N., Enkin, R., Farquharson, C., Fayol, N., Feltrin, L., Feng, J., Gaillard, N., Gleeson, S., Gouiza, M., Grenon, C., Guffey, S., Guilmette, C., Guo, K., Hart, C., Hattori, K., Hollings, P., Joyce, N., Kamal, D., King, J., Kyser, K., Layton-Matthews, D., Lee, R., Lesage, G., Leybourne, M., Linnen, R., Lypaczewski, P., McGaughey, J., Mitchinson, D., Milkereit, B., Mir, R., Morris, W., Oldenburg, D., Olivo, G., Perrouty, S., Piercey, S., Piette-Lauzière, N., Raskevicius, T., Reman, A., Rivard, B., Ross, M., Samson, I., Scott, S., Shamsipour, P., Shi, D., Smith, R., Sundaralingam, N., Taves, R., Taylor, C., Valentino, M., Vallée, M., Wasyliuk, K., Williams-Jones, A., Winterburn, P., 2017, Integrated Multi-Parameter Exploration Footprints of the Canadian Malartic Disseminated Au, McArthur River-Millennium Unconformity U, and Highland Valley Porphyry Cu Deposits: Preliminary Results from the NSERC-CMIC Mineral Exploration Footprints Research Network. In “Proceedings of Exploration 17: Sixth Decennial International Conference on Mineral Exploration” edited by V. Tschirhart and M.D. Thomas, 2017, 325–347

Lindsay, M.D., Spratt, J., Occhipinti, S.A., Aitken, A.R.A., Dentith, M.C., Hollis, J.A. and Tyler, I.M., 2018. Identifying mineral prospectivity using 3D magnetotelluric, potential field and geological data in the east Kimberley, Australia. Geological Society, London, Special Publications, 453(1), pp.247-268., https://doi.org/10.1144/SP453.8

MATLAB and MATLAB Classification Learner, https://au.mathworks.com/help/stats/classification-learner-app.html, MATLAB 2018a, The MathWorks, Inc.

Malehmir, A., Durrheim, R., Bellefleur, G., Urosevic, M., Juhlin, C., White, D.J., Milkereit, B. and Campbell, G., 2012. Seismic methods in mineral exploration and mine planning: A general overview of past and present case histories and a look into the future. *Geophysics*, *77*(5), pp.WC173-WC190.

Massimo Chiaradia, Alexey Ulianov, Kalin Kouzmanov & Bernardo Beate, Why large porphyry Cu deposits like high Sr/Y magmas? Scientific Reports volume 2, Article number: 685 (2012)

McLean, M.A., and Betts, P.G., 2003, Geophysical constraints of shear zones and geometry of the Hiltaba Suite granites in the western Gawler craton, Australia: Australian Journal of Earth Sciences, v. 50, p. 525–541.

McFarlane, C.R.M., 2006. Palaeoproterozoic evolution of the Challenger Au deposit, South Australia, from monazite geochronology. *Journal of Metamorphic Geology*, *24*(1), pp.75-87.

Merdith, A. S., Landgrebe, T. C., & Müller, R. D. (2015). Prospectivity of Western Australian iron ore from geophysical data using a reject option classifier. Ore Geology Reviews, 71, 761-776.

Motta, J.G.; Betts, P.G.; Souza-Filho, C.R.; Thiel, S.; Curtis, S.; Armit, R.J. Proxies for basement structure and its implications for mesoproterozoic metallogenic Provinces in the Gawler Craton. J. Geophys. Res. Solid Earth 2019, 124.

Mountrakis, G., Im, J. and Ogole, C., 2011. Support vector machines in remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing, 66(3), pp.247-259.

Payne, J., Hand, M., Barovich, K. and Wade, B., 2008, Temporal constraints on the timing of high-grade metamorphism in the northern Gawler Craton: implications for assembly of the Australian Proterozoic: Australian Journal of Earth Sciences, v. 55, pp. 623–640.

Payne, J.L., Ferris, G., Barovich, K.M. and Hand, M., 2010, Pitfalls of classifying ancient magmatic suites with tectonic discrimination diagrams: An example from the Palaeoproterozoic Tunkillia Suite, southern Australia: Precambrian Research, v. 177, pp. 227–240.

Pedregosa et al. 2001, Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011

Pohl, W.L., 2011. Economic Geology Principles and Practice: Metals, Minerals, Coal and Hydrocarbons — Introduction to Formation and Sustainable Exploitation of Mineral Deposits. Wiley-Blackwell, West Sussex.

