

Original Paper

Maximum Entropy and Random Forest Modeling of Mineral Potential: Analysis of Gold Prospectivity in the Hezuo–Meiwu District, West Qinling Orogen, China

Shuai Zhang ^{1,5}, Keyan Xiao,^{2,5} Emmanuel John M. Carranza,³ and Fan Yang ^{1,4}

Received 13 September 2018; accepted 25 October 2018
Published online: 30 October 2018

This study tested and compared the mineral potential mapping capabilities of the random forest (RF) and maximum entropy (MaxEnt) algorithms using gold deposit occurrences within the Hezuo–Meiwu district, West Qinling Orogen, China. Eighteen orogenic gold deposits in this district and associated regional exploration datasets were used to construct data-driven predictive models to identify locations prospective for gold mineralization. The 18 orogenic gold deposits used in the modeling can be divided into magmatic-hydrothermal gold deposits and mesothermal gold deposits in terms of metallogenic characteristics and nine evidential maps associated with Au deposit occurrences (i.e., distance to intrusions and faults; Au, As, Ag, Cu, and Sb singularity indices; and principal component scores (PC1 and PC2) based on isometric logratio-transformed geochemical data were selected as inputs to the models). The PC1 represents a primary geochemical signature of tectonic process or their products (i.e., fault system), whereas PC2 represents a secondary geochemical signature. Both RF and MaxEnt models were then used to quantitatively rank the importance and identify the sensitivity of the evidential maps based on their spatial relationships to the known gold deposits in the study area. The two groups of populations in the response curves and marginal effect curves indicate that the mineral potential mapping should be performed by zones in consideration of different metallogenic characteristics of gold deposits. The accuracy of the resulting models was then assessed, and the results of the mineral potential mapping were examined using receiver operating characteristic (ROC) analysis, capture-efficiency curve, and success rate curve. Both mineral potential mapping by zones with RF and MaxEnt models have higher area under the ROC curve (AUC) values than the models performed in the study area and delineate 19% of the study area containing > 88% of the known deposit occurrences. Finally, according to the concentration–area (C-A) thresholds for prospectivity maps, two ternary prospectivity maps were generated for further mineral exploration. The results indicate that the RF and MaxEnt algorithms can be used effectively

¹School of Earth Sciences and Resources, China University of Geosciences Beijing, 29 Xueyuan Road, Beijing 100083, China.

²MIR Key Laboratory of Metallogeny and Mineral Resource Assessment, Institute of Mineral Resources, Chinese Academy of Geological Sciences, Beijing 100037, China.

³School of Agricultural, Earth and Environmental Sciences, University of KwaZulu-Natal, Westville Campus, Durban, South Africa.

⁴Department of Earth Sciences, School of Physical Sciences, The University of Adelaide, Adelaide, SA 5005, Australia.

⁵To whom correspondence should be addressed; e-mail: kyanxiao@sohu.com

for mineral potential mapping and represent machine learning algorithms that can be used in areas with a few known mineral occurrences.

KEY WORDS: Random forest, MaxEnt model, Gold deposits, West Qinling Orogen, Mineral potential mapping.

INTRODUCTION

Mineral potential mapping is a tool used in mineral exploration that involves establishing a conceptual model of the targeted type of deposits or mineral system, translating mineralization-related processes into exploration criteria, deriving evidential or predictive maps by combining these criteria with mineral exploration datasets, and finally obtaining and integrating weighted spatial evidential maps to delineate exploration targets (Bonham-Carter 1994; Carranza 2008; Porwal and Carranza 2015). The formulation of appropriate targeting criteria and the application of innovative and robust techniques for the derivation and weighting of evidential features associated with exploration criteria are the key points in mineral potential mapping (Joly et al. 2012).

Previous research on mineral potential mapping has used traditional mineral deposit models (Cox and Singer 1986) for the formulation of exploration criteria to be used in the construction of input evidential layers. However, the fact that mineral deposit models focus mostly on deposit-scale characteristics means that larger but probably prospective regional- or camp-sized targets that lack deposit-scale features will be identified as non-prospective (McCuaig et al. 2007; McCuaig and Hronsky 2014). In addition, the significant differences between regional- and deposit-scale datasets mean that mineral deposit models are not ideally suited for use in mineral potential mapping (Sillitoe 2004; Simmons et al. 2005; Sillitoe and Thompson 2006). These problems led to the development of the mineral systems approach to mineral potential mapping (Wyborn et al. 1994), which emphasizes that mineral deposits form foci of much larger mineral systems that involve energy and mass transfer as a result of various relevant Earth processes that operate in time and space (Wyborn et al. 1994; Hronsky and Groves 2008; McCuaig and Hronsky 2014). This concept shifts the focus from deposit-scale features to those present within generic mineral systems that are based on mineralizing

processes that operate at different scales, suggesting that mineral system models can be used to define predictive maps that portray processes at certain scales (Knox-Robinson and Wyborn 1997; McCuaig et al. 2010; Porwal and Kreuzer 2010). In practice, the formulation of mineral system models involves identifying the following components: sources of energy, fluids, ligands, and metals, pathways for focused fluid flow, physical throttles involved in trapping of fluids, and chemical scrubbers needed for the precipitation of metals. Although the formulation of exploration criteria for each component and the generation of evidential map(s) per criterion comprise an approach that has been widely (or even unconsciously) used previously in mineral potential mapping, McCuaig et al. (2010) translated this mineral system approach into an effective exploration targeting system.

Therefore, mineral potential mapping is based on the identification and derivation of geologically representative evidential maps based on a mineral system model, meaning that the first step in this modeling is the development of a thorough understanding of the geology of the mineral system in question as well as geographical information system (GIS) and statistical skills (Porwal and Carranza 2015). The generation of evidential maps can be complemented by empirical analysis, especially in brownfields areas. This approach provides objective measures of spatial associations between evidential features and mineral occurrences, as well as new insights into conceptual mineral system modeling (Porwal and Carranza 2015). The widely used techniques in this field include fractal and multifractal analysis (Cheng 1999, 2007; Agterberg 2007), principal component analysis, and factor analysis (Carranza 2010; Wang et al. 2015), all of which can advance our understanding of empirical spatial associations between mineral occurrences and evidential features. In addition to the derivation of evidential maps, another important step is the weighting and integration of evidential maps based on models of the deposit or mineral system (Yousefi and Nykänen 2017).

There are two general approaches to mineral potential mapping, namely data-driven and knowledge-driven approaches. The former empirically assigns weights to evidential features using training datasets, whereas the latter is based on expert judgment (Bonham-Carter 1994; Carranza 2008). Both data- and knowledge-driven methods have disadvantages. For example, the use of known mineral occurrences as training sites in data-driven approaches introduces stochastic bias and error. In comparison, knowledge-driven methods are subjective and require an in-depth understanding of mineralizing processes and the relationships between evidential maps and mineral occurrences. Hybrid approaches that combine both known mineral occurrences and expert knowledge in assigning evidential weights have also been used, although these approaches suffer from the disadvantages of the data- and knowledge-driven methods in terms of assigning weights to evidential maps (Yousefi and Nykänen 2017). Numerous methods that employ machine learning have recently been developed for use in data-driven modeling (Lewkowski et al. 2010; Oh and Lee 2010). The most widely used of these methods include decision trees (DTs) (Breiman 2017; Elith et al. 2008), artificial neural networks (ANNs) (Brown et al. 2000; Porwal et al. 2003), support vector machines (SVMs) (Zuo and Carranza 2011; Abedi et al. 2012), and classification tree ensembles such as random forest (RF) (Breiman 2001; Rodriguez-Galiano et al. 2014; Carranza and Laborte 2015a; Gao et al. 2016; Zhang et al. 2016).

This paper assesses and compares the mineral potential mapping capabilities of the MaxEnt model and the random forest algorithm by applying these methods to define prospective areas for gold exploration within the Hezuo–Meiwu district of China. These machine learning methods are increasingly used in Earth Science; therefore, it is important to compare the usefulness of these methods to make the users aware of their strengths and weaknesses.

GEOLOGICAL SETTING AND GOLD MINERALIZATION

The study area—the Hezuo–Meiwu district—is located in the western part of the west Qinling Orogen, part of the Qinling–Qilian–Kunlun orogenic belt that stretches across central China (Meng and Zhang 2000). The west Qinling Orogen contains more than 100 gold deposits, with proven reserves

of > 1200 tons of gold (Mao et al. 2002; Chen and Santosh 2014). Of these 100 gold deposits, 18 are located in the study area. The gold mineralization within the west Qinling Orogeny is controlled by NW-trending striking faults and folds that developed during Triassic orogenic deformation related to the convergence of the South China Block and the North China Craton. The gold mineralization is generally hosted by Paleozoic to early Triassic clastic and carbonate rocks (Mao et al. 2002) and was most likely generated as a result of the metamorphic devolatilization of Paleozoic sedimentary units (Mao et al. 2002; Chen et al. 2004). The West Qinling Orogeny has a complex geological history that records the opening, subduction, and closure of the proto- and paleo-Tethys, and the subsequent Late Triassic continental collision between the South China Block and the North China Craton (Kröner et al. 1993; Lerch et al. 1995; Zhang et al. 2004; Dong et al. 2011). The area contains voluminous early Paleozoic to early Mesozoic marine sedimentary rocks that have been intruded by numerous Triassic granitoid intrusions, the majority of which yield zircon U–Pb ages of 247–200 Ma (Zhang et al. 2008; Dong et al. 2011). These intrusions formed from magmas generated in either subduction or post-collision tectonic settings (Guo et al. 2012; Li et al. 2013).

The Hezuo–Meiwu district is dominated by Carboniferous to Triassic marine sedimentary units (Fig. 1) and contains NW–SE-trending structures as well as the Xiahe–Hezuo thrust that divides the study area into eastern and western zones. The eastern zone contains Carboniferous to Permian clastic rocks and carbonates that cover parts of the granitoid plutons. These plutons include the Meiwu and Dewulu intrusions that were emplaced into Permian or Carboniferous rocks, in an unknown tectonic setting. A total of 12 deposits in this zone are more closely related to magmatic rocks and spatially generated within or near the contact zone of intermediate-acidic rocks mass, such as Laodou, Jili, and Labuzaika gold deposits. The western zone contains a variety of dioritic to granodioritic and granitic stocks and dikes that were emplaced into Triassic marine clastic rocks (Sui et al. 2017). Another six deposits are from the western zone and represented by the Zaozigou deposit, which is related to shallow intermediate-acidic magma intrusion. Although there are some difference in the main metallogenic geological characteristics of the two zones, there are still many similarities, such as the

