Natural Resources Research, Vol. 25, No. 2, June 2016 (© 2015)
DOI: 10.1007/s11053-015-9274-z

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

G. McKay[1] and J. R. Harris[2,3]

Received 2 April 2015; accepted 15 June 2015
Published online: 15 July 2015

This paper outlines the process taken to create two separate gold prospectivity maps. The first was created using a combination of several knowledge-driven (KD) techniques. The second was created using a relatively new classification method called random forests (RF). The purpose of this study was to examine the results of the RF technique and to compare the results to that of the KD model. The datasets used for the creation of evidence maps for the gold prospectivity mapping include a comprehensive lake sediment geochemical dataset, interpreted geological structures (form lines), mapped and interpreted faults, lithology, topographic features (lakes), and known Au occurrences. The RF method performed well in that the gold prospectivity map created was a better predictor of the known Au occurrences than the KD gold prospectivity map. This was further validated by a fivefold repetition using a subset of the input training areas. Several advantages to the use of RF include (1) the ability to take both continuous and/or categorical data as variable inputs, (2) an internal, unbiased estimation of the mapping error (out-of-bag error) removing the need for a cross-validation of the final outputs to determine accuracy, and (3) the estimation of importance of each input variable. Efficiency of prediction curves illustrates that the RF method performs better than the KD method. The success rate is significantly higher for the RF method than for the KD method.

KEY WORDS: Mineral prospectivity, Random forest, Classification, Gold.

INTRODUCTION

Exploration for gold deposits has been a major focus for mineral exploration companies across Canada for much of the last century. There are four main phases involved in mineral exploration: (1) area selection, (2) target generation, (3) resource evaluation, and (4) reserve definition. Traditionally, area selection in mineral exploration starts by outlining geological units that are suitable for specific types of mineral deposits based on knowledge of the near surface geological environment, where geological processes are favorable for mineral deposition. Next, target generation starts with grassroots prospecting and/or geochemical sampling of various media (till, soil, rock, lakes, streams, etc.) over those favorable geological units/belts, and is designed to discover anomalous geochemical concentrations of various elements and preferential structures within the rock units. If anything of interest is discovered by prospecting, a resource evaluation may be carried out which can include systematic drilling programs and/or geophysical surveys in an attempt to locate

[1] 927-1695 Playfair Drive, Ottawa, ON K1H 8J6, USA.
[2] Geological Survey of Canada, 601 Booth St, Ottawa, ON K1A 0E8, USA.
[3] To whom correspondence should be addressed; e-mail: harris@nrcan.gc.ca

1520-7439/16/0600-0125/0 © 2015 International Association for Mathematical Geosciences

and delineate high-grade mineral deposits as well as to determine a rough estimate of the grade and size of the deposits. If these deposits are large enough and with high enough grade to be mined economically, reserve definition can begin. This involves precise definition drilling and sampling to determine and classify areas of the deposit as ore or waste rock based on grade, density, and location.

For various reasons, economic mineral deposits are becoming increasingly difficult to locate. As incredibly large volumes of geoscience data are being collected by industry and government, new methods and tools for archiving, managing, manipulating, integrating, and visualizing these data are being created. Geographic Information Systems (GIS) are an excellent tool for all stages of gold exploration. The creation of *mineral prospectivity maps* using GIS and various geoscience datasets can be a very cost- and time-effective way of executing the first two phases of mineral exploration.

The term mineral prospectivity is defined as the probability or likelihood that mineral deposits of the type sought can be found in a specific area. The process of creating a mineral prospectivity (mineral potential) map for a study area includes defining an exploration model, preparing evidence maps, outlining the modeling method, creating the prospectivity map, and evaluating the map. A good model for mineral prospectivity makes two assumptions. Firstly, an area is deemed highly prospective if it contains or is characterized by the same attributes that are found in conjunction with known mineral deposits or occurrences of the type sought and secondly, the attributes used to determine the mineral prospectivity of an area are present, independent of each other, and varying across the study area.

There are two main approaches for creating mineral prospectivity maps, each with several methods: (1) data-driven models and (2) knowledge-driven models for which reviews can be found in Bonham-Carter (1994), Wright and Bonham-Carter (1996), and Carranza (2009a, b). The preference of one method over the other is generally decided by the amount of available geoscience data and whether existing mineral occurrences are present for the area under study. Each method has advantages and disadvantages; therefore, the use of one particular method is usually decided based on the specific needs of the user. Data-driven approaches use known locations of mineral prospects, occurrences, or deposits as controls and use spatial statistical methods to determine the weight of importance for each data layer. This approach has the advantage of needing very little geologist input, and potentially biased human input is kept at a minimum. Methods such as logistic regression (Chung and Agterberg 1980), weights of evidence (WofE) (Bonham-Carter 1994), decision tree analysis (Reddy and Bonham-Carter 1991), neural networks (Brodeur et al. 1992; Singer and Kouda 1996; Harris and Pan 1999; Brown et al. 2000; Porwal et al. 2003), and support vector machine (Zuo and Carranza 2011; Abedi et al. 2012) are examples of data-driven approaches.

Knowledge Driven (KD) approaches do not require any data on mineral deposits or occurrences within the study area. They rely on the inputs of a geologist with reasonable knowledge and experience to determine the weight of importance for each evidence map. This approach may bring up human bias but has the advantage of using the knowledge of the geologist on all aspects of the model, and not needing a dataset of known mineral prospects, occurrences, or deposits as controls. Examples of KD approaches include Boolean logic, index overlays (Harris 1989), analytical hierarchy process (An et al. 1992; Harris et al. 2001), fuzzy logic (An et al. 1991), and evidential belief (An et al. 1994a, b; Carranza et al. 2005; Carranza 2014).

This paper outlines the process taken to create two separate Au prospectivity maps. The first was created using a combination of different KD techniques. The second was created using a random forests (RF) classification technique. The purpose of this study was to examine the results of the RF technique and to compare the results to that of the KD model.

RF is a data-driven method for classification created by Breiman (2001). It creates a large ensemble of decision trees created by randomly sampling a small portion of the input evidence maps (variables) and creates a bootstrap random sampling of 2/3 of the available training data for classification and remaining 1/3 for validation. The final classification map is created through a majority vote from all trees on a pixel-to-pixel basis. While this is not a new method of producing mineral potential maps (Rodriguez-Galiano et al. 2014; Carranza and Laborte 2015; Harris et al. 2015), the strength of RF, in this case, comes from its ability to also create a probability (confidence) map for the classification of each pixel. It is this probability map, as opposed to the classified map, that is used as a measure of prospectivity (Harris et al. 2015). The KD technique we employ involves a combination of basic index overlay and fuzzy-weighted index overlay approaches.

OBJECTIVES

This paper compares the results of a relatively new classification algorithm, RF, to a commonly used KD, method for the production of gold prospectivity maps of a large territory in southern Nunavut, Canada. Interpolated lake sediment, geologic structural, and lithological data were used as evidence maps (predictors) for both methods. Known gold-bearing mineral occurrences were used to train the RF algorithm to produce a prospectivity map. Due to the relatively small number of gold prospects in the area (i.e., 16), all of the prospects were used to train the RF classification algorithm. However, to study the effect of the small number of Au occurrences used for training the RF classifier, we use a fivefold repetition of the RF classifier using a random selection of eight Au occurrences and eight non-Au occurrences for each repetition. Thus, five RF classification and probability maps were produced each with an associated RF out-of-bag (*oob*) and classification error value. The KD approach used the same evidence maps, some of which were combined, but the known gold prospects were not used to derive the Au prospectivity map. The RF Au prospectivity (probability) map and the KD prospectivity maps were evaluated to see how well they predicted the known Au occurrences using efficiency of prediction curves (Chung and Fabbri 2003; Agterberg and Bonham-Carter 2005; Harris et al. 2006).

STUDY AREA

We undertake our experiments in the southern Hearne geologic province (Fig. 1). The study area is bounded by the latitudes 60 and 61 degrees, and longitudes −102 and −96 degrees. The study area has a perimeter of 907.2 km and an area of 36,723.84 km^2.

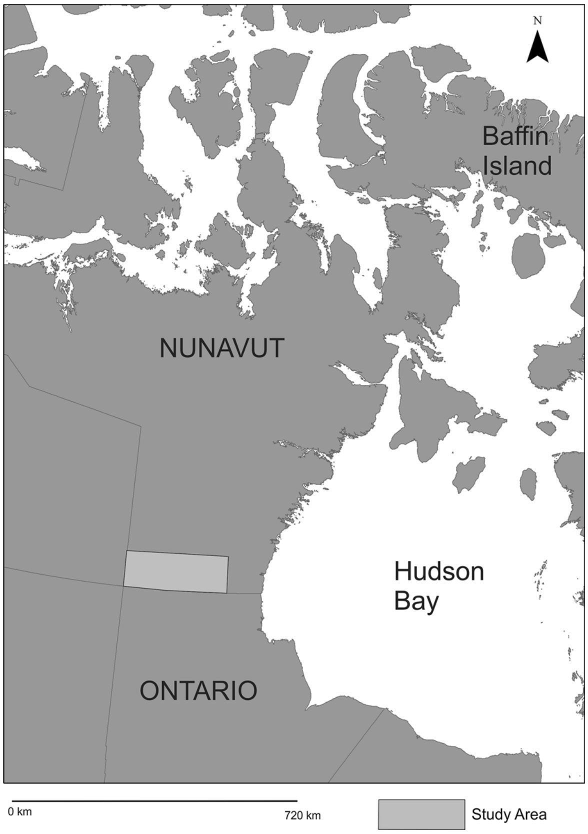

Figure 1. Location of study area—Hearn Geologic Province, Nunavut, Canada.