Qi, Y., Klein-Seetharaman, J., Bar-Joseph, Z.: Random forest similarity for protein–protein interaction prediction from multiple sources. In: Proceedings of the Pacific Symposium on Biocomputing (2005)

Reid, A., Hand, M., Jagodzinski, E., Kelsey, D. and Pearson, N.J., 2008a, Palaeoproterozoic orogenesis within the southeastern Gawler Craton, South Australia: Australian Journal of Earth Sciences, v. 55, pp. 449– 471

Reid, A.J.; Hand, M. Mesoarchean to Mesoproterozoic evolution of the southern Gawler Craton, South Australia. Episodes 2012, 35, 216–225.

Reid, A., 2019. The Olympic Cu-Au Province, Gawler Craton: a review of the lithospheric architecture, geodynamic setting, alteration systems, cover successions and prospectivity. *Minerals*, *9*(6), p.371.

Richards, J.P., 2013. Giant ore deposits formed by optimal alignments and combinations of geological processes. Nature geoscience, 6(11), pp.911-916.

Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M. and Chica-Rivas, M.J.O.G.R., 2015. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. Ore Geology Reviews, 71, pp.804-818.

Sillitoe, R. H. Major gold deposits and belts of the North and South American cordillera: distribution, tectonomagmatic settings, and metallogenic considerations. Econ. Geol. 103, 663–687 (2008).

Singer, D., V. Berger, and B. Moring (2008), Porphyry copper deposits of the world: Database and grade and tonnage models, U.S. Geol. Surv. Open File Rep., 2008–1155

Skirrow, R.G.; Walshe, J.L. Reduced and oxidized Au-Cu-Bi iron oxide deposits of the Tennant Creek Inlier, Australia: An integrated geologic and chemical model. Econ. Geol. 2002, 97, 1167–1202.

Skirrow, R.G.; Bastrakov, E.N.; Barovich, K.; Fraser, G.L.; Creaser, R.A.; Fanning, C.M.; Raymond, O.L.; Davidson, G.J. Timing of iron oxide Cu-Au-(U) hydrothermal activity and Nd isotope constraints on metal sources in the Gawler Craton, South Australia. Econ. Geol. 2007, 102, 1441–1470.

Tassara, S., González-Jiménez, J.M., Reich, M. et al. Plume-subduction interaction forms large auriferous provinces. Nat Commun 8, 843 (2017). https://doi.org/10.1038/s41467-017-00821-z

Tomkins, A.G. and Mavrogenes, J.A., 2002. Mobilization of gold as a polymetallic melt during pelite anatexis at the Challenger deposit, South Australia: A metamorphosed Archean gold deposit. *Economic Geology*, *97*(6), pp.1249-1271.

Wade, C.E., Reid, A.J., Wingate, M.T.D., Jagodzinski, E.A. and Barovich, K., 2012, Geochemistry and geochronology of the c. 1585 Ma Benagerie Volcanic Suite, southern Australia: Relationship to the Gawler Range Volcanics and implications for the petrogenesis of a Mesoproterozoic silicic large igneous province: Precambrian Research, v. 206–207, pp. 17-35.

Wise, T.; Katona, L. Gravity anomalies as trap sites in prospectivity modelling of the eastern Gawler copper-gold belt. Aust. Soc. Explor. Geophys. Ext. Abstr. 2015, 1, 1–4.

Wise, T, Prospectivty modelling of the Olympic Cu-Au Province, MESA Journal 90, pages 36-41, 2019.

Yousefi, M., Kreuzer, O. P., Nykänen, V., & Hronsky, J. M. (2019). Exploration information systems–A proposal for the future use of GIS in mineral exploration targeting. Ore Geology Reviews, 111, 103005.

Zuo, R., and E. J. M. Carranza (2011), Support vector machine: A tool for mapping mineral prospectivity, Comput. Geosci., 37(12), 1967–1975, doi:10.1016/j.cageo.2010.09.014.

## **Dataset References**

SA 2020, Department for Energy and Mining, the Government of South Australia, Data and Products, accessed on 28th June 2020,<https://catalog.sarig.sa.gov.au/geonetwork/srv/eng/catalog.search>

ASTER Mineral Maps for SA from ExploreSA: The Gawler Challenge additional pre-package data from Cudahy, Tom; Caccetta, Mike; Lau, Ian; Rodger, Andrew; Laukamp, Carsten; Ong, Cindy; Chia, Joanne; Collings, Simon; Rankine, Terry; Fraser, Ryan; Woodcock, Robert; Vote, Josh; Warren, Peter; Thomas, Matilda; Tyler, Ian; Mauger, Alan; Close, Dorothy; Jones, Mal; Abrams, Mike (2012): Satellite ASTER Geoscience Map of Australia. v3. CSIRO. Data Collection.<https://doi.org/10.4225/08/51400D6F7B335> (<https://catalog.sarig.sa.gov.au/geonetwork/srv/eng/catalog.search#/metadata/c9039762-e4e0-48d5-a6b8-7ddfe7f736b8>), accessed 28th June 2020.

Depth to crystalline basement, Cowley, W.M.; Katona, L.F.; Gouthas, G.; Hough, L.P.; Menpes, S.A.; Foss, C.A. SA Department for Energy and Mining; Geological Survey of South Australia. Geoscientific Information Strategy Team; CSIRO Earth Science and Resource Engineering, 2018, (<https://catalog.sarig.sa.gov.au/geonetwork/srv/eng/catalog.search#/metadata/216146e3-7cb0-410a-bcb2-68e5e6bcf9f4>), accessed 28th June 2020.

SA GRAVITY 2016, South Australian regional gravity images, (<https://catalog.sarig.sa.gov.au/geonetwork/srv/eng/catalog.search#/metadata/f1c5ab7f-21c8-4ed3-85fa-14f6dba40ca6>), accessed 28th June 2020.

SA Radiometrics, 2016, South Australian regional radiometric images, created by Gary Reed on 2011-04-06 (<https://catalog.sarig.sa.gov.au/geonetwork/srv/eng/catalog.search#/metadata/193db970-4f6c-4c15-b8a9-87f746dc7b9b>), accessed 28th June 2020

SA TMI, 2016, South Australian regional total magnetic intensity images, (<https://catalog.sarig.sa.gov.au/geonetwork/srv/eng/catalog.search#/metadata/1a797550-8d6b-4ab1-81f2-be6ed15f06d0>), accessed 28th June 2020.

Hutchinson, M.F., Stein, J.L., Stein, J.A., Anderson, H., Tickle, P.K. 2008. GEODATA 9 second DEM and D8: Digital Elevation Model Version 3 and Flow Direction Grid 2008. Record DEM-9S.v3. Geoscience Australia, Canberra.<http://pid.geoscience.gov.au/dataset/ga/66006>

100K Geology - surface geology, (<https://catalog.sarig.sa.gov.au/geonetwork/srv/eng/catalog.search#/metadata/e27f9a25-b749-4dba-bfb3-ca90baf04d79>), accessed 28th June 2020.

7M Geology, Generalised regional surface of South Australia, derived from 1:2 million (Geology 2M),<https://catalog.sarig.sa.gov.au/geonetwork/srv/eng/catalog.search#/metadata/6f93ccae-e52d-4b91-ba16-ee7c0798daeb>, accessed 28th June 2020.

Solid Geology: Archaean - Early Mesoproterozoic geology, (<https://catalog.sarig.sa.gov.au/geonetwork/srv/eng/catalog.search#/metadata/79d50974-a57e-4754-91aa-1f0be6489a8c>), accessed 28th June 2020.

Solid Geology: Archaean - Early Mesoproterozoic faults, (<https://catalog.sarig.sa.gov.au/geonetwork/srv/eng/catalog.search#/metadata/8e517d3f-8255-4159-8bcf-3e101622a2c2>), accessed 28th June 2020.

Solid Geology: Neoproterozoic - Ordovician faults, (<https://catalog.sarig.sa.gov.au/geonetwork/srv/eng/catalog.search#/metadata/7e5ff842-36cf-4b91-87c2-1e715ef62f0d>), accessed 28th June 2020.

Solid Geology: Gairdner Dolerite, (<https://catalog.sarig.sa.gov.au/geonetwork/srv/eng/catalog.search#/metadata/a53f799b-1f4e-45b3-9ad6-d27b9b38ea5e>) accessed 28th June 2020.

AusLAMP, 2020. Robertson Kate, Thiel Stephan Heinson Graham (2018) Evolving 3D lithospheric resistivity models across southern Australia derived from AusLAMP MT. ASEG Extended Abstracts 2018, 1-5.<https://doi.org/10.1071/ASEG2018abM2_1G> and (<https://catalog.sarig.sa.gov.au/geonetwork/srv/eng/catalog.search#/metadata/1cab41af-eddf-40e4-b091-a6e733e5701a>) and<http://energymining.sa.gov.au/minerals/geoscience/geological_survey/gssa_projects/auslamp> accessed 28th June 2020.

South Australian all mines and mineral deposits, (<https://catalog.sarig.sa.gov.au/geonetwork/srv/eng/catalog.search#/metadata/a0e4b62c-ec88-44b8-a530-b4e744a6b414>), accessed 27th June 2020.

Stamen, 2020. Stamen Design, under CC BY 3.0. Data by OpenStreetMap, under ODbL

Gawler Craton Outline, (<https://catalog.sarig.sa.gov.au/geonetwork/srv/eng/catalog.search#/metadata/e27df3f0-4a96-4330-b4d7-b4b282ee0d11>), accessed 28th June 2020.