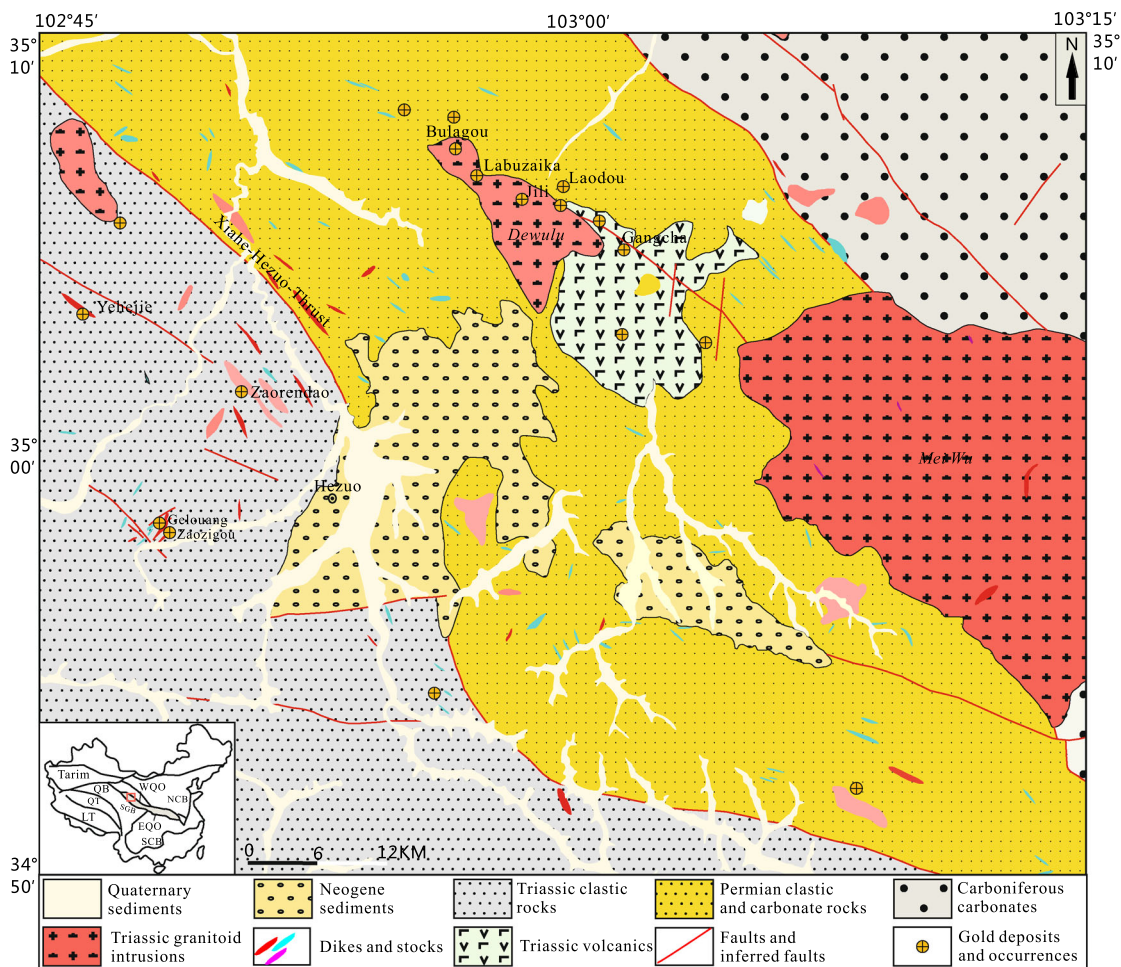


Fig. 1. Geological map of the Hezuo–Meiwo district showing the locations of major Au deposits. (Modified from Sui et al. 2017).

distribution of ore deposits and the occurrence of ore bodies are strictly controlled by the regional deep great fault zones and its secondary faults (Liu 2011).

From north to south in the study area, there are obvious group zoning characteristics, that is, the transition from medium–high temperature to medium–low temperature (Qi et al. 2013). The north belt elements association is: Cu, As, W, Sn; the south belt elements association is: Pb, Zn, Ag, Au, As–Hg, Sb. The anomalies of Au, Ag, As, Sb are well developed in the Xiahe–Hezuo fault zone, and its distributions are closely related to N–E-trending faults which reflect that the fault zone is a channel for low-temperature hydrothermal activity. The elements of Au, As, Sb, Bi have strong differentiation degree, high dispersion and high metallogenic probability;

arsenic anomaly exists in the study area because arsenopyrite is ubiquitous in the study area (Qi et al. 2013).

Orogenic gold deposits are associated with intermediate to felsic intrusions, but it remains unclear whether the gold in these deposits and the hydrothermal fluids that formed this mineralization were derived from the associated igneous rocks (Goldfarb et al. 2005). Research into the Dewulu quartz diorite pluton led Sui et al. (2017) to suggest that the sediment-hosted disseminated and magmatic-hosted vein-type gold deposits in the study area formed broadly contemporaneously with the Dewulu Au–Cu skarn deposit. This result, combined with the fact that the disseminated and vein gold deposits are thought to be genetically related to reduced granitoid intrusions, means that the deposits

in the study area represent intrusion-related gold systems.

METHODS

MaxEnt Method

The maximum entropy (MaxEnt) approach is based on statistical mechanics (Jaynes 1957) and is a general-purpose method that can be used to make predictions or inferences from incomplete information. This approach is widely used in modeling the geographical distribution of biological species using presence-only data within environmentally variable spaces (Elith et al. 2011). The MaxEnt approach estimates the probability of target variables with maximum entropy and is controlled by a set of constraints that represent the incomplete information available about the target distribution (Elith et al. 2011). The incomplete information available refers to a set of real-valued variables (in this study evidential maps were derived from exploration datasets), whereas the constraints mean that the expected values of each feature (i.e., the empirical average values of an evidential map) should match those of a set of sample points taken from the target distribution. The most important feature of the MaxEnt approach is that this method can fit highly complex response functions by integrating several function types (linear, quadratic, product, threshold, hinge, and category indicators) (Phillips and Dudík 2008). The algorithm used can also be interpreted from a machine learning perspective (Phillips et al. 2006). Liu et al. (2018) demonstrated its application to mineral potential mapping.

Random Forest Method

The random forest (RF) method is an ensemble algorithm that represents an extension of classification and regression trees, and can be used to classify or predict the value of a target variable based on a number of evidential variables. It is sequentially applied from a root node to a terminal node (leaf) to make repeated predictions (Breiman 2001). Classification and regression trees are the basic classifiers used in the RF method, which uses a bagging technique to ensure that training subsets are randomly chosen, with each subset forming a decision tree (Breiman 1996). This bagging technique means that

roughly one-third of the available training samples are not used in the construction of RF trees; instead, they are used to validate the prediction accuracy (also referred to as “out-of-bag” or OOB samples). The resulting OOB error is an unbiased estimate of the generalization error during RF analysis (Breiman 2001). The evidential variables used for each node in the decision tree are also randomly chosen. The outcome of RF modeling is dependent on the average prediction of all of the trees involved in the model (Cutler et al. 2007).

The RF algorithm begins with splitting parent nodes (i.e., evidential features) into binary pieces, where child nodes are purer than the parent node. Searching through all of the candidate splits yields optimal splits that maximize the “purity” of the resulting trees. The RF algorithm uses the Gini impurity index to calculate the information purity of child nodes compared with their parent nodes, with splitting thresholds determined from the maximum reduction in purity values (Breiman 2001). This splitting process is repeated until a stop condition is reached.

The advantages of the RF algorithm include the fact that the bagging technique, which involves random resampling and replacement, yields different training subsets that can be subsequently used to generate decision trees, thereby increasing the diversity within the model and avoiding correlations between trees during the RF process. This allows greater stability and prediction accuracy, as some of the input are not used, avoiding certain variations.

The best evidential features are used as splitting points to enable tree growth during the RF process. The random selection of evidential features to be used as part of the overall set of input evidential features also reduces correlations between trees, lessening the generalization error within RF models. The RF method has been demonstrated for mineral potential mapping by Carranza and Laborte (2015b, 2016), McKay and Harris (2016) and Hariharan et al. (2017).

Evaluation of Results

The results of MaxEnt and RF modeling of mineral potential were assessed using receiver operating characteristic (ROC) curves and area under the curve (AUC). The ROC is both robust and threshold-independent, meaning that they are ideal for determining the accuracy of the results of clas-

sification modeling (Chung and Fabbri 1999; Lee and Pradhan 2007). This method plots sensitivity (true-positive rate) against “1-specificity” (i.e., the false-positive rate), and the AUC is then calculated for all possible probability thresholds. The AUC values range from 0.5 to 1, where 0.5 is analogous to a completely random prediction and 1 implies perfect prediction (Lee and Dan 2005); AUC values of > 0.9 denote very good model performance (McCune et al. 2002).

The performance of the RF and MaxEnt models in producing predictive maps with floating values ranging from 0 to 1 was further assessed using success rate curve analysis. This approach was described by Agterberg and Bonham-Carter (2005) and involves the classification of a series of prospective pixels based on 5-percentile intervals of probability values within the modeled predictive map. The highest (100th percentile) cutoff probability relates to the minimum proportion of the prospective parts of the study area, whereas the lowest (0 percentile) cutoff probability contains the maximum prospective area. Other percentile intervals have success rates that represent the proportion of gold deposit occurrences contained within the associated prospective area. The proportion of gold occurrences was also compared with cumulative probability proportion values, where the equally divided proportions of cumulative probability values were plotted from highest to lowest against the cumulative proportions of gold deposits contained within each interval.

DATASETS AND APPLICATION

Spatial Datasets

The spatial datasets used in this study include a geological map, a map showing the locations of faults, fractures, and Au deposit locations (Qi et al. 2013) (Fig. 1). These spatial datasets were processed using a grid with a pixel size 150×150 m to prepare the data for analysis. This pixel size was objectively determined based on the spatial pattern of known Au deposits and the distribution of related faults and intrusions to ensure the pixels adequately represent the spatial resolution of the datasets being used and that only one deposit exists in any given pixel (Carranza 2009). A dataset of geochemical concentrations of 13 trace elements (Au, As, Sb, Bi, Hg, Ba, Co, Cu, Pb, Zn, Ag, W, and Mo) derived from 9041

stream sediment samples was also used. The analytical method used is inductively coupled plasma mass spectrometry (ICP-MS) with relative standard deviation (RSD) 6.1%.

Due to the fact that geochemical data are compositional data, standard statistical treatments are unable to deal with such datasets that are represented as summing to a unit constant (Filzmoser and Hron 2008; Carranza 2011; Zuo 2014). Restricted by the force of a constant sum, geochemical information carried by compositions is trading off with each other. In practice, logratio transformations including additive logratio transformation (alr) (Aitchison et al. 1982), centered logratio transformation (clr) (Aitchison et al. 1982), and isometric logratio transformation (ilr) (Egozcue et al. 2003) are commonly used in geochemical data processing to address the constant sum (i.e., closure) problem. However, if compositions do not add to a constant sum, they are considered as sub-compositions and closed by adding them to the undetermined parts or by forcing them to sum up to a constant depending on the unit of measurement (Otero et al. 2005). In our case study, the data of sub-compositions of 13 trace elements determined from the samples were closed by forcing them to sum up to 100%. Then, principal component analysis was applied to ilr-transformed geochemical data using the R package “robCompositions” version 2.0.8.

In the family of logratio transformations that are commonly used to open a closed system, only the ilr-transformed variables lie in orthogonal system and standard statistics designed for Euclidean space are consequently applicable to ilr-transformed variables (Aitchison et al. 2000; Buccianti 2013; Filzmoser et al. 2009, 2010; Pawlowsky-Glahn and Egozcue 2006). However, ilr transformation reduces the number of resulting variables bringing about the difficulty in interpretation of statistical results (Pawlowsky-Glahn and Egozcue 2006). For the sake of ease of interpretation, the loadings and scores from PCA based on ilr-transformed variables are back-transformed to clr space (Filzmoser et al. 2009).

Target Variables

Mineral deposit occurrences comprise a dichotomous target variable that is used in data-driven mapping of mineral potential. This variable is represented by values of 1 and 0 for pixels contain-

ing deposits and no mineralization, respectively. The mineral potential mapping here used the locations of the 18 gold deposits and 18 barren locations (here termed non-deposits) in the study area. The latter was generated using the following selection criteria.