The geology of the Hearne study area (Fig. 2) derived from Paul et al. (2002) is dominated by Archean quartz-feldspathic granitoid rocks, ranging from granites to tonalites with septa of supracrustal rocks which form part of the Ennadai belt (Hanmer et al. 2004). These rocks are overlain by the rocks of the Hurwitz Group (Aspler et al. 2001). Paleoproterozoic plutons comprise the Hudson granitoid suite (ca. 1.83 Ga) and the younger Nueltin granite intrusive suite (ca. 1.75 Ga) (Van Breeman et al. 2005; Scott et al. 2012). Hanmer et al. (2004) provide a detailed description of the Hearne domain.

Economic mineral prospects within the Nueltin Suite include uraninite and uranium-bearing silicate minerals and secondary REE carbonates (Charbonneau and Swettenham 1986; Scott 2012). Scott et al. 2012 have determined that the mineral occurrences are hosted by pegmatitic phases within Nueltin granite. Gold occurrences often associated

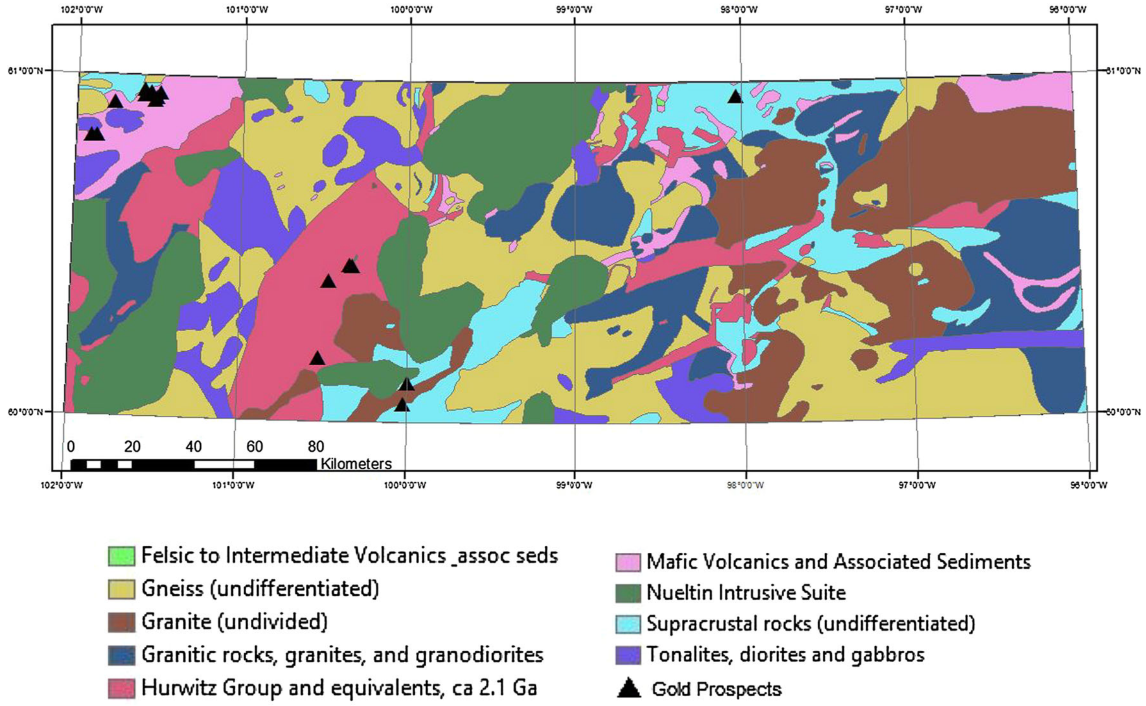
102°0'0"W
101°0'0"W
100°0'0"W
99°0'0"W
98°0'0"W
97°0'0"W
96°0'0"W
61°0'0"N
60°0'0"N
0
20
40
60
80
Kilometers
Felsic to Intermediate Volcanics _assoc seds
Gneiss (undifferentiated)
Granite (undivided)
Granitic rocks, granites, and granodiorites
Hurwitz Group and equivalents, ca 2.1 Ga
Mafic Volcanics and Associated Sediments
Nueltin Intrusive Suite
Supracrustal rocks (undifferentiated)
Tonalites, diorites and gabbros
Gold Prospects

Figure 2. Generalized geology of the study area (from Paul et al. 2002).

with various base metals also occur within the study area and are the focus of our study.

DATA

The data used in the creation of the evidence maps for the Au mineral prospectivity maps included a comprehensive lake sediment geochemical dataset, interpreted geological structures (form lines), mapped and interpreted faults, lithology, topographic features (lakes), and known Au occurrences.

The lake sediment geochemical data used for the study were published in GSC Open File 6986 (McCurdy et al. 2012). This dataset comprises new analytical data for 60 elements from the reanalysis of lake sediment samples collected from 2377 sites within the study area during the Federal Uranium Reconnaissance Program. Inductively Coupled Plasma Mass Spectrometry (ICP-MS) was used to analyze 53 elements, whereas Atomic Absorption Spectroscopy (AAS) was used to analyze Zn, Cu, Pb, Ni, Co, Ag, Mn, Fe, and Cd. More details regarding the survey, analysis, and quality control measures can be found in McCurdy et al. (2012). The lake sediment data used can give an indication of potential gold mineralization based on spatially distributed high concentrations and are commonly used in mineral exploration programs.

The interpreted structures and the mapped and interpreted faults used in the study were published in GSC Open File 7649 (Behnia et al. 2013). This comprehensive structural dataset includes compiled and newly interpreted form lines, faults, dykes, fractures, and lineaments for Canada's North. Faults and form lines, especially those that mark potential contact zones between different lithologies, act as conduits for mineralized fluids and are commonly used in exploration projects.

The lithological data were published in the GSC Open File 7649 in map 2159A (Behnia et al. 2013). The lithology polygons were interpreted using data from GSC Open File 4236 (Paul et al. 2002). Lithology is obviously important in exploration as Au, depending on the tectonic regime, commonly occurs in specific rock types (e.g., sediments) that are associated with faults and shear zones.

The topographic data were published in the GSC Open File 7649 in map 2159A (Behnia et al. 2013). This dataset was used to create a lake mask, as described below. The known Au occurrences were collected and archived by the Canada-Nunavut Geoscience Office and published in the GSC Open File 7649 (Behnia et al. 2013).

DATA PROCESSING

The RF classification was completed using the EnMap add-on for ENVI 5.0, and the rest of the processing was done using ArcMap 10.2. The geoprocessing environments were set up in ArcMap such that all of the evidence maps created were clipped to the extents of the study area defined above (Fig. 1). Additionally, each evidence raster map was created with a cell size of 400 meters. This cell size was chosen because it creates a sufficient number of cells within the study area (183,599) and it nearly matches the size of a single mining claim unit in Nunavut (1500 ft., 457.2 m).

A fuzzy evidence map is created for each dataset used in the analysis. Fuzzification is defined by Carranza (2009a, b) as the process of converting individual sets of spatial evidence into fuzzy sets. A fuzzy set is defined as a collection of objects whose grades of membership in that set range from complete (=1) to incomplete (=0). The grade of membership of a fuzzy set is defined using one of several mathematical functions, the most common of which are linear, small, large, MS small, MS large, and near (Fig. 3). Each evidence map in this paper was fuzzified using either a linear function (where there is a linear relationship between the input and output values, e.g., geochemical data, faults) or a near function (where the maximum value of an output is at a specific value of the input, and output values taper-off away from that input value, e.g., form lines, geology). The lake mask was binary in format. In this case, the evidence maps were fuzzified (based on geologic and exploration knowledge) in order to normalize each evidence map so as not to assume a higher importance of one evidence map over another.

Table 1 presents the evidence maps and masks used for both the RF and KD Au prospectivity maps. The evidence maps comprised 10 lake sediment interpolated, and fuzzified maps, two structural maps (fuzzy fault and form lines) and a fuzzy geology map, creating 13 evidence maps for the RF and KD prospectivity mapping. However, as described below, we combined the geochemical, structural, and lithological data into three evidence maps to produce the KD prospectivity map.

Lake Sediment Geochemical Data

The density of the sediment sample points was sufficient to warrant interpolation into continuous surface maps (Grunsky et al. 2014). Ten elements were used to create evidence maps for the models; Ag, As, Au, Co, Cu, Fe, Hg, Ni, Pb, and Zn. The unit used for the measurement of the element in the lake sediment samples was either percentage (Fe) or ppm (all other elements). The raw elements listed were interpolated using an inverse distance-weighted interpolation algorithm (IDW) available in the ArcGIS Geostatistical Analyst extension. The 'Power' setting for the IDW algorithm, which controls the influence of known values on the interpolated values based on their distance from the predicted point, was optimized to minimize the RMS error for each interpolated element. The 10 interpolated and fuzzy geochemical evidence maps are presented in Figure 4.

In a previous study (Harris et al. 2015), the same geochemical dataset was used; however, the geochemical data were corrected for closure (Aitchison 1986) and censored values. They found that when RF was applied to the corrected or raw geochemical data, there was little difference in the output classification results. Thus, we used the raw data in this study.

Interpreted Structures and Faults Data

The interpreted structures dataset includes linear features comprising dykes, faults, fold axis, form

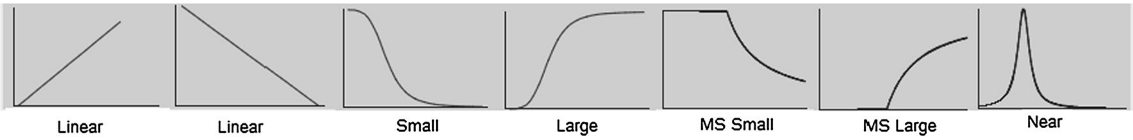

Figure 3. Different look-up-tables (algorithms) for fuzzification of evidence maps.

