

# Assessment of Exploration Bias in Data-Driven Predictive Models and the Estimation of Undiscovered Resources

Mark F. Coolbaugh,<sup>1,3</sup> Gary L. Raines,<sup>2</sup> and Richard E. Zehner<sup>1</sup>

Received 15 February 2007; accepted 15 February 2007  
Published online: 16 May 2007

The spatial distribution of discovered resources may not fully mimic the distribution of all such resources, discovered and undiscovered, because the process of discovery is biased by accessibility factors (e.g., outcrops, roads, and lakes) and by exploration criteria. In data-driven predictive models, the use of training sites (resource occurrences) biased by exploration criteria and accessibility does not necessarily translate to a biased predictive map. However, problems occur when evidence layers correlate with these same exploration factors. These biases then can produce a data-driven model that predicts known occurrences well, but poorly predicts undiscovered resources.

Statistical assessment of correlation between evidence layers and map-based exploration factors is difficult because it is difficult to quantify the “degree of exploration.” However, if such a degree-of-exploration map can be produced, the benefits can be enormous. Not only does it become possible to assess this correlation, but it becomes possible to predict undiscovered, instead of discovered, resources.

Using geothermal systems in Nevada, USA, as an example, a degree-of-exploration model is created, which then is resolved into purely explored and unexplored equivalents, each occurring within coextensive study areas. A weights-of-evidence (WofE) model is built first without regard to the degree of exploration, and then a revised WofE model is calculated for the “explored fraction” only. Differences in the weights between the two models provide a correlation measure between the evidence and the degree of exploration.

The data used to build the geothermal evidence layers are perceived to be independent of degree of exploration. Nevertheless, the evidence layers correlate with exploration because exploration has preferred the same favorable areas identified by the evidence patterns. In this circumstance, however, the weights for the “explored” WofE model minimize this bias. Using these revised weights, posterior probability is extrapolated into unexplored areas to estimate undiscovered deposits.

**KEY WORDS:** Weights-of-evidence, GIS, geothermal, resources, undiscovered, data-driven, exploration.

## INTRODUCTION

Data-driven modeling techniques such as weights-of-evidence and logistic regression have

proven useful for modeling the spatial distribution of a variety of natural resources and natural phenomena (Bonham-Carter, 1994; Raines, 1999). Two issues may not be explicitly dealt with in data-driven models that derive from the utilization of known resource occurrences as training sites. The first issue is that bias inevitably occurs during the selective search and exploration for training sites (usually referred to here as “deposits”). Sources of exploration bias are many, and include accessibility factors such as the location of outcrops, roads, lakes,

<sup>1</sup>Great Basin Center for Geothermal Energy, MS 172, University of Nevada, Reno, Reno, NV 89557, USA.

<sup>2</sup>United States Geological Survey, MS 176, Mackay School of Earth Sciences and Engineering, University of Nevada, Reno, Reno, NV 89557, USA.

<sup>3</sup>To whom correspondence should be addressed; e-mail: sereno@dim.com

swamps, property boundaries, political boundaries, etc., as well as perceived exploration criteria, such as faults, alteration, geochemical anomalies, etc. The result is a set of training sites accumulated after years of exploration that is not likely to be spatially random in regards to the full set of all occurrences, known and unknown. The second issue is that data-driven models do not by themselves quantify the undiscovered resource base, which usually is their ultimate objective.

The use of training sites biased by exploration criteria and accessibility does not necessarily translate to a biased predictive map, as long as the exploration search is unbiased in regards to the evidence layers used to predict the resource. An example would be a training set of raptor nests compiled by searching trees within 500 m of accessible roads. As long as this search corridor is unbiased relative to patterns or classes in the evidence layers used for modeling, the biases may not be transmitted to the predictive map. The road corridors might thus resemble the random walk of an individual. But if the roads are biased with respect to the occurrence of wetlands, for example, and wetlands are used as evidence, the resulting model might be biased. And of course, proximity to roads should not be used as a predictive layer in the model if the roads themselves were used as a basis for searching for training sites.

Common sense can be used to avoid obvious bias between training sites and evidence layers, but the sources of bias may be subtle. For example, some geologic formations are more resistant to weathering and thus more likely to form outcrops where mineral deposits would be discovered. Another source of bias occurs when evidence layers correspond with the same exploration criteria used by explorers in the past, with the result that training points have been searched more extensively for in the areas identified by the evidence. This seems to be the source of a systematic moderate bias that has affected all four of the evidence layers used in the geothermal model presented below.

If it were possible to build a map showing where resources (training sites) have been searched for, and how effective that search has been, it would be possible to assess possible spatial correlations between the degree-of-exploration and potential evidence layers. Importantly, it also would make it possible to estimate undiscovered resources, if the degree-of-exploration map is combined with other models that predict where such

resources are most likely to occur. Building a degree-of-exploration map is anything but easy, however, which is probably why there are not many examples of such maps. Exploration for mineral deposits has taken place over hundreds, if not thousands of years, using a multitude of techniques of varied effectiveness, each with its own spatial bias.

