15

Application of the PROSPECTOR system to geological exploration problems†

J. Gaschnig
Artificial Intelligence Center
SRI International, USA

Abstract

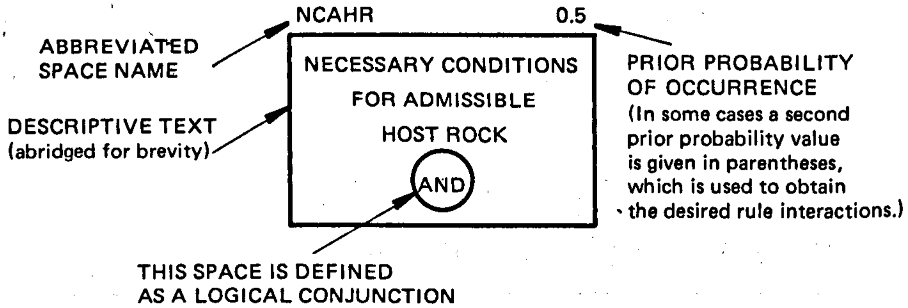
A practical criterion for the success of a knowledge-based problem-solving system is its usefulness as a tool to those working in its specialized domain of expertise. Here we describe several applications of the PROSPECTOR consultation system to mineral exploration tasks. One was a pilot study conducted for the National Uranium Resource Estimate program of the U.S. Department of Energy. This application estimated the favourability of several test regions for occurrence of sandstone uranium deposits. For credibility, the study was preceded by a performance evaluation of the relevant portion of PROSPECTOR's knowledge base, which showed that PROSPECTOR's conclusions agreed very closely with those of the model designer over a broad range of conditions and levels of detail. A similar uranium favourability evaluation of an area in Alaska was performed for the U.S. Geological Survey. Another application involved measuring the value of a geological map. We comment on characteristics of the PROSPECTOR system that are relevant to the issue of inducing geologists to use the system.

1. INTRODUCTION

This paper describes an evaluation and several applications of a knowledge-based system, the PROSPECTOR consultant for mineral exploration. PROSPECTOR is a rule-based judgmental reasoning system that evaluates the mineral potential of a site or region with respect to inference network models of specific classes of ore deposits. Knowledge about a particular type of ore deposit is encoded in a computational model representing observable geological features and the relative significance thereof.

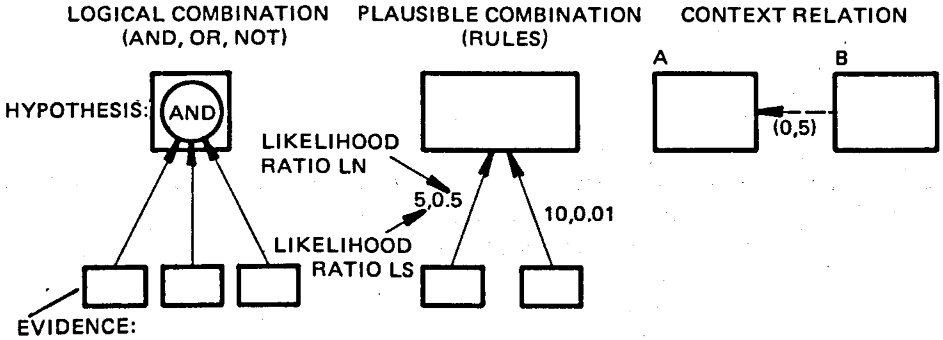
† Any opinions, findings, and conclusion or recommendations expressed in this report are those of the author and do not necessarily reflect the views of the U.S. Geological Survey.
A shorter version of this paper appeared as (Gaschnig 1980b). Parts of this paper are excerpted from (Gaschnig 1980a).

ASSERTION SPACES:



NOTE: If box is dashed rather than solid, then its complete definition (including subnetwork, if any) appears on another page.

NETWORK LINKS:



INTERPRETATION:

In the case of an "AND" connection, all pieces of evidence must be present to establish the hypothesis. In the case of an "OR" connection, the hypothesis is established by any piece of evidence.

