

Editor's Note: The U.S. Geological Survey and the Geological Survey of Canada, with funding from five mining companies, has developed an ArcView GIS extension called Arc-WofE that uses a quantitative method known as weights of evidence for mapping potential mineral sites using GIS. This article describes the weights-of-evidence method and how it was used to develop the Arc-WofE extension. See page 48 for a glossary of the terms used in this article.

Predictive Probabilistic Modeling Using ArcView GIS

By Gary L. Raines, Graeme F. Bonham-Carter, and Laura Kemp

Weights of Evidence

Weights-of-evidence methodology combines spatial data from diverse sources to describe and analyze interactions, provide support for decision makers, and make predictive models. The method was originally developed for a nonspatial application in medical diagnosis. In this application, the evidence consisted of a set of symptoms, and the hypothesis was "this patient has disease x." For each symptom, a pair of weights was calculated, one for presence of the symptom and one for absence of the symptom. The magnitude of the weights depended on the measured association between the symptom and the occurrence of disease in a large group of patients. The weights could then be used to estimate the probability that a new patient would get the disease, based on the presence or absence of symptoms.

The weights-of-evidence methodology was adapted for mineral potential mapping with GIS. The Arc-WofE extension uses the statistical association between a training points theme such as a theme showing mineral sites and evidential themes showing rock types, geochemistry, or geophysical measurements to determine the weights.

The weights-of-evidence method is based on the application of Bayes' Rule of Probability, with an assumption of conditional independence. The model is stated in loglinear form so the weights from the evidential themes can be added. The Arc-WofE extension uses the weights-of-evidence methodology to produce a response theme. The response theme is an output map that combines the weights of predictor variables from the evidential themes to express the probability that a unit cell (small unit of area) will contain a training point. Evidential themes may have categorical values (e.g., the classes on geology or soil maps) or ordered values (e.g., geochemical concentrations or distance to linear and other spatial objects).

Weight values are easy to interpret. A positive weight for a particular evidential-theme value indicates that more training points occur on that theme than would occur due to chance, whereas

the converse is true for negative weights. A weight of zero indicates that the training points are spatially uncorrelated to the theme. The range-in-weight values for a particular evidential theme, known as the contrast, gives an overall measure of how important the theme is in the model. Uncertainties due to variances of weights and missing data allow the relative uncertainty in posterior probability to be estimated and mapped. Because conditional independence is never satisfied completely, the posterior probabilities are usually overestimated in absolute terms. However, the relative variations in posterior probability (as seen in spatial patterns on the response map) are usually not much affected by violations of this assumption.

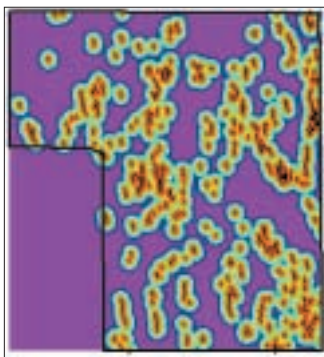
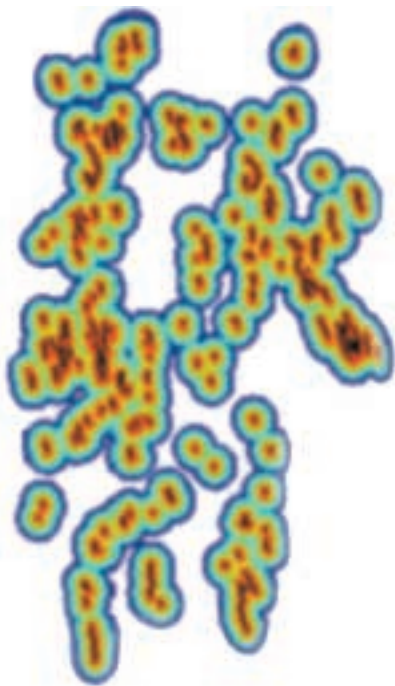
Mapping Quantitative Mineral Potential

The GIS-based mineral potential mapping process can be broken down into four main steps.