1. The number of non-deposits should be equal to the number of Au deposit occurrences to ensure optimal regression (Breslow and Cain 1988; Schill et al. 1993). If the number of non-deposits exceeds the number of deposits, then the information derived from the evidential maps input to the model is diminished (King and Zeng 2001).
2. Non-deposits should be distal from any known Au deposit to avoid having similar multivariate spatial data signatures to areas of known mineralization.
3. Non-deposit should be randomly spatially distributed.

These criteria were used during the selection of the 18 non-deposits used during RF analysis.

Evidential Variables

Analysis of the mineral system associated with the mineralization in the study area indicates that it is worth assessing the spatial association of the target variables in terms of distances to (1) faults and (2) intrusions.

The geochemical data used in this study were assessed using the local singularity analysis approach (Cheng 2007), which can discriminate between weak anomalies and background concentrations of mineralization-related or pathfinder elements. Here, we use the concentrations of Au, As, Sb, Ag, and Cu as indicators for singularity analysis (Fig. 2), where red zones indicate areas containing accumulations of these elements that are considered to be either genetically related to mineralization or directly indicate the location of Au deposits. Local singularity indices were produced in GeoDAS (Cheng 2000) with singularity indices inversely proportional to the geochemical anomalies identified using the techniques above.

The final step was a principal component analysis (PCA) of the geochemical dataset, yielding PC1 and PC2 scores that were used as evidential variables. The association elements and its correlations in PC1 and PC2 are depicted in the biplot (Fig. 3).

According to the geochemical field characteristics and biplot of the first principal component (PC1) and second principal component (PC2) of the data (Fig. 3), the association of Au and As in the third quadrant indicates that the Au deposit occurrences are closely related to As anomaly. The association of Hg, Sb, Au, and As in the second and third quadrants, resulting in negative value of PC1, corresponds to the geochemical signature of tectonic process or their products (i.e., fault system). The separation of Au from the association of Hg and Sb, which is represented by PC2, might imply extraordinary immobility of Au in the study area, as it may not always coexist with the more mobile Sb and Hg in the surficial environment.

MaxEnt Modeling

This study used MaxEnt software version 3.4.1 to construct a MaxEnt model including nine evidential maps, with default parameters set to a maximum of 500 iterations, a maximum convergence threshold of 0.00001, which are conditions used to stop training. Because MaxEnt is calculated over the set of pixels, large number of pixels will increase processing time without a significant improvement in modeling performance, random samples of maximum 10,000 “background” pixels were used to compute the MaxEnt distribution over the union of “background” pixels and samples for deposits being modeled; also, a regulation parameter of $\beta = 1$, which depends on the sample size was used. The larger sample size generally leads to a smaller value of β in terms of features types, and the logistic output format was selected for ease of interpretation. Detail of the parameters can refer to help for Maximum Entropy Species Distribution Modeling in the software.

Random Forest Modeling

The RF modeling used the Random Forest package within the R statistical environment (Liaw and Wiener 2002; R Development Core Team 2008). The parameters set for this model were the number of trees (k) and the number of evidential maps (m) that were randomly sampled at each split. The m value can be empirically determined by calculating the fraction of the total number of evidential maps represented by the square root of the total number

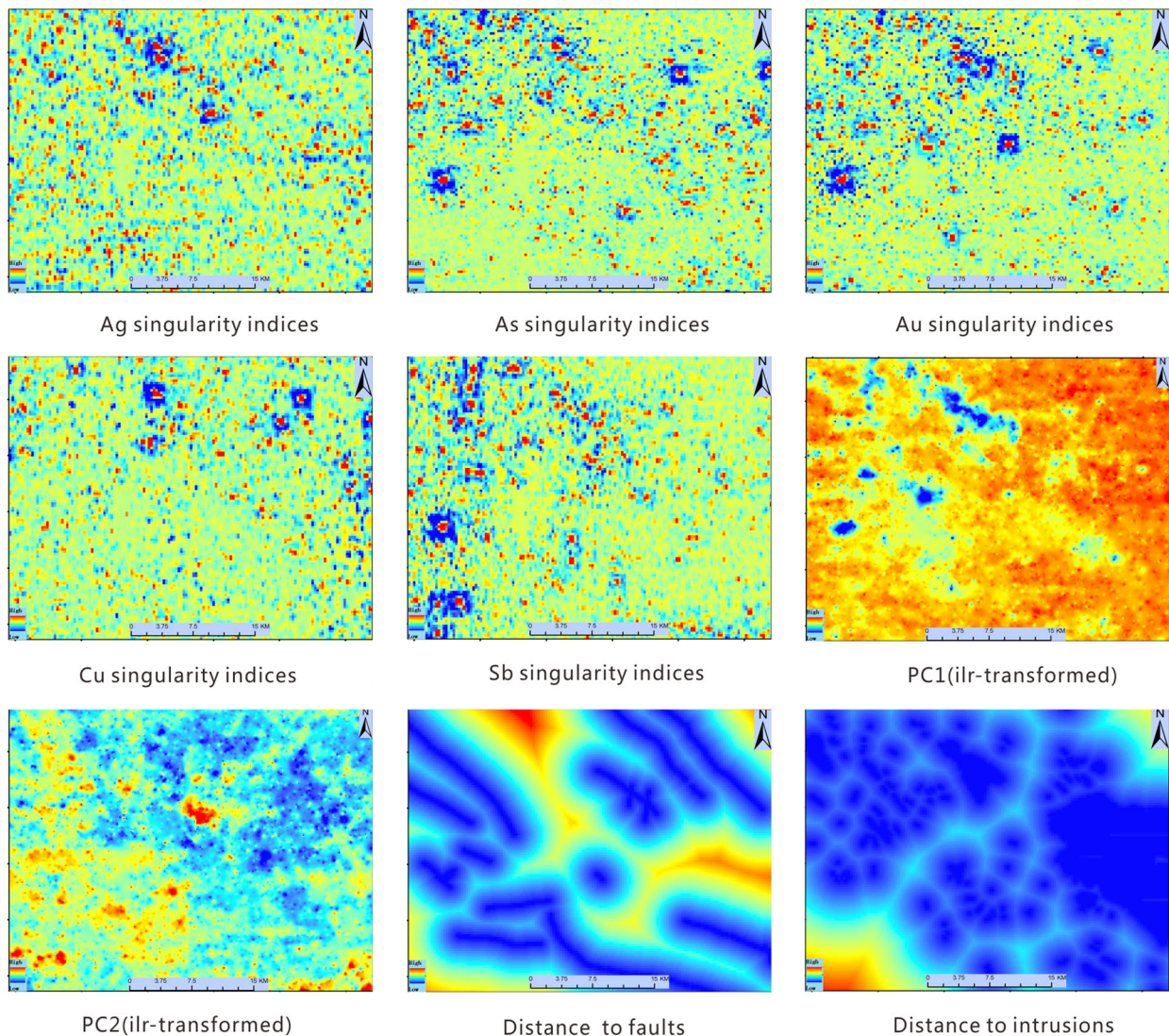


Fig. 2. Maps showing the distributions of evidential variables used in MaxEnt and RF modeling during this study.

of evidential maps ($m = \sqrt{n}$, where n indicates total number of evidential maps). Although Breiman (2001) and Liaw and Wiener (2002) indicated that an m value as low as 1 can yield accurate results, Grömping (2009) reported that the m value needs to include at least two evidential variables.

This study used the “tuneRF” function to determine optimal parameters. Multiple experiments indicated that the m parameter is consistent with the empirical value outlined above and the minimum k value of 1000 yields both the lowest prediction errors and the most stable predictions. The suitable values of these parameters ensure that

the RF algorithm will find a fit between the targets (i.e., deposits and non-deposits) and evidential maps and that all of the evidential maps input to the model can then be applied to the model to compute probabilities for all locations.

RESULTS AND DISCUSSION

Results of MaxEnt Modeling

The MaxEnt modeling outputs were mapped using ArcGIS software with continuous logistic

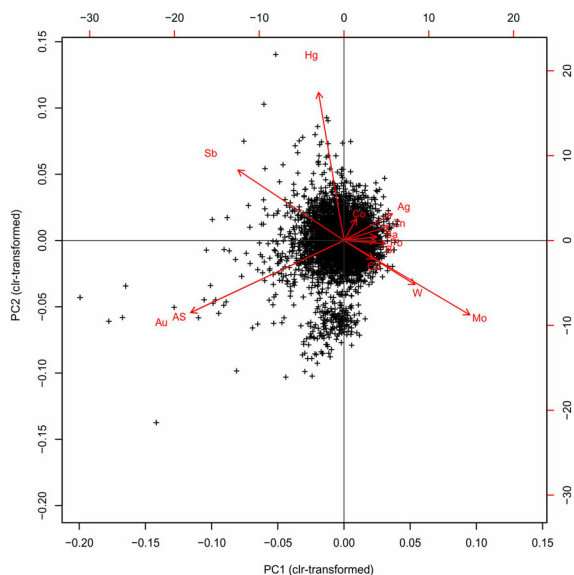


Fig. 3. Biplot of the first two PCs.

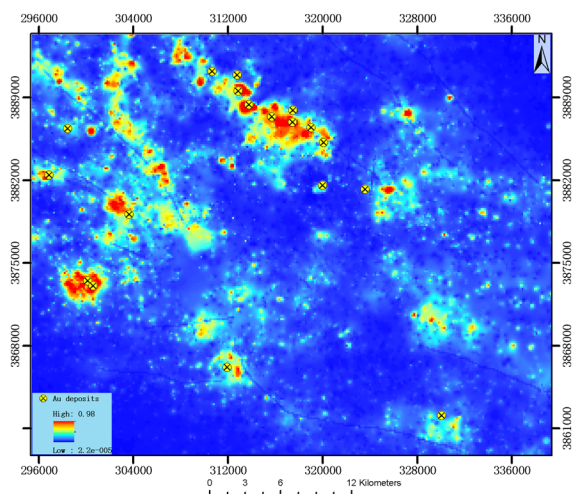


Fig. 4. MaxEnt-model-derived gold prospectivity map for the Hezuo-Meiwu district.

probabilities ranging from 0.000022 to 0.98. The probability map shown in Figure 4 indicates a good relationship between areas with high-probability values and the locations of known Au deposits.

A jackknife analysis of the results of the MaxEnt model provides an indication of the relative importance of each evidential map used in the model, as well as guidance for Au exploration in the study area. Jackknife analysis successively excludes an evidential map from the analysis before re-run-

ning the MaxEnt model using the rest of the evidential maps and a separate model using the excluded evidential map only. This allows the determination of the relative importance of each evidential map to the final probabilities determined by the model (Fig. 5). The results show that the PC1 scores, the singularity values calculated using Ag and As concentrations, and the distance to intrusions are more important evidential maps than the rest in the MaxEnt model. The PC1 scores and the Ag singularity indices derived from the geochemical datasets are the two most important evidential maps in terms of relationships to the locations of known gold deposit occurrences, reflecting the importance of using geochemical anomalies during exploration in this region. The third most important evidential map is the distance to individual intrusions, which is consistent with the fact that the deposits in the study area are intrusion-related gold systems (Sui et al. 2017).