Table 1. Summary of evidence maps

Evidence map	Description
Geochemical data	
Fuzzy Ag	Fuzzy evidence maps where the value of each cell is based on the weighted average of lake bed geochemical samples in the immediate area of the cell
Fuzzy As	
Fuzzy Au	
Fuzzy Co	
Fuzzy Cu	
Fuzzy Fe	
Fuzzy Hg	
Fuzzy Ni	
Fuzzy Pb	
Fuzzy Zn	
Structure	
Fuzzy faults	A fuzzy evidence map, where cells close to a mapped fault, based on Table 2, are assigned a higher value than values further from a mapped fault
Fuzzy form lines	A fuzzy evidence map, where cells close to a mapped form line, based on Table 2, are assigned a higher value than values further from a mapped form line
Fuzzy geology	A fuzzy evidence map where cells within or near a potential gold deposit hosting geology are assigned a high value, and all other cells are assigned a value of 0, based on Table 3. Cells around the edges of geological units are assigned values of the highest values
Masks	
Lakes mask	A mask layer where any cell, more than 50% covered by a lake larger than 160,000 km^2 is assigned a value of 0, and all other cells are assigned a value of 1

lines, lineaments, and shear zones interpreted from Landsat and airborne magnetic survey data derived from Behnia et al. (2013). The form lines and faults were selected from this dataset to create form line and fault datasets. Fuzzy evidence maps were created from the form lines and faults using a 'Euclidean Distance' tool to measure the distance from the lines and reclassifying the raster cells as summarized in Table 2. The fuzzy evidence maps for faults and form lines are presented in Figure 5a and b, respectively.

Lithology Data

The lithological units (Fig. 2) are outlined in Paul et al. (2002), which also list the Eon, Era, setting, lithology and metamorphism information. From the dataset, three geological units were deemed 'favorable' for gold exploration based on the lithology and common understanding of the genesis of gold deposits. The favorable lithology types (Fig. 2) are as follows: (1) Hurwitz Group and equivalents, (2) mafic volcanics and associated sediments, and (3) supracrustal rocks (undifferentiated). A fuzzy evidence map was created from the 'favorable' geology dataset using a 'Euclidean Distance' tool to measure from the outline of the geology, and reclassifying the raster cells using a 'near' fuzzy mathematical function according to Table 3. Within favorable lithologies the value of each cell never drops below 0.8. Outside of the favorable geologies the cell values decreases with increasing distance from the favorable geologic units until a value of 0 is reached at a distance of 800 m and greater. The highest raster values were assigned close to the boundaries of the geological units where the rock is typically more fractured allowing for the intrusion of epithermal gold deposits. The graded boundaries around the geological units also partially account for the inherent estimation of geological boundaries from field surveys and mapping. The fuzzy evidence map for lithology is presented in Figure 5c.

Topographic Data

A 'Large Lakes' evidence map was created and used as a mask for both the RF and KD prospectivity maps. It was determined that mineral exploration and mining would be too difficult logistically under large lakes for several reasons including access, exploration viability, and permitting issues. To create a fuzzy lakes evidence raster, the first step was to select large lakes (lakes greater than 1 cell size, >160,000 m^2). A raster was created from this selection where raster cells covered more than 50%

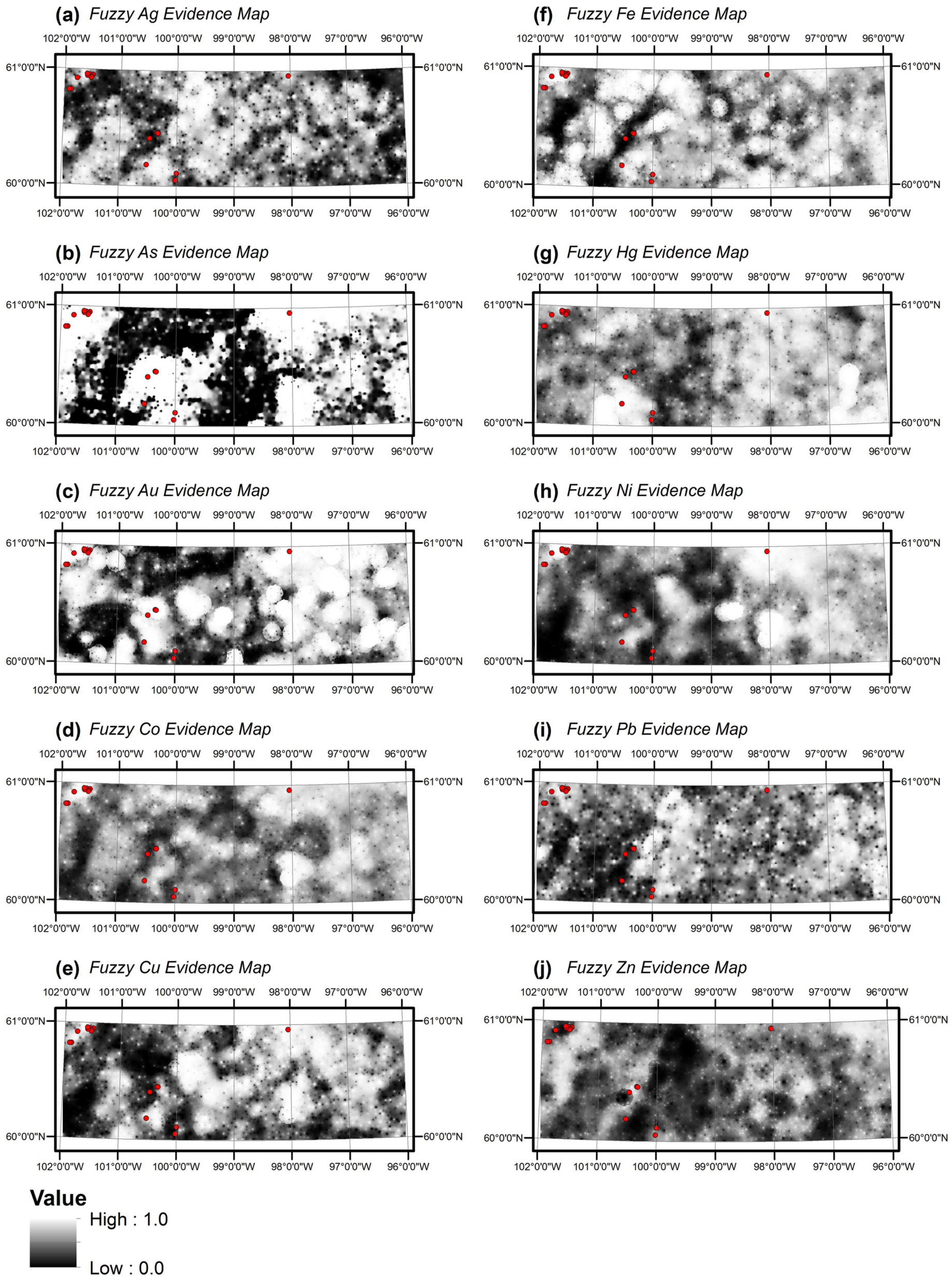
(a) Fuzzy Ag Evidence Map
(b) Fuzzy As Evidence Map
(c) Fuzzy Au Evidence Map
(d) Fuzzy Co Evidence Map
(e) Fuzzy Cu Evidence Map
(f) Fuzzy Fe Evidence Map
(g) Fuzzy Hg Evidence Map
(h) Fuzzy Ni Evidence Map
(i) Fuzzy Pb Evidence Map
(j) Fuzzy Zn Evidence Map
102°0'0"W
101°0'0"W
100°0'0"W
99°0'0"W
98°0'0"W
97°0'0"W
96°0'0"W
61°0'0"N
60°0'0"N
Value
High : 1.0
Low : 0.0

Figure 4. Geochemical evidence maps—data were interpolated and fuzzified—see text for description.

Table 2. Fuzzification values of linear spatial datasets into raster evidence maps

Form lines evidence map		Faults evidence map	
Distance (m)	Raster value	Distance (m)	Raster value
0–250	0.80	0–400	1.0
250–500	0.90	400–800	0.95
500–750	1.00	800–1200	0.90
750–1000	0.90	1200–1600	0.85
1000–1250	0.80	1600–2000	0.80
1250–1500	0.70	2000–2400	0.70
1500–1750	0.60	2400–2800	0.60
1750–2000	0.50	2800–3200	0.50
2000–2250	0.40	3200–3600	0.40
2250–2500	0.35	3600–4000	0.30
2500–2750	0.30	>4000	0.20
2750–3000	0.25		
>3000	0.20		

by lakes were given a value of 0, and the remaining were assigned a value of 1. A final correction was made to the raster to ensure that raster cells covering a point representing a 'known gold deposit' were assigned a value of 1 regardless of whether the lake mask covered more than 50% of the cell. Two of the known gold deposits were covered by raster values of 0 because they were discovered on, or near a rocky shoreline. The fuzzy evidence map for large water bodies is presented in Figure 5d.

Mineral Deposits

The mineral deposit data comprised 399 known mineral deposits within Nunavut. Of this dataset, 31 deposits are within the study area, and 16 of these are Au occurrences some of which also contain base metals (Fig. 2). A list of the known mineral deposits within the study area containing gold is presented in Table 4, and Figure 2 shows their spatial distribution.