INTERPRETATION:

LS measures the degree of sufficiency or suggestiveness of the evidence for establishing the hypothesis. (A larger value of LS means greater sufficiency.) LN measures the degree of necessity of the evidence for establishing the hypothesis. (A smaller value of LN means greater necessity.) The value LS = 1 (LN = 1) indicates that the presence (absence) of the evidence is irrelevant to the hypothesis. For example, if LS > 1 and LN = 1, then the presence of the evidence is suggestive of the hypothesis; its absence does not lower the probability of the hypothesis.

INTERPRETATION:

Do not attempt to establish space B unless and until space A has been established with certainly greater than zero and less than or equal to 5. Context interval [–5, 0] indicates A must have negative certainty before attempting to establish B. Context interval [–5, 5] indicates simply that one should ask about A (regardless of the answer) before asking about B. Omitted context interval indicates (0, 5].

Fig. 1 – Schematic key to PROSPECTOR model diagrams.

The collection of assertations and rules comprising an inference network are easiest to understand when presented in a graphical format. Figure 1 presents a schematic key for interpreting PROSPECTOR model diagrams. Figure 2 depicts the top level of a PROSPECTOR model, called RWSSU, for a class of 'Western States' sandstone uranium deposits. Dashed boxes in that diagram indicate sections of the model that are defined on other pages of the complete diagram.