In the geothermal example presented here, fuzzy logic and expert guidance are used to build a "degree-of-exploration" map. This map then is intersected with a WofE predictive model of geothermal systems to estimate the number of undiscovered geothermal systems. In the process, correlations between evidence layers and degree of exploration are assessed and minimized.

### **INITIAL WEIGHTS-OF-EVIDENCE (WOFE) MODEL**

An initial WofE model of geothermal potential was constructed for the state of Nevada, USA, without considering any factors related to the degree of exploration. The geothermal systems used as training points are subsurface circulation zones (0–4 km below the land surface) of thermal groundwater with temperatures  $\geq 100^{\circ}\text{C}$ . By virtue of their temperatures, these groundwaters have the potential to generate electricity when fed through turbines. A total of 69 such geothermal systems in Nevada were known to exist as of the preparation of this paper, and all were used as training sites.

Several types of geological, geophysical, and geochemical evidence are predictive of geothermal potential (Koenig and McNitt, 1983; Coolbaugh and Bedell, 2006). For the current model, evidence was selected carefully for its ability to predict geothermal potential independently of the degree of exploration. For example, a map of watertable depth was not used directly as an evidence layer, even though areas of shallow groundwater correlate with known geothermal activity. Areas with shallow groundwater tend to have surface indications of geothermal activity that would attract exploration efforts.

Four evidence layers were used, derived from (1) earthquakes catalogs (Coolbaugh and others, 2005), (2) crustal strain rates from global positioning system station velocities and slip rates from Quaternary faults (Coolbaugh and others, 2005), (3) the isostatically corrected gravity field (Singer, 1996),

**Table 1.** Weights-of-Evidence Statistics

Evidence layer	Pattern 1 Weight	Pattern 1 Stan. Dev.	Pattern 2 Weight	Pattern 2 Stan. Dev.	Pattern 3 Weight	Pattern 3 Stan. Dev.	Contrast	Confidence
(a) Initial Weights of Evidence, without Degree of Exploration								
Gravity gradient	1.3739	0.2899	0.5895	0.2299	-0.3775	0.1623	1.75	5.27
Crustal strain	0.8362	0.2005	0.3338	0.2003	-0.7510	0.2295	1.59	5.21
Earthquakes	0.3030	0.1350	-0.7065	0.2674			1.01	3.37
Isostatic gravity	0.2558	0.1669	-0.2201	0.1742			0.48	1.97
(b) Revised Weights of Evidence, with Degree of Exploration								
Gravity gradient	1.1760	0.2926	0.4692	0.2309	-0.3407	0.1627	1.52	4.53
Crustal strain	0.6473	0.2016	0.2828	0.2011	-0.6700	0.2299	1.32	4.31
Earthquakes	0.2603	0.1356	-0.6420	0.2678			0.90	3.01
Isostatic gravity	0.1967	0.1675	-0.1782	0.1747			0.37	1.55
(c) Percent Change, Explored vs. Initial Model								
Gravity gradient	-14.4	0.9	-20.4	0.4	-9.7	0.2	-13.4	-14.0
Crustal strain	-22.6	0.5	-15.3	0.4	-10.8	0.2	-17.0	-17.3
Earthquakes	-14.1	0.4	-9.1	0.1			-10.6	-10.8
Isostatic gravity	-23.1	0.4	-19.0	0.3			-21.2	-21.5

Confidence equals contrast divided by its standard deviation. The term "evidence layer pattern" is equivalent to the term "evidence layer class."

and (4) the total horizontal derivative of gravity. The training site unit cell size was 9 km<sup>2</sup> and the conditional independence (CI) ratio (Bonham-Carter, 1994) was 0.95; weights of evidence are listed in Table 1a. The posterior probability map (Fig. 1A) was moderately successful in classifying the training sites; 77% fell within the upper 55% of the probability rankings (weighted by area) and 40% fell in the upper 90% of the probability rankings.