The RF model was designed to classify areas into prospective and non-prospective areas so the 16 known Au occurrences (400 m grid cell) were used as training data for the prospective areas, and 16 randomly assigned points, in lithologies not prospective for gold, were used as non-prospective areas. Ideally for a more accurate dataset, and to minimize error, field work should be conducted in order to properly classify these areas as "non-prospective." As mentioned above, a random selection of both the occurrences and non-occurrences was used for the fivefold repletion of RF.

METHODOLOGY

Knowledge-Driven Method

There are many different KD methods for producing a mineral prospectivity map referenced in the introduction section. The method used for the creation of the gold prospectivity map in this paper combined the lake sediment sample data (10 evidence maps—Fig. 4) into a single raster, and the structural and geological data (3 evidence maps—Fig. 5) into another single raster, then the two single rasters were combined. This method provides equal weighting to the lake sediment sample data as a whole, and the structural and geological data as a whole. The KD method implemented in this paper is summarized in Figure 6.

Each of the IDW rasters created from the 10 lake sample elements was normalized between 0 and 1. They were then combined using a weighted sum method (multiplying the values from each raster by a 'weight' value, then adding the results together), to create a final soil sample prospectivity raster ("Weighted Soils Evidence Map"). The weights for each element used in the weighted sum method, as well as some statistics of the output weighted soils evidence map are presented in Table 5.

The "Form Lines Evidence Map," "Faults Evidence Map," and "Geology Evidence Map" were all combined into a single "GeoStruct Evidence Map" using a series of fuzzy overlay methods, which overlay the input rasters and return an output raster (Fig. 6). The form lines and faults were first combined using a 'Fuzzy Or' overlay method which

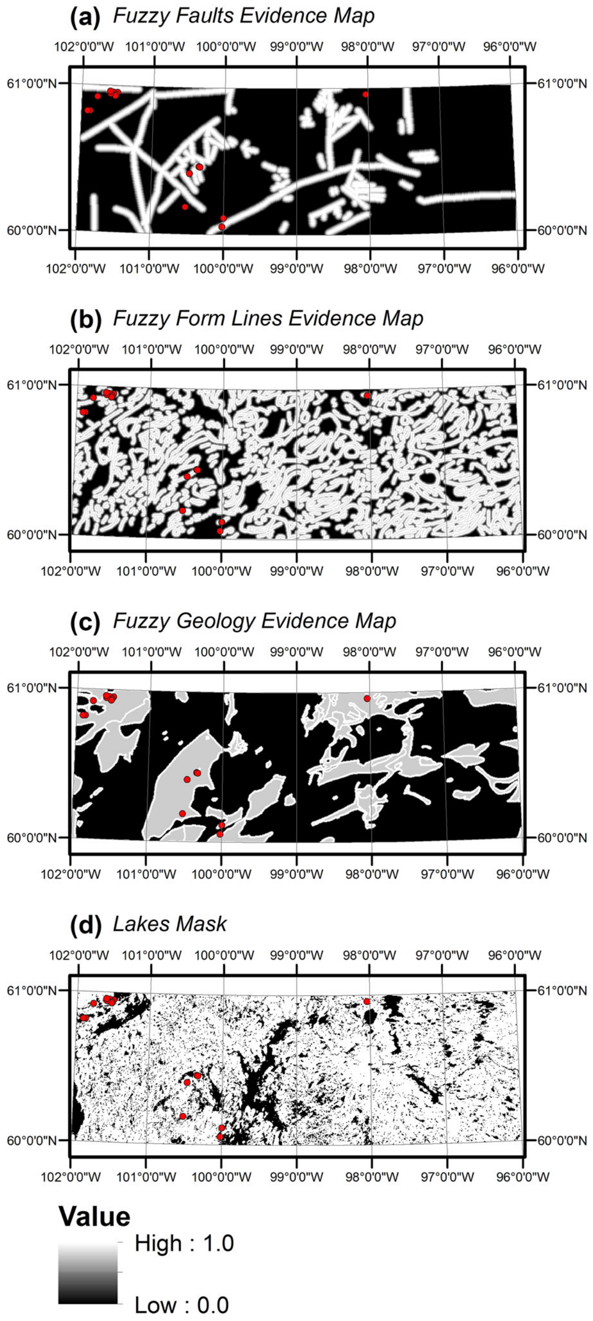

Figure 5. Geologic evidence maps: **a** fuzzified faults, **b** fuzzified form lines, **c** fuzzified lithology, **d** lake mask.

kept only the maximum value from the two (assuming that the proximity to either faults, or form lines would be deemed prospective). That output raster was then combined with the geology using a 'Fuzzy And' overlay method, which kept only the minimum value (assuming that proximity to faults or form lines would only be ideal if it was also spatially associated with ideal geology/lithology, and vice versa).

The "GeoStruct Evidence Map" raster and the "Weighted Soils Evidence Map" raster were then multiplied by each other using a 'Fuzzy Product' overlay method (Fig. 6). The multiplication was done to emphasize the highly prospective areas and to de-emphasize the poorly prospective areas. This raster was then stretched to values between 0 and 100. Finally, a simple mask of lakes over 160,000 m^2 was applied to create the final KD prospectivity map. The mask of lakes was applied on the assumption that gold deposits under large lakes are either nearly impossible to explore for and/or nearly impossible to mine (either for mine structure feasibility or permitting reasons).

Random Forests Classification Algorithm

RF is an ensemble (multiple) decision tree classifier, which does not assume normal distribution of the input data. RF was originally developed by L. Breiman and A. Cutler at the University of California, Berkley (Breiman 2001). Training data/classes (in this case 16 locations of mineral occurrences, and 16 non-prospective locations) are required for this approach, similar to other data-driven approaches.

For each tree (the number of trees is determined by the operator, in this case 1000 trees were chosen), a random selection of the input variables (i.e., geochemical, structural, lithological datasets, in this case the 13 evidence maps) is made. The number of variables selected for each tree is a fraction of the total number of variables; the square root of the number of variables is often used (in this case it was 4). Each tree employs a bagging process (i.e., bootstrap sample), where approximately 2/3 of the training areas (pixels) are used to create a prediction (referred to as in-bag) and 1/3 to validate the accuracy of the prediction (referred to as out-of-bag, or *oob*). This random sampling with replacement of the training dataset is undertaken for every tree. In-bag data are used to create multiple decision trees which are applied to produce independent classifications. At each node of the individual decision tree, the best split is chosen from a random sample of variables. Each tree is grown to the maximum extent with no pruning. Thus, an ensemble of trees (e.g., forest) is created and a voting procedure is employed to assign the majority class from all trees to each pixel in the final prediction map. RF is not sensitive to noise or over fitting and there is no need for cross-validation or a separate test training dataset to get an unbiased estimate of overall classification accuracy as it is tested internally (Rodriguez-Galiano et al. 2014). Additionally, the probability of membership to each class is also generated which can be used to access the uncertainty of the RF classification or the predictive power of the RF classifier for a specific class. This probability map is what we used for our gold prospectivity map as opposed to the actual RF classification, which is a 2-class thematic map showing areas favorable and unfavorable for gold exploration.

Another very useful aspect of RF is that it calculates the importance (predictive power) of each variable in the classification process. This is accomplished by producing Mean Decrease *Accuracy* and Mean Decrease *Gini index* plots (Gislason et al. 2006; Waske and Braun 2009). In this study, we employed the *Gini index* because it is more stable and provides more robust results than the Mean Decrease Accuracy index (Menze et al. 2009; Calle and Urrea 2010). The *Gini index* plot shows the mean decrease in accuracy caused by the input band, which is determined during the oob error calculation. The higher the value of a band on the Gini Index plot, the more useful the band is in performing the classification.

Table 3. Fuzzification values of the lithology spatial dataset into raster format

Geology evidence map	
Distance from edge of favorable unit (m)	Raster value
Within unit	
0–400	1.00
>400	0.80
Outside unit	
0–400	1.00
400–800	0.60
>800	0.00

Specifically, the *Gini index* is calculated by

- For each tree, the oob training samples are put down the tree and the number of correct classifications are calculated (nC).
- Randomly permute the values of variable m in the oob cases and put these cases down the tree (i.e., original data (nP)).
- Calculate nC − nP; (number of votes for the correct class in the variable m permuted oob data) − (number of votes for the correct class in the untouched oob data).

The average of all trees gives the predictive power (importance) of each variable. The score is normalized by the standard deviation of these differences. Features that produce large values for this score are ranked as more important than features which produce small values. Figure 7 provides an overview of RF classification.