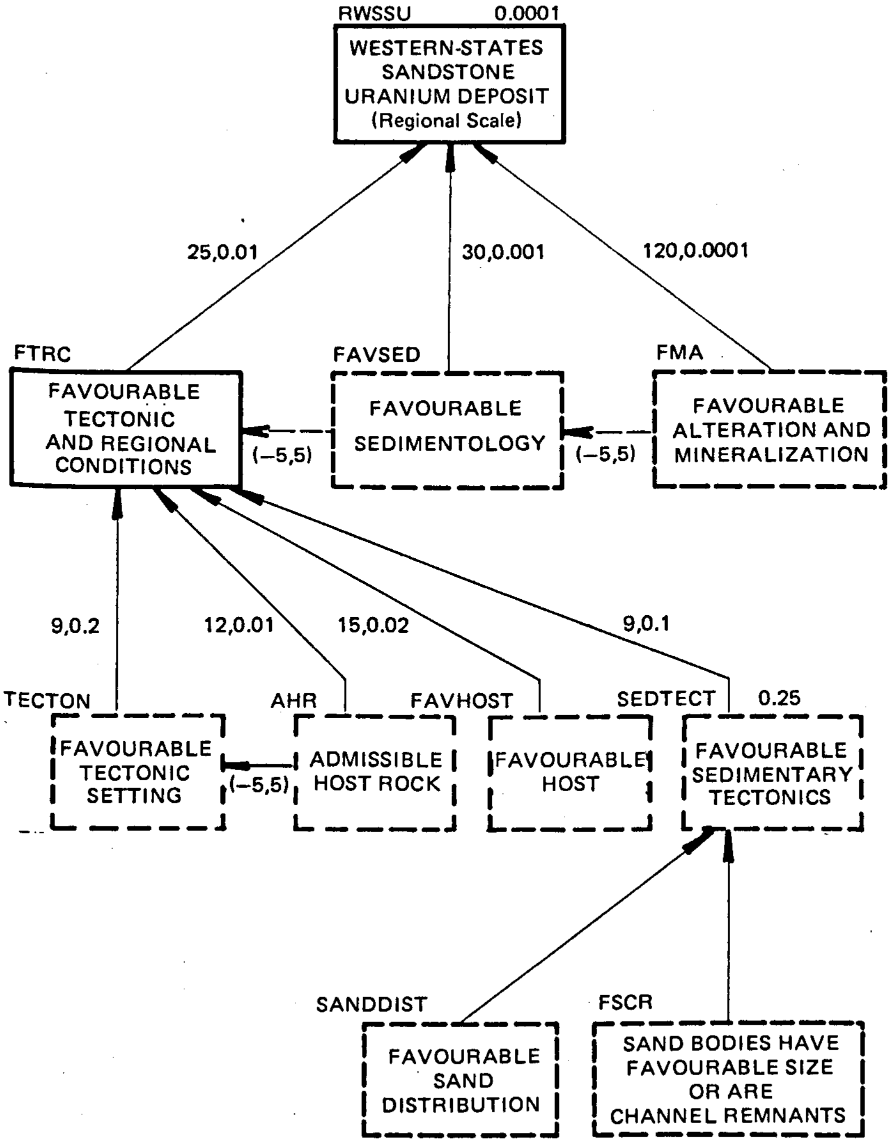


Fig. 2 – Top levels of inference network for regional-scale Western-States sandstone uranium model (RWSSU).

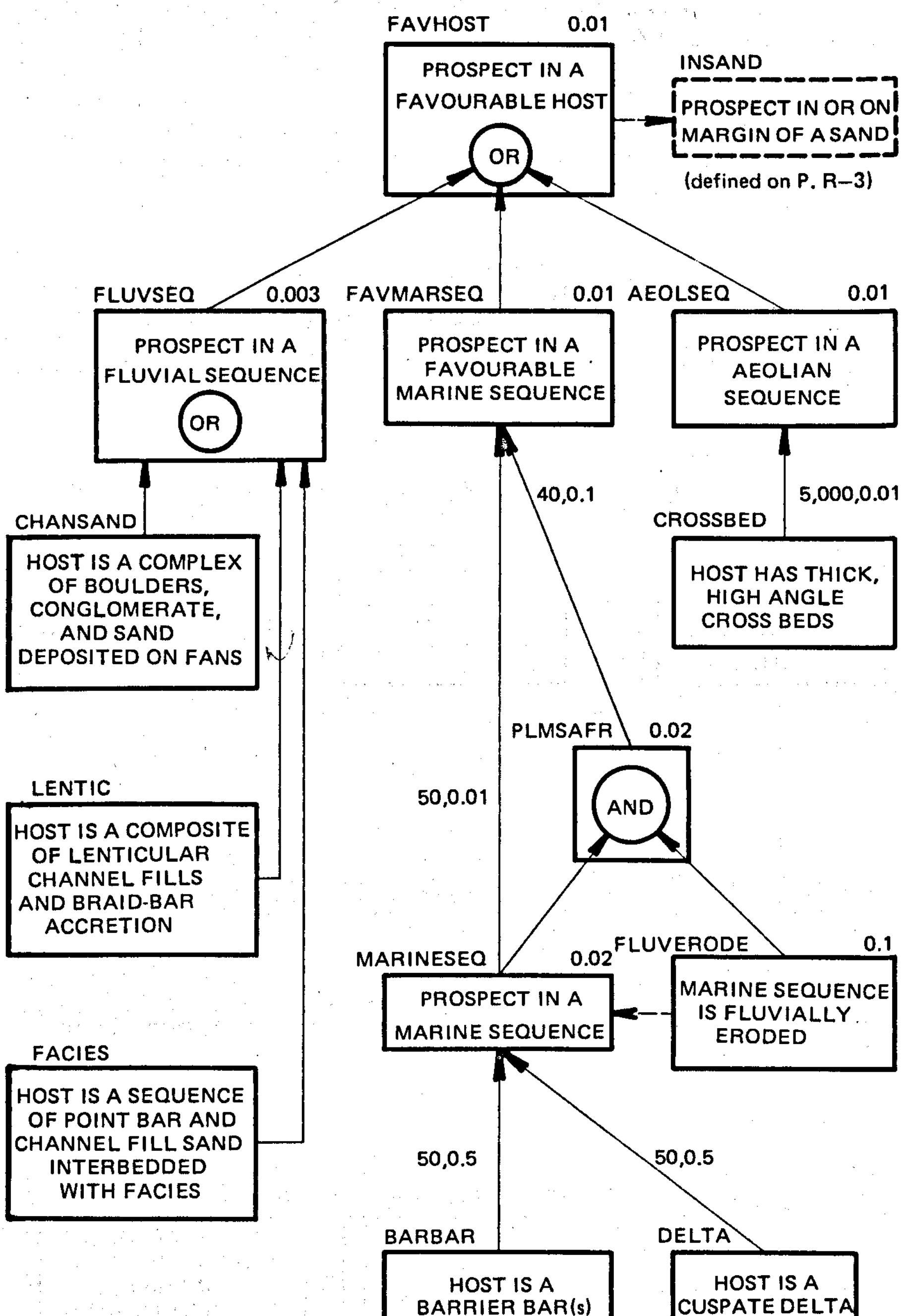


Fig. 3 – A portion of the inference network for the RWSSU model.