1. Building a spatial digital database
2. Extracting predictive evidence for a particular deposit type based on an exploration model
3. Calculating weights for each predictive map or evidential theme
4. Combining the evidential themes to predict mineral potential

The Arc-WofE extension assumes that a spatial digital database exists. Step 2 may have been partially completed prior to using the Arc-WofE extension but some of Step 2 and all of Step 3 are done within the extension. The examination of spatial relationships between training points and evidential themes is exploratory. It yields measurements of spatial association that are often unexpected. The process of generalization, grouping classes of evidential themes, and identifying breaks in continuous variables that maximize spatial associations often leads to a better understanding of the data. In Step 4 a map is generated that shows the combined evidence and is valuable for identifying areas for subsequent exploration. The map also gives a measure of confidence in the results, based on the uncertainty due to the variances of weights and/or missing data.

Continued on page 46



Extract Contacts creates a grid theme with a boundary between user-specified attributes of a polygon shape or grid theme.



The Buffer Features menu choice operates on shape or grid themes producing multiple buffer zones.

Predictive Probabilistic Modeling Using ArcView GIS

Continued from page 45



Weights-of-evidence methodology is normally applied to exploration situations in which there is an adequate number of mineral deposits or other located data that can be used as training points for calculating weights. However, the extension has been designed so that the user can calculate weights with training point data. The training set can be imaginary, and the overlap relationships between points and the selected evidential themes is decided arbitrarily by the user.

This leads to weights governed by expert opinion. In all other respects the treatment of the problem is the same. If a user decides that the available training points do not represent an adequate sample of the deposits actually present in the area, the weighting of the evidential themes can be based on expert judgment instead of statistical associations. This is advantageous for developing a variety of scenarios for different weights schemes, reflecting the differing opinions of experts.

Creating a Model with Arc-WofE

The Arc-WofE extension was developed by the U.S. Geological Survey and the Geological Survey of Canada with funding from the Western Mining Corporation, BHP, Barrick Gold Corporation, Inco Limited, Placer Dome Inc., and PT Freeport Indonesia. The extension was made available to the public in May 1999. It can be downloaded at no charge from the Arc-WofE home page (gis.nrcan.gc.ca/software/arcview/wofe). The Arc-WofE extension requires ArcView Spatial Analyst as well as ArcView GIS.

Modeling typically proceeds in three phases—specification, prediction, and testing. The specification of the model begins with the definition of a set of sites where some phenomenon has been observed such as mineral deposits, earthquake epicenters, pack rat middens, or other point objects. These sites are training points. The assumption is that a collection of training points will, in aggregate, have common characteristics that will allow their presence in other similar sites to be predicted. Other specifications required for the model include a defined study area, preparation of data for use in evidential themes, exploration of spatial associations between potential evidential themes and training points, and the generalization of evidential themes. The tools provided in the Arc-WofE extension are particularly valuable for spatial data exploration and generalization.

Arc-WofE Tools

Loading the Arc-WofE extension modifies the View GUI to include the Weights-of-Evidence menu. This menu contains six menu items

groupings. The top four groupings reflect the normal sequence of operations in using the Arc-WofE tools. The first grouping includes the Set Analysis Parameters... and Select Point by Location... menu choices that are used to select the study area, training set, and model parameters. The second grouping contains tools for exploring spatial associations between training points and evidential themes. Choices in the third grouping specify the model that will be used to produce the response map. Tools in the fourth menu grouping are used to check assumptions and evaluate predictions. Pre-processing tools designed primarily for geological applications are available from the fifth menu grouping. The user's manual for the Arc-WofE extension, available from the last menu grouping, gives descriptions of the use of all menu choices.

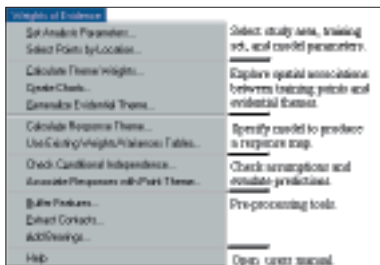
Calculating Weights

Reducing the number of classes, often to just two classes, improves the statistical robustness of the weights. However, multiclass evidential themes are allowable and sometimes desirable. Three approaches to calculating weights are possible—categorical, cumulative ascending, and cumulative descending.

Categorical weights are used for themes where the attribute values are unordered (e.g., names) and for themes with ordered values, a case in which midrange values are most predictive.