Response curves indicate how each evidential map influences the resulting probability when all variables are used to build a full model and reflect the spatial relationships between the evidential maps and areas containing gold deposits. High values along the Y axis of a response curve indicate that these areas have a higher logistic probability of containing a known gold deposit (Liu et al. 2017). The PC1 scores are the most important evidential variable and show an increase along the X axis that leads to a decrease in the probability of gold deposit occurrences. Variations in distance to intrusions record an increase along the X axis, leading to a decrease in the probability of gold deposit occurrences with increasing distance. This result is expected, as the mineralization in the study area is genetically related to intrusions. The response curves for the Ag, As, and Cu singularity indices reflect the nature of these indices, where singularity index values of > 2 indicate the dispersion of the corresponding element. The response curves for these elements indicate that all known deposit occurrences are closely related to areas containing accumulations of Ag, As, and Cu (Fig. 6). In comparison, the response curves for Au singularity indices, distance to faults, PC2 scores and Sb singularity indices are different from the rest. According to the spatial relationship between Au deposits occurrence and values of Au singularity indices, most of Au deposits related to magmatic rocks (Dewulu) have high accumulation of Au (red zone, low values of Au singularity in Figure 2); however, the gold deposits in the western

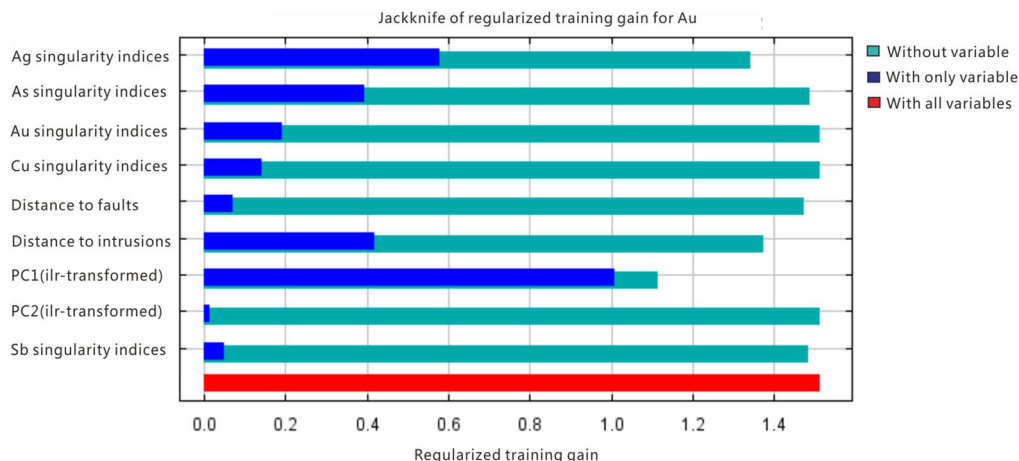


Fig. 5. Jackknife analysis indicating the relative importance of the individual evidential maps used in the MaxEnt model.

zone especially the Zaozigou deposit have higher values (dispersion) of Au singularity indices. As to the distance to faults, only three gold deposits that near to the Dewulu have values of distance to faults larger than 6000 m and 14 gold deposits are less than 4000 m from the faults.

Finally, although distances to faults are of lesser importance in the MaxEnt model, the response curve for distance to faults (< 4 km and > 6 km from faults) suggests that these faults in the study area might have an influence over a larger scale (e.g., from a mineral systems viewpoint).

Results of Random Forest Modeling

The RF modeling yielded a less distinct relationship between high-probability areas and areas containing known Au deposits. The red zones within the RF model (Fig. 7) indicate areas of higher probability are much larger than those within the MaxEnt model. The RF algorithm also ranks the importance of evidential maps using mean decrease accuracy and mean decrease Gini indices (Fig. 8). The first measure is computed from OOB data, reflecting the decrease in accuracy within the entire forest model, whereas the Gini importance index measures the average gain of purity by using splits of a given variable. The sensitivity of each evidential map was determined using marginal effects (Fig. 9) on target variables while holding all other evidential maps constant.

The results of RF modeling indicate that the three most important evidential maps are PC1 scores, As singularity indices, and the distance to intrusions (Fig. 8). The positive spatial associations between known gold deposit occurrences and the locations of intrusions and faults mean that the former are located proximal to the latter. This is confirmed by the RF modeling, which yields an optimal positive spatial association within ~ 1000 m (accounting for 15 gold deposits) and ~ 4000 m (accounting for 14 gold deposits) of intrusions and faults, respectively (Fig. 9). The positive spatial association between known gold deposits and PC1 scores also indicates that the locations of known Au deposits are characterized by negative PC1 scores.

The spatial associations of gold deposit occurrences with geochemical singularity indices obtained during the RF modeling are slightly different from the response curves obtained from the MaxEnt modeling. The in-depth analysis of singularity indices distribution shows that gold deposits related to Dewulu intrusions in the eastern zone usually have low values of Au, As singularity and deposits associated with dikes and stocks in the western zone usually have high value of Au, As singularity which reflects two groups of deposits in the study area. All of the evidential maps can be interpreted from geological viewpoints to some extent, and none of these data have flat responses, indicating that all of the evidential values have some spatial association with the known gold deposit occurrences.

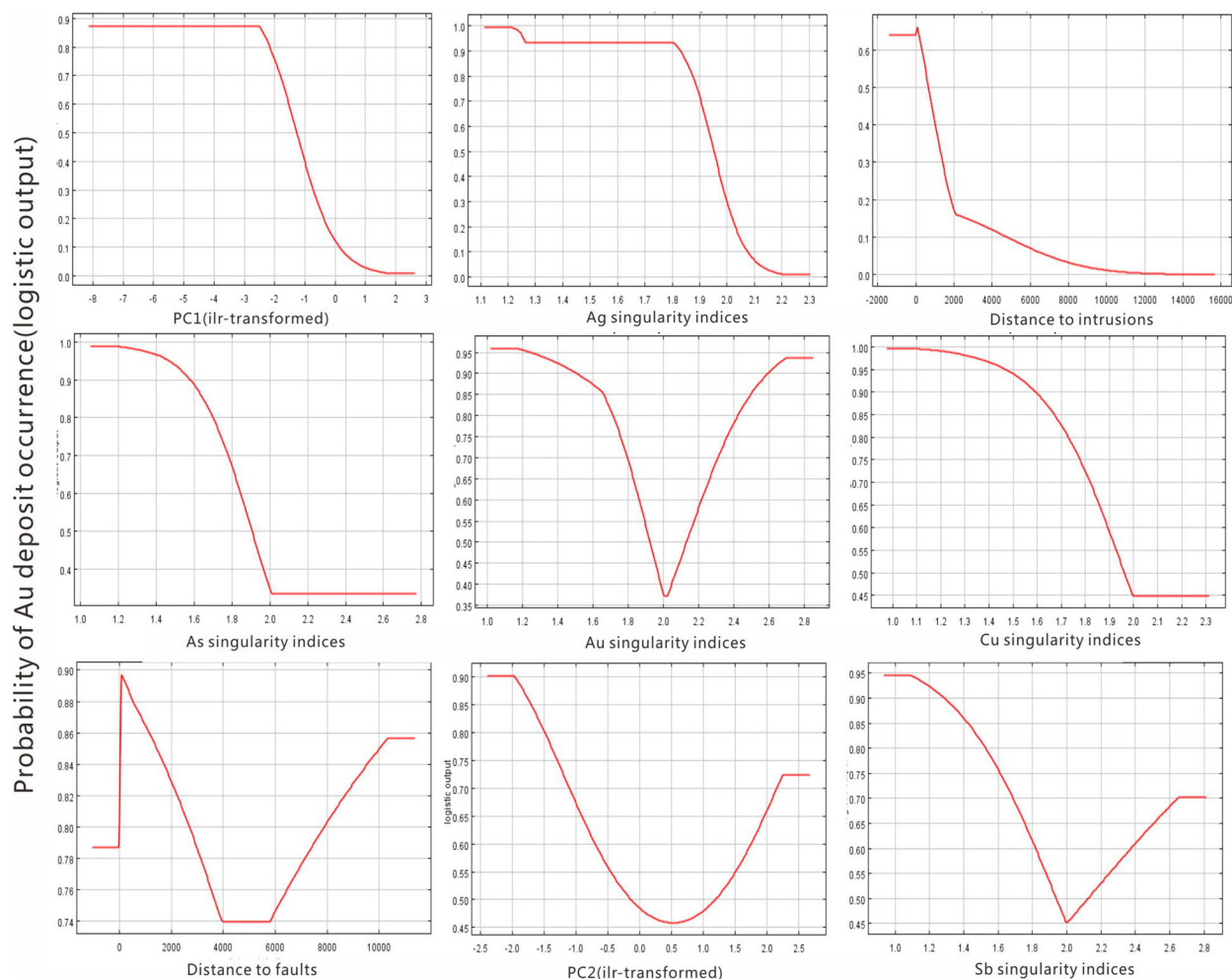


Fig. 6. Response curves for the evidential variables used during MaxEnt modeling.

Mineral Potential Mapping by Zones

Rationale

Considering the two groups of deposits and two populations in the response curves and marginal effects (Figs. 6 and 9), it is necessary to investigate uncertainty in mineral potential mapping in the study area by zones. Therefore, the 12 deposits in the eastern zone were used to train the MaxEnt and RF models, and the results were cross-validated using the six deposits in the western zone. Then, the MaxEnt and RF models were trained using the six deposits in the western zone, and the results were cross-validated using the 12 deposits in the eastern zone.

Firstly, two groups of deposits used as training points for MaxEnt modeling to assess the changes of

regularized training gain of evidential maps (Fig. 10) corresponding to deposits in the eastern zone (E) and western zone (W). Compared to the jackknife of regularized training gain for deposits in the study area, the great increase gained from distance to intrusions and distance to faults in the western zone shows that the fault and intrusions (dikes and stocks) play an important role in the formation of gold deposit. However, in the eastern zone, the PC1 scores and Ag, As singularity indices indicate that geochemical anomalies would contribute significantly in prospecting for gold deposits.

Secondly, apart from the jackknife analysis of regularized training gain, the performance of training models of MaxEnt and RF on the test data helps to clarify the difference between the groups of gold deposits. Figure 11 shows that the trained MaxEnt models give different values of AUC. The training

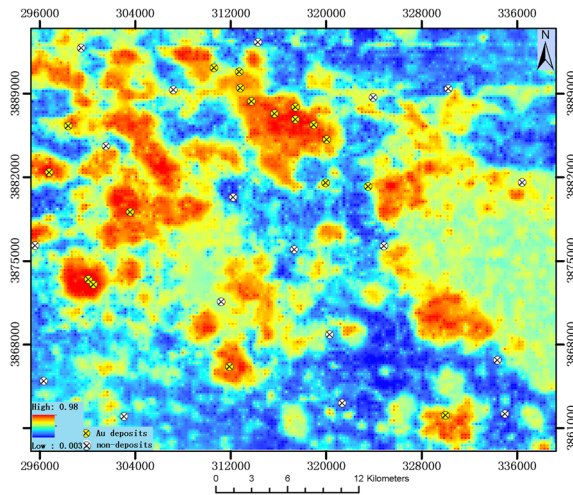


Fig. 7. RF-modeling-based gold prospectivity map for the Hezuo–Meiwu district.

model in the eastern zone works well with the western test data, which can be explained by the jackknife of regularized training gain. That is because the types of evidential maps in the eastern zone are the same as the type of evidential maps in the western zone, although the evidential maps in the western zone lack the Ag and As singularity maps (Fig. 10) leading to poor performance of MaxEnt model trained on the western test data. The RF model produces the confusion matrices (Table 1) of training data and test data. The large class errors also imply that training should indeed be di-

vided according to two zones for mineral potential mapping, respectively.

Analysis and Results

In order to reduce the uncertainty in mineral potential mapping, the study area was divided into two zones by the Xiahe–Hezuo thrust. The mineral potential mapping using MaxEnt and RF models was then performed for each zone. Figure 12 shows the importance of evidential maps to deposits in the corresponding zones. The jackknife analysis of regularized training gains in the MaxEnt modeling shows that the most important evidential maps in the eastern and western zones are PC1 scores and distance to intrusions, respectively, which are also confirmed by the mean decrease accuracy and mean decrease Gini in the RF modeling (Fig. 13). Moreover, the appearance of PC2 in the western zone (Fig. 12) and the disappearance of Ag singularity index in the eastern zone (Fig. 12) corresponding to Figure 10 imply that the efficacy of the evidential variables likely depends on a certain scale.

Discussion

The AUC values for the RF modeling by zones (RF-EW), RF, MaxEnt modeling by zones (MaxEnt-EW) and MaxEnt models generated during this

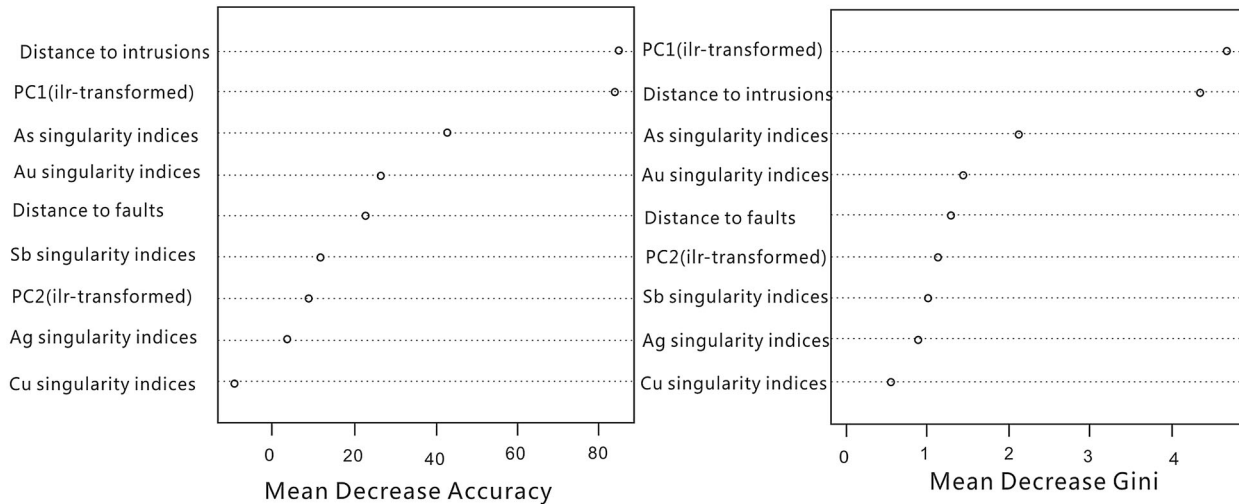


Fig. 8. Relative importance of evidential variables used during RF modeling.

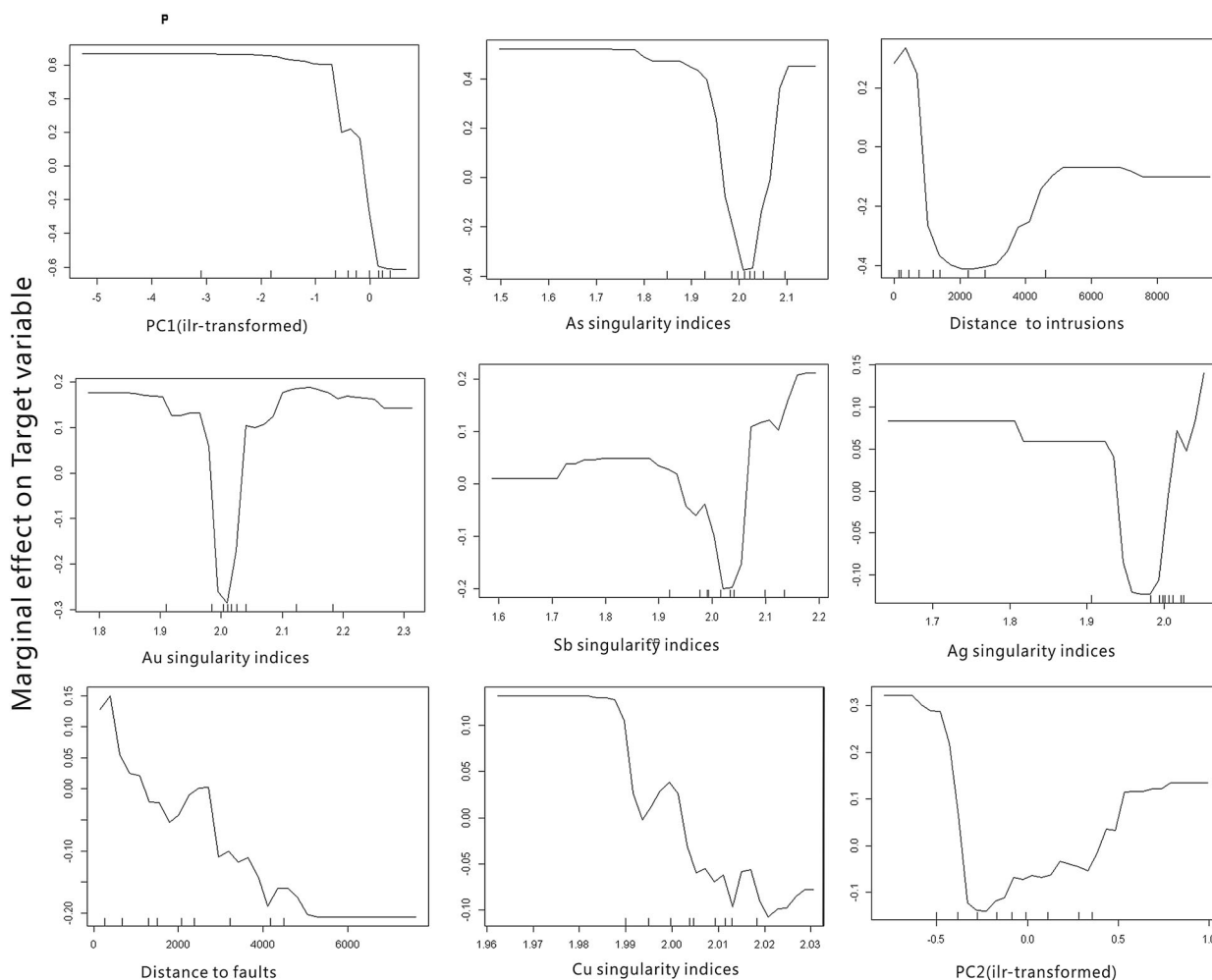


Fig. 9. Marginal effects of evidential variables on the target variables during RF modeling.

study are 0.958, 0.926, 0.914 and 0.864, respectively (Fig. 14). These results indicate that the mineral potential mapping by zones have much larger AUC value, indicating that in-depth analysis of evidential maps at different scales based on metallogenic characteristics will help to reduce uncertainty in mineral potential mapping.

The evidential map importance analysis and ROC curves discussed above provide insights into the predictive quality of the output of each of the four models generated during this study. The performance of the RF and MaxEnt models in producing predictive maps with floating values ranging from 0 to 1 was further assessed using success rate curve analysis. The success rate curve is a capture-efficiency curve that indicates the relationship be-

tween the probability distribution and Au deposit locations. The derived capture-efficiency curves (Fig. 15) indicate the following. All of known gold deposits are located within the top 35% (i.e., high probability) part of the RF modeling by zones and 94% of gold deposit within top ~ 40% (i.e., high probability) in RF modeling in the whole study area. Less than 45% of the known gold deposits lie within the same top 35% section of the MaxEnt modeling by zones and only ~ 39% of gold deposits within top 35% probability in MaxEnt modeling in the whole study area. These results, combined with the fact that both RF models contain an area representing 19% of the total study area but that contains more than 88% of the known gold occurrences (Fig. 16), suggest that the RF model produces a more rea-

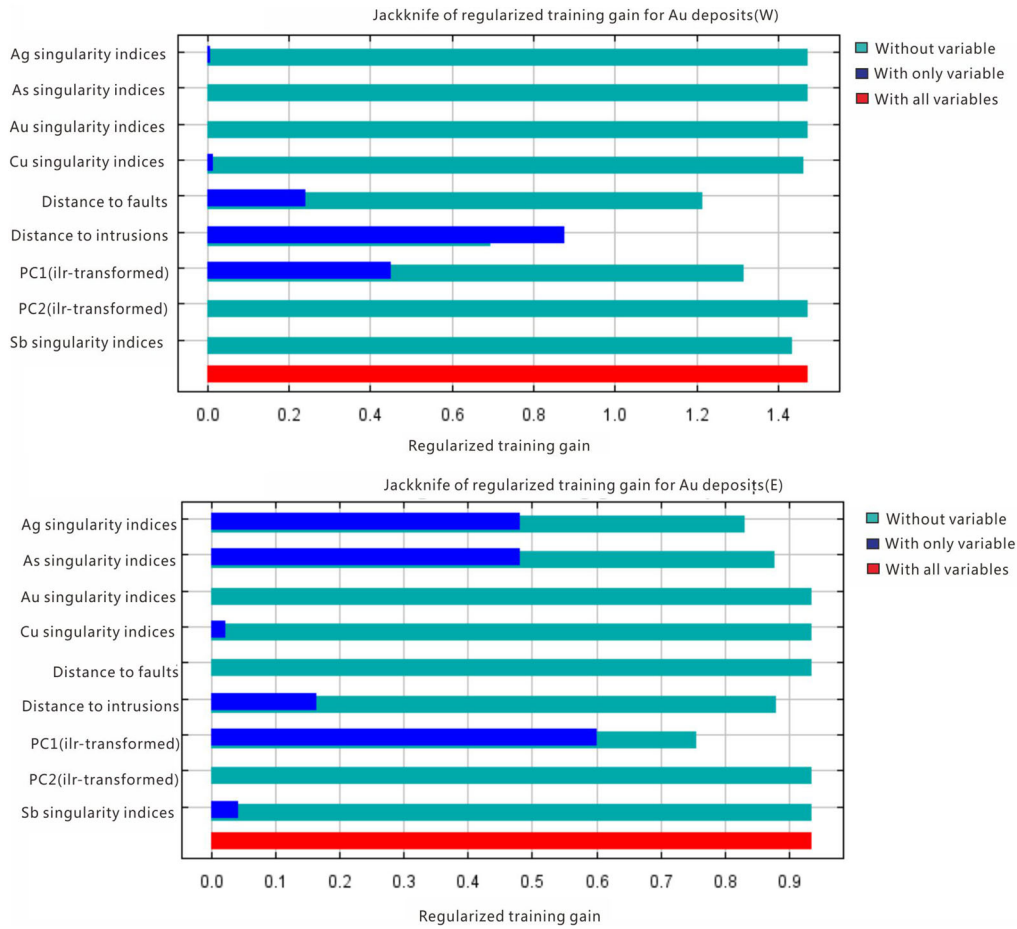


Fig. 10. Jackknife analysis indicating the relative importance of the individual evidential maps used in the MaxEnt model with the western and eastern deposits, respectively.

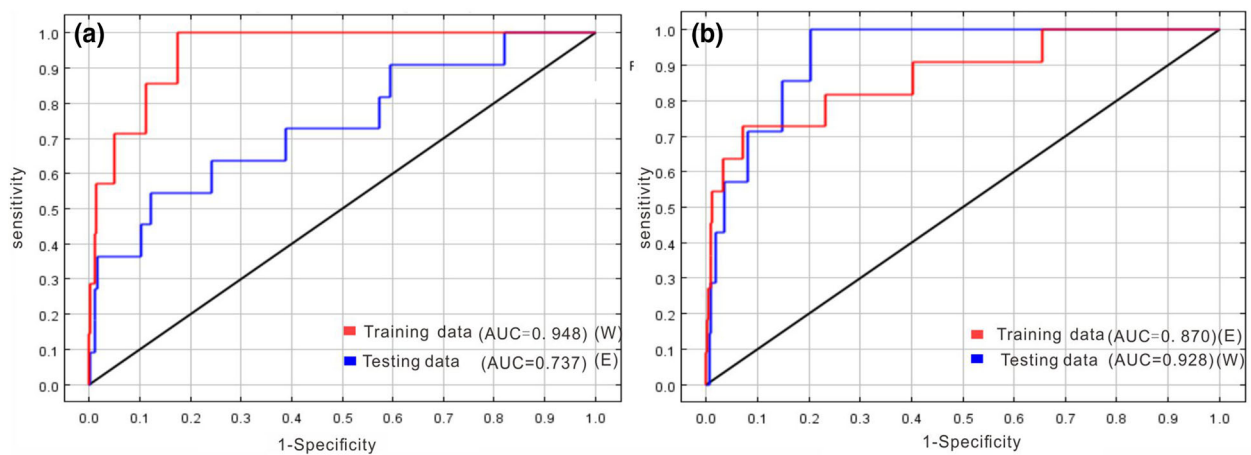


Fig. 11. The test AUC for the eastern and western zones.

Table 1. Confusion matrices of training data and test data in RF

	0	1	Class error
Eastern training	0	11	1
	1	0	12
Western test	0	4	2
	1	1	5
Western training	0	5	1
	1	0	6
Eastern test	0	12	0
	1	5	7
			0.083
			0
			0.333
			0.167
			0.167
			0
			0
			0.41

sonable probability distribution outcome than the MaxEnt model, given the locations of known gold deposits in the study area.

The success rate and capture-efficiency curves provide some insights in probability and area dis-

tribution corresponding to proportion of Au deposit occurrences. Here, we use the correlation indices of probabilities (Table 2) in the prospectivity map generated by the four models to investigate the spatial relations of the prospective area and provide more reliable results for further exploration.

The RF modeling by zones (RF-EW) with highest AUC value of 0.958 has a higher correlation index of 0.82 with RF modeling in the study area, indicating that the PC1 scores and distance to intrusions dominate the classification efficacy. Given that both MaxEnt and RF models performed differently in the whole study area in the two zones indicate that those two evidential variables play different roles in each of group of deposits. The correlation index of probabilities between MaxEnt by zones (MaxEnt-EW) and RF by zones is as high as 0.71, which means that the high potential area are spatially related, and visual ternary class maps from

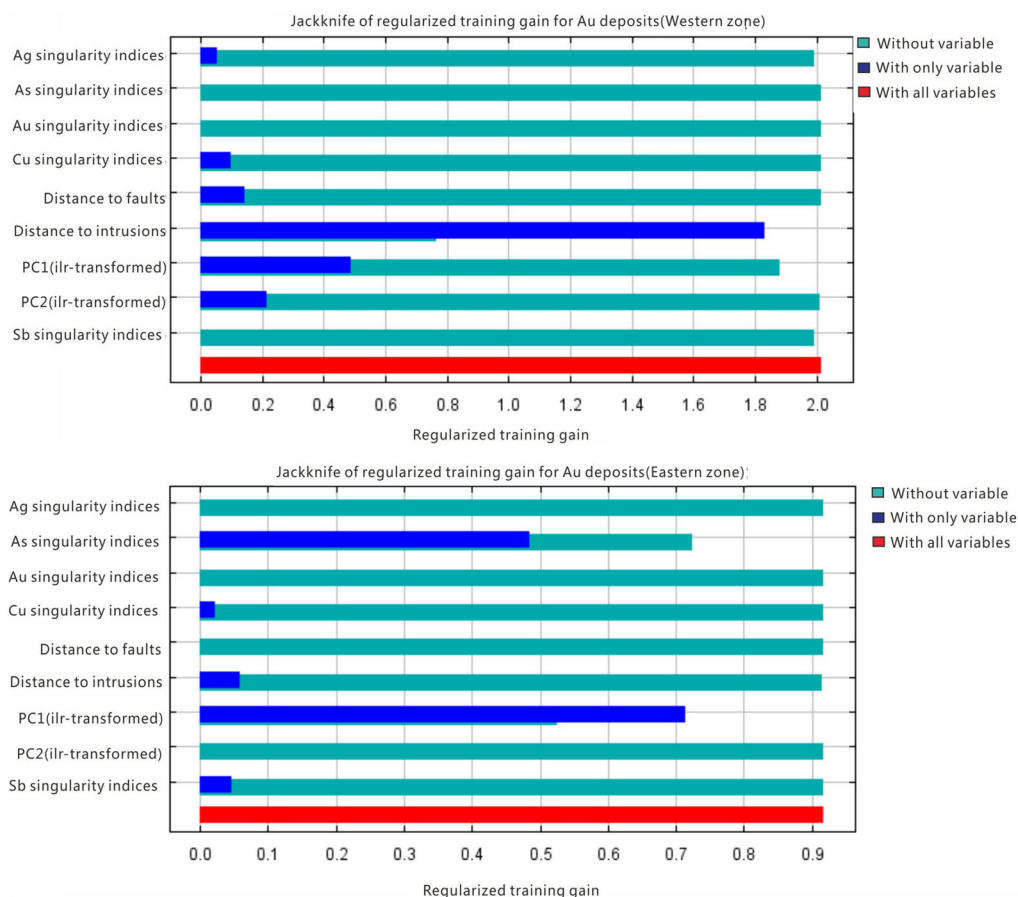


Fig. 12. Jackknife analysis indicating the relative importance of the individual evidential maps used in the MaxEnt model by zones.

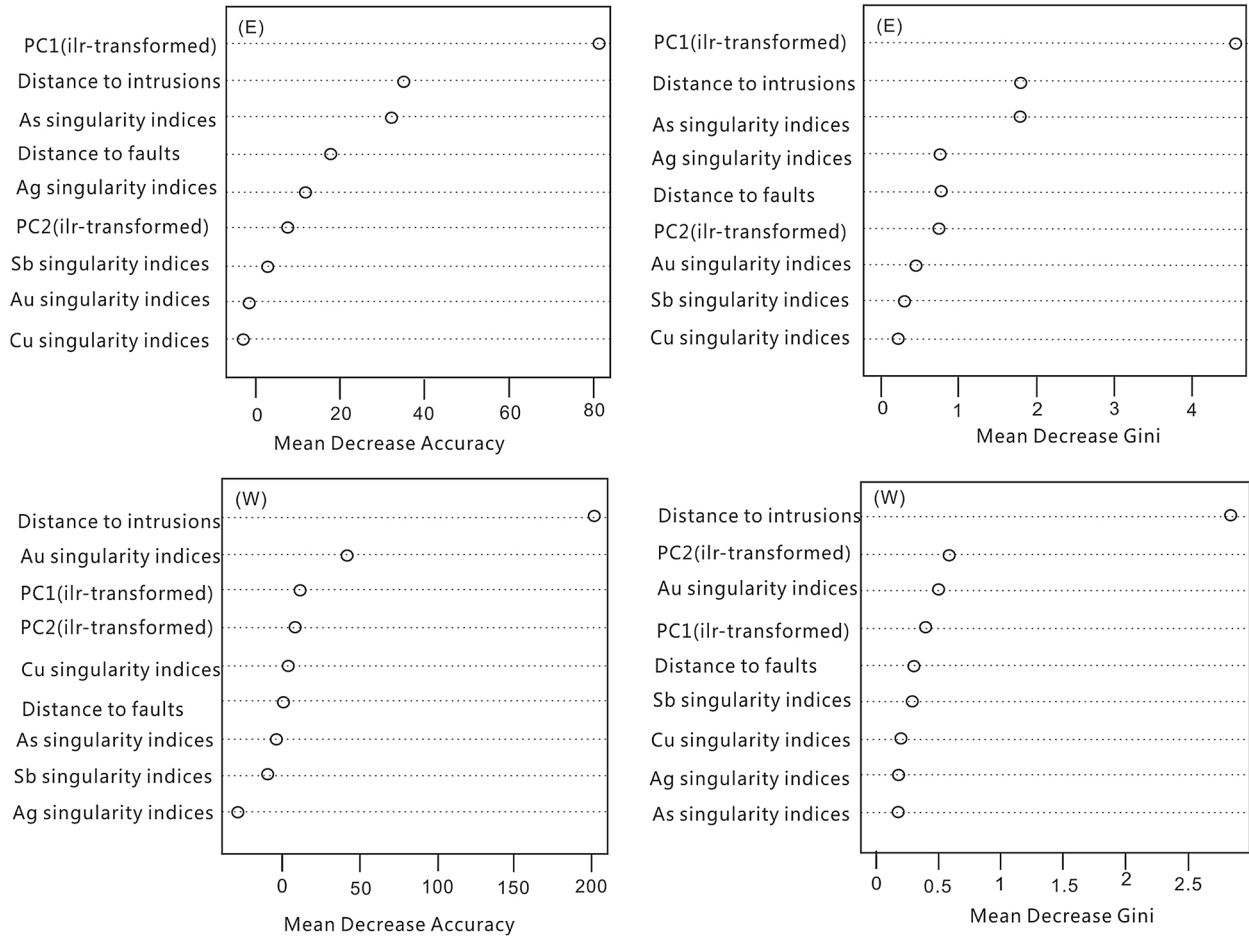


Fig. 13. Relative importance of evidential variables in the eastern (E) and western (W) zones determined by RF modeling.

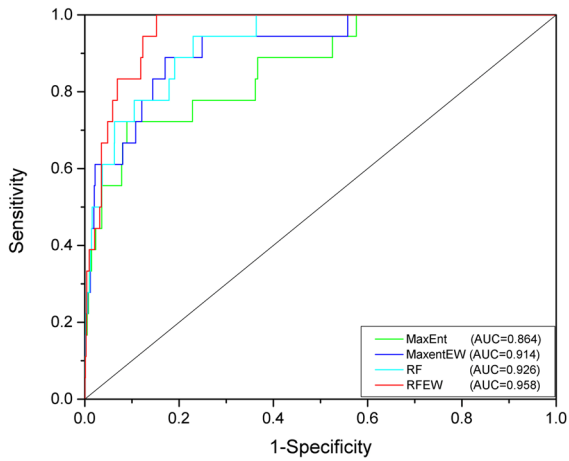


Fig. 14. ROC curves for the MaxEnt and RF models generated during this study.

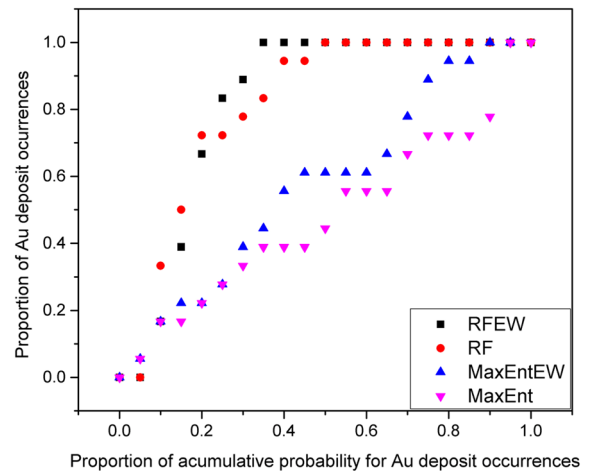


Fig. 15. Capture-efficiency curves for the MaxEnt and RF models.

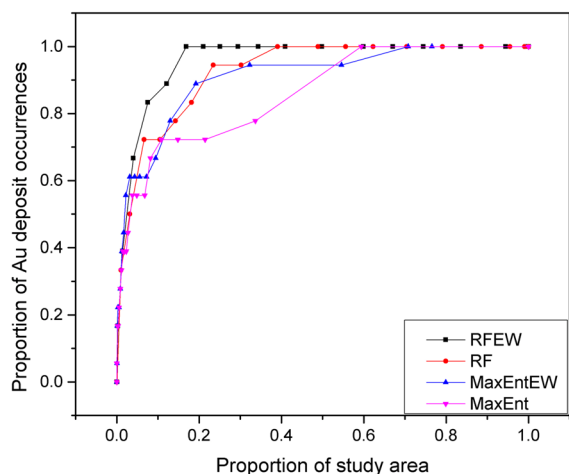


Fig. 16. Success rate curves for the MaxEnt and RF models.

Table 2. Correlation of probabilities in the mineral potential maps

Correlation	MaxEnt	MaxEnt-EW	RF	RF-EW
MaxEnt	1	0.59	0.68	0.63
MaxEnt-EW	0.59	1	0.49	0.71
RF	0.68	0.49	1	0.82
RF-EW	0.63	0.71	0.82	1

the MaxEnt-EW and RF-EW would facilitate the decision of exploration.

In order to distinguish the levels of potential for gold deposits of the study area, the concentration–area (C-A) fractal analysis (Cheng 1999) was adopted

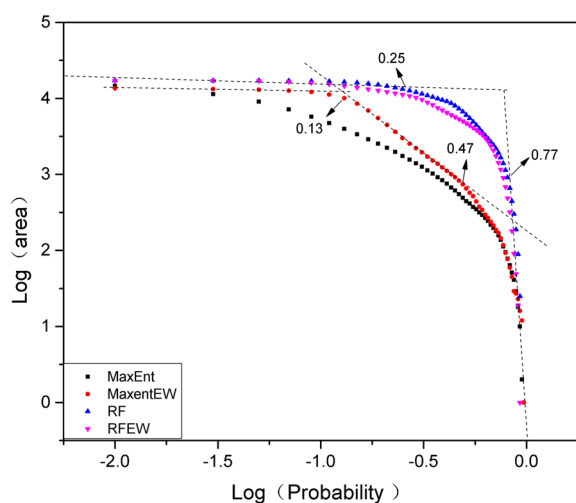


Fig. 17. C-A thresholds for prospectivity maps.

to delineate three probability populations corresponding to high potential, moderate potential, and low potential (Fig. 17).

Future mineral exploration in this area can be guided by ternary class gold prospectivity maps obtained from the MaxEnt-EW (Fig. 18) and RF-EW (Fig. 19) models. According to C-A thresholds for prospectivity maps, the thresholds of 0.47 and 0.13 were used to divide the prospectivity map into high potential, moderate potential and low potential area in the MaxEnt modeling by zones. The thresholds of 0.77 and 0.25 were used in RF modeling by zones. The area delineated by the threshold 0.77 in the RF modeling should be prioritized because it outlines 7.5% of the study area as prospective with 83% success rate.

SUMMARY AND CONCLUSIONS

Mineral potential mapping represents an important tool for mineral exploration in both brownfield and greenfield environments. This study evaluated the applicability and compared the performance of the random forest (RF) and the maximum entropy (MaxEnt) models to mineral potential mapping, both of which have been widely used in environmental and ecological modeling but so far have only had limited use in mineral potential mapping. This contribution investigated the potential use of the RF and MaxEnt models in identifying prospective areas for mineral exploration and in

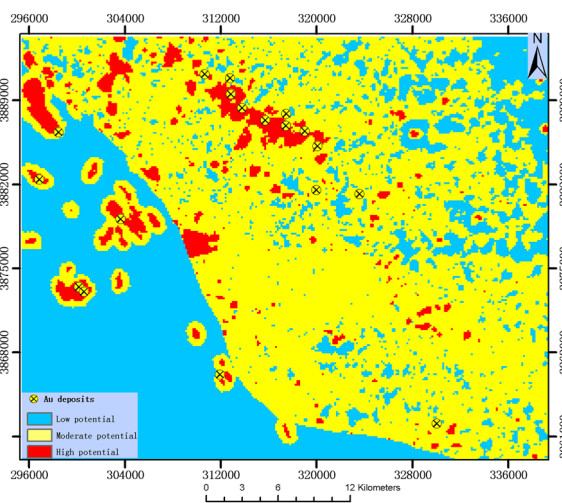


Fig. 18. Ternary class gold prospectivity map obtained from the MaxEnt modeling by zones.

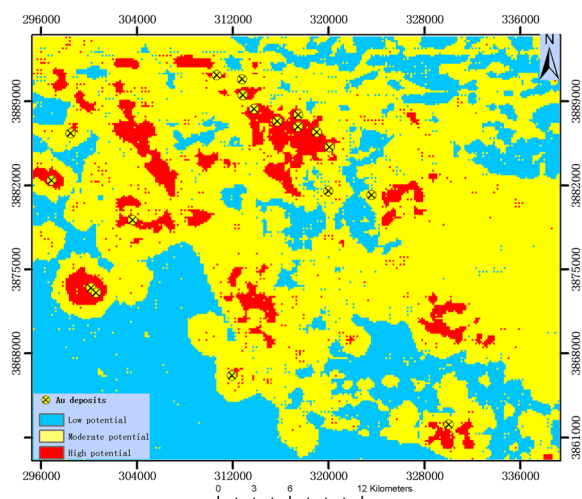


Fig. 19. Ternary class gold prospectivity map obtained from RF modeling by zones.

predicting the spatial distribution of potential gold deposit occurrences within the Hezuo–Meiwu area of China.

Mineral potential maps were generated by selecting and preparing evidential maps relevant to the specific type of the mineral deposits present in the study area as well as the associated larger mineral systems. These data were then examined using innovative methods such as the singularity index analysis of individual geochemical elements and the commonly used principal component analysis, yielding useful information that could be used during prospectivity modeling. Finally, these data were combined with RF and MaxEnt modeling to produce mineral potential maps in a GIS environment.

The accuracy of the models indicates that the RF models in the study area or performed by zones are slightly better than the MaxEnt model. These two models also allow a sensitivity analysis, yielding quantitative evaluations of the effect of evidential maps on the resulting mineral potential maps. These sensitivity analyses indicate that there are two groups of deposits, and mineral potential mapping should be carried out in terms of metallogenic characteristics. In order to reduce the uncertainty in prospectivity maps, both of MaxEnt and RF are performed by zones. According to the sensitivity analysis, the mineral potential mapping in the western zone is most sensitive to distance to intrusions, and the eastern zone is dominated by PC1 scores.

In addition to their high AUC values, the MaxEnt and RF modeling by zones can more precisely identify areas containing known gold deposits, with both models identifying areas occupying 19% of the study area and containing > 88% of the known gold deposits. The results enable a comprehensive evaluation of the mineral potential of the study area and provide a case study for machine-learning-based data-driven predictive modeling in an area containing only a few (< 20) known mineralized occurrences.

ACKNOWLEDGMENTS

This research was financially supported by Project No. 2017YFC0601501 from the National Key Research and Development Program of China, Project Nos. 1212010733806 and 1212011120140 from China National Mineral Resources Assessment Initiative.

REFERENCES

- Abedi, M., Norouzi, G. H., & Bahroudi, A. (2012). Support vector machine for multi-classification of mineral prospectivity areas. *Computers and Geosciences*, 46(3), 272–283.
- Agterberg, F. P. (2007). Mixtures of multiplicative cascade models in geochemistry. *Nonlinear Processes in Geophysics*, 14(3), 201–209.
- Agterberg, F. P., & Bonham-Carter, G. F. (2005). Measuring the performance of mineral-potential maps. *Natural Resources Research*, 14(1), 1–17.
- Aitchison, J. (1982). The statistical analysis of compositional data. *Technometrics*, 30(1), 120–121.
- Aitchison, J., Barcelóvidal, C., Martín-Fernández, J. A., & Pawlowsky-Glahn, V. (2000). Logratio analysis and compositional distance. *Mathematical Geology*, 32(3), 271–275.
- Bonham-Carter, G. F. (1994). Geographic information systems for geoscientists-modeling with GIS. *Computer Methods in the Geosciences*, 13, 398.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123–140.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Breiman, L. (2017). *Classification and regression trees*. New York: Routledge.
- Breslow, N. E., & Cain, K. C. (1988). Logistic regression for two-stage case-control data. *Biometrika*, 75(1), 11–20.
- Brown, W. M., Gedeon, T. D., Groves, D. I., & Barnes, R. G. (2000). Artificial neural networks: a new method for mineral prospectivity mapping. *Australian Journal of Earth Sciences*, 47(4), 757–770.
- Bucciantini, A. (2013). Is compositional data analysis a way to see beyond the illusion? *Computers and Geosciences*, 50(1), 165–173.