The main point of ensemble classifiers, such as RF, is that the process learns from not just one prediction (decision tree) but from many predictions that are then combined (Doan and Foody 2007;

Table 4. Known mineral deposits within the study area

Showing name	Alias	Commodities	Latitude	Longitude
Hurwitz Lake West	PP 126	Au, As	60.9583	−98.0333
Swamp	Don 1; Little Huey	As, Ag, Cu, Pb	60.9508	−101.502
Nigel	Bob 1	Au, Pb	60.9201	−101.7792
Wish Zone	Moon 6	Au, As, Cu	60.9519	−101.5598
4600 Vein		Au, Cu	60.9367	−101.5232
Sample 34306	Dawn 1	Au, Cu	60.9283	−101.5331
East of Ronchon Lake		Au, As	60.8214	−101.8776
Ronchon Lake	Moon 5	Cu, Zn, Au	60.8211	−101.9109
ENN 8		Au, As	60.9433	−101.5960
ENN 8-9		Au, As	60.9577	−101.6015
Nueltin Project-1	PP 132	Cu, Au, Ag	60.1856	−100.5236
Gold Point		Au, As	60.4128	−100.4723
Cobalt		Au, Co As	60.4610	−100.3457
Airstrip		Au, U, Co	60.4580	−100.3305
LES-1	Nueltin Lake Prop	Au, Ag, Bi, Cu, Co, Ni, Mo, U, W	60.1133	−99.9972
Raven Au–Co Showing	AI 1-2, Don 1-3	Au, Co	60.0535	−100.0181

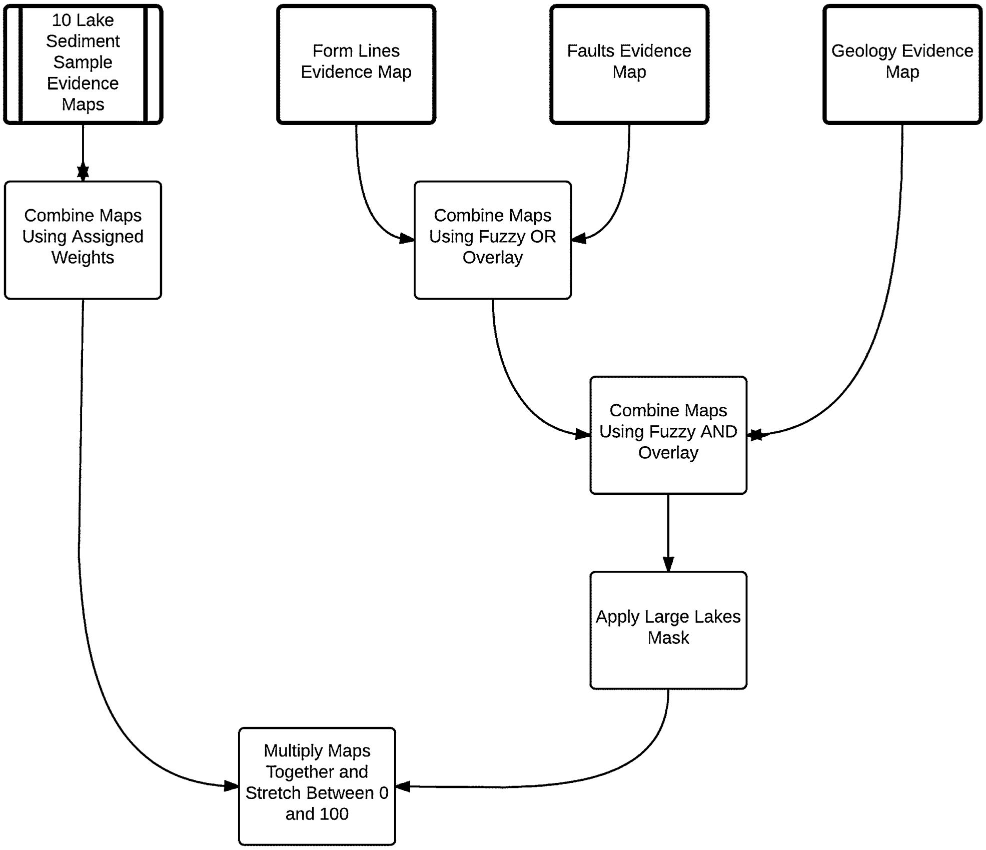
10 Lake Sediment Sample Evidence Maps
Form Lines Evidence Map
Faults Evidence Map
Geology Evidence Map
Combine Maps Using Assigned Weights
Combine Maps Using Fuzzy OR Overlay
Combine Maps Using Fuzzy AND Overlay
Apply Large Lakes Mask
Multiply Maps Together and Stretch Between 0 and 100

Figure 6. Knowledge-driven modeling methodology.

Table 5. Lake sediment sample combination weights, and final output raster statistics

Element	Assigned weight
Arsenic	9
Gold	10
Cobalt	5
Copper	4
Iron	5
Mercury	2
Nickel	2
Lead	1
Zinc	1
Silver	7
Output raster statistics	**Raster cell value**
Low value	6.55
High value	43.25
Mean value	24.68

Harris et al. 2015). This is extremely beneficial as this process helps to reduce the variance as the results are less dependent on peculiarities of a single training dataset. Furthermore, a more robust estimate of the overall classification accuracy is achieved. In addition to the classification map generated by RF, a probability map can also be generated, which shows the strength of membership for each mineral prospect class.

We undertook a number of experiments involving the RF classification (Table 6) in part to study the effect of a limited number of Au occurrences to use as training points for the RF classification. Firstly, all the training points, comprising 16 Au occurrences and 16 non-occurrences, were used to produce a baseline RF Au prospectivity (probability) map. Then, a fivefold repetition of the RF classifier was undertaken using a random selection of eight occurrences and eight non-occurrences for each repetition of the RF classifier. The *oob* error and classification and cross-classification errors were calculated for all experiments for comparison purposes and to establish an estimate of overall classification error. Furthermore, the Au prospectivity maps were statistically compared and averaged,

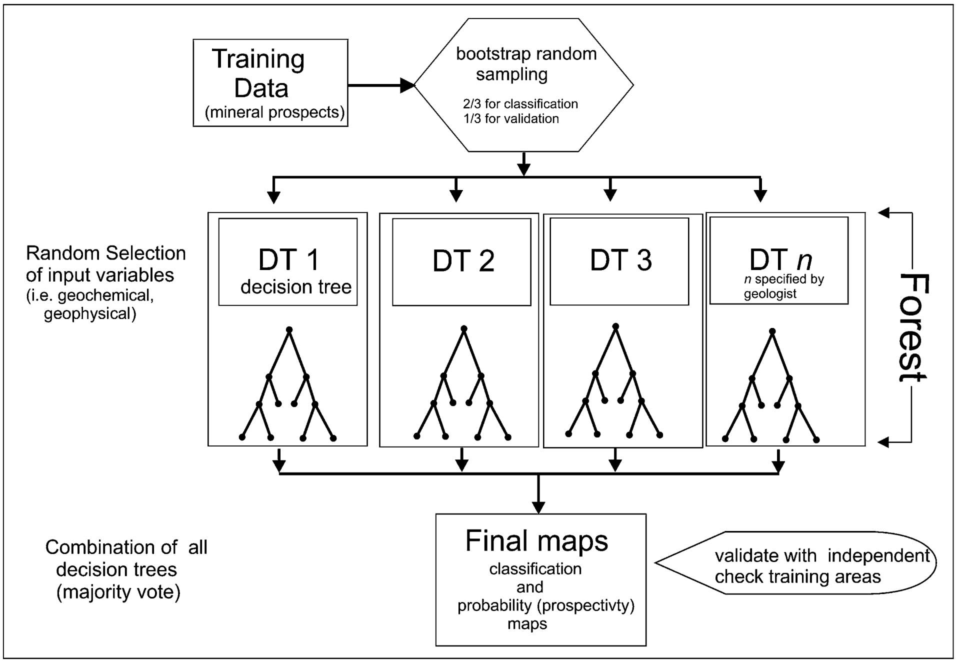
Training Data (mineral prospects)
bootstrap random sampling
2/3 for classification
1/3 for validation
Random Selection of input variables (i.e. geochemical, geophysical)
DT 1 decision tree
DT 2
DT 3
DT n
n specified by geologist
Forest
Combination of all decision trees (majority vote)
Final maps
classification and probability (prospectivty) maps
validate with independent check training areas

Figure 7. Random Forest classification methodology.

producing an average estimate of Au prospectivity. The prospectivity maps (probability) were stretched from values between 0 and 1 to values between 0 and 100 for display purposes, and the mask of large lakes was applied.

RESULTS

Table 7 presents a summary of classification results for the experiments listed in Table 6. Figure 8a, b shows the RF classification and associated Au prospectivity (probability) map produced using all 16 Au training and non-occurrence points (experiment # 1, Table 6). The *oob* error for this classification map is 84.3%, whereas the average *oob* for the fivefold repetition of RF (experiments 2–5, Table 7) is 79.6%. Figure 8c shows the average prospectivity (probability) of the fivefold repetition of RF. The overall accuracy of these classification maps is 100% as might be expected as the overall accuracy is simply calculated by how many Au occurrences used to produce the map are predicted. The *oob* and cross-correlation accuracies (Table 7) are a better representation of the actual accuracy as they are established using Au occurrences that were not used to produce the actual classification map. It has been demonstrated by Breiman (2001) that the *oob* error is a good estimator of the prediction error when the number of trees is large enough. In this case, we used 1000 trees and the *oob* accuracy stabilized after approximately 100 trees. The *oob* accuracy is sufficiently high to trust the results. Table 7 indicates that there is natural variability with respect to accuracy in the RF classification process as the number of evidence maps and training areas are randomly chosen for each decision tree.

The Pearson correlation between all six RF prospectivity (probability) maps produced by RF (Table 6) ranges from a low of 0.77 (T1 vs T4) to a high of 0.96 (T1 vs T3), indicating spatial variability based on the RF random selection process. However, and importantly, the high potential areas are spatially consistent between the RF prospectivity maps produced by the fivefold repetition of the classification process.