Cumulative ascending weights calculations are used in proximity analysis. In this type of analysis the zones near the training sites are most predictive, and the area and number of points are determined cumulatively starting with the first buffer zone around a feature, then the first plus the second, and so on. For each cumulative distance the weights and contrast are determined, assuming that this distance is a cut point between binary presence (close to feature) and absence (far away from feature). This allows decisions about reclassification to be based on the resulting statistics. Cumulative descending weights calculations work in the opposite direction, and are applicable for themes where the points are mainly associated with high values (e.g., in defining geochemical anomalies with high metal concentrations).

Calculated weights tables can be inspected directly or charted using the Create Charts menu selection. After considering the standard deviations of weights, the contrasts in the tables, and making subjective judgments based both on statistical results and interpretations based on real-world knowledge, the evidential themes can be generalized. The simple rule for reclassifying an ordered evidential theme



The Arc-WofE extension adds the Weights of Evidence menu to the View GUI.



The Evidential Theme Weights dialog for managing and performing weight calculations.



The Define Groups tool uses a query process that takes advantage of the standard ArcView GIS query tool for generalizing a theme.

binary “pattern present” and “pattern absent” is to select the maximum contrast as a threshold. However, in many cases the contrast curve is not simple due to statistical noise and factors unrelated to physical processes. In these cases, consider interpretations that are more complex or allow multiple classes or discard an evidential theme that is not acceptable.

Generalizing the Evidential Theme

After considering the weights and deciding on the appropriate breaks to generalize the evidential theme, the Generalize Evidential Theme menu provides several tools that facilitate this process. The Define Groups tool uses a query process that takes advantage of the standard ArcView GIS query tool for generalizing a theme. The generalization created can be inspected by symbolizing the evidential theme with the new reclassification or its descriptor field. If the pattern is acceptable, then generalization is complete. If it is not acceptable, then alternative generalizations can be generated and stored. New themes are not produced—the process simply adds new fields to the theme’s attribute table. This spatial data exploration and generalization step is repeated for each theme that will be used as evidence. This process completes the specification of the model.

Creating a Response Theme

After specifying the model, the user selects the evidential themes for prediction using the Calculate Response menu. For each evidential theme, the attribute field containing the desired generalization is selected. The complete weights-of-evidence calculations create a response theme, symbolized by posterior probability values (default), that is added to the view. Although in previous operations point evidential themes could be either grids or features, the response theme is a grid. It contains the intersection of all of the input themes in a single integer theme. Each row of the attribute table contains a unique vector (tuple) of evidential-theme values, the number of training points, area in unit cells, sum of weights, posterior logit, posterior probability, and the measures of uncertainty. The variances of the weights and variance due to missing data are summed to give the total variance of the posterior probability. The response theme may be

symbolized by any of the fields in the attribute table. Various measures to test the conditional independence assumption are also reported. Results are stored as tables in .dbf format for later inspection.

Once an acceptable model is created it can be tested in a variety of ways. If pairwise conditional independence tests indicate that some theme pairs are strongly correlated, one theme should be discarded or they should be combined. If overall conditional independence is a significant problem, the posterior probabilities should be thought of as relative favorabilities. The user can associate the calculated response values with both the training and test point subsets as another evaluation of the model. The user will want to select meaningful categories of posterior probabilities for symbolization of the response theme. The Arc-WofE extension provides several tools to help define meaningful symbolization that is easily interpreted.

This extension also provides an expert approach to weighting. This approach can be used when no training points are available or to modify the calculated weights using expert judgment after the data-driven model is created. The details of using the expert weighting are explained in the user’s manual.

Concluding Remarks

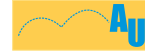
Use of the Arc-WofE extension with various types of minerals has successfully predicted the location of these deposits. There has been a high degree of agreement between response maps and expert-opinion maps. The evolution of this extension continues with further enhancements to the weights-of-evidence method and the addition of logistic regression, fuzzy logic, and neural network modeling tools. A new enhanced extension, supported by a consortium of industry and government organizations, has been completed and will be in the public domain in January 2001. [AU](#)

About the Authors

Gary L. Raines is a research geologist for the U.S. Geological Survey in Reno, Nevada. His research focuses on the integration of geoscience information for predictive modeling in mineral resource and environmental applications.