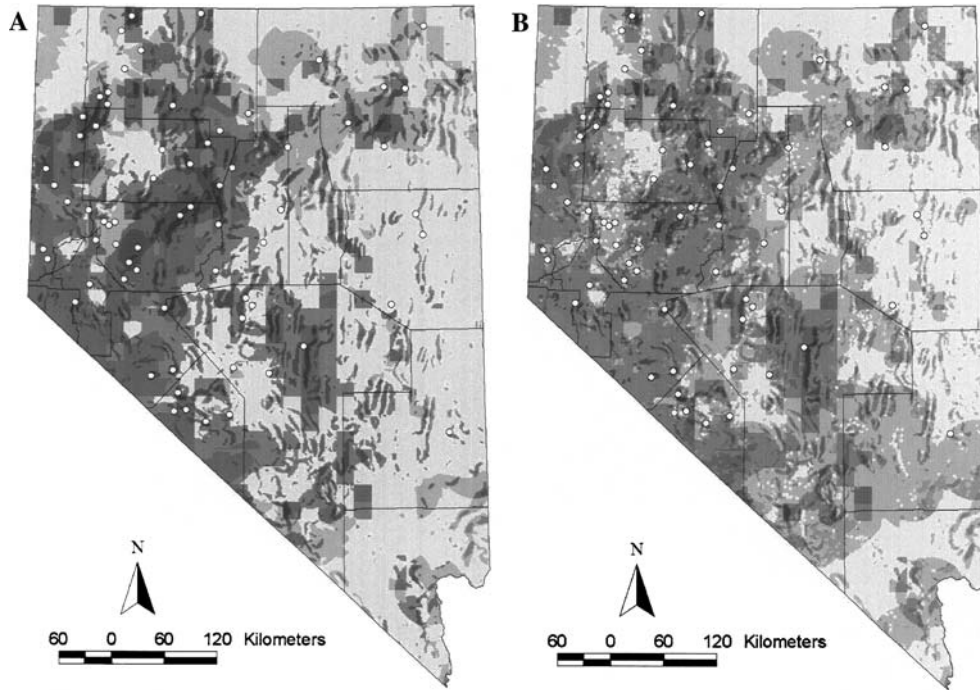
## DEGREE-OF-EXPLORATION MODEL

Exploration for geothermal systems in Nevada is far from complete, partly because the entire state of Nevada is permissive for the occurrence of geothermal systems (Coolbaugh and others, 2005), and partly because many geothermal systems have no surface expression. The presence of deep watertables, cold water aquifers, and near-surface impermeable cap rocks may prevent thermal groundwaters from reaching the surface, and where they do reach the surface, they may have been cooled or diluted with near-surface groundwaters that disguise the geothermal signature. Of the 69 geothermal systems used as training points, 24 (35%) are not associated with hot springs and thus can be considered concealed.

There are many ways to search for geothermal systems, including searching for hot and warm springs, water geochemical sampling, geologic mapping, gravity, magnetic, and seismic surveys, and

well drilling. The effectiveness of each of these techniques, and where they have been employed in the state, is a matter for debate. A variety of approaches for building a degree-of-exploration model are possible, employing a variety of statistical relationships. The fairly simple method presented here is not a unique solution, but serves as an example of how such a model might be built.

Fuzzy logic and expert knowledge were used to build a degree-of-exploration model scaled from 0 representing 0% efficiency (no exploration and no geothermal systems located), to 1 representing 100% efficiency (all geothermal systems discovered). Four types of evidence were used: (1) temperature gradient and geothermal wells, (2) other (nongeothermal) wells, (3) depth to the watertable, and (4) presence of a carbonate aquifer. Temperature gradient and geothermal wells were compiled from databases at Southern Methodist University (<http://www.smu.edu/geothermal/>) and the Nevada Division of Minerals (<http://minerals.state.nv.us/>) and total 6,671 in number. Nongeothermal wells were compiled from the United States Geological Survey National Water Information System (NWIS) database (<http://waterdata.usgs.gov/nwis/>), the Nevada Division of Water Resources well log database (<http://water.nv.gov/Engineering/wlog/wlog.cfm>), and a Nevada Bureau of Mines and Geology oil and gas well database (Hess, 2001), and total 161,753 wells. Tim Minor of the Desert Research Institute, Reno, Nevada, generated a depth-to-watertable map using approximately 40,000 NWIS water well records.



**Figure 1.** Geothermal potential for state of Nevada, USA. Initial weights-of-evidence posterior probability map A was created without consideration of degree of exploration; B, depicts potential for undiscovered geothermal systems after intersecting initial weights-of-evidence model with degree-of-exploration model (Fig. 2). Darker colors on both maps represent progressively higher probability levels, using seven natural breaks. White circles are geothermal training sites.

Prudic, Harrill, and Burbey (1995) provided a carbonate aquifer map.

Well-drilling is one of the more effective methods of geothermal exploration. Deep wells are more likely than shallow wells to encounter ther-

mal waters, and consequently a higher degree of exploration was assigned to areas with deeper wells (Table 2). The presence of geothermal or temperature gradient wells is believed more indicative of serious geothermal exploration than the presence of

**Table 2.** Degree-of-Exploration Estimated for Areas Outside the Carbonate Aquifer Under the Specified Conditions of Distance to Wells, Type and Depth of Drilling and Watertable Depth

Distance from well (km)	Drilling depth (ft)	Watertable depth (ft)		
		0-50	50-200	> 200
<b>(a) Geothermal and Temperature Gradient Wells</b>				
>2	None	0.30	0.25	0.18
≤ 2	0-50	0.70	0.60	0.50
≤ 2	50-200	0.70	0.60	0.50
≤ 2	200-1000	0.80	0.70	0.65
≤ 2	> 1000	0.90	0.85	0.81
<b>(b) Non-Geothermal Wells</b>				
>2	None	0.30	0.25	0.18
≤ 2	0-50	0.35	0.30	0.23
≤ 2	50-200	0.40	0.38	0.28
≤ 2	200-1000	0.45	0.43	0.34
≤ 2	> 1000	0.50	0.48	0.42

Values for the carbonate aquifer were estimated at 10% less than the values shown here.

nongeothermal wells, because geothermal drilling may be accompanied by other types of exploration, and also because hot water usually is not reported when encountered in nongeothermal wells. For this reason, for a given well depth and watertable depth, higher degrees of exploration were assigned to geothermal-related wells than to nongeothermal wells (compare equivalent cells in Table 2a and 2b).

All wells were assigned a 2-km circular radius of influence. Five of seven newly recognized geothermal systems in Nevada (unpublished data, 2005, Great Basin Center for Geothermal Energy (GBCGE), Reno, Nevada) occur within 2 km of existing wells, suggesting that at greater distances, the presence of a well is not an effective exploration guide.

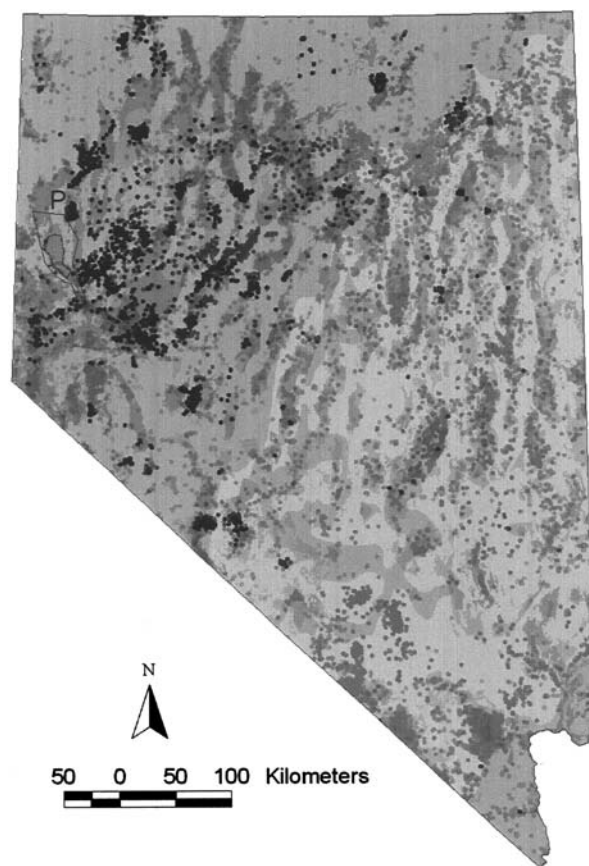
Geothermal systems are less likely to be concealed, and consequently will be better explored for, in areas where the watertable is shallow. Koenig and McNitt (1983) and Coolbaugh and others (2002) have documented a correlation between shallow watertables and the location of hot springs and known geothermal systems in the Great Basin. Logically, surface exploration techniques (such as looking for hot springs) are more effective when the watertable is shallow.

Complicating this relationship is the fact that geothermal systems also correlate with low topographic elevations (which in turn are associated with active crustal tectonics). Shallow watertables and low topographic elevations may occur in the same areas and it was difficult to separate the effects of the two quantitatively. Instead, a more qualitative method based on observed field relationships in known geothermal areas was used. The watertable map was classified into three categories: 0–50 ft, 50–200 ft, and >200 ft (Table 2). For mountain ranges, where water wells are lacking, water depths were assumed to fall within the “>200 ft” category. Shallow groundwaters are present locally in mountain ranges, but they may occur in perched water zones that do not provide useful information on the deeper geothermal potential. The WofE contrast statistic was useful in picking a threshold depth of 50 ft, at which a maximum statistical distinction occurs between shallow waters that correlate with geothermal systems, and deeper waters that do not.

Fewer known geothermal systems than expected occur in areas underlain by regional aquifers, such as the carbonate aquifer in Nevada (Coolbaugh and others, 2005). It is hypothesized that aquifers may capture and entrain rising thermal fluids before

they reach the surface. Consequently, these areas are considered less well explored than nonaquifer areas, when other exploration factors are equal. For the carbonate aquifer, the degree of exploration was reduced by 10% relative to equivalent categories outside the aquifer.

To produce a degree-of-exploration map, the exploration evidence was combined together to form a unique conditions map grid, and for each unique condition, degree-of-exploration values (exploration efficiency) were assigned as shown in Table 2 (Fig. 2). Unique conditions include the presence or absence of drilling (i.e.,  $\leq 2$  km from a well or  $>2$  km from a well), the type of drilling, depth of drilling, depth to the watertable, and presence or absence of the carbonate aquifer. A fuzzy “OR” statement was used when multiple types and depths of wells are present, such that the well with the highest ranked degree of exploration was used.