# **Figure and Table Captions**

Figure 1. Geological map of South Australia with simplified basement geology, as in the legend, for the Gawler craton (Archaean-Early Mesoproterozoic Geology, 2020). Two mineral provinces, the Central Gawler Gold Province and the Olympic Iron-Oxide-Copper-Gold Province, overlain as dashed white lines.Bases on Reid (2019). Thin black lines show locations of Archean-Early Mesorprotezoic faults. Faults and geology are used as model inputs.

Figure 2. Currently known deposits for a) Gold (Au) and b) Copper (Cu) from SA Mines and Minerals (2020). Deposits labelled as being “significant” in the dataset are indicated with triangles. Four representative examples of the 94 geophysical data layers used as model features (a) Digital elevation model (Hutchinson et al., 2008) (b) bouguer gravity anomaly (SA GRAVITY, 2016) (c) Basement elevation (Cowley et al., 2018) (d) Magnetics (SA TMI, 2016). Note: the resolution and coverage can change between datasets. Figure 1 also includes two additional input data layer examples, specifically, Archean-Early Mesoproterozoic Geology and Faults.

Figure 3: Prospectivity results across the Gawler target region (dotted black outline) showing the predicted classification probabilities from low (0.0, blue) to high (1.0, red), including known deposits (red dots) and random non-deposit training points (blue dots). (a) **Cu** Prospectivity Model I. (b) **Cu** Prospectivity Model II, with only known “significant” deposit locations (triangles Fig. 2; see text for further discussion) and an equal number of non-deposit points used for training. (c) **Au** Prospectivity Model I. (d) **Au** Prospectivity Model II.

Figure 4: (a) The top ten ranked important geophysical layers partitioned from the geological and geophysical datasets for Cu Model I (see section 3). Y-axis shows the label of important features, X-axis is the relative importance with values between ~2% (0.02) and ~5% (0.055). This highlights the relatively low importance of any one single feature, indicating minimal feature bias, and informs us that an interplay of geophysical parameters are correlated with known mineral deposits. **(b)** same as (a) for **Cu** Model II **(c)** same as (a) for **Au** Model I **(d)** same as (a) for **Au** Model II.

Figure 5: (a) Filtered **Cu** Model II medium (light red) to high (red) prospectivity probability classification (Fig. 3b) plotted over regional geological domains (Fig. 1) and the two major metallogenic provinces after Reid (2019) as black and white dotted lines. Green dots denote locations of recorded “significant” **Cu** ore deposits. PH: Prominent Hill, OD: Olympic Dam, CH: Cairn Hill. CP: Carrapateena. (b) Same as (a) but for **Au** Model II. Gold points represent locations of recorded “significant” **Au** ore deposits. CL: Challenger Mine, CH: Cairn Hill, PH: Prominent Hill, TC: Tarcoola, OD: Olympic Dam, CP: Carrapateena.

Figure 6: (a) Preferred **Cu** model prospectivity targets (white outline, see section 5) overlaying published IOCG prospectivity maps for Prominent and Cairn Hill in the northern Gawler (Wise and Katona, 2015) (b) same as (a) but for preferred **Au** model ( Fig. 3d) (c) same as (a) overlaying previously generated prospectivity maps of the Olympic Dam IOCG Province (Wise, 2019). (d) Same as (c) but for preferred **Au** Model. Significant mines labelled OD: Olympic Dam, CP: Carrapateena, PH: Prominent Hill, CH: Cairn Hill.

Figure 7. Histograms of example geophysical parameters (a) Basement Elevation, (b) Gravitational Bouguer Anomaly, and (c) ASTER MgOH composition Index. The different histograms on each panel represent different sampling of the parameter. As in the legend, SA refers to the entire geophysical grid, Gawler Target Region is the grid just within the Gawler, Non-Deposits count the geophysical values associated with the “non-deposit” data for the Cu Model I set, and Deposits count the geophysical values associated with “deposit” data for the Cu Model I set. The similarities or differences between these dataset can inform us of how important, or at least distinct, the feature might be in correlating with a mineral deposit.

Table 1. List of datasets used to train and test the model. Full references included at end of report. Visual examples of some of the datasets are plotted in Figure 1 and Figure 2.

Table 2. The 10-fold cross-validation accuracy score for the example Cu and Au datasets and both deposit selections Models. This score provides a reasonable idea for the accuracy of what the model has been trained to classify. A score of 1 implies it is successfully classifying all training points as “deposits” or “non-deposit”. A score of 0.5 implies random performance.