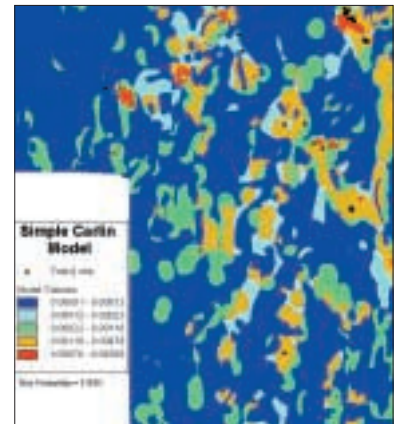
Graeme F. Bonham-Carter is a research geologist working in the Mineral Deposits Division of the Geological Survey of Canada, Ottawa. He is interested in applications of GIS to mineral exploration and environmental problems. He is editor-in-chief of *Computers & Geosciences*, a journal devoted to all aspects of computing in the geosciences.

Laura Kemp is an Ottawa, Ontario, based GIS technologist specializing in ArcView GIS application development.

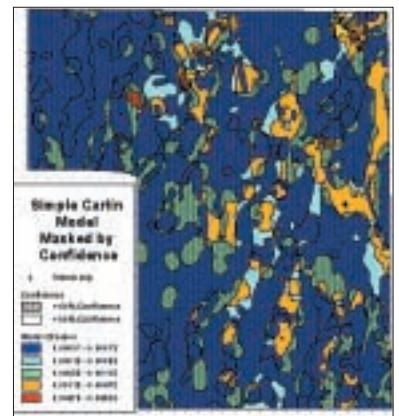


For a list of references, including books and papers by the authors, visit *ArcUser Online*.

See page 48 for a glossary of terms relating to this article.



The extension supplies various methods for generalizing evidential themes.



A response map showing five classes of posterior probability defined by natural breaks shown with and without a confidence mask.

Predictive Probabilistic Modeling Using ArcView GIS



Coming to Terms

A glossary of terms relating to predictive probabilistic modeling as described in the previous article

Model—The characteristics of a set of training points (e.g., mineral deposits), and the relationships of the training points to a collection of evidential themes (e.g., exploration data sets).

Study Area—A grid theme that acts as a mask to define the area where the model is developed and applied. It may be irregular in outline and may contain interior holes.

Training Points—The set of spatial point objects whose locations are to be predicted. In mineral exploration, these are the sites of known mineral deposits. Points are either present or absent. Size or other attributes of these points are not modeled.

Evidential Theme—A theme whose attributes are associated with the occurrence of the training points.

Missing Data—Values of an evidential theme within the study area that are unknown. For example, areas containing younger rocks that conceal the rocks of interest may be treated as unknown in a geological theme.

Weight—A measure of an evidential-theme class (value) as a predictor of training points. A weight is calculated for each theme class. For binary themes, these are often labeled W+ and W-. For multiclass themes, each class can also be

described by a W+ and W- pair, assuming presence/absence of this class versus all other classes. Positive weights indicate that more points occur on the class than due to chance, and the inverse for negative weights. The weight for missing data is zero. Weights are approximately equal to the proportion of training points on a theme class divided by the proportion of the study area occupied by theme class, approaching this value for an infinitely small unit cell.

Contrast—W+ minus W-, which is an overall measure of the spatial association of an evidential theme with the training points.

Unit Cell—A small unit of area used for counting. Each training point is assumed to occupy a unit cell. All area measures are transformed to unit cells. The number of unit cells with training points is unaffected by changes in unit cell size; whereas, area measurements are affected. The unit cell should be small—normally smaller than the minimum resolution of evidential themes. Weight values are relatively insensitive to unit cell if unit cell is small.

Prior Probability—The probability that a unit cell contains a training point before considering the evidential themes. Normally it is assumed to be a constant over the study area equal to the training point density (total number of training points divided by total study area in unit cells).

Posterior Probability—The probability that a unit cell contains a training point after consideration of the evidential themes. This measurement changes from location to location depending on the values of the evidence, being larger than the prior where the sum of the weights is positive. The mathematical model is log-linear in form and uses a log-odds (logit) scale that is transformed to probability. The posterior logit is determined from the sum of the prior logit and the applicable weights of evidential themes.

Confidence—A measure based on the ratio of posterior probability to its estimated standard deviation that is used as an informal test of whether the posterior probability is greater than zero.

Conditional Independence (CI)—The Arc-WofE model assumes that evidential themes are independent, not correlated conditional on the locations of training points. If this assumption is true, the summation of area times predicted posterior probability over the study area should equal the number of training points. The CI ratio, the number of observed training points divided by the number of predicted points, gives one measure of violation of this assumption. ■