- Carranza, E. J. M. (2008). *Geochemical anomaly and mineral prospectivity mapping in GIS* (Vol. 11). Amsterdam: Elsevier.
- Carranza, E. J. M. (2009). Objective selection of suitable unit cell size in data-driven modeling of mineral prospectivity. *Computers and Geosciences*, 35(10), 2032–2046.
- Carranza, E. J. M. (2010). Improved wildcat modelling of mineral prospectivity. *Resource Geology*, 60(2), 129–149.
- Carranza, E. J. M. (2011). Analysis and mapping of geochemical anomalies using logratio-transformed stream sediment data with censored values. *Journal of Geochemical Exploration*, 110(2), 167–185.
- Carranza, E. J. M., & Laborte, A. G. (2015a). Data-driven predictive mapping of gold prospectivity, Baguio district, Philippines: Application of Random Forests algorithm. *Ore Geology Reviews*, 71, 777–787.
- Carranza, E. J. M., & Laborte, A. G. (2015b). Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines). *Computers and Geosciences*, 74, 60–70.
- Carranza, E. J. M., & Laborte, A. G. (2016). Data-driven predictive modeling of mineral prospectivity using random forests: a case study in Catanduanes Island (Philippines). *Natural Resources Research*, 25(1), 35–50.
- Chen, Y. J., & Santosh, M. (2014). Triassic tectonics and mineral systems in the Qinling Orogen, central China. *Geological Journal*, 49(4–5), 338–358.
- Chen, Y. J., Zhang, J., Zhang, F. X., Pirajno, F., & Li, C. (2004). Carlin and Carlin-like gold deposits in Western Qinling Mountains and their metallogenic time, tectonic setting and model. *Geological Review*, 50(2), 134–152.
- Cheng, Q. (1999). Spatial and scaling modelling for geochemical anomaly separation. *Journal of Geochemical Exploration*, 65(3), 175–194.
- Cheng, Q. (2000). *GeoDAS Phase I: User's guide and exercise manual* (p. 298). York University, Toronto, (**Unpublished notes**).
- Cheng, Q. (2007). Mapping singularities with stream sediment geochemical data for prediction of undiscovered mineral deposits in Gejiu, Yunnan Province, China. *Ore Geology Reviews*, 32(1–2), 314–324.
- Chung, C. J. F., & Fabbri, A. G. (1999). Probabilistic prediction models for landslide hazard mapping. *Photogrammetric Engineering and Remote Sensing*, 65(12), 1389–1399.
- Cox, D. P., & Singer, D. A. (1986). *Mineral deposit models* (Vol. 1693). Washington: US Government Printing Office.
- Cutler, D. R., Edwards, T. C., Jr., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., et al. (2007). Random forests for classification in ecology. *Ecology*, 88(11), 2783–2792.
- Dong, Y., Zhang, G., Neubauer, F., Liu, X., Genser, J., & Hauzenberger, C. (2011). Tectonic evolution of the Qinling orogen, China: Review and synthesis. *Journal of Asian Earth Sciences*, 41(3), 213–237.
- Egozcue, J. J., Pawłowsky-Glahn, V., Mateu-Figueras, G., & Barcelo-Vidal, C. (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology*, 35(3), 279–300.
- Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4), 802–813.
- Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, 17(1), 43–57.
- Filzmoser, P., & Hron, K. (2008). Outlier detection for compositional data using robust methods. *Mathematical Geosciences*, 40(3), 233–248.
- Filzmoser, P., Hron, K., & Reimann, C. (2009). Principal component analysis for compositional data with outliers. *Environmetrics: The Official Journal of the International Environmetrics Society*, 20(6), 621–632.
- Filzmoser, P., Hron, K., & Reimann, C. (2010). The bivariate statistical analysis of environmental (compositional) data. *Science of the Total Environment*, 408(19), 4230–4238.
- Gao, Y., Zhang, Z., Xiong, Y., & Zuo, R. (2016). Mapping mineral prospectivity for Cu polymetallic mineralization in southwest Fujian Province, China. *Ore Geology Reviews*, 75, 16–28.
- Goldfarb, R., Baker, T., Dube, B., Groves, D. I., Hart, C. J., & Gosselin, P. (2005). Distribution, character and genesis of gold deposits in metamorphic terranes. *Society of Economic Geologists*.
- Grömping, U. (2009). Variable importance assessment in regression: Linear regression versus random forest. *The American Statistician*, 63(4), 308–319.
- Guo, X., Yan, Z., Wang, Z., Wang, T., Hou, K., Fu, C., et al. (2012). Middle Triassic arc magmatism along the northeastern margin of the Tibet: U-Pb and Lu-Hf zircon characterization of the Gangcha complex in the West Qinling terrane, central China. *Journal of the Geological Society*, 169(3), 327–336.
- Hariharan, S., Tirodkar, S., Porwal, A., Bhattacharya, A., & Joly, A. (2017). Random forest-based prospectivity modelling of greenfield terrains using sparse deposit data: An example from the Tanami region, Western Australia. *Natural Resources Research*, 26(4), 489–507.
- Hronsky, J. M., & Groves, D. I. (2008). Science of targeting: definition, strategies, targeting and performance measurement. *Australian Journal of Earth Sciences*, 55(1), 3–12.
- Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, 106(4), 620.
- Joly, A., Porwal, A., & McCuaig, T. C. (2012). Exploration targeting for orogenic gold deposits in the Granites-Tanami Orogen: Mineral system analysis, targeting model and prospectivity analysis. *Ore Geology Reviews*, 48, 349–383.
- King, G., & Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis*, 9(2), 137–163.
- Knox-Robinson, C. M., & Wyborn, L. A. I. (1997). Towards a holistic exploration strategy: Using geographic information systems as a tool to enhance exploration. *Australian Journal of Earth Sciences*, 44(4), 453–463.
- Kröner, A., Zhang, G. W., & Sun, Y. (1993). Granulites in the Tongbai area, Qinling belt, China: Geochemistry, petrology, single zircon geochronology, and implications for the tectonic evolution of eastern Asia. *Tectonics*, 12(1), 245–255.
- Lee, S., & Dan, N. T. (2005). Probabilistic landslide susceptibility mapping in the Lai Chau province of Vietnam: Focus on the relationship between tectonic fractures and landslides. *Environmental Geology*, 48(6), 778–787.
- Lee, S., & Pradhan, B. (2007). Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. *Landslides*, 4(1), 33–41.
- Lerch, M. F., Xue, F., Kröner, A., Zhang, G. W., & Tod, W. (1995). A middle Silurian-Early Devonian magmatic arc in the Qinling Mountains of central China. *The Journal of Geology*, 103(4), 437–449.
- Lewkowsky, C., Porwal, A., & González-Álvarez, I. (2010). Genetic programming applied to base-metal prospectivity mapping in the Aravalli Province, India. In *EGU general assembly conference abstracts* (Vol. 12, p. 523).
- Li, X. W., Mo, X. X., Yu, X. H., Ding, Y., Huang, X. F., Wei, P., et al. (2013). Petrology and geochemistry of the early Mesozoic pyroxene andesites in the Maixiu Area, West Qinling, China: Products of subduction or syn-collision? *Lithos*, 172, 158–174.
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2(3), 18–22.
- Liu, X. L. (2011). A study on the geology feature and minerals exploration sign of structure-metamorphism-type gold-bearing

- ing deposit in Gansu-Xiahe-Hezuo region. *Gansu Metallurgy*, 2, 33.
- Liu, Y., Zhou, K., & Xia, Q. (2018). A MaxEnt model for mineral prospectivity mapping. *Natural Resources Research*, 27(3), 299–313.
- Liu, Y., Zhou, K., Zhang, N., & Wang, J. (2017). Maximum entropy modeling for orogenic gold prospectivity mapping in the Tangbale-Hatu belt, western Junggar, China. *Ore Geology Reviews*, 100, 133–147.
- Mao, J., Qiu, Y., Goldfarb, R. J., Zhang, Z., Garwin, S., & Fengshou, R. (2002). Geology, distribution, and classification of gold deposits in the western Qinling belt, central China. *Mineralium Deposita*, 37(3–4), 352–377.
- McCuaig, T. C., Beresford, S., & Hronsky, J. (2010). Translating the mineral systems approach into an effective exploration targeting system. *Ore Geology Reviews*, 38(3), 128–138.
- McCuaig, T. C., & Hronsky, J. M. (2014). The mineral system concept: the key to exploration targeting. *Society of Economic Geologists Special Publication*, 18, 153–175.
- McCuaig, T. C., Kreuzer, O. P., & Brown, W. M. (2007). Fooling ourselves—dealing with model uncertainty in a mineral systems approach to exploration. In *Mineral exploration and research: Digging deeper: Proceedings of the 9th biennial SGA Meeting*, Dublin (pp. 1435–1438).
- McCune, B., Grace, J. B., & Urban, D. L. (2002). Analysis of ecological communities. *Gleneden Beach* (p. 302).
- McKay, G., & Harris, J. R. (2016). Comparison of the data-driven Random Forests model and a knowledge-driven method for mineral prospectivity mapping: a case study for gold deposits around the Huritz Group and Nueltin Suite, Nunavut, Canada. *Natural Resources Research*, 25(2), 125–143.
- Meng, Q. R., & Zhang, G. W. (2000). Geologic framework and tectonic evolution of the Qinling orogen, central China. *Tectonophysics*, 323(3–4), 183–196.
- Oh, H. J., & Lee, S. (2010). Application of artificial neural network for gold–silver deposits potential mapping: a case study of Korea. *Natural Resources Research*, 19(2), 103–124.
- Otero, N., Tolosana-Delgado, R., Soler, A., Pawlowsky-Glahn, V., & Canals, A. (2005). Relative versus absolute statistical analysis of compositions: A comparative study of surface waters of a Mediterranean river. *Water Research*, 39(7), 1404–1414.
- Pawlowsky-Glahn, V., & Egozcue, J. J. (2006). Compositional data and their analysis: An introduction. *Geological Society, London, Special Publications*, 264(1), 1–10.
- Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3–4), 231–259.
- Phillips, S. J., & Dudík, M. (2008). Modeling of species distributions with MaxEnt: New extensions and a comprehensive evaluation. *Ecography*, 31(2), 161–175.
- Porwal, A., & Carranza, E. J. M. (2015). Introduction to the special issue: GIS-based mineral potential modelling and geological data analyses for mineral exploration. *Ore Geology Reviews*, 71, 477–483.
- Porwal, A., Carranza, E. J. M., & Hale, M. (2003). Artificial neural networks for mineral-potential mapping: A case study from Aravalli Province, Western India. *Natural Resources Research*, 12(3), 155–171.
- Porwal, A. K., & Kreuzer, O. P. (2010). *Introduction to the special issue: Mineral prospectivity analysis and quantitative resource estimation*.
- Qi, J. H., Li, Z. C., & Wang, X. W. (2013). *Report on the prospect survey of mines in Hezuo–Meiwu district. Gansu: Third institute geological and mineral exploration of Gansu provincial bureau of geology and mineral resources*, Gansu, China, (p. 155). (in Chinese).
- R Development Core Team. (2008). *The R project for statistical computing*. <http://www.R-project.org>. Accessed 28 Oct 2018.
- Rodriguez-Galiano, V. F., Chica-Olmo, M., & Chica-Rivas, M. (2014). Predictive modelling of gold potential with the integration of multisource information based on random forest: A case study on the Rodalquilar area, Southern Spain. *International Journal of Geographical Information Science*, 28(7), 1336–1354.
- Schill, W., Jöckel, K. H., Drescher, K., & Timm, J. (1993). Logistic analysis in case-control studies under validation sampling. *Biometrika*, 80(2), 339–352.
- Sillitoe, R. H. (2004). *Musings on future exploration targets and strategies in the Andes. Andean metallogeny: New discoveries, concepts, and updates. Society of Economic Geologists, Boulder* (pp. 1–14).
- Sillitoe, R. H., & Thompson, J. F. H. (2006). Changes in mineral exploration practice: Consequences for discovery. *Special Publication-Society of Economic Geologists*, 12, 193.
- Simmons, S. F., White, N. C., & John, D. A. (2005). Geological characteristics of epithermal precious and base metal deposits. *Economic Geology*, 100, 485–522.
- Sui, J. X., Li, J. W., Wen, G., & Jin, X. Y. (2017). The Dewulu reduced Au–Cu skarn deposit in the Xiahe-Hezuo district, West Qinling orogen, China: Implications for an intrusion-related gold system. *Ore Geology Reviews*, 80, 1230–1244.
- Wang, W., Zhao, J., Cheng, Q., & Carranza, E. J. M. (2015). GIS-based mineral potential modeling by advanced spatial analytical methods in the southeastern Yunnan mineral district, China. *Ore Geology Reviews*, 71, 735–748.
- Wyborn, L. A. I., Heinrich, C. A., & Jaques, A. L. (1994). Australian Proterozoic mineral systems: Essential ingredients and mappable criteria. In *Proceedings of the AusIMM annual conference* (vol. 1994, pp. 109–115), Darwin, AusIMM.
- Yousefi, M., & Nykänen, V. (2017). *Introduction to the special issue: GIS-based mineral potential targeting*.
- Zhang, G., Dong, Y., Lai, S., Guo, A. L., Meng, Q. R., Liu, S. F., et al. (2004). Mianlue tectonic zone and Mianlue suture zone on southern margin of Qinling–Dabie orogenic belt. *Science in China Series D Earth Sciences-English Edition*, 47(4), 300–316.
- Zhang, C. L., Wang, T., & Wang, X. X. (2008). Origin and tectonic setting of the Early Mesozoic granitoids in Qinling orogenic belt. *Geological Journal of China Universities*, 14(3), 304–316.
- Zhang, Z., Zuo, R., & Xiong, Y. (2016). A comparative study of fuzzy weights of evidence and random forests for mapping mineral prospectivity for skarn-type Fe deposits in the southwestern Fujian metallogenic belt, China. *Science China Earth Sciences*, 59, 556–572.
- Zuo, R. (2014). Identification of geochemical anomalies associated with mineralization in the Fanshan district, Fujian, China. *Journal of Geochemical Exploration*, 139, 170–176.
- Zuo, R., & Carranza, E. J. M. (2011). Support vector machine: A tool for mapping mineral prospectivity. *Computers and Geosciences*, 37(12), 1967–1975.