**Figure 2.** Degree-of-exploration model for geothermal systems in Nevada. Darker colors represent progressively greater degrees of exploration. Pyramid Lake Paiute Reservation is outlined in black with label “P”.

### INTERSECTION OF THE WEIGHTS-OF-EVIDENCE MODEL WITH THE DEGREE-OF-EXPLORATION MODEL

The simplest situation of intersecting a degree-of-exploration model with a data-driven predictive model occurs when the exploration map is binary; that is, composed of perfectly explored and unexplored areas. In this situation, a revised WofE model can be calculated using the reduced area of the perfectly explored area, and then the revised prior probability and revised weights of the explored area can be extrapolated into the unexplored area using the unique conditions table of evidence patterns to estimate undiscovered deposits.

A more general situation occurs when degree of exploration is not binary but is instead scaled from 0 to 1. The approach adopted here is to resolve the scaled degree-of-exploration model into perfectly explored and perfectly unexplored equivalent fractions, both of which fall into coextensive study areas. For example, a polygon considered 40% explored (0.40 in Table 2) would have 40% of its area assigned to a “completely explored” study area, and 60% assigned to a “completely unexplored” study area, even though it is not possible to determine exactly which cells in the polygon are the explored ones and which are the unexplored ones. All training sites belong to the “completely explored” fraction, because it would be impossible to discover them without some type of exploration. In these circumstances, the total area of the initial study,  $A_I$ , (i.e., the state of Nevada in the geothermal example), would equal the sum of the areas of the completely explored fraction,  $A_E$ , and completely unexplored fraction,  $A_U$ , study areas:

$$A_I = A_E + A_U, \quad (1)$$

where values for  $A_E$  and  $A_U$  can be determined by summing the explored and unexplored fractions of each grid cell in the total study area  $A_I$ .

The calculation of a revised WofE posterior probability for the explored fraction study area ( $P_{Epost}$ ) then can proceed using only the “completely explored” study area in the calculations. There are some tricks involved with this computation, however, because the average degree of exploration for the entire study area will inevitably differ from the average degree of exploration associated with each evidence layer class or pattern. An example calculation using geothermal systems is provided below.

It ends up being convenient, for purposes of assessing exploration bias, to compare the computation of  $P_{Epost}$  to the computation of posterior probability of the initial WofE model ( $P_{Ipost}$ ), which was done without consideration to exploration. The first step involves calculation of prior probability. The prior probability of the explored model ( $P_{Eprior}$ ) will differ from the prior probability of the initial model ( $P_{Iprior}$ ) because the explored area is smaller than the total area:

$$P_{Iprior} = (N_T \times \text{unit cell})/A_I, \text{ and,} \quad (2)$$

$$P_{Eprior} = (N_T \times \text{unit cell})/A_E, \quad (3)$$

where  $N_T$  = total number of training sites and the unit cell = unit area assigned to each training site. Using Equations (2) and (3),  $P_{Eprior}$  can be reexpressed in terms of  $P_{Iprior}$ :

$$P_{Eprior} = P_{Iprior}/(A_E/A_I). \quad (4)$$

In other words, the prior probability in the explored fraction of the study area is equal to the initial prior probability divided by the ratio of the “explored” study area to the initial study area.

Similarly, the weights for each pattern of each evidence layer in the explored model can differ from the weights in the initial model. The relationship between the weights in the initial and the explored models is formulated here using a simplified expression of weights of evidence in terms of a normalized density function (Mihalasky and Bonham-Carter, 2001; Coolbaugh and Bedell, 2006):

$$W_{Ii,j} \cong \ln[(N_i/N_T)/(A_{Ii}/A_I)], \text{ and,} \quad (5)$$

$$W_{Ei,j} \cong \ln[(N_i/N_T)/(A_{Ei}/A_E)], \quad (6)$$

where  $W_{Ii,j}$  = the weight for pattern “ $i$ ” on evidence map “ $j$ ” in the initial WofE model,  $W_{Ei,j}$  = the weight of pattern “ $i$ ” on evidence map “ $j$ ” in the explored model,  $N_i$  = number of deposits or geothermal systems associated with pattern “ $i$ ” on evidence map “ $j$ ”,  $A_{Ii}$  = area of pattern  $i$  (on the evidence map  $j$ ) in the initial study area, and  $A_{Ei}$  = explored fraction of the area of pattern  $i$  (on the evidence map  $j$ ).

Mihalasky and Bonham-Carter (2001, appendix) show that when the unit cell is reduced in area, the weights approach the natural logarithm of the ‘normalized density’ [equivalent to Eqs. (5) and (6)] and are identical when the unit cell has zero area. Coolbaugh and Bedell (2006) demonstrate for geo-

thermal systems in Nevada, that errors associated with representing weights-of-evidence as density functions were typically less than 1/2 of 1% (Coolbaugh, oral comm., 2006). It is not necessary here to formulate the weights in terms of a density function, but it is done for simplicity and to better illustrate the relationship between the weights of the explored and initial WofE models.

$W_{Ei,j}$  can be expressed in terms of the initial  $W_{Ii,j}$  by expanding Equation (6) as follows:

$$W_{Ei,j} \cong \ln \left\{ \left[ \frac{(A_{Ii}/A_I)}{(A_{Ei}/A_E)} \right] \times \left[ \frac{(N_i/N_T)}{(A_{Ii}/A_I)} \right] \right\}_j. \quad (7)$$

With substitution of Equation (5) and rearranging of terms, Equation (7) becomes:

$$W_{Ei,j} \cong W_{Ii,j} + \ln \left[ \frac{(A_E/A_I)}{(A_{Ei}/A_{Ii})} \right]_j. \quad (8)$$

Using the revised weights and prior probability, the posterior probability  $P_{Epost}$  can be determined using the weights-of-evidence formulas of Bonham-Carter, Agterberg, and Wright (1988). Because  $P_{Epost}$  has been calculated only for the fully explored fractions of areas, it can be considered equal to the total density or frequency of occurrence of the resources or deposits being modeled. The probability of locating an undiscovered resource or deposit then depends on the degree of exploration, as follows:

$$P_U = P_{Epost} \times (1 - f_E) \quad (9)$$

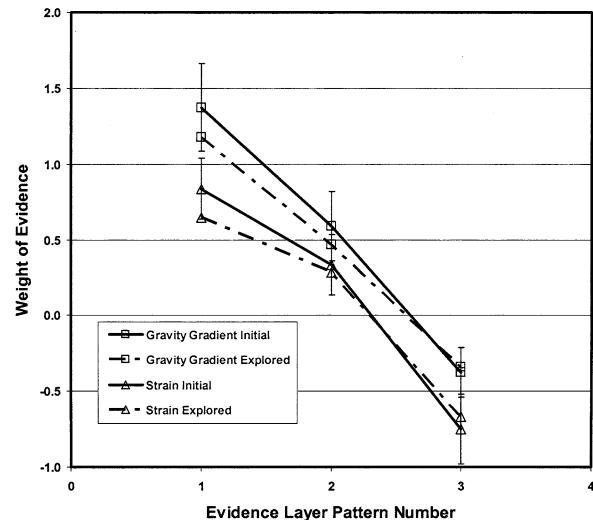
where  $P_U$  = probability of an undiscovered deposit and  $f_E$  = the degree of exploration. When the degree of exploration is 1 (100%), the probability of an undiscovered deposit is 0, and when the degree of exploration is 0, the probability of an undiscovered deposit =  $P_{Epost}$ .

## RESULTS – MEASURE OF CORRELATION

Several methods are available for measuring correlation between maps, including Pearson’s and Spearman’s correlation coefficients, the chi-square statistic, the coefficient of agreement kappa, and the odds ratio (Bonham-Carter, 1994). An alternative method in the current context is suggested by Equation (8), which shows that a weight for the explored fraction of the total study area  $W_{Ei,j}$  will not differ from the initial weight of the total study area,  $W_{Ii,j}$ , unless the average degree of exploration for a given pattern (as measured by  $A_{Ei}/A_{Ii}$ ) differs from the average degree of exploration for the entire study

area (as measured by  $A_E/A_I$ ). The explored weight  $W_{Ei,j}$  can either go up or down, relative to the corresponding initial weight  $W_{Ii,j}$ , depending on whether the area of a given pattern is better or worse explored compared to the average degree of exploration of the entire study area. Consequently, the difference in the weights calculated with and without the effects of degree-of-exploration, is in itself a measure of the exploration bias.

This weight-difference measure reveals a systematic moderate exploration bias in all evidence layers of the geothermal model (Table 1c, Fig. 3). The magnitudes of the weight differences range from 9% to 23% and the contrast differences range from 11% and 21%, and in each situation the magnitudes of the weights and contrasts decrease in the exploration model relative to the initial model. This is a direct result of historical exploration being biased towards areas of higher favorability as defined in the current model. For the isostatic gravity layer, the confidence of the contrast (Studentized contrast) dropped from 1.97 to 1.55. Although this layer was retained in the final WofE model, it could be argued that the contrast is no longer significant, in which example this evidence layer would be removed from the model, and the exploration analysis would have played the role of eliminating evidence that initially (falsely) seemed to have a significant correlation with training sites.



**Figure 3.** Weights-of-evidence for gravity gradient and strain evidence layers before and after degree-of-exploration. Error bars for preexploration weights shown for comparison. Although change in weights is overlapped by error, weight changes caused by exploration bias are systematic and real.

The data values used to create the four geothermal evidence layers are believed to be independent of degree-of-exploration. If this is true, then the observed correlation between the evidence and the degree-of-exploration is presumed to be the result of a natural tendency of the model to identify, at least in part, the same favorable conditions recognized by explorers when they searched for geothermal systems. Under these circumstances, the revised weights of the explored study area correct for this exploration bias by effectively eliminating the “exploration variable” from the equations (with the obvious caveat of course, that the degree-of-exploration model is qualitative and not quantitative in nature). A more complicated situation is produced when the data used to create evidential layers is itself biased by geothermal exploration, in which case the modeling methods presented here do not compensate for such bias, and more complex analyses would be required before such layers could be used in an unbiased manner.

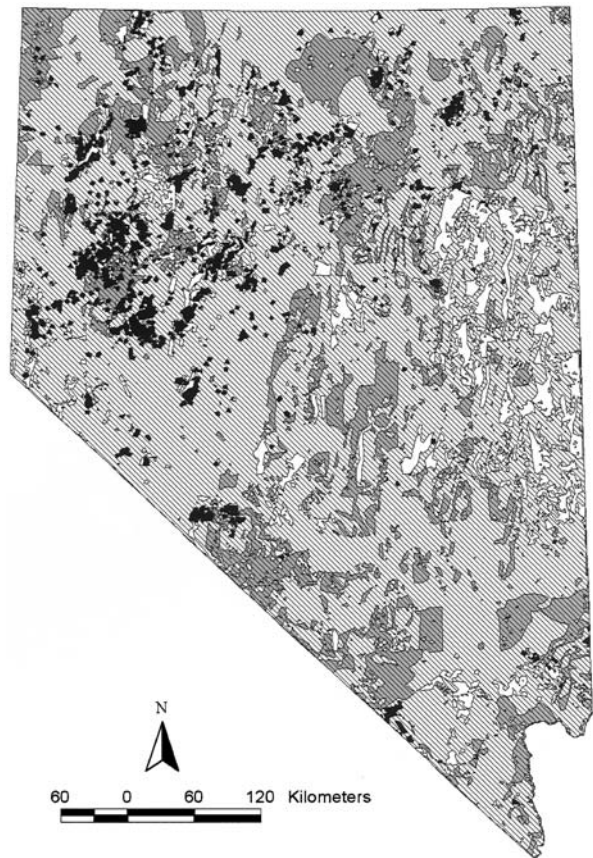
## RESULTS – UNDISCOVERED DEPOSITS AND NEW DISCOVERIES

Using the methodology as described, a total of 170 undiscovered geothermal systems (fitting the criteria of the training points) are predicted in the state of Nevada. This number is, of course, entirely subject to the qualitative nature of the degree-of-exploration model. The same model also predicts 75 known geothermal systems in the state, slightly higher than the 69 training sites used. The C.I. ratio of the revised explored model therefore is 0.92, slightly lower than the 0.95 of the initial model. This ratio could be used to downward-revise the number of estimated undiscovered deposits from 170 to 157, on the assumption that moderate conditional dependencies exist in the model.

There are three primary differences between the posterior probabilities of the initial (Fig. 1A) and final (Fig. 1B) models. First, posterior probabilities generally are higher in the final model compared to the initial model because of the increase in the prior probability in the former. Second, areas that are well explored have low probabilities for undiscovered resources in the final model, as expected. Finally, in a more subtle effect, the contrast between high-probability and low-probability areas in the initial model is somewhat more subdued in the final model, a consequence of the final model’s

compensation for the initial model’s tendency to over-predict probabilities in high-favorability areas (because of exploration bias). As a consequence, the final model tends to predict relatively higher favorability in peripheral areas; peripheral areas that were defined initially as being relatively unfavorable based on the evidence layers alone (Fig. 4).

An earlier attempt at estimating undiscovered geothermal resources in Nevada (Coolbaugh and Shevenell, 2004) highlighted an area in west-central Nevada, including the Pyramid Lake Paiute Reservation (P, Fig. 2) as having high potential. The new model similarly predicts good potential in this area, and estimates 1.5 undiscovered systems within the 1400 km<sup>2</sup> of the reservation. In the past year, two and possibly three geothermal systems were discovered (Coolbaugh and others, 2006) on the res-



**Figure 4.** Ranked difference between initial weights-of-evidence posterior probability map (Fig. 1A) and the undiscovered posterior probability map of Figure 1B. The input maps (Fig. 1A and B) were reclassified into 20 equal-area ranks prior to subtraction. Initial WofE map was subtracted from undiscovered map. Black = -11 to -1; white = 0, diagonals with white background 1 to 3; diagonals with gray background = 4 to 6.



ervation during an exploration program conducted by the University of Nevada, Reno. The three systems fall within the top 99th, 91st; and 66th percentiles of cumulative-area probability rankings of the undiscovered geothermal model. The lower ranking of the 3rd system (66th) may be because it occurs at lake level (e.g., a high watertable). The good exploration results can be attributed partly to the fact that the reservation is less well explored than the rest of the state, for cultural and economic reasons that were not included in the degree-of-exploration model.

## DISCUSSION

The methodology presented here provides a method for assessing the effects of exploration bias on data-driven predictive models, and also provides a way of using that bias to predict undiscovered resources. The concept of a “degree-of-exploration model” also is promoted, because however difficult building such a model is, it provides an additional tool for exploring the complexities of predicting undiscovered resources (and only some of those complexities have been described herein).

Although weights-of-evidence was used as an example, the methodology could be adapted easily to, and may be more appropriate for, other data-driven methods such as logistic regression that are not as dependent on assumptions of conditional independence. The accuracy of the correlation measures and estimates of undiscovered resource potential are wholly dependent on the degree-of-exploration model. In spite of the qualitative nature of this degree-of-exploration model, it provides a framework within which potential exploration bias can be investigated. The geothermal example suggests that such exploration bias may be widespread in natural resource modeling. The methodology also facilitates the iterative evaluation of multiple scenarios wherein different degree-of-exploration models can be evaluated for their impact on the implied undiscovered resource base.

## ACKNOWLEDGMENTS

This paper has benefited from advice and discussions with Colin Williams and Marshall Reed of the Menlo Park office of the United States Geological Survey. We also want to extend our appreciation to Lisa Shevenell, director of the Great Basin

Center for Geothermal Energy, for her encouragement to pursue this investigation. Funding for this research was made possible by a grant from the U.S. Department of Energy under instrument number DE-FG07-02ID14311.

## REFERENCES

- Bonham-Carter, G. F., 1994, *Geographic Information Systems for geoscientists, modelling with GIS*: Pergamon Press, Oxford, 398 p.
- Bonham-Carter, G. F., Agterberg, F. P., and Wright, D. F., 1988, Integration of geological datasets for gold exploration in Nova Scotia: *Photogrammetric Engineering and Remote Sensing*, v. 54, no. 11, p. 1585–1592.
- Coolbaugh, M. F., and Bedell, R., 2006, A simplification of weights of evidence using a density function, fuzzy distributions, and geothermal systems, *in* Harris, J. R., ed., *GIS for the Earth Sciences*: Geol. Assoc. Canada, Spec. Publ. 44, p. 115–130.
- Coolbaugh, M. F., and Shevenell, L. A., 2004, A method for estimating undiscovered geothermal resources in Nevada and the Great Basin: *Proc. Ann. Meeting (Palm Springs, CA)*, Geothermal Resources Council Trans., v. 28, p. 13–18.
- Coolbaugh, M. F., Taranik, J. V., Raines, G. L., Shevenell, L. A., Sawatzky, D. L., Minor, T. B., and Bedell, R., 2002, A geothermal GIS for Nevada: defining regional controls and favorable exploration terrains for extensional geothermal systems: *Geothermal Resources Council Trans.*, v. 26, p. 485–490.
- Coolbaugh, M., Zehner, R., Kreemer, C., Blackwell, D., Oppliger, G., Sawatzky, D., Blewitt, G., Pancha, A., Richards, M., Helm-Clark, C., Shevenell, L., Raines, G., Johnson, G., Minor, T., and Boyd, T., 2005, Geothermal potential map of the Great Basin, western United States: Nevada Bur. Mines and Geology Map 151.
- Coolbaugh, M. F., Faulds, J. E., Kratt, C., Oppliger, G. L., Shevenell, L. A., Calvin, W. M., Ehni, W. J., and Zehner, R. E., 2006, Geothermal potential of the Pyramid Lake Paiute Reservation, Nevada, USA: Evidence of at least two previously unrecognized moderate-temperature (150–170°C) geothermal systems: *Geothermal Resources Council Trans.*, v. 30, p. 59–67.
- Hess, R., 2001, Nevada oil and gas well database map: Nevada Bur. Mines and Geology Open File Rept. 2001–7, unpaginated.
- Koenig, J. B., and McNitt, J. R., 1983, Controls on the location and intensity of magmatic and nonmagmatic geothermal systems in the Basin and Range province: *Geothermal Resources Council Spec. Rept. No. 13*, 93 p.
- Mihalasky, M. J., and Bonham-Carter, G. F., 2001, Lithodiversity and its spatial association with metallic mineral sites, Great Basin of Nevada: *Natural Resources Research*, v. 10, no. 3, p. 209–226.
- Prudic, D. E., Harrill, J. R., and Burbey, T. J., 1995, Conceptual evaluation of regional groundwater flow in the carbonate-rock province of the Great Basin, Nevada, Utah, and adjacent states: *U.S. Geol. Survey Prof. Paper 1409-D*, 102 p.
- Raines, G. L., 1999, Evaluation of weights of evidence to predict epithermal-gold deposits in the Great Basin of the western United States: *Natural Resources Research*, v. 8, no. 4, p. 257–276.
- Singer, D. A., 1996, ed., *An analysis of Nevada’s metal-bearing mineral resources*; Nevada Bur. Mines and Geology Open-File Rept. 96–2, unpaginated.