Figure 9 shows the KD Au prospectivity map. The area defined as “Prospective” and “Highly Prospective” is significantly smaller in the KD map than in the RF map. The Pearson correlation be-

Table 6. Summary of RF experiments

Experiment	# of evidence maps (see Table 1)	Training data
T1	13	16 Au occurrences and 16 non-Au occurrences randomly selected from non-prospective lithologies
T2	13	8 Au and 8 non-Au occurrences randomly selected
T3	13	8 Au and 8 non-Au occurrences randomly selected
T4	13	8 Au and 8 non-Au occurrences randomly selected
T5	13	8 Au and 8 non-Au occurrences randomly selected
T6	3	8 Au and non-Au occurrences randomly selected

Table 7. Summary of RF classification results

Experiment	*Oob* (%)	Overall accuracy (%)	Cross-correlation accuracy

All training data (16 Au occurrences, 16 non-occurrences)	84.3	100	T1 = 93.7%
			T2 = 100%
			T3 = 93.7%
			T4 = 93.7%
			T5 = 100%
			Average = 96.2%

T1—8 random selected Au occurrences and 8 non-occurrences	68.2	100	T2 = 87.5%
T2—8 random selected Au occurrences and 8 non-occurrences	95.1	100	T1 = 87.5%

T3—8 random selected Au occurrences and 8 non-occurrences	66	100	T2 = 100%
			T1 = 87.5%

T4—8 random selected Au occurrences and 8 non-occurrences	77	100	T1 = 87.5%
			T2 = 93.7%
			T3 = 87.5%

T5—8 random selected Au occurrences and 8 non-occurrences	91.6	100	T1 = 87.5%
			T2 = 100%
			T3 = 87.5%
			T4 = 93.7%

tween the RF and KD map is 0.73, moderate correlation at best, indicating the difference between the two methods. Figure 10 shows an agreement map between the KD and RF prospectivity models, which only displays the top 5% of the most prospective areas. Areas A, C, and D are areas of high spatial overlap, whereas Area B is labeled as high prospectivity only on the RF map. The KD map shows a number of areas of high prospectivity in the eastern portion of the study area that are not present on the RF map.

Figure 11 shows the success rate in the estimation of the known Au occurrences according to the percentage of prospective area at or below the raster prospectivity value at the location of the Au occurrence (e.g., efficiency of prediction curves). In the KD map, 10 of 16 (62.5%) gold prospects are within the top 11% of the most prospective study area (value of 56 or higher). For the RF map, 15 of 16 (93.75%) gold prospects are within the top 11% of the most prospective study area (value of 81 or higher). The single outlier (seen with an Au prospectivity value at 58; Fig. 11) is located 36 m from the boundary of a preferable lithology. Since RF can use nominal datasets as inputs, a simple geology dataset was used, with crisp edges (as opposed to the gradational edges to the preferential geology units described in the ‘‘Lithology’’ section in Data Processing). This allows the RF classifier to determine preferential geology type, instead of using the fuzzy geology evidence map created for the KD model. This resulted in 29.2% of the area being classified as ‘‘prospective’’ by the RF classifier. Using that value, the RF model correctly classifies 16 of 16 (100%) of the gold prospects, and the KD model correctly classifies 15 of 16 (93.75%) of the gold prospects.

Table 8 summarizes the predictive power of each evidence map used in the RF classification. The best predictor, as might be expected is the lithology,

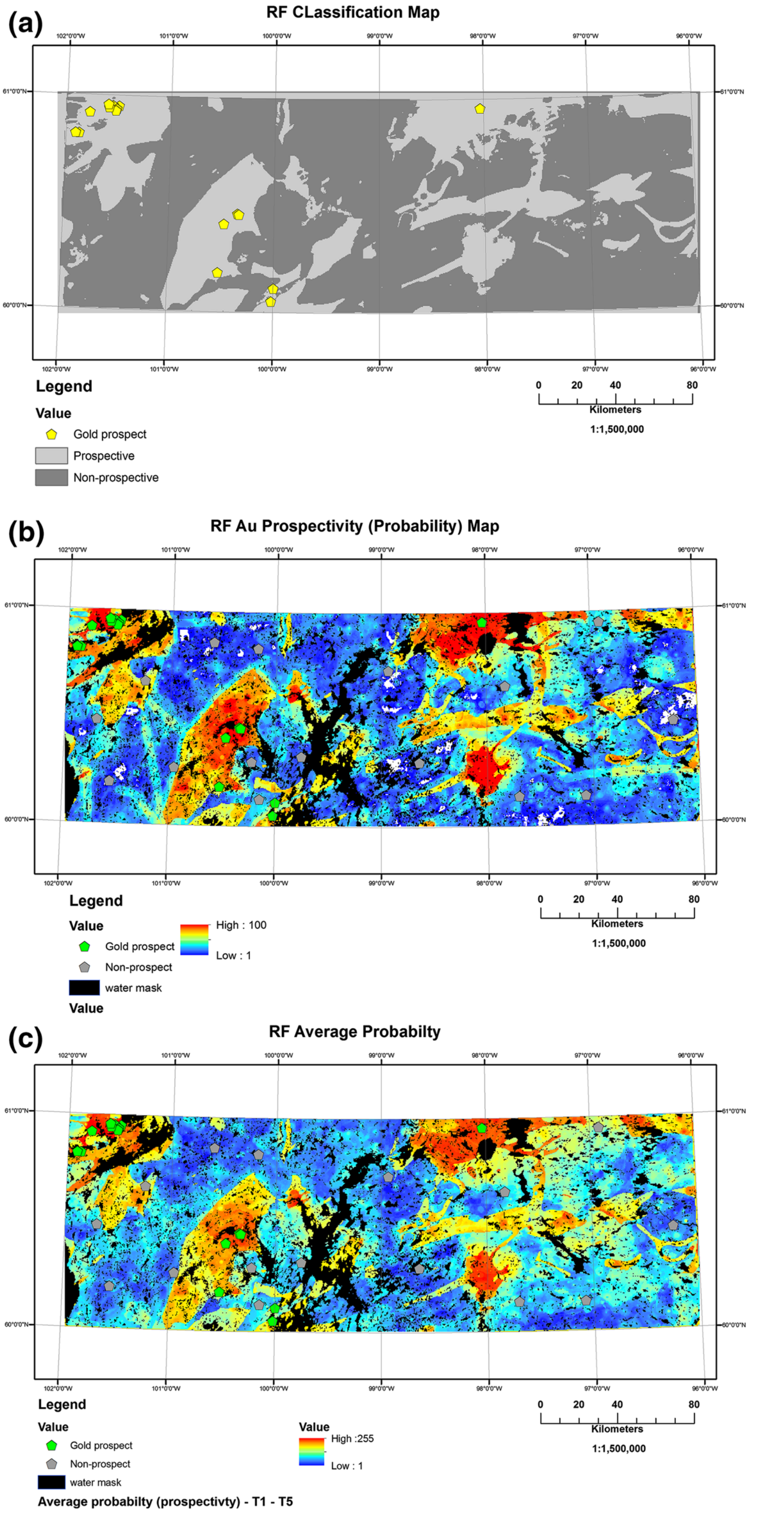

Figure 8. **a** RF classification map showing prospective and non-prospective areas as well as the 16 Au prospects. **b** RF Au prospectivity (probability) maps showing Au and non-Au prospects used for training. **c** Average Au prospectivity (probability) map derived from the fivefold repetition of RF (see text for description).

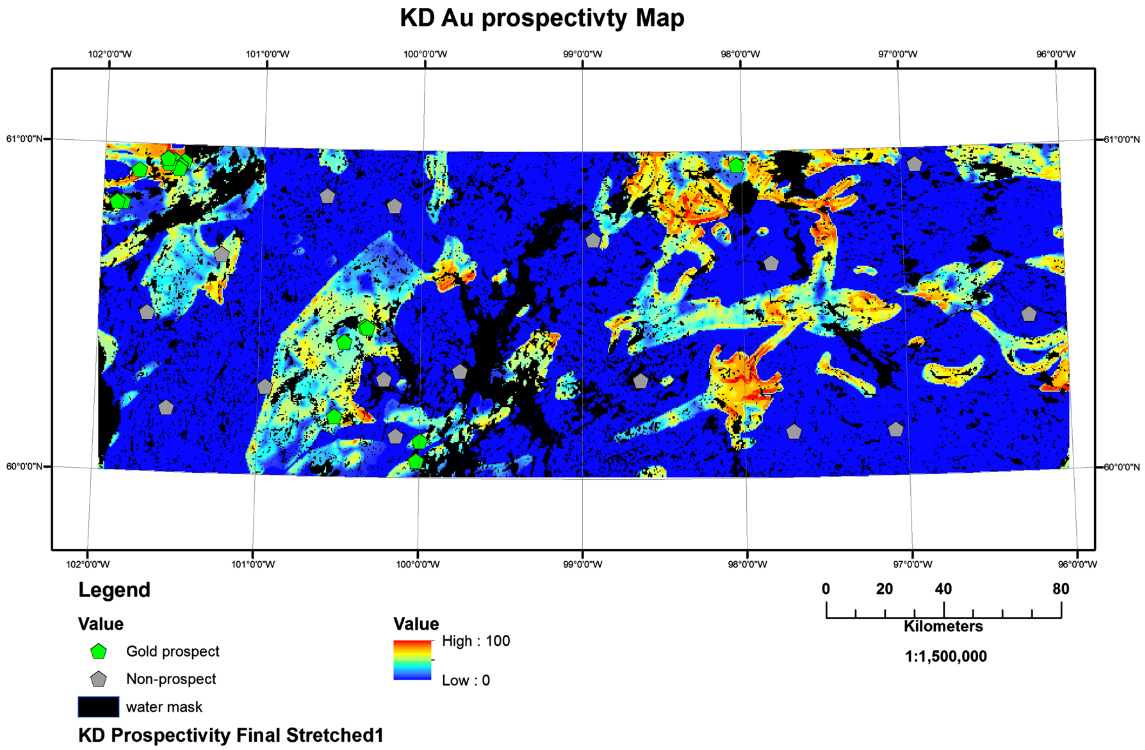

Figure 9. Knowledge-driven (KD) Au prospectivity map with Au and non-Au prospects.

followed by As, Co, Ni, and distance to faults using the normalized *Gini index* score as a measure of importance. It is interesting to note that Au in this case study was only predictive in one of the fivefold repetition of the RF but was in the top 5 predictors when all training areas were used.