For example, Fig. 3 defines the FAVHOST section appearing in a dashed box in in Fig. 2. The complete diagram of the RWSSU model spans 31 pages (Gaschnig 1980a). An overview of the PROSPECTOR system and its inference network methodology is provided in Duda, Gasching, Hart & 1979.

Here we focus on the RWSSU model, and report the results of extensive quantitative tests measuring how faithfully it captures the reasoning of its designer across a set of specific sites (used as case studies in fine-tuning the model), and with respect to the detailed subconclusions of the model as well as its overall conclusions.

Having so validated the performance of the RWSSU model, we then describe a pilot study performed in conjunction with the National Uranium Resource Evaluation (NURE) program of the U.S. Department of Energy. The pilot study applied the RWSSU model to evaluate and compare five target regions, using input data provided by DoE and USGS geologists (using the medium of a model-specific questionnaire generated by PROSPECTOR. The results of the experiment not only rank the test regions, but also measure the sensitivity of the conclusions to more certain or less certain variations in the input data.

One interesting facet of this study is that several geologists provided input data independently about each test region. Since input data about each region varies among the responding geologists, so do the conclusions; we demonstrate how PROSPECTOR is used to identify and resolve the disagreements about input data that are most significantly responsible for differences in the resulting overall conclusions.

The paper concludes with brief descriptions of other recent practical applications of PROSPECTOR.

2. VALIDATING PROSPECTOR MODELS

2.1. Methodology

The practical usefulness of an expert system is limited if those working in its domain of expertise do not or will not use it. Before they will accept and use the system as a working tool, such people (we shall call them the 'domain users') usually expect some evidence that the performance of the system is adequate for their needs (e.g., see Yu *et al.* 1978). Accordingly, considerable effort has been devoted to evaluating the performance of the PROSPECTOR system and its various models (Duda *et al.* 1978, Gaschnig 1979). In the present case, we first needed to validate the performance of the uranium model to be used in the pilot study for the U.S. Department of Energy.

The methodology used to evaluate PROSPECTOR's performance is discussed in detail elsewhere (Duda *et al.* 1978, Gaschnig 1979). For brevity, here we outline a few relevant factors. The PROSPECTOR knowledge base contains a distinct inference network model for each of a number of different classes of ore deposits, and a separate performance evaluation is performed for each model. Here we are concerned with one such model, called the regional-scale 'Western

States' sandstone uranium model (RWSSU), designed by Mr R. Rackley. Since there exist no objective quantitative measures of the performance of human geologists against which to compare that of PROSPECTOR, we instead use a relative comparison of the conclusions of a PROSPECTOR model against those of the expert geologist who designed it. To do so, first a number of test regions are chosen, some being exemplars of the model and others having a poor or less good match against the model. For each such case, a questionnaire is completed detailing the observable characteristics that the model requests as inputs for its deliberation. PROSPECTOR evaluates each such data and derives its conclusion for that test case, which is expressed on a scale from −5 to 5. As a basis of comparison, we also independently elicit the model designer's conclusion about each test case, based on the same input data, and expressed on the same − 5 to 5 scale. Then we compare PROSPECTOR's predictions against the target values provided by the model designer.

2.2. Comparing PROSPECTOR with the Expert

Table 1 compares the top-level conclusions of PROSPECTOR (using the RWSSU model) against the corresponding target values provided by the model designer for eight test regions.

Table 1 – Comparison of RWSSU model with designer for eight cases.

Test region	Target value	Prospector score	Difference
Black Hills	3.50	4.33	−0.83
Crooks Gap	4.70	4.26	0.44
Gas Hills	4.90	4.37	0.53
Shirley Basin	4.95	4.13	0.82
Ambrosia Lake	5.00	4.39	0.61
Southern Powder River	4.40	4.40	0.00
Fox Hills	1.50	2.17	−0.67
Oil Mountain	1.70	3.32	−1.62

Table 1 indicates that the average difference between the PROSPECTOR score and the corresponding target value for these eight cases is 0.69, which is 6.9% of the −5 to 5 scale.

One feature of the PROSPECTOR system is the ability to explain its conclusions at any desired level of detail. Besides the overall conclusions reported above, quite detailed information about PROSPECTOR's conclusions was collected for each test case. In its normal interactive mode, the user can interrogate PROSPECTOR's conclusions by indicating which conclusions or subconclusions he wishes to see more information about. The same sort of information is presented in Table 2 (using the Gas Hills region as an example), in the form of PROSPECTOR's overall evaluation, the major conclusions on which the overall evaluation is based, and the subconclusions that support each major conclusion. For brevity, each section of the RWSSU model represented in Table 2 is identified by a symbolic name, which is indented to show its place in the hierarchy of the model. A key describing each symbolic name follows Table 2. For comparison, we first elicited from the model designer his target values for each section of the model listed in Table 2; these values are included in Table 2.

Table 2 – Detailed comparison of RWSSU model with designer for Gas Hills region.

Section of model	Target value	Prospector score	Difference
RWSSU	4.90	4.37	0.53
FTRC	4.80	4.64	0.16
TECTON	4.50	4.50	0.00
AHR	5.00	4.95	0.05
FAVHOST	4.80	5.00	−0.20
SEDTECT	4.80	4.88	−0.08
FAVSED	4.90	4.68	0.22
FLUVSED	4.90	4.68	0.22
MARINESED	−3.50	−2.07	−1.43
AEOLSED	−2.50	−2.10	−0.40
FMA	4.95	4.41	0.54
RBZONE	5.00	4.60	0.40
AIZONE	4.00	4.77	−0.77
MINZONE	5.00	5.00	0.00

Average difference = 0.36 (average of absolute values).

Key to assertion names:

RWSSU:	the region is favourable for 'Western States' sandstone uranium deposits
FTRC:	there are favourable tectonic and regional conditions
TECTON:	the prospect lies in a favourable tectonic setting
AHR:	there is admissible host rock
FAVHOST:	the prospect is in a favourable host
FSCR:	the sand bodies are of favourable size or are channel remnants

FAVSED:	there is a favourable sedimentology
FLUVSED:	there is a favourable fluvial sedimentology
MARINESED:	there is a favourble marine sedimentology
AEOLSED:	there is favourable aeolian sedimentology
FMA:	there are favourable mineralization and alteration
RBZONE:	there is a favourable remote barren zone
AIZONE:	there is a favourable altered interior zone
MINZONE:	there are indications of a possible mineralized zone

The data in Table 2 indicate that PROSPECTOR not only reaches essentially the same numerical conclusions as its designer, but does so for similar reasons.

The type of detailed comparison shown in Table 2 was repeated for each of the eight test cases listed in Table 1, resulting in 112 distinct comparisons between PROSPECTOR's prediction and designer's target value (i.e., 8 test regions times 14 sections of the model). Table 3 shows the results. Each column in Table 3 corresponds to the column labeled "Difference" in Table 2. The letters A through H in Table 3 identify the eight test cases in accordance with the key immediately following this table. Hence the data in Table 3 under column C is taken from Table 2.

As indicated by the datum in the lower right corner of Table 3, the grand average of the RWSSU model's error in predicting Mr Rackley's conclusions across the 112 combinations of 8 test regions and 14 major sections of the model is 0.70, which represents 7.0% of the −5 to 5 scale.

The rightmost column in Table 3 lists the averages over the eight test regions for each of the 14 major sections of the RWSSU model. Hence this column ranks these model sections according to their predictive abilities. This information suggