*Original Paper*

# Mineral Occurrence Target Mapping: A General Iterative Strategy in Prediction Modeling for Mineral Exploration

## C.-J. Chung[1] and A. G. Fabbri [2,3]

Procedures are discussed to construct target maps ranking the likelihood of future discoveries: for instance, of gold occurrences, knowing location and spatial context, of a set of genetically related gold vein deposits. A *favorability modeling* process is iterated with a subset of the known occurrences, i.e., the locations of the deposits. The resulting *prediction patterns* are cross-validated with the distribution of the left-out occurrences, considered as representing the future discoveries. The target map originates from integration of all *prediction patterns* from the iterations. Rank-based statistics related to the target maps provides measures of quality, robustness and uncertainty of the classification of a study area into likelihood of discovery. Much of this is a relatively new area of research, so that to interpret such uncertainty is still a challenge. Four critical questions are formulated that identify areas in need of extensive research for any modeling procedure. They relate to the quality of *prediction patterns*, their associated uncertainty of class membership, their sensitivity to redundancy and to congruity within the database. A spatial database developed for advanced training is used to generate target maps. It comes from a study in the Red Lake area in northern Ontario, Canada. It contains information on 37 gold vein deposits. Their neighborhood distribution is instrumental to establishing spatial relationships with the units of categorical thematic maps and continuous field maps. The experimental results point at the extraction of significant properties of the spatial data that cannot be ignored but that we have yet to master to substantiate the reliability of *prediction patterns*.

## INTRODUCTION

It is now common practice in the geosciences to use spatially distributed data and model mineral potential maps to guide in further resource exploration. Within a study area, the data relate to the spatial distribution of mineral occurrences of known and uniform genetic type and possibly of the size of

mineral deposits, whose locations we have termed as occurrences. Moreover, they relate to the corresponding geologic mapping units and geochemical/geophysical continuous field values hopefully representing the spatiotemporal conditions of the occurrences.

Within a conveniently selected and delimited study area, spatial relationships between the occurrences, sets of mapping units and continuous field values are established and integrated by mathematical models. Seldom can the occurrences be separated into time intervals to allow using an older set to obtain a map with classes ranking the likelihood

[1]Spatialmodels Inc., Ottawa, Canada.
[2]University of Milano-Bicocca, Milan, Italy.
[3]To whom correspondence should be addressed; e-mail: andrea.fabbri@unimib.it

of future occurrence. The map is validated by the distribution of the younger set. Instead, convenient partitioning of the known occurrences is used to have a subset for modeling and another subset for validation as the "next" occurrences. By validation, we mean that the distribution of the "next" occurrences is used for capturing the corresponding likelihood ranks in the modeled map (Fabbri and Chung 2008). The higher is the rank containing the occurrences, the better is the validation. To do this, the data should preferably be part of a digital database with consistent resolution.

Modeling the future mineral occurrence discoveries, however, is a complex task. At present, much needs to be done to satisfy such a task. This contribution points at its complexities to propose ways to resolve them.

In our analyses, a spatial database developed for advanced training is used to generate target maps. It comes from a study in the Red Lake area in northern Ontario, Canada. The processing of the database is described with the purpose of identifying issues that are generally ignored or bypassed in many modeling applications. The following section introduces definitions and assumptions implicit in favorability modeling. It is followed by a brief description of the study area database. Next, experiments leading to a set of questions are discussed that are critical to the interpretation of the modeling results. Conclusions follow on the relevance of providing answers to the questions.

## DEFINITIONS AND ASSUMPTIONS FOR FAVORABILITY FUNCTION MODELING

What we term as *favorability function modeling* is the generation of digital maps or better *prediction patterns*. They are to show the classification of all points of a study area into ranks of similarity with the neighborhoods of mineral occurrences, in terms of the corresponding presence or absence of units and continuous field values. Such a classification within the study area is usually based on the empirical likelihood ratios between the distribution of units and continuous field values in the presence of the occurrence neighborhoods and those in their absence. Commonly, the occurrences are a few tens to a few hundred in number, while their neighborhoods cover a few picture elements or pixels (say $3 \times 3$ or $5 \times 5$) of a given uniform resolution (e.g.,

10 m, 30 m or 50 m). The resolution depends on the assumed spatial support of the occurrences. The entire study area generally consists of hundreds of thousands to millions of pixels. It means that the data corresponding to a few hundreds of pixels are used to classify possibly millions of pixels. Hopefully, the relationships established in the presence of the occurrences are contrasting those established outside the occurrence neighborhoods. This holds even if we expect or assume that more occurrences will be found there.

The task of spatial prediction modeling the location of new discoveries is far from simple insofar that it requires, besides a relevant database, a variety of definitions, assumptions, mathematical models for integrating spatial relationship, procedural strategies and validations of the results of the modeling, the *prediction patterns*. The complexity of generating and interpreting the patterns has motivated the generation of target mapping as cautious general-purpose modeling based on rank-based statistics and iterative cross-validation. Target maps are ranking the likelihood of future discoveries: for instance, of discoveries like gold occurrences, knowing location and spatial context, of a set of genetically related gold vein deposits.

A unified mathematical framework for *favorability function modeling* was introduced (Chung and Fabbri 1993; Chung 2003; Chung et al. 2003) indicating several spatial prediction models, their interpretation and estimation procedures. In particular, for the two models applied here, the empirical likelihood ration function, ELR, and the logistic discriminant function, LOG, theoretical and application backgrounds were discussed by Chung (2006) (see also Chung 1978; Davis et al. 2006). The ELR model is based on the multivariate frequency distribution function under the assumption of conditional independence of the spatial context. If the empirical distribution function appears non-normal, the LOG model can be used. It is based on a logistic estimation under the assumption of conditional independence.

A spatial prediction model is defined as a logical or analytical procedure to construct a *prediction image*. A *prediction pattern* is a thematic representation of the *prediction image* to visualize and perceive it. Let us consider a proposition, i.e., a mathematical statement that can be supported as true or false by spatial evidence:

$T_p$: that a point $p$ in the study area contains a mineral deposit of specific type|the presence of spatial evidence.

To support such a proposition, we need a spatial database that provides such evidence and a mathematical model to convert spatial data into evidence by establishing and integrating spatial relationships. The spatial database relates to a study area and contains the distribution of mineral deposits of a distinct and uniform genetic type. Along with it, it contains the distribution of categorical map units and continuous field value maps related to the mineral deposits, i.e., representing the conditions of deposition. We have termed as direct supporting pattern, DSP, of the proposition the binary map (presence/absence) of the mineral deposits. We have also termed indirect supporting patterns, ISPs, the other maps that hopefully represent the conditions or settings of the deposits. The DSP occupies a very small part of the study area, SA, and is the known part of the *target pattern*, i.e., of all the existing deposits in the SA. While the SA is defined as the area in which we wish to predict the unknown part of the *target pattern*, the training area, TA, is the area where the quantitative characteristics and properties of the *target pattern* will be established in the spatial prediction analysis. In an ideal TA, the *target pattern* should have been completely discovered, but in a real situation only a portion, possibly most, of the *target pattern* is known.

We are after estimating the likelihood of discovering the remaining deposits. We use the term pattern because in the modeling all spatial data are transformed from the initial mapping units and values into spatial evidence, thus becoming measures of spatial relationships between DSP and ISPs. Such terminology will be used in the remainder of this contribution.

For modeling, a favorability function at each point in the SA must have two properties:

(1) be able to measure a relative level of likelihood that point $p$ contains a part of undiscovered deposits of the given type, so that the *prediction image* can be generated; and

(2) be able to provide measures of uncertainty associated with the function by using the known part of the *target pattern* in the TA, i.e., the DSP.

Furthermore, three assumptions must be reasonably satisfied (Chung and Fabbri 2004): (a1) the known deposits, the DSP, are a random selection of all existing ones, known and unknown (allowing to extend the function to the entire SA from the TA); (a2) the ISPs are correlated with the *target pattern* (allowing to estimate the function using the known part of the *target pattern* in the TA); and (a3) the process of deposition is not random and follows a certain rule (allowing to model the favorability function).

The next section shows how the Red Lake database provides the inputs to the modeling process.

## THE RED LAKE STUDY AREA AND SPATIAL DATABASE

The spatial database used in the prediction experiments to follow was developed at the Geological Survey of Canada. It was used for the purpose of training professionals in advanced courses on spatial data integration for mineral exploration target mapping (GSC 2000; Chung and Fabbri 2008; Hillary et al. 2008). The database uses information collected for the Red Lake study area in northwestern Ontario, whose location is shown in Figure 1a.

The database contains digital maps and their interpretation for mineral potential evaluation. One map contains the locations of 37 gold deposits. Their distribution is shown in Figure 1b as black and red dots on a background of the study area in light gray. The darker gray areas fall outside the study area. The remainder of the database that consists of three categorical and twelve continuous field maps is shown in Figure 2. Their description is given in Table 1. All 16 (1 + 3 + 12) digital maps use the resolution of 40 m and are spatially co-registered so as to maintain a point-to-point or pixel-to-pixel correspondence. This is to facilitate the computation of spatial relationships.

The database also contains additional magnetic and electromagnetic geophysical data not used in the present analyses: total magnetic field, resistivity, horizontal gradient, analytical signal, susceptibility, tilt and vertical gradient. Of these, only resistivity shows a relatively strong support, however, masked by those of the three categorical maps.

While the mineral deposit map in Figure 1b is indicating the presence of mineral occurrences, the other 15 digital maps are to represent the geologic
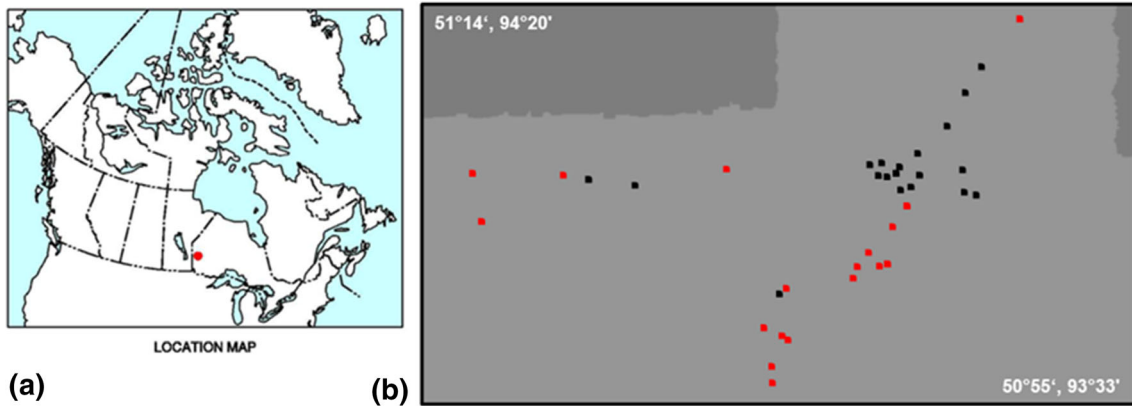
**Figure 1.** Location of the Red Lake study area, northern Ontario, Canada, in (**a**), and distribution of the 37 gold deposit occurrences in (**b**), as black and red dots of exaggerated size to represent 3 × 3 pixel neighborhoods for visibility. For modeling, they have been termed direct supporting **pattern**, DSP. See "How Congruous are the Mineral Occurrences Used in Modeling?" section for the separation of black and red occurrences.
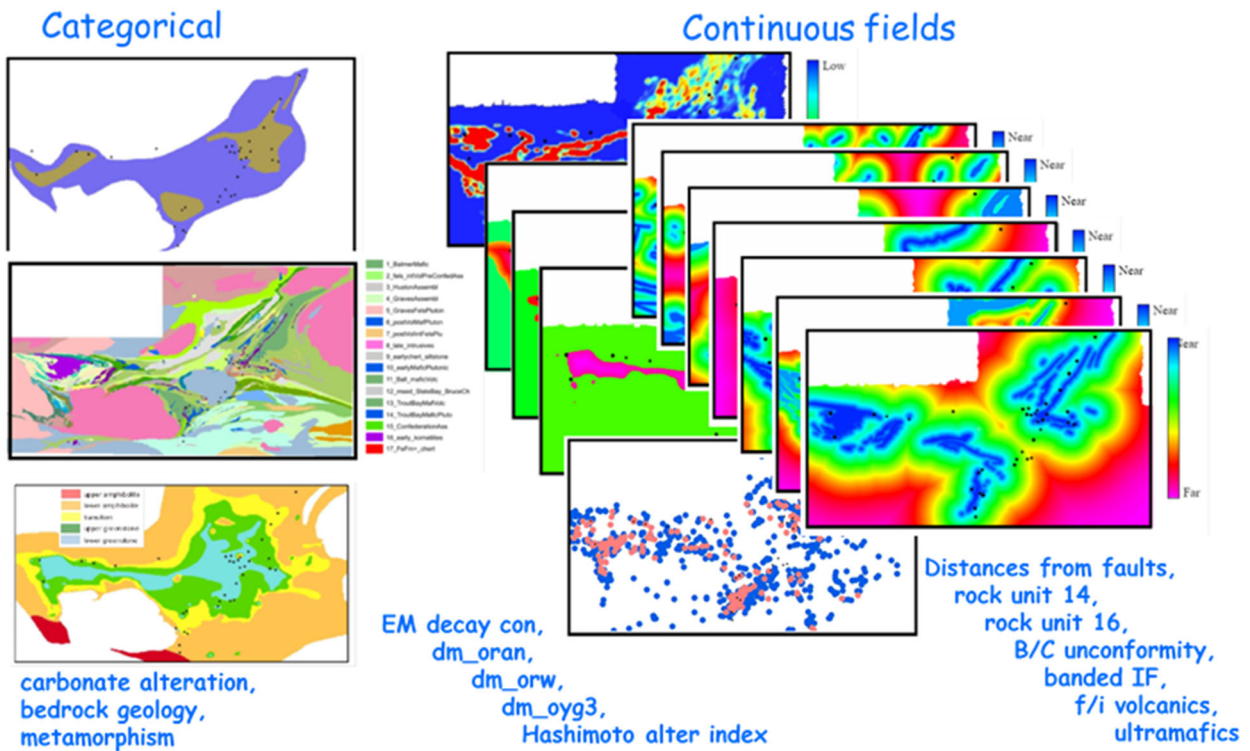


**Figure 2.** Categorical and continuous field maps converted into indirect supporting patterns, ISPs, by their spatial relationships with the distribution of the 37 Au occurrence neighborhoods used for spatial prediction modeling, shown in Figure 1b.

settings of the occurrences. The 37 gold occurrences comprise vein gold deposits, namely gold-bearing "cherty quartz with banded Fe-carbonate"; "sulfide-poor veins," "sulfide-rich replacement zones," high-grade zones; strata-bound dissemination of gold in sulfide-bearing tuffaceous rocks and gold-bearing veins in granitoid intrusions. Tentatively, they will be considered as a uniform type of deposit. For subsequent modeling, the locations were converted

**Table 1.** Categorical and continuous field ISPs, $\mathbf{a}_{1-3}$, $\mathbf{g}_{1-17}$, $\mathbf{m}_{1-5}$ and $\mathbf{CNT}_{1-12}$, respectively, in the Red Lake area

| Carbonate alteration, $\mathbf{a}_{1-3}$ | Continuous field maps, $\mathbf{CNT}_{1-12}$ |
|---|---|
| 1. **Proximal alteration** (brown dominated by ferroan dolomite) | 1. **dalter**, lithogeochemical data: Hashimoto alteration_index (pink_> 0.58, blue < 0.58) |
| 2. Distal alteration (blue dominated by calcite) | 2. dbif, distance from iron formation |
| 3. No alteration | 3. **decay**, electromagnetic decay constant: the time required for an electromagnetic field to decay to a value of 1/e of the original value. In time- domain electromagnetic data, the time constant is proportional to the size and conductance of a tabular conductive body, also called the time constant |
| **Geology,** $\mathbf{g}_{1-17}$ | |
| 1. **Balmer Mafics** | |
| 2. **Early Felsic-Intermediate Volcanics** | |
| 3. **Huston Assemblage** | 4. dfault, distance from faults (unsubdivided) |
| 4. Graves Assemblage | 5. dfvol, distance from felsic and Intermediate volcanic rocks (un-subdivided) |
| 5. Graves Felsic Pluton | |
| 6. PostVolcanics Mafic Plutonics | 6. **dg14,** distance from rock unit 14 (granitoid rocks, circa 2720 Ma) |
| 7. PostVolcanics Felsic Pluton (unit 14) | 7. dg16, distance from rock unit 16 (late granitoid rocks, circa 2700 Ma) |
| 8. Late Intrusive (unit 16) | |
| 9. Early Cherty + Siltstone | 8. **dm_oran**, distance from lower amphibolite vs. transition zone isograd as measured from higher-grade rocks |
| 10. Early Mafic Plutonics | |
| 11. **Ball MaficVolcanics** | 9. **dm_oryw**, distance from upper greenschist vs. transition zone iso-grad as measured from higher-grade rocks |
| 12. Slate Bay Metasediments | |
| 13. Trout Bay Mafic Volcanics | 10. **dm_oyg3**, distance from upper greenschist vs. lower greenschist isograd as measured from higher-grade rocks |
| 14. Trout Bay Mafic Pluton | |
| 15. Confederation Assemblage | 11. **dultra**, distance from ultramafic rocks |
| 16. Early Komatiite Flows | 12. **dunconf**, distance from the unconformity between Balmer and Confederation assemblage |
| 17. Fe Fm + Chert | |
| **Metamorphism,** $\mathbf{m}_{1-5}$ | |
| 1. **Lower greenschists;** | |
| 2. **Upper greenschists;** | |
| 3. Transition zone; | |
| 4. Lower amphibolite; | |
| 5. Upper amphibolite. | |

In bold are the more supportive ones as indicated by the ELR values ≥ 2 in Table 2

into 3 × 3 neighborhoods of pixels with the initial pixel location at the center.

The mineral deposit map with the distribution of the 37 gold deposits, considered as occurrences, used initially single pixels to represent their location. The 37 3 × 3 pixel neighborhoods of gold occurrences were used as DSP in the modeling and termed **au37**. Using single pixels or 3 × 3, 5 × 5 or 7 × 7 pixels neighborhoods led to identical spatial relationships. Their locations are shown in Figure 1b. The purpose is to establish their spatial relationships with the 15 ISPs, termed **agmCNT12** (carbonate **a**lteration, bedrock **g**eology, **m**etamorphism and the 12 continuous field maps) shown in Figure 2. While the study area, SA, is contained within a rectangular raster of 1250 pixels × 700

lines, it occupies only 746,085 pixels of 40 m resolution. The 37 gold occurrence neighborhoods instead occupy 333 pixels of the same resolution. It means that the SA covers an area three orders of magnitude larger than that of the DSP. In the analyses to follow, the entire study area was considered as TA. In other situations, when assuming that a wider neighborhood of the occurrences has been thoroughly explored, the TA could be selected as larger neighborhood, say 25 × 25 or 51 × 51 pixels.

For the Red Lake study area, the spatial relationships between the DSP and the ISPs are represented by empirical likelihood ratios that are summarized in Table 2. For instance, an ELR value of 2 indicates that the normalized frequency in the

**Table 2.** Empirical likelihood ratios to be used in predictions with the distribution of the 37 gold occurrences, **au37** as DSP and with **agmCNT12** as ISPs

| DSP and ISPs | Empirical likelihood ratios |
|---|---|
| **au37** | |
| c. alteration | $a_2 = $ **6.29**, $a_2 = 1.67$, $a_3 = 0.19$; |
| geology | $g_1 = $ **2.21**, $g_2 = $ **2.00**, $g_3 = $ **3.27**, $g_7 = 1.65$, $g_{10} = 1.78$, $g_{11} = $ **2.32,** $g_{12} = 0.65$, $g_{15} = 0.38$; |
| metamorphism | $m_1 = $ **4.05**, $m_2 = $ **2.29**, $m_3 = 0.89$, $m_4 = 0.43$; |
| dalter (*0.1–268.0*) | ≥ 2: *0.10–23.62,* (max **5.31**); |
| dbif (*1–398*) | ≥ 1: *1.00–51.44,* (max 1.83); *120.95–174.58,* (max 1.63); |
| decay (*32–698*) | ≥ 1: *61.97–207.10,* (max 1.76); |
| dfault (*1–261*) | ≥ 2: *1.00–12.98,* (max **2.54**); |
| dfvol (*1–250*) | ≥ 1: *1.00–63.52,* (max 1.76); |
| dg14 (*1–243*) | ≥ 2: *1.00–6.32,* (max **2.05**); |
| dg16 (*1–250*) | ≥ 1: *12.97–41.42,* (max 1.59); |
| dm_oran (*1–125*) | ≥ 2: *25.98–47.84,* (max **2.90**); *68.60–78.91,* (max **2.28**); *94.93–98.29* (max **2.12**) |
| dm_oryw (*1–109*) | ≥ 2: *14.42–41.80,* (max **2.69**); *61.07–86.93,* (max **4.41**); |
| dm_oyg3 (*1–60*) | ≥ 2: *13.41–30.29,* (max **7.79**); |
| dultra (*1–436*) | ≥ 2: *1.00–34.95,* (max **3.43**); |
| dunconf (*1–336*) | ≥ 2: *1.00–39.15,* (max **2.36**). |

Abbreviations for categorical ISPs are as in Table 1: $a_{1–3}$ carbonate alteration units, $g_{1–17}$ bedrock geological units and $m_{1–5}$ metamorphism units. The continuous field ISPs, **CNT12**, are also listed in Table 1. Next to abbreviation of continuous ISPs is, in italics, the value range within brackets. Values are bold if ELR ≥ 2.00. In italics is the corresponding range of classes, with maximum ratio in brackets

presence of the occurrences is twice that in the remainder of the SA. For categorical ISPs, the ELR values are ratios between normalized frequencies of units within the presence of the **au37** with the ratios in the remainder of the SA. For continuous ISPs, we have the corresponding density functions and ELR functions. Two examples are shown in Figure 3. The ratios provide a measure of contrast between the areas of **au37** and the remainder of the SA assumed to be devoid of known occurrences. Identical ratios were obtained for the different neighborhoods from 1 × 1 to 7 × 7 pixels.

The Red Lake study area database was also used to document SPM, a spatial prediction modeling software (Fabbri and Chung 2012), which was later developed further into the spatial target mapping program, STM, used here (Fabbri et al. 2017), and to highlight the likely effect of conditional dependence (Chung and Fabbri 2013).

## FOUR CRITICAL QUESTIONS IN PREDICTION MODELING

Of the many mathematical models used to establish and integrate spatial relationships to gen-erate *prediction patterns*, we have used two in our study: the empirical likelihood ratio function, ELR, and the logistic discriminant analysis function, LOG, mentioned in "Definitions and Assumptions for Favorability Function Modeling" section. The four critical questions discussed here, however, do apply to any kind of model and modeling procedure. They are as follows: (1) "How 'good' is my *prediction pattern*?"; (2) "How uncertain is the rank member-ship in the *prediction pattern*?"; (3) "What is the impact of existing redundancy in the indirect sup-porting patterns?"; and (4) "How congruous are the settings of mineral occurrences used as direct sup-porting patterns?".

Those questions are considered as fundamental for realistic evaluation of *prediction patterns* and for their use in decision making in further mineral exploration. We will discuss each one in the analysis of the database for the Red Lake study area antici-pated in the previous section.

### How "Good" is My Prediction Pattern?

Mathematical models integrate empirical like-lihood ratios for each pixel of all ISPs into prediction
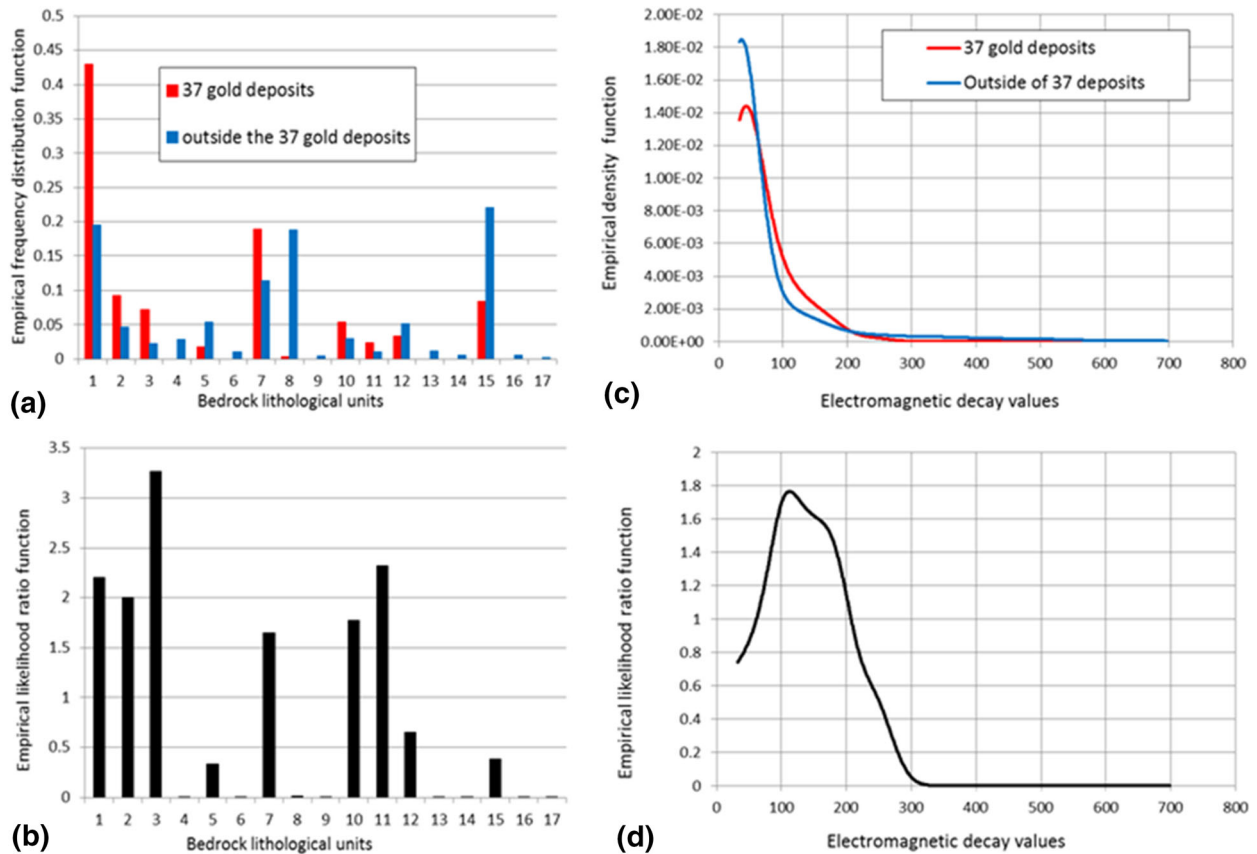
**Figure 3.** Two examples of empirical frequency distribution and density functions in (**a**) and (**c**) for bedrock lithologies and electromagnetic decay values, respectively, and corresponding likelihood ratio functions in (**b**) and (**d**).

scores that make up a *prediction image*. The image itself and its scores are not interpretable as such. The scores generated by different models, here ELR and LOG, are also not directly comparable due to their different meanings. However, comparability and interpretability are obtained at the level of equal area ranking of the scores once sequenced in decreasing order from maximum to minimum. Such ranks provide what we have termed a *prediction pattern*. We have found it convenient to generate 200 equal area ranks, each of 0.5% of the SA, i.e., of approximately 3730 pixels (0.5% of 746,085 ≈ 3730). Figure 4 shows the relationships between the integrated scores in the *prediction image* and the 200 corresponding equal area ranks of the *prediction pattern*.

A fixed set of rank classes permits the visual comparison of *prediction patterns* from the two models as shown in Figure 5. In the illustration, the classes are narrower for the higher and more inter-

esting classes: top 1%, next 1.5%, next 2.5%, …, next 12.5%. We can observe spatial variations in the equal area classes and in the corresponding location of the **au37** shown as black points. Some occurrences fall on high ranks, while others fall on low ranks.

There are indeed some differences in the spatial configurations of classes of the two *prediction patterns* in Figure 5. However, we do not know yet the effectiveness of the patterns as predictors, i.e., as representing the likelihood of future occurrences. For that, we need to know the distribution of future gold occurrences. This is done by pretending not to know the location of one of the 37 occurrences in order to use the remaining 36 to generate another *prediction pattern* and use the 37th to validate its classification obtaining its prediction rank. Iterating such sequential exclusion of one occurrence provides 37 *prediction patterns* and the respective 37 prediction ranks, one for each validation occurrence (each representing $1/37 = 2.7\%$ in proportion of
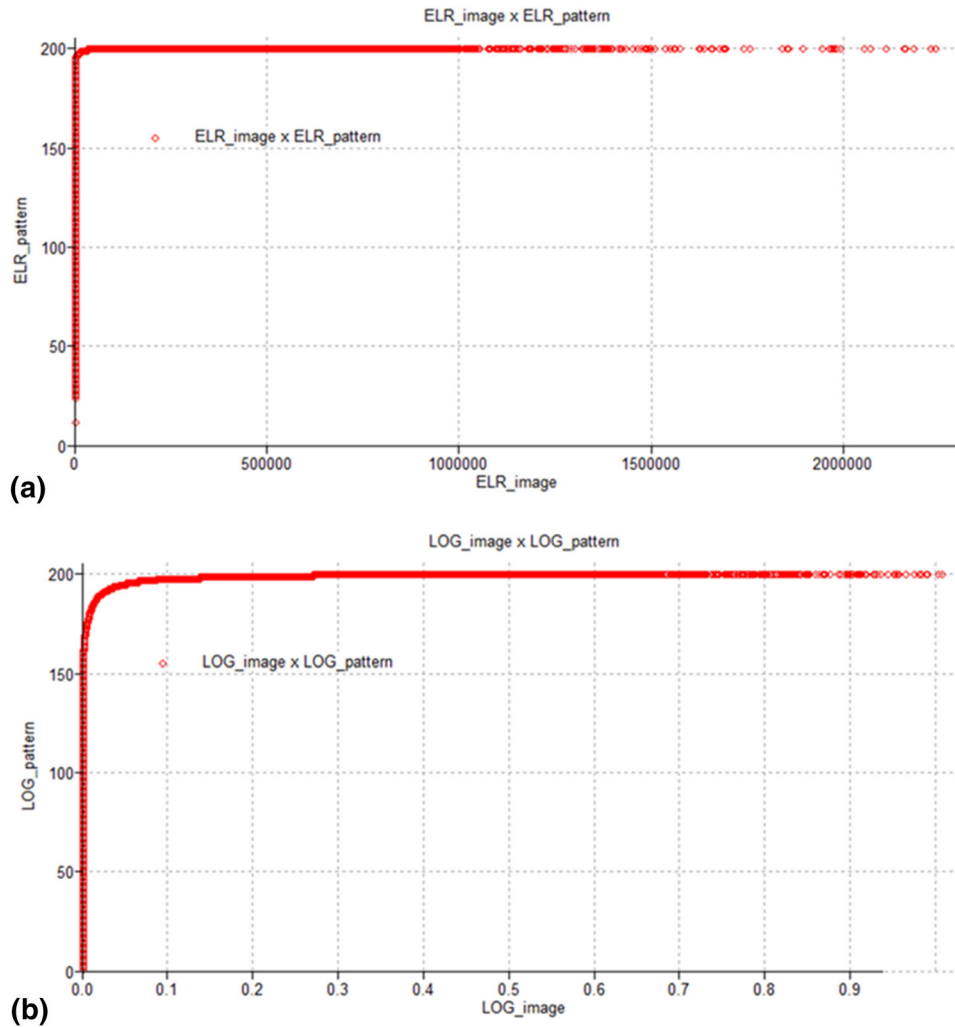
**Figure 4.** Relationships between 746,085 *prediction image* scores and *prediction pattern* ranks for ELR modeling in (**a**) and LOG modeling in (**b**).

occurrence neighborhoods). Prediction rate curves are obtained by plotting the ordered cumulative equal area ranks on the horizontal axis vs. the corresponding cumulative proportion of occurrence pixels in each rank.

Figure 6a shows such curves for the patterns resulting from the two models using the same iterative procedure of sequential exclusion of one occurrence and the same 15 ISPs. The patterns are short named as **ELR_au37m1_agmCNT12** and **LOG_au37m1_agmCNT12**, respectively, where **m1** signifies minus one. The LOG curve shows a somewhat better prediction, particularly for the top rank intervals from 5% to 15%. The top 5, 10, 15, 20 and 30% ranks contain 32, 46, 62, 78 and 92% of the gold

occurrences for the ELR curve, while the LOG curves show 38, 57, 64, 81 and 92%, respectively.

The two curves show meaningful parts for the top 30% ranks, and we can explore further those ranks looking at the respective histograms shown in Figure 6b and c. There, each short column corresponds to one individual occurrence or 2.7% in area proportion of the 333 pixels. Only 17 of those are located within the top 10% in the ELR pattern, while there are 20 in the LOG pattern. That explains the higher values in the LOG curve of Figure 6a. The top 30% ranks contain 34 of the 37 occurrences in either *prediction pattern* so that altogether the curves show a strong similarity of ''goodness'' of patterns.
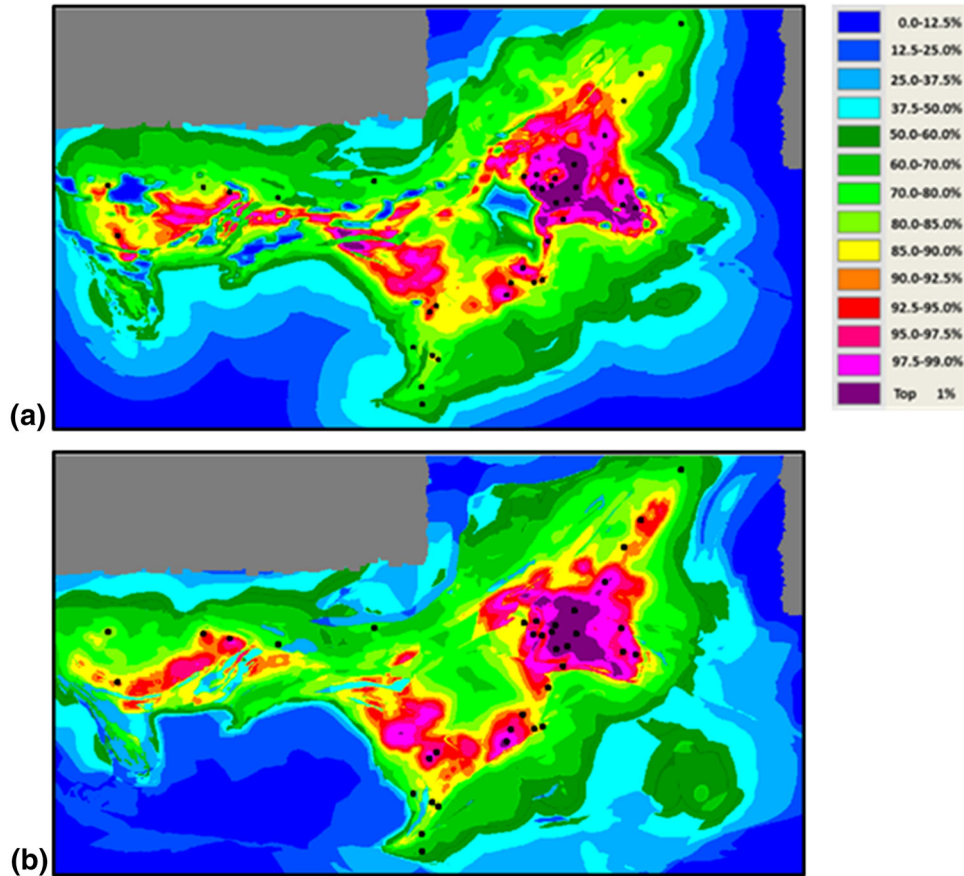
**Figure 5.** *Prediction patterns* of mineral occurrences in the Red Lake study area. In (**a**) using the ELR model, the **au37** occurrences as DSP and the 15 ISPs, **agmCNT12**; in (**b**), using the LOG model and the same ISPs. The size of the 37 3 × 3 pixel neighborhoods, in black, is exaggerated for visibility.

Another aspect of interpreting the *prediction patterns* is that of generating by aggregation equal area ranks to represent monotonically decreasing histograms, considering the monotonic property as implicit assumption in the classification. An example is shown in Figure 7 for the two *prediction patterns* using classes of 7.5% of the SA for the top 30%. The property is visible for the ELR pattern's histogram with a gentler decreasing sequence, but it is not obtained for the steeper LOG pattern histogram's decreasing sequence.

The ''goodness'' of the patterns is only a relative measure of predictability within the database, and it remains to be interpreted not only in numerical terms but also in geological terms. How many classes can we identify in our *prediction patterns*? How to name them as very high, high, intermediate, low, etc.? These are still open questions requiring specific and extensive research going be-

yond the arbitrariness of our selection of classes and their evaluation used here.

In addition, the differences in spatial expressions that we can observe in the *prediction patterns* of Figure 5 easily lead to suspect the existence of spatial instability caused by the uncertainty in rank membership. We are going to consider this uncertainty in the next subsection.

### How Uncertain is the Rank Membership in the Pattern?

To obtain the ELR and the LOG prediction rate curves of Figure 6a, 37 *prediction patterns* were generated by the **au37m1** procedure of sequential exclusion of one occurrence. The two curves show some small differences but also show an overall similarity of prediction capability. We can explore further the properties of the two sets of patterns to
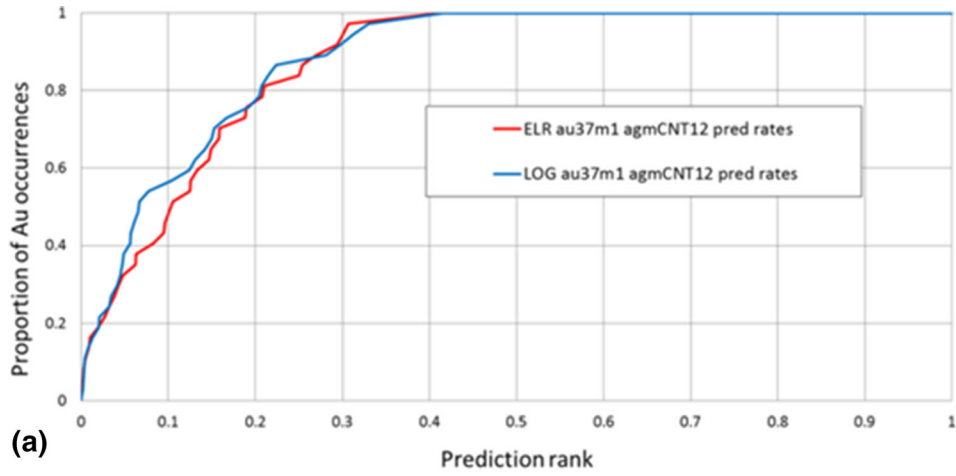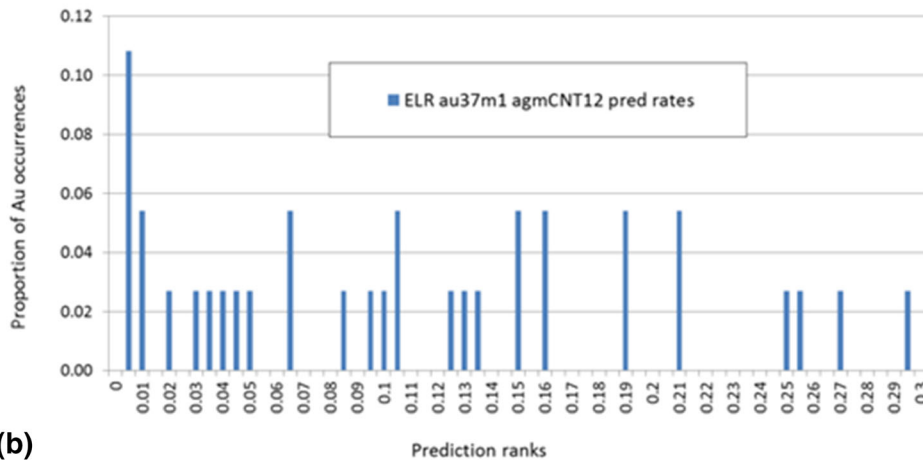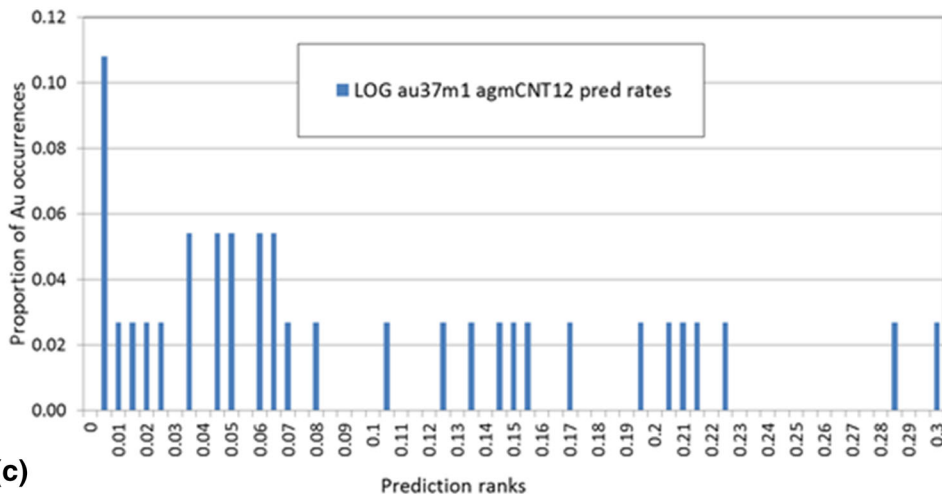
**Figure 6.** Prediction rate curves and histograms. In (**a**) are the curves for ELR and LOG modeling with **au37m1** and **agmCNT12** as ISPs. In (**b**) is the top 30% rank histogram for ELR and in (**c**) for LOG modeling.
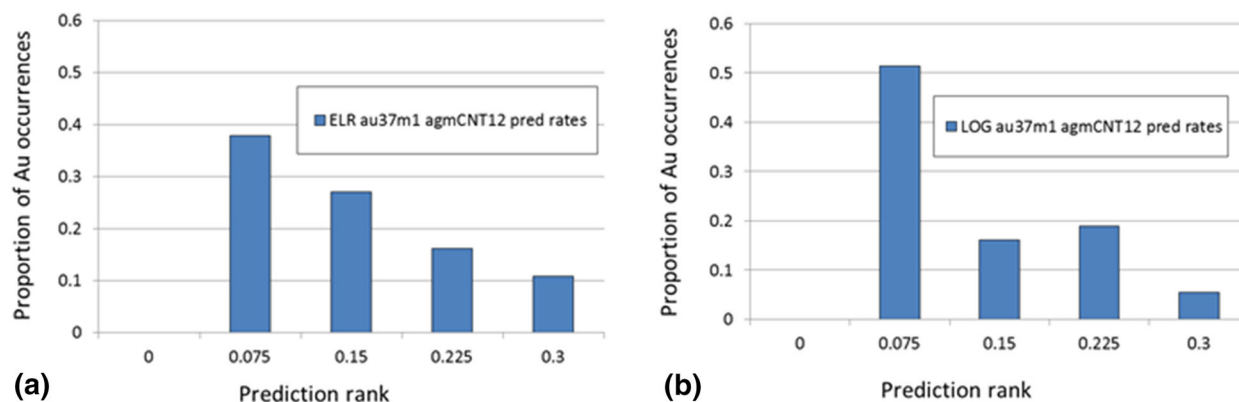
**Figure 7.** Prediction rate histograms of 7.5% classes for the top 30% ranks. In (**a**) is the histogram for ELR and in (**b**) that for the LOG modeling.

assess the uncertainty of rank membership. By statistical analysis, we can produce what we have named as *target* and *uncertainty patterns*. The *target pattern* is the combination of the *prediction patterns* for each set of 37 iterations. Using rank-based statistics, for instance, the *target patterns* contain for each pixel the median rank of the 37 ranks. The corresponding ELR and LOG *uncertainty patterns* rank the ranges around the median of the 37 ranks. Rank-based statistics is very robust, and the corresponding *uncertainty patterns* are shown in Figure 8a and c. The two *target patterns* are not shown here because they are undistinguishable from the *prediction patterns* of Figure 5 that uses the same legend and threshold colors. Figure 8b and d shows the 50% *combination patterns* of *uncertainty* and *target patterns*. It is one way to interpret and classify the uncertainty of rank membership: We arbitrarily selected a percentage of lower uncertainty ranks, say 50%, to visualize the ranks in the *target pattern*.

In Figure 8b and d, we have 50% *combination patterns* so that we can see that several gold occurrences are located in areas of uncertainty higher than the lower 50% ranks of the *uncertainty pattern*. *Prediction*, *target* and *uncertainty pattern* ranks for the 37 gold occurrences are shown in Table 3 in units of 1000*.

In our case, while 18 occurrences rank ≥ 900 (top 10% ranks) for the ELR modeling in column 2, PELR, 20 occurrences are in that rank in the LOG modeling in column 5, PLOG. The corresponding target ranks (median) show 25 occurrences with ranks ≥ 900 for both TELR and TLOG in columns 3 and 6, respectively.

As to the uncertainty ranks, UELR and ULOG in columns 4 and 7, we have that 15 occurrences are in lower uncertainty zones in the ELR *target pattern*, ≤ 500, and 27 in the LOG *target pattern*, respectively. Figure 9 compares visually as histograms the same ranks for the two models, ELR and LOG, using the values in Table 3. It appears that the *target pattern* from the LOG prediction is affected by less uncertainty than that from the ELR prediction.

It is not possible, however, to compare such uncertainties due to the relative significance of these measures. Effectively, it becomes a new area of research to identify ways to measure such differences. For instance, if instead of using rank-based statistics with median rank for *target* and rank of range for *uncertainty patterns*, we use basic statistics such as average and variance, the 50% *combination pattern* would show that all the 37 occurrences are located in areas of higher uncertainty. It means that, although there are similarities between the two *combination patterns* in Figure 8b and d for the ELR and the LOG models, we can also detect the spatial instability of the less uncertain ranks.

## What is the Impact of Strong Data Redundancy (Conditional Dependence)?

The ISPs are to represent the geologic conditions of mineralization. Thus, they all relate to the process leading to the occurrences so that, naturally and potentially, they contain redundant information. The
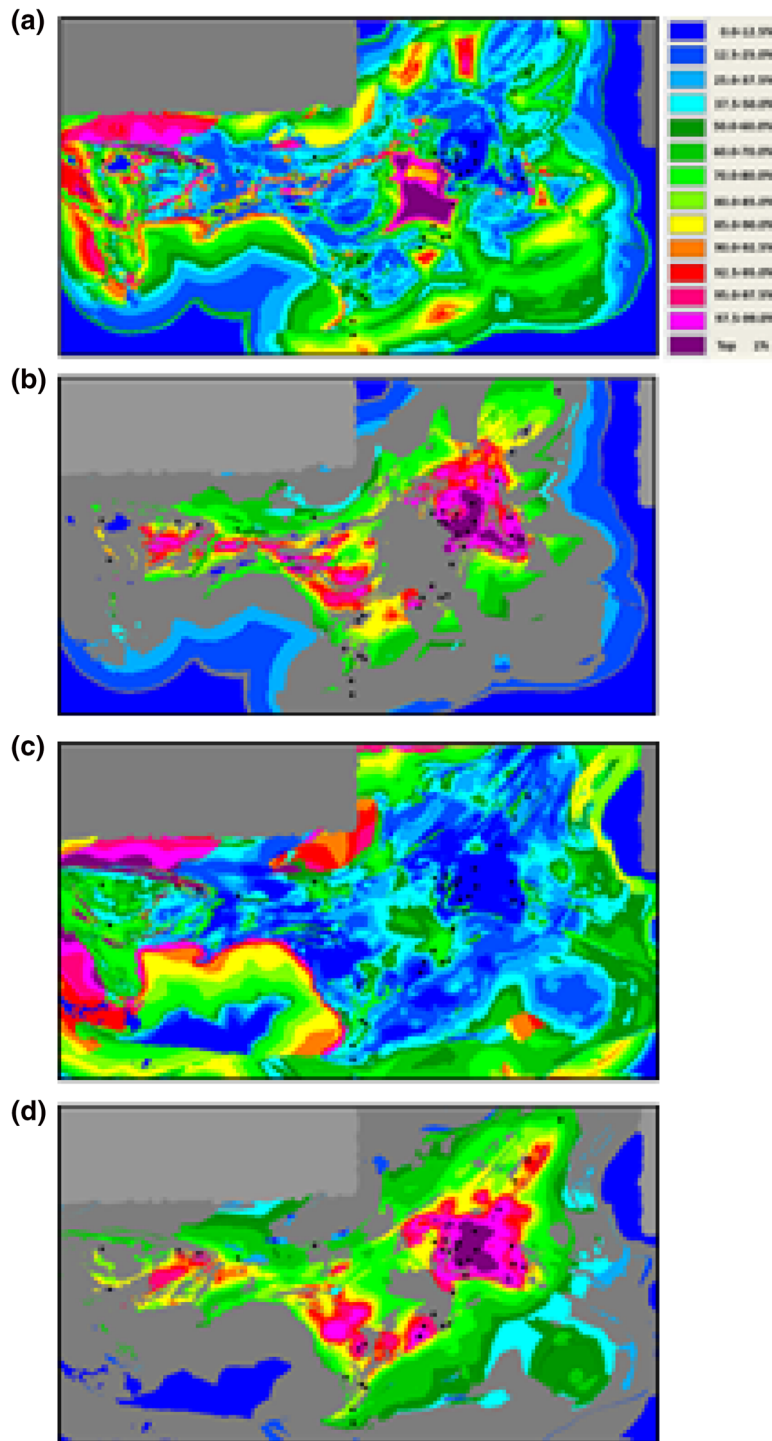
**Figure 8.** *Uncertainty patterns and 50% combination patterns* of mineral occurrences in the Red Lake study area in (**a**), for ELR using the **au37m1** iterative cross-validation and the 15 ISPs, **agmCNT12**. In (**b**), the corresponding 50% *combination pattern* of uncertainty and target. In (**c**) is *uncertainty* as in (**a**) but for LOG and in (**d**) *combination* as in (**b**) but for LOG. The size of the 37 3 × 3 pixel neighborhoods in black is exaggerated for visibility.

presence of a particular geologic unit, for instance, whose normalized frequency within the occurrence neighborhoods is well higher than that in the remainder of the SA cannot be conditionally independent from the distance from its contacts, the boundary of the same unit. Consider the empirical likelihood ratio for categorical unit $g_7$ in Table 2, 1.65, and continuous value field dg14 with ratios $\geq 2$ between distances 1.00 and 6.32, with a maximum value of 2.05.

Such redundancy helps in the recognition of important features in the classification, but it might also put anomalous weight on the intensity of the spatial relationship between the DSP and the ISP. On the other hand, the ISPs integrated in the modeling can also compensate one another, thus contributing to a better classification of the SA.

The experiments made are exploring how extremely exaggerated redundancy can affect the

**Table 3.** Ranks in 1000* units for prediction, target and uncertainty for the 37 gold occurrences, Red Lake area

| Au occurrences | Rank-based statistics: rank, median rank and rank of range | | | | | |
|---|---|---|---|---|---|---|
| 1<br>ID | 2<br>PELR | 3<br>TELR | 4<br>UELR | 5<br>PLOG | 6<br>TLOG | 7<br>ULOG |
| 1 | 694 | 734 | 673 | 688 | 712 | 425 |
| 2 | 842 | 880 | 670 | 924 | 944 | 177 |
| 3 | 843 | 893 | 768 | 870 | 895 | 359 |
| 4 | 958 | 985 | 406 | 980 | 986 | 48 |
| 5 | 998 | 999 | 3 | 988 | 991 | 39 |
| 6 | 982 | 990 | 132 | 956 | 967 | 103 |
| 7 | 905 | 947 | 706 | 834 | 861 | 416 |
| 8 | 939 | 955 | 273 | 997 | 998 | 30 |
| 9 | 707 | 732 | 463 | 671 | 709 | 666 |
| 10 | 953 | 976 | 371 | 954 | 964 | 76 |
| 11 | 854 | 928 | 854 | 794 | 847 | 757 |
| 12 | 974 | 985 | 137 | 993 | 994 | 34 |
| 13 | 747 | 844 | 926 | 797 | 939 | 930 |
| 14 | 996 | 997 | 4 | 998 | 998 | 28 |
| 15 | 963 | 992 | 241 | 969 | 980 | 77 |
| 16 | 992 | 996 | 6 | 944 | 961 | 83 |
| 17 | 906 | 940 | 635 | 934 | 957 | 230 |
| 18 | 731 | 753 | 468 | 703 | 725 | 343 |
| 19 | 1000 | 1000 | 1 | 997 | 998 | 29 |
| 20 | 998 | 999 | 3 | 999 | 999 | 27 |
| 21 | 969 | 987 | 220 | 945 | 958 | 97 |
| 22 | 991 | 997 | 9 | 960 | 974 | 99 |
| 23 | 938 | 976 | 592 | 936 | 962 | 300 |
| 24 | 875 | 942 | 842 | 851 | 909 | 775 |
| 25 | 793 | 972 | 981 | 877 | 909 | 563 |
| 26 | 896 | 977 | 865 | 897 | 939 | 676 |
| 27 | 876 | 949 | 853 | 848 | 909 | 796 |
| 28 | 811 | 891 | 879 | 859 | 896 | 681 |
| 29 | 852 | 909 | 811 | 952 | 967 | 119 |
| 30 | 918 | 968 | 737 | 981 | 989 | 53 |
| 31 | 868 | 909 | 699 | 940 | 966 | 313 |
| 32 | 900 | 942 | 691 | 967 | 979 | 90 |
| 33 | 751 | 785 | 661 | 777 | 803 | 451 |
| 34 | 813 | 865 | 778 | 786 | 843 | 725 |
| 35 | 791 | 833 | 729 | 809 | 839 | 455 |
| 36 | 700 | 765 | 842 | 720 | 754 | 520 |
| 37 | 592 | 652 | 830 | 586 | 612 | 352 |
| # $\geq$ 900 | 18 | 25 | | 20 | 25 | |
| # $\leq$ 500 | | | 15 | | | 27 |

Iterative prediction process **au37m1** with ELR and LOG models: P is for prediction, T is for target and U for uncertainty ranks. Uncertainty values $\leq 500$ indicate the 50% lower ranks
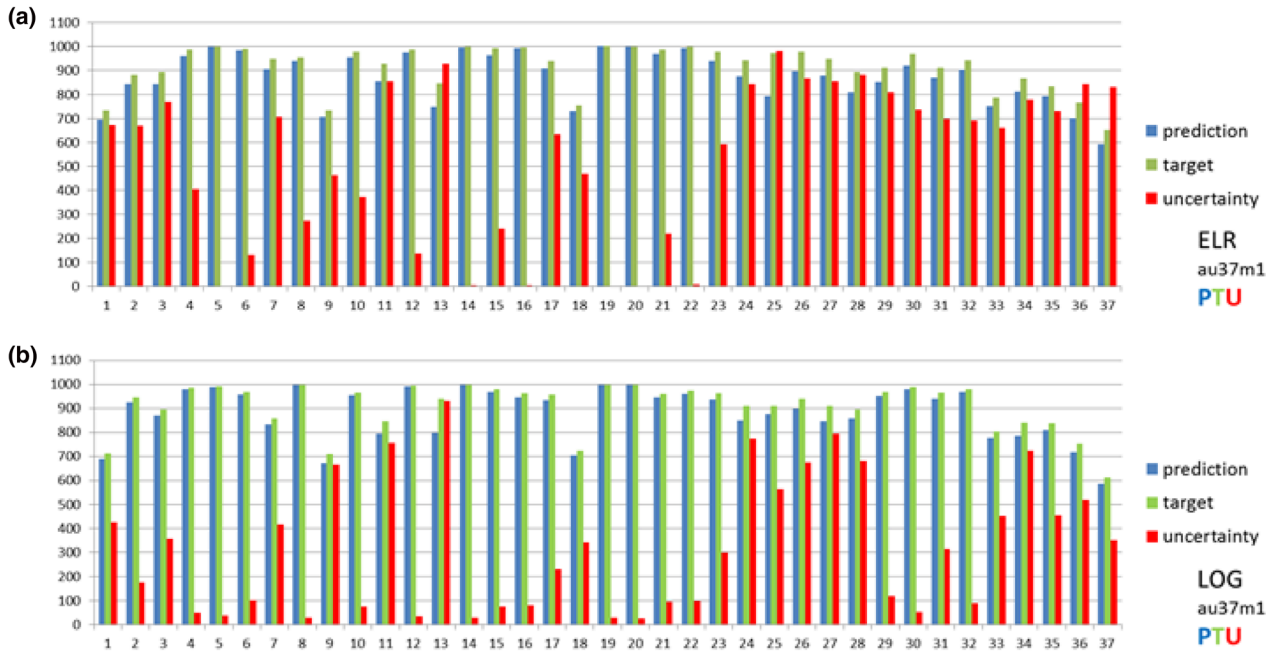
**Figure 9.** Prediction, target and uncertainty ranks in 1000* units for the **au37** occurrences using iterative cross-validation **au37m1** with model ELR in (**a**) and with model LOG in (**b**).

rankings of high and low classes of *prediction patterns* and whether the redundancy observed leads to still acceptable *prediction patterns*. The categorical ISPs, **agm**, show high ELR values for units $a_1$, $g_1$ to $g_3$, $g_{11}$, $m_1$ and $m_2$ (i.e., 6.29, 2.21, 2.00, 3.27, 2.32, 4.05 and 2.29, respectively, as shown in Table 2). ELR and LOG *target patterns* with iterating process **au37m1** were generated adding duplicate categorical ISPs, **ag2 m**, **ag2m2** and **a2g2m2**, with the continuous field ISPs **CNT12**. Figure 10 shows the prediction rate curves for the corresponding top 30% ranks. In the illustrations, the threshold ranks used for pattern visualization are shown as red tick marks. As can be seen, the impact in those curves is minimal, hardly noticeable in the histograms from both the sets of cross-validations in Figure 11a, ELR, and b, LOG, that can be compared with the corresponding histograms of Figure 6b and c. As it turns out, the two corresponding top 30% histograms for predictions using **a2g2m2** show slight shifts for individual 0.5% ranks, but the overall ranking sequence is little affected, particularly for higher ranks. The authors have noticed such a weak impact in several studies (cf. Chung 2006; Fabbri et al. 2014). Possibly, this occurs with some generality within earth science databases and is certainly

something worth some research efforts. Much work on attempting to lessen or to eliminate conditional dependence might prove ineffective or counterproductive in spatial modeling.

## How Congruous are the Mineral Occurrences Used in Modeling?

We have assumed in our modeling until now that the 37 3 × 3 pixel neighborhoods used as DSP all belong to a congruous or cohesive type of setting. However, we know by now that out of the iterative cross-validation process, **au37m1**, part of the occurrences are distributed in high-rank classes and part in low-rank classes in the *prediction patterns* from either model. Do they belong to a broad type of setting or do they represent a mixture of more than one setting?

One way to explore such a situation is of applying the iterative process of sequential selection of one occurrence neighborhood. The result of such a modeling strategy is a set of 37 *prediction patterns* each using only one occurrence neighborhood as DSP and generating the prediction ranks for the remaining 36. This produces a 37 × 37 array of
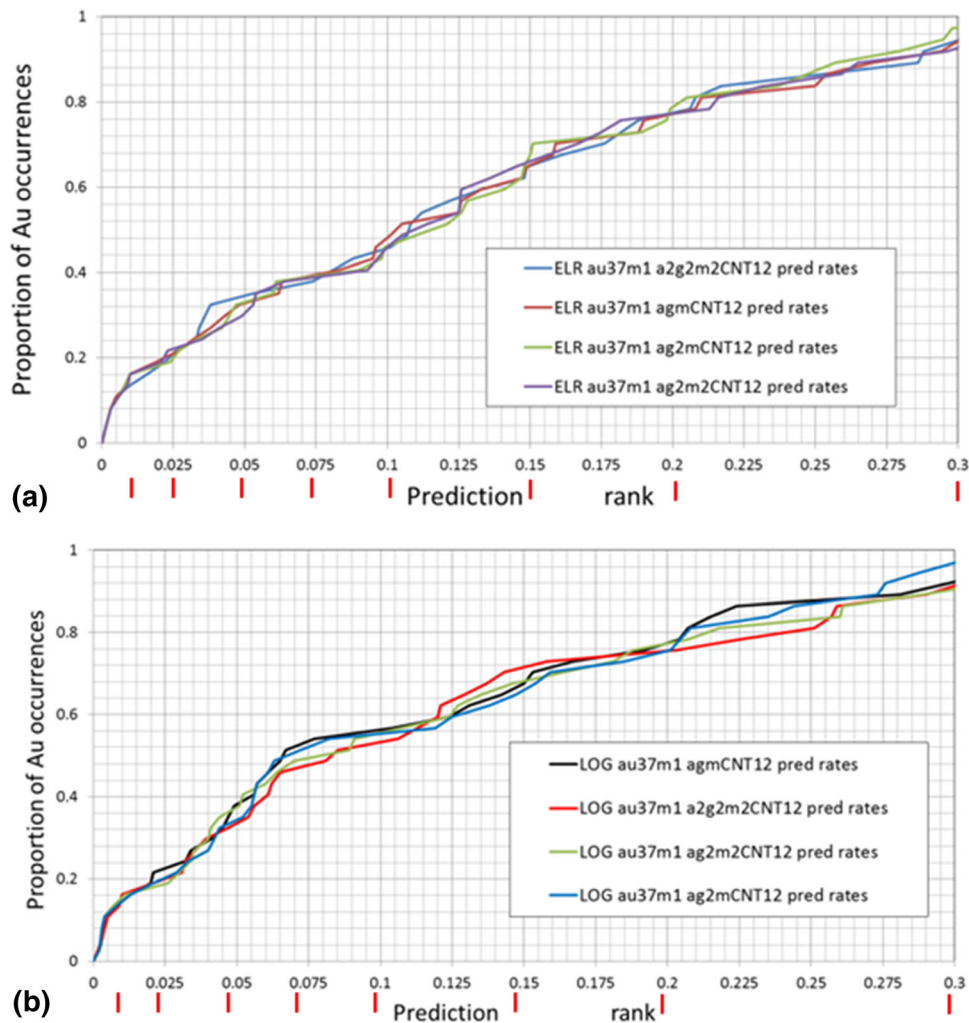
**Figure 10.** Prediction rate curves for patterns generated with and without duplicate categorical ISPs in addition to the continuous fields. In (**a**) the ones from the ELR model and in (**b**) those from the LOG model. The red tick marks indicate the threshold values for visualizing the *prediction pattern* top 30% ranks.

ranks with null values in the diagonal. By arbitrarily choosing a threshold rank in units of 1000*, say 900 (i.e., the top 10%), we can count how many occurrences are predicted well with rank $\geq 900$ and at the same time we can predict well occurrences with rank $\geq 900$.

We have used the ELR function model for this and considered as well predicted/predicting, **Wpp**, those predicted well by at least eight as DSP and predicting well at least five occurrences. The remaining ones were considered as poorly predicted/

predicting, **Ppp**. They are termed **19Wpp** and **18Ppp**, respectively, occupying 171 and 162 pixels. Their characteristic signature in terms of ELR values is given in Tables 4 and 5, while their distribution in the SA is shown in Figure 2b as black dots and red dots, respectively. For **ELR_au19Wpp-m1_agmCNT12** and **ELR_au18Pppm1_agmCNT12**, the *target patterns* are shown in Figure 12 and their prediction rate curves are shown in Figure 13.

As we can see in Table 4, most of the categorical support is due to units $a_1$, $g_1$ and $m_1$ (10.43, 3.24
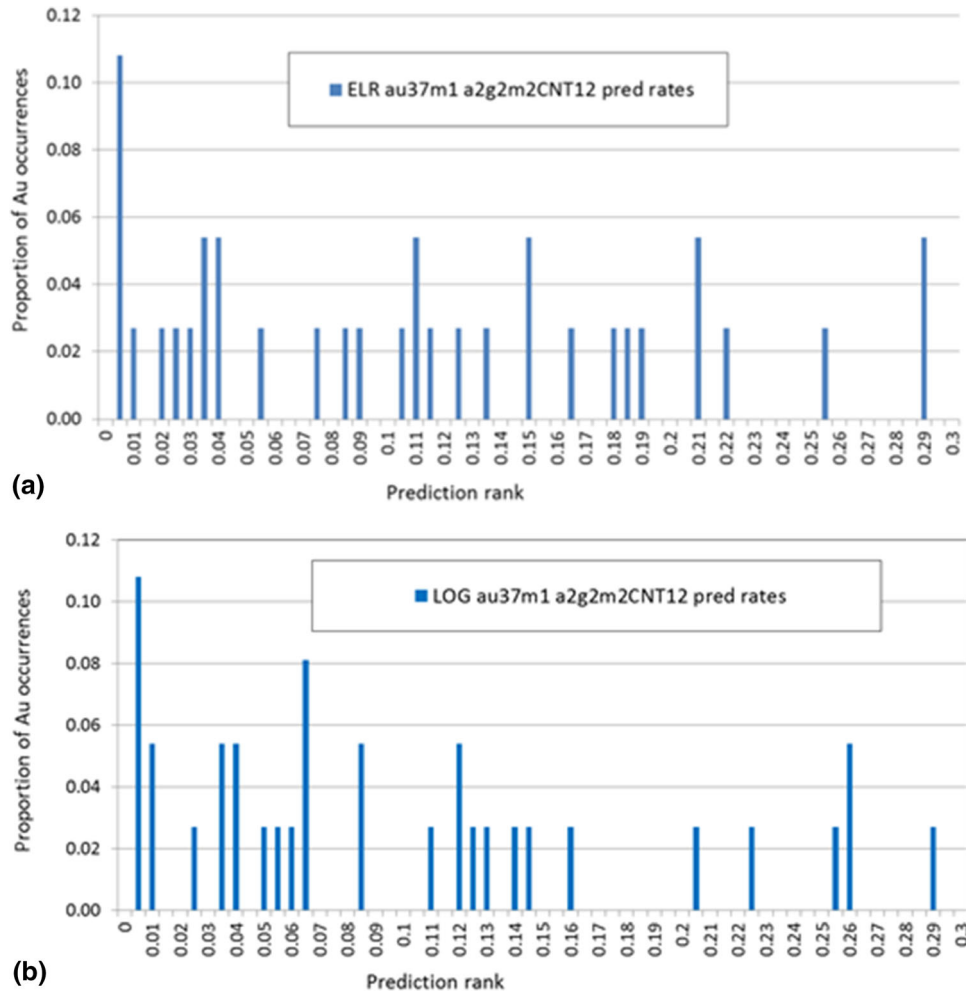
**Figure 11.** Prediction rank histogram for patterns with duplicated categorical ISPs: **a2g2m2CNT12**, for ELR model in (**a**) and LOG model in (**b**). Only the top 30% ranks are shown. Histograms are to be compared with the ones in Figure 6b and c.

and 6.20) using **au19Wpp**. From **au18Ppp** in Table 5, we can see that the categorical support is from the units: $a_1$, $a_2$, $g_2$, $g_3$, $g_7$, $g_{10}$, $g_{11}$ and $m_2$.

## DISCUSSING THE RESULTS

The iterative process **au37m1** used in ''How 'good' is My Prediction Pattern?'' section had been selected after a set of sequential exclusions of 1, 3, 5, 7, 9 and 18 occurrences, all providing cumulative prediction rate curves running below the first one. Given their relatively small number, choosing **au37m1** appeared a robust way of evaluating the prediction rates of all 37 occurrences by cross-validation. Instead of having hundreds of occurrences, it

is preferable to have a greater number of sequentially eliminated occurrences. Additional iterative analyses with random selection of 24 occurrences of the 37 iterated 1 and 3 times showed higher prediction rate curves than those of **au37m1** but when iterated 9, 17 and 37 times showed similar curves for either model.

Concerning the prediction rate curves in Figure 6a, model curves could be fitted or a threshold selected to identify the rank at the start of an inconvenient cost–benefit situation in which the distribution of occurrences becomes erratic, or directly proportional to the area increment (cf. Chung and Fabbri 2003). Furthermore, the strong similarity of the two prediction rate curves for ELR and LOG in Figure 6 is suggestive of a stronger influence of

**Table 4.** ELR values are shown for predictions using as DSP the distribution of the **au19Wpp** gold occurrences, with **agmCNT12** as ISPs

| DSP and ISPs | Empirical likelihood ratios |
|---|---|
| **au19Wpp** | |
| c. alteration | $a_1$ = **10.43**, $a_2$ = 0.82, $a_3$ = 0.88; |
| geology | $g_1$ = **3.24**, $g_2$ = 1.51, $g_3$ = 1.06, $g_7$ = 0.92, $g_{10}$ = 1.54, $g_{11}$ = 0.00, $g_{12}$ = 0.57, $g_{15}$ = 0.25; |
| metamorphism | $m_1$ = **6.20**, $m_2$ = 1.71, $m_3$ = 0.87, $m_4$ = 0.14; |
| dalter *(0.1–268.0)* | ≥ 2: *0.10–28.48,* (max **5.43**); |
| dbif *(1–398)* | ≥ 2: *1.00–39.53,* (max **3.24**); |
| decay *(32–698)* | ≥ 1: *61.31–217.10,* (max 1.89); |
| dfault *(1–261)* | ≥ 2: *1.00–21.31,* (max **2.95**); |
| dfvol *(1–250)* | ≥ 1: *8.95–68.23,* (max 1.88); |
| dg14 *(1–243)* | ≥ 2: *6.80–16.45,* (max **2.10**); |
| dg16 *(1–250)* | ≥ 1: *8.98–39.17,* (max 1.65); *62.12–69.36,* (max 1.01); *91.82–142.22,* (max 1.43); |
| dm_oran *(1–125)* | ≥ 2: *18.28–50.73,* (max **4.30**); |
| dm_oryw *(1–109)* | ≥ 2: *14.12–44.50,* (max **4.42**); |
| dm_oyg3 *(1–60)* | ≥ 2: *7.36–8.83,* (max **2.22**); *12.95–31.60,* (max **10.94**); |
| dultra *(1–436)* | ≥ 2: *1.00–39.32,* (max **4.46**); |
| dunconf *(1–336)* | ≥ 2: *15.72–44.49,* (max **2.65**). |

Abbreviations for categorical ISPs are as in Table 1. The continuous field ISPs, **CNT12**, are also listed in Table 1. Next to abbreviation of continuous ISPs is in italics the value range within brackets. Values are in bold if ELR ≥ 2.00. In italics is the corresponding range of classes, with maximum ratio in brackets

**Table 5.** Empirical likelihood ratios are shown for predictions using as DSP the distribution of the **au18Ppp** gold occurrences, with **agmCNT12** as ISPs

| DSP and ISPs | Empirical likelihood ratios |
|---|---|
| **au18Ppp** | |
| c. alteration | $a_1$ = **2.21**, $a_2$ = **2.53,** $a_3$ = 0.32; |
| geology | $g_1$ = 1.11, $g_2$ = **2.52**, $g_3$ = **5.60**, $g_7$ = **2.42**, $g_{10}$ = **2.03**, $g_{11}$ = **4.77,** $g_{12}$ = 0.73, $g_{15}$ = 0.53; |
| metamorphism | $m_1$ = 1.79, $m_2$ = **2.89**, $m_3$ = 0.92, $m_4$ = 0.74; |
| dalter *(0.1–268.0)* | ≥ 2: *0.10–16.71,* (max **5.18**); |
| dbif *(1–398)* | ≥ 2: *123.35–173.39,* (max **3.10**); |
| decay *(32–698)* | ≥ 1: *63.33–197.94,* (max 1.89); |
| dfault *(1–261)* | ≥ 2: *1.00–3.82,* (max **2.12**); |
| dfvol *(1–250)* | ≥ 1: *1.00–56.12,* (max 1.67); |
| dg14 *(1–243)* | ≥ 2: *1.00–6.32,* (max **2.39**); |
| dg16 *(1–250)* | ≥ 1: *16.99–42.97,* (max 1.64); *71.69–89.93,* (max 1.17); *115.16–164.62,* (max 1.85); |
| dm_oran *(1–125)* | ≥ 2: *58.92–103.19,* (max **5.05**); |
| dm_oryw *(1–109)* | ≥ 2: *54.23–89.57,* (max **9.73**); |
| dm_oyg3 *(1–60)* | ≥ 2: *14.61–26.00,* (max **4.49**); |
| dultra *(1–436)* | ≥ 2: *1.00–18.85,* (max **2.46**); |
| dunconf *(1–336)* | ≥ 2: *1.00–27.76,* (max **3.38**). |

Abbreviations for categorical ISPs are as in Table 1. The continuous field ISPs, **CNT12**, are also listed in Table 1. Next to abbreviation of continuous ISPs is in italics the value range within brackets. Values are in bold if ELR ≥ 2.00. In italics is the corresponding range of classes, with maximum ratio in brackets

the database characteristics over that of the type of prediction model. The problem of interpreting the cross-validation and the prediction rate curve remains open to date.

As to the *uncertainty patterns* of range of ranks obtained, they revealed 15 and 27 occurrences with uncertainty rates below 500 in Table 3. The associated 50% *combination pattern*s in Figure 8 showed much of the *target patterns* of median ranks with uncertainty rate below 500. However, when using different statistics, for example basic statistics, sample average and variance, or sample median and
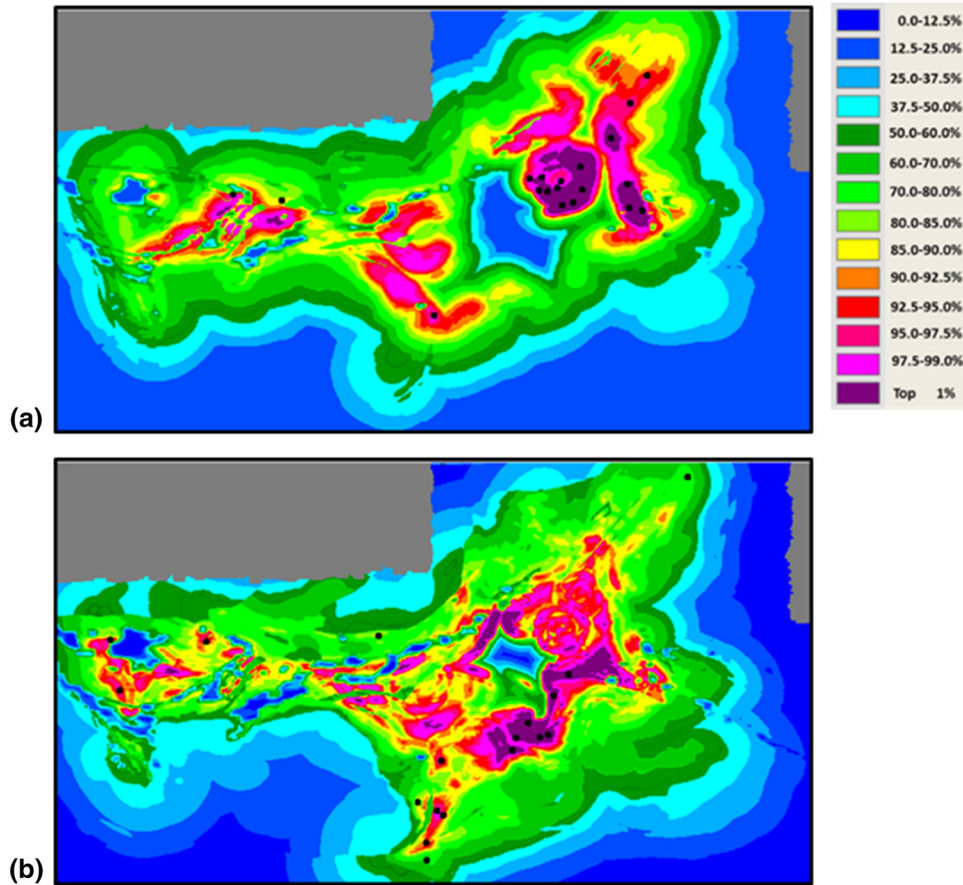
**Figure 12.** *Target patterns* of mineral occurrences in the Red Lake study area. In (**a**) using the ELR model, the **au19Wppm1** occurrences as DSP and the 15 ISPs, **agmCNT12**; in (**b**), using the **au18Pppm1** as DSP and the same ISPs. Size of $3 \times 3$ pixel neighborhoods in black is exaggerated.

range (or even Jackknife statistics with average and variance), the resulting *uncertainty patterns* showed very high ranks all well above 500 so that all 37 occurrences fall on high uncertainty rates. Such a condition has been found also with *target patterns* of natural hazard, in spite of the difference in meaning they have. The future mineral occurrences are yet to take place differently from the mineral deposit occurrences, which have been deposited but are not yet discovered (if not destroyed by erosion or other processes). Nevertheless, one would be inclined to believe that uncertainty of class membership in natural hazard prediction modeling is less than that in modeling mineral deposit discoveries. A situation to be studied should such a comparison be possible! Can we model a target pattern minimizing the uncertainty of class membership or is the uncertainty a property of the database? How to manage such

uncertainty? These additional questions need attention.

The prediction rate curves in Figure 10 and the histograms in Figure 11 reveal a minimal impact of data redundancy or strong conditional dependence on the *prediction pattern* (cf. Chung 2006). This kind of result has encountered objections by Schaeben (2014) in his analyses of artificial databases in which the impact of conditional dependence could not be linked to the ranking of natural categorical or continuous map units. As to our findings, instead, could it be that the stronger the dependence the weaker is the effect on the *prediction pattern* and vice versa the weaker is the dependence the stronger is the effect? Such a suspicion would be worth exploring and experimenting on. Should it be possible to hide or to lessen the conditional dependence within a database, we may wonder if the *prediction pattern* ob-
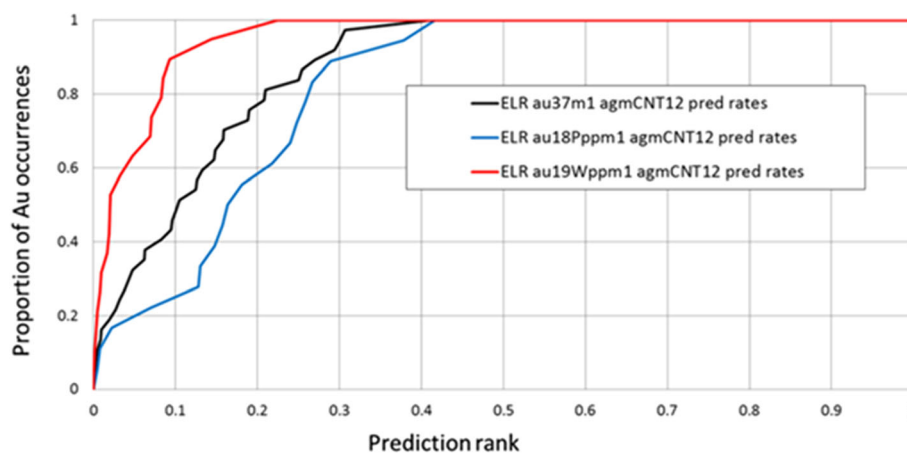
**Figure 13.** ELR prediction rate curves for iterative cross-validations using **au37m1**, **au19Wppm1** and **au18Pppm1** all with **agmCNT12** as ISPs.

tained would be preferable to the one generated with conditionally dependent data. This comparison was made by Fabbri et al. (2014) on a database for landslide hazard prediction. They observed that using also the conditionally dependent supporting pattern improved the geological interpretability and quality of the *prediction pattern*.

The separation of ELR values in Tables 4 and 5 shows that the support of ISPs differs strongly between the two groups of occurrences, **au19Wpp** and **au18Ppp**, in terms of the specific database characteristics, so do the two corresponding *prediction patterns*, **ELR_au19Wppm1_agmCNT12** and **ELR_au18Pppm1_agmCNT12**, shown in Figure 12. This could lead to reconsider the soundness of the group of **au37** occurrences as a uniform genetic type or style of deposit. Other groupings could be formulated either by expert's knowledge or by the modeling statistics. This we deem as a worthwhile analytical challenge.

We believe that the four questions proposed concern critical aspects of prediction modeling using spatially distributed data. In order to offer additional grounds for recommending further focused research, let us consider just two examples of recent studies revealing the actual status of prediction modeling practice in the field of natural hazards. In that field, studies have probably intensified more than in mineral resource exploration. A comparative analysis of 565 peer-review papers published between 1983 and 2016, made by Reichenbach et al. (2018), discovered many inconsistent approaches, databases, models and modeling result evaluations.

A chaotic situation was found that affects the credibility of prediction modeling. Furthermore, a study by Vakhshoori and Zare (2018) showed serious doubts about the use of ROC curves and AUC measures to evaluate and compare the results of prediction modeling. Unfortunately, such a use is common practice in most applications to date.

## CONCLUDING REMARKS

A *favorability modeling* process is iterated with a subset of the known occurrences, and the resulting *prediction patterns* are cross-validated with distribution of the left-out occurrences. The target map, we termed it *target pattern*, originates from integration of all *prediction patterns* from the iterations. Rank-based statistics related to the target maps provides measures of quality, robustness and uncertainty of the classification of a study area into likelihood of discovery. We consider this a general analytical strategy. Much of this is a relatively new area of research, so that to interpret that uncertainty is still a challenge.

The four questions proposed point at fundamental problems of spatial prediction modeling. The application example discussed here exposes what is poorly known or even unknown about *prediction patterns*, and it demands further research. In conclusion, the experiments performed point at the extraction of significant properties of the spatial data that cannot be ignored but that we have yet to

master to substantiate the reliability of *prediction patterns*.

# REFERENCES

Chung, C. F. (1978). Computer program for the logistic model to estimate the probability of occurrence of discrete events. Geological Survey of Canada, GSC paper 78–11 (pp. 1–23).

Chung, C. F. (2003). Use of airborne geophysical surveys for constructing mineral potential map. In W. D. Goodfellow, S. R. McCutcheon, & M. Peter Jan (Eds.), *Massive sulfide deposits of the bathurst mining camp, and northern Maine* (pp. 879–891). Economic Geology Monograph 11, Society of Economic Geology Inc., Littleton, CO.

Chung, C. F. (2006). Using likelihood ratio functions for modeling the conditional probability of occurrence of future landslides for risk assessment. *Computers & Geosciences, 32,* 1052–1068.

Chung, C. F., & Fabbri, A. G. (1993). The representation of geoscience information for data integration. *Nonrenewable Resources, 2,* 122–139.

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

Chung, C.-J., & Fabbri, A. G. (2004). Systematic procedures of landslide hazard mapping for risk assessment using spatial prediction models. In T. Glade, M. G. Anderson, & M. J. Crozier (Eds.), *Landslide hazard and risk* (pp. 139–174). New York, NY: Wiley.

Chung, C.-J., & Fabbri, A. G. (2008). Modular course in quantitative methods for mineral exploration, Department of Earth Sciences, University of Ottawa, Canada, 16–23 February, 2008 (pp. 1–52). Unpublished manuscript.

Chung, C.-J., & Fabbri, A. G. (2013). Modeling target maps of future gold occurrences with combination of categorical and continuous conditionally-dependent supporting patterns. In *Proceedings of 12th SGA biennial meeting*, 2 (pp. 476–479), Uppsala, Sweden, August 12–15.

Chung, C. F., Fabbri, A. G., & Chi, K. H. (2003). A strategy for sustainable development of nonrenewable resources using spatial prediction models. In A. G. Fabbri, G. Gáal, & R. B. McCammon (Eds.), *Geoenvironmental deposit models for resource exploitation and environmental security* (pp. 101–118). Dordrecht: NATO ASI Volume, Kluwer.

Davis, J. C., Chung, C. F., & Ohlmacher, G. C. (2006). Two models for evaluating landslide hazards. *Computers & Geosciences, 32,* 1120–1127.

Fabbri, A. G., & Chung, C.-J. (2008). On blind tests and spatial prediction models. *Natural Resources Research, 17,* 107–118.

Fabbri, A. G., & Chung, C.-J. (2012). A spatial prediction modeling system for mineral potential and natural hazard mapping. In *Proceedings of EUREGEO2012, II* (pp. 756–757), Bologna, Italy, June 12–15, 2012.

Fabbri, A. G., Patera, A., & Chung, C.-J. (2017). Spatial target mapping: An approach to susceptibility prediction based on iterative cross-validations. In E. Alonso, J. Corominas & H. Hürlimann (Eds.), *Proceedings of IX Simposio Nacional sobre Taludes y Laderas Instables* (pp. 468–479). Santander Spain 27–30, June 2017, Barcelona, Centre Internacional de Mètodes Numèrics en Enginyeria (CIMNE).

Fabbri, A. G., Poli, S., Patera, A., Cavallin, A., & Chung, C.-J. (2014). Estimation of information loss when masking conditional dependence and categorizing continuous data. Further experiments on a database for spatial prediction modeling in Northern Italy. In E. Pardo-Igúzquiza, C. Guardiola-Albert, J. Heredia, L. Moreno-Merino, J. J. Duran & J. A. Vargas-Guzmán (Eds.), *Mathematics of planet earth*, *proceedings of the 15th annual conference of the international association for mathematical geosciences* (pp. 291–294). Madrid, 2–6 September 2013. Berlin: Springer.

GSC 2000. Panagapko, D. A. (2000). Preliminary release of geoscience data Red Lake greenstone belt, northwestern Ontario, Geological Survey of Canada Open File 3751.

Hillary, B., Kerswill, J., Keating, P., Harris, J., & Chung, C.-J. (2008). Description and illustration of data used in the Red Lake case study. In *Modular Course in Quantitative Methods for Mineral Exploration, Department of Earth Sciences, University of Ottawa, Canada, 16–23 February, 2008* (pp. 1–28). Unpublished manuscript.

Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., & Guzzetti, F. (2018). A review of statistically based landslide susceptibility models. *Earth Sciences Reviews, 180,* 60–91.

Schaeben, H. (2014). Potential modeling: Conditional dependence matter. *GEM – International Journal of Geomathematics, 5,* 99–116.

Vakhshoori, V., & Zare, M. (2018). Is the ROC curve a reliable tool to compare the validity of landslide susceptibility maps? *Geomatics, Natural Hazards and Risks, 9,* 249–266.