It is visually apparent that the most prospective areas for both maps are mostly controlled by the lithology (Figs. 8, 9, 10; Table 8). It is important to note that a decision based on Au exploration in Canada's North was made to select only favorable lithologies before producing the Au prospectivity maps. Additionally, all the known Au occurrences fall within these selected lithologies, or the fuzzy boundaries of the lithologies, comprising Hurwitz sediments, mafic volcanics, and associated sediments and supracrustal rocks.

With respect to the most prospective areas (Fig. 10), Area A, predicted as highly prospective by both the RF and KD methods, is situated along an E-W trending contact between supracrustals and mafic volcanics. A large E-W trending fault, or possible shear zone, also occurs within this Area A and may have acted as a conduit for mineralized Au-bearing fluids. Seven of the 16 known Au occurrences occur in this area. Area B within the Hurwitz sediments, in the vicinity of the Nueltin intrusive suite, is designated as a high potential area only on the RF prospectivity map. Fault density is high within Area B and is bracketed to the north by a major NE trending fault (shear) that has a strong linear magnetic expression on the airborne magnetic data of the area. Three known Au occurrences are found within Area B. Area C is situated within supracrustals and Hurwitz group sediments and is delineated as highly prospective on both the RF and KD maps. This area contains 1 known Au occurrence at the contact of the supracrustal rocks and a small body of mafic volcanics. Area D, prospective on both RF and KD maps, is situated within supracrustals and Hurwitz sediments and is transected by E-W trending faults. This area is surrounded by Hudson granites, a possible heat engine, and does not contain any known Au occurrences. Recall that the Hudson granites are older than the Nueltin intrusive suite (1.83 vs. 1.75 Ga), so its role as a heat engine may be reduced when compared to the Nueltin suite. The KD prospectivity map indicates a large number of smaller prospective zones in the eastern portion of the study area, all of which occur

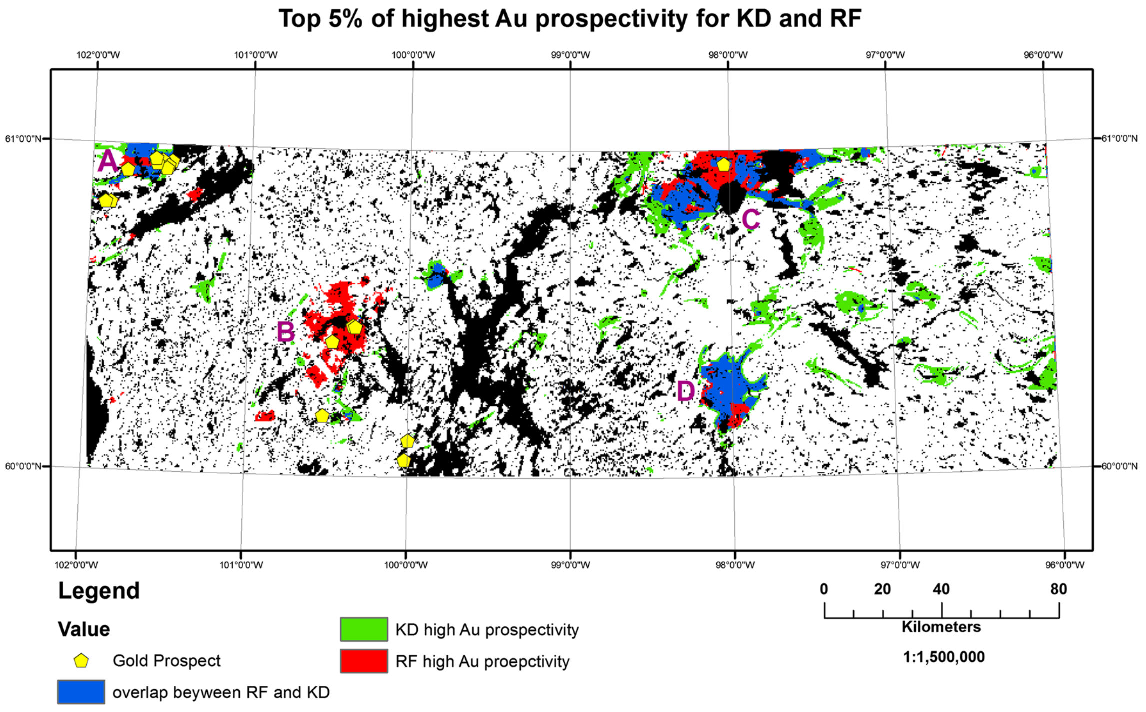

Figure 10. Agreement map between the RF and KD maps showing only ~5% of the top prospective areas—*a* to *d*—discussed in text.

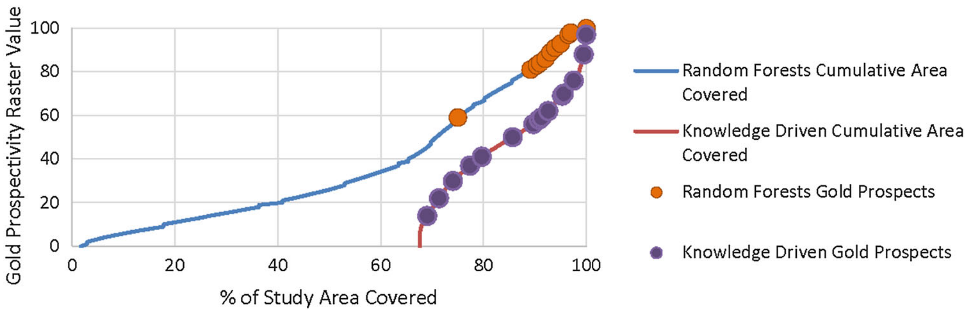

Figure 11. Efficiency of prediction curves for the RF and KD Au prospectivity maps.

in either Hurwitz sediments or supracrustal rocks. Many of these areas co-occur with linear magnetic anomalies potentially representing fault zones.

The late Paleoproterozoic Nueltin suite outcrops across a large portion of the western Churchill Province and marks the last period of extensive igneous activity in the Kivalliq Region of Nunavut. Occurrences of uranium, rare earth elements (REE), and precious metals (Au–Ag), found in association with Nueltin suite rocks implicate the suite as prospective for exploration and economic potential (Scott 2012).

SUMMARY AND CONCLUSIONS

This study evaluated the performance of the RF method in the creation of prospectivity maps and compared the results of the RF method to those of a simple KD method. The RF method performed well

Table 8. Top 5 evidence maps in terms of RF predictive power (ranked from 1 to 5)

Experiment (see Table 6)	Ag	As	Au	Co	Cu	Hg	Ni	Pb	Zn	Faults	Form lines	Lithology
All		4	5	2			3					1
T1		2	2	4		3				5		1
T2		2		4			5				3	1
T3		4					3	5			2	1
T4		2		5		4				3		1
T5		4		3			5			2		1
Total		2		3			4			5		1

in that the Au prospectivity map created was a better predictor of the known Au occurrences than the KD Au prospectivity map. There are several advantages to the use of RF including (1) the ability to take both continuous and/or categorical data as variable inputs, (2) an internal, unbiased estimation of the mapping error (oob error) removing the need for a cross-validation of the final outputs to determine accuracy, and (3) the estimation of importance of each input. One major difference between the KD and RF models in this study comes from the use of two different geological evidence maps. A classified 'preferential geology' evidence map with gradational boundaries was used for the KD model and a crisp nominal geology evidence map was used for the KD method.

One of the concerns in this study is the limited training dataset (16 Au occurrences and 16 non-Au occurrences) used for the RF classification. A recent paper by Carranza and Laborte (2015) achieved satisfactory RF classification results using a limited training dataset (<20). Oshiro et al. (2012) also achieved good results using RF applied to 10 evidence maps and 12 prospects and 12 non-prospects. In this paper, we increased the number of trees to 1000, as recommended by Carranza and Laborte (2015) to add stability to the results as well as employing a fivefold repletion of the RF classifier, again to add stability and bracket classification errors. As mentioned above, the *oob* error stabilized after only 60 trees. Rodriguez-Galiano et al. (2014) recommended 1000 trees with the use of nine evidence maps (predictor variables), 46 deposits and 57 non-deposits. The average *oob* error for the fivefold repletion of RF in this study was 80% and average accuracy using independent check occurrences in the fivefold repletion of RF was 92% (Table 7), both acceptable accuracies.

Efficiency of prediction curves (Fig. 11) illustrates that the RF method performs better than the KD method even with limited training areas. The success rate (rate at which the areas containing gold prospects were classified as highly prospective) is significantly higher for the RF method than the KD method. Although the RF method classified a significantly larger area as prospective and very prospective than the KD method, Figure 11 illustrates that more gold prospects are located within a smaller classification area in the RF results than the KD results.

ACKNOWLEDGMENTS

This work was supported and funded under the Geo-Mapping for Energy and Minerals (GEM) program at the Geological Survey of Canada (GSC). We would also like to thank Algonquin College located in Ottawa, Canada for their support.

REFERENCES

Abedi, M., Norouzi, G. H., & Bahroudi, A. (2012). Support vector machine for multi-classification of mineral prospectivity areas. *Computers Geosciences, 46*, 272–283.

Agterberg, F. P., & Bonham-Carter, G. F. (2005). Measuring the performance of mineral-potential maps. *Natural Resources Research, 14*, 1–17.

Aitchison, J. (1986). *The statistical analysis of compositional data*. New York: Chapman and Hall.

An, P., Moon, W.M., & Bonham-Carter, G.F. (1992). On a knowledge-based approach of integrating remote sensing, geophysical and geological information. In *Proceedings IGARSS'92* (pp. 34–38).

An, P., Moon, W. M., & Bonham-Carter, G. F. (1994a). An object-oriented knowledge representation structure for exploration data integration. *Nonrenewable Resources, 3*, 132–145.

An, P., Moon, W. M., & Bonham-Carter, G. F. (1994b). Uncertainty management in integration of exploration data using the belief function. *Nonrenewable Resources, 2*, 60–71.

An, P., Moon, W., & Rencz, A. N. (1991). Application of fuzzy theory for integration of geological, geophysical and remotely sensed data. *Canadian Journal of Exploration Geophysics, 27*, 1–11.

Aspler, L. B., Wisotzek, I. E., Chiarenzelli, R., Losonczy, M. F., Cousens, B. L., McNicoll, V. J., & Davis, W. J. (2001). Paleoproterozoic intracratonic basin processes, from breakup of Kenora to assembly of Laurentia: Hurwitz basin, Nunavut, Canada. *Sedimentary Geology, 141–142*, 287–318.

Behnia, P., Harris, J. R., Harrison, C., St-Onge, M., Okulitch, A., Irwin, D., & Gordy, S. (2013). *Geo mapping frontiers: Compilation and interpretation of geologic structures North of 60°*. Geological Survey of Canada Open File: Canada. **7649**.

Bonham-Carter, G. F. (1994). *Geographic information systems for geoscientists: Modeling with GIS*. New York: Pergamon, Elsevier Science Inc.

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32.

Brodeur, G., El-Hihi, S., & Jebrak, M. (1992). Applications of neural network computing to mineral exploration in the southern Abitibi greenstone belt. In *Program with Abstracts-Geological Association of Canada, Mineralogical Association of Canada, Canadian Geophysical Union: Joint Annual Meeting* (p 17).

Brown, W. M., Gedeon, T. D., Groves, D. L., & Barnes, R. G. (2000). Artificial neural networks; A new method for mineral prospectivity mapping. *Australian Journal of Earth Sciences, 47*, 757–770.

Calle, M., & Urrea, V. (2010). Letter to the editor: Stability of random forest measures. *Brief Bioinformatics, 12*, 86–89.

Carranza, E. J. M. (2009a). Chapter 7: Knowledge-driven modeling of mineral prospectivity. In E. J. M. Carranza (Ed.), *Handbook of exploration and environmental geochemistry* (pp. 189–247). Amsterdam: Elsevier Science B.V.

Carranza, E. J. M. (2009b). Chapter 8: Data-driven modeling of mineral prospectivity. In E. J. M. Carranza (Ed.), *Handbook of exploration and environmental geochemistry* (pp. 249–310). Amsterdam: Elsevier Science B.V.

Carranza, E. J. M. (2014). Data-driven evidential belief modeling of mineral potential using few prospects and evidence with missing values. *Natural Resources Research,*. doi: 10.1007/s11053-014-9250-z.

Carranza, E. J. M., & Laborte, A. G. (2015). Random Forest predictive modeling of mineral prospectivity with small numbers of prospects and data with missing values. *Computers & Geosciences, 74*, 60–70.

Carranza, E. J. M., Woldai, T., & Chikambwe, E. M. (2005). Application of data-driven evidential belief functions to prospectivity mapping for aquamarine-bearing pegmatites, Lundazi District, Zambia. *Natural Resources Research, 14*, 47–63.

Charbonneau, B.W., & Swettenham, S.S. (1986). Gold occurrence in radio-active calc-silicate float at Sandy beach Lake, Nueltin lake area, District of Keewatin. In: Current Research 1986-A; Geological Survey of Canada, pp. 803–808.

Chung, C. F., & Agterberg, F. P. (1980). Regression models for estimating mineral resources from geological map data. *Mathematical Geology, 12*, 473–488.

Chung, C. F., & Fabbri, A. G. (2003). Validation of spatial prediction models for landslide hazard mapping. *Natural Hazards, 30*, 451–472.

Doan, H. T., & Foody, G. M. (2007). Increasing soft classification accuracy through the use of an ensemble classifier. *International Journal of Remote Sensing, 28*, 4609–4623.

Gislason, P., Benediktsson, J., & Sveinsson, J. (2006). Random Forests for land cover classification. *Pattern Recognition Letters, 27*, 294–300.

Grunsky, E. C., Mueller, U. A., & Corrigan, D. (2014). A study of the lake sediment geochemistry of the Melville Peninsula using multivariate methods: Applications for predictive geological mapping. *Journal of Geochemical Exploration, 141*, 15–41.

Hanmer, S., Sandeman, H. A., Davis, W. J., Aspler, L. B., Rainbird, R. H., Ryan, J. J., et al. (2004). Neoarchean tectonic setting of the Central Hearne supracrustal belt, western Churchill Province, Nunavut, Canada. *Precambrian Research, 134*, 63–83.

Harris, J. R. (1989). Data integration for gold exploration in eastern Nova Scotia using a GIS. *Proceedings of Remote Sensing for Exploration Geology* (pp. 285–292). Alberta: Calgary.

Harris, J. R., Grunsky, E., Behnia, P., & Corrigan, D. (2015). Data-and knowledge-driven mineral prospectivity maps for Canada's North. *Ore Geology Reviews,*. doi: 10.1016/j.oregeorev.2015.01.004.

Harris, D. A., & Pan, R. (1999). Mineral favourability mapping: a comparison of artificial networks, logistic regression and discriminant analysis. *Natural Resources Research, 8*, 93–109.

Harris, J. R., Sanborn-Barrie, M., Panagapko, D. A., Skulski, T., & Parker, J. R. (2006). Gold prospectivity maps of the Red Lake greenstone belt: application of GIS technology. *Canadian Journal of Earth Sciences, 43*, 865–893.

Harris, J. R., Wilkinson, L., Heather, K., Fumerton, S., Bernier, M. A., Ayer, J., & Dahn, R. (2001). Application of GIS processing techniques for producing mineral prospectivity maps—A case study: Mesothermal Au in the Swayze greenstone belt, Ontario, Canada. *Natural Resources Research, 10*, 91–124.

McCurdy, M.W., McNeil, R. J., Day, S.J.A., & Pehrsson, S.J. (2012). Regional lake sediment and water geochemical data, Nueltin lake area, Nunavut (NTS 65A, 65B and 65C), Geological Survey of Canada Open File 6986, 1 CD-ROM.

Menze, B. H., Kelm, B. M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W., & Hamprecht, F. A. (2009). A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics, 10*, 213.

Oshiro, T. M., Perez, P. S., & Baranauskas, J. A. (2012). How many trees in a random forest? In P. Perner (Ed.), *Machine learning and data mining in pattern recognition. Lecture notes in computer science* (Vol. 7376, pp. 154–168). Berlin: Springer.

Paul, D, Hanmer, S., Tella, S., Peterson, T.D., & Lecheminant, A.N. (2002). Geological Survey of Canada, Open File 4236, 1 sheet; 1 CD-ROM.

Porwal, A., Carranza, E. J. M., & Hale, M. (2003). Artificial neural networks for mineral-prospectivity mapping: A case study from Aravalli Province, Western India. *Natural Resources Research, 12*, 156–171.

Reddy, R. K. T., & Bonham-Carter, G. F. (1991). A decision-tree approach to mineral potential mapping in Snow Lake area. *Manitoba, Canadian Journal of Remote Sensing, 17*, 191–200.

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*, 1336–1354.

Scott, J.M.J. (2012). Paleoproterzoic (1.75 Ma) granitoid rocks and uranium mineralization on the Baker Lake-Thelon Basin region, Nunavut. B.Sc. thesis, Department of Earth Science, Carleton University, unpublished.

Scott, J.M.J, Peterson, T.D., & McCurdy, W.W. (2012). U, Th and REE occurrences within the Nueltin granite suite at Nueltin lake, Nunavut, recent observations, Geological Survey of Canada, Current Research 2012-1.

Singer, D. A., & Kouda, R. (1996). Application of feed forward neural network in search for Kuroko deposits in the Hokuroku district, Japan. *Mathematical Geology, 28*, 1017–1023.

Van Breeman, O., Peterson, T. D., & Sanderman, H. A. (2005). U-Pb zircon geochronology and Nd isotope geochemistry of

Proterozoic granitoids in the western Churchill Province: Intrusive age pattern and Archean source domains. *Canadian Journal of Earth Science, 42*, 339–377.

Waske, B., & Braun, M. (2009). Classifier ensembles for land cover mapping using multitemporal SAR imagery. *ISPRS Journal of Photogrammetry and Remote Sensing, 64*, 450–457.

Wright, D.F., & Bonham-Carter, G.F. (1996). VHMS favourability mapping with GIS-based integration models, Chisel Lake-Anderson Lake Area. In G. F. Bonham-Carter, A. G. Galley, & G. E. M. Hall (Eds.), EXTECH I: A multidisciplinary approach to massive sulphide research in the Rusty Lake-Snow Lake Greenstone Belts. Manitoba: Geological Survey of Canada, Bulletin 426, pp. 339–376 and pp. 387–401. Special Publication 44: 21.

Zuo, R., & Carranza, E. J. M. (2011). Support vector machine: A tool for mapping mineral prospectivity. *Computers Geosciences, 37*, 1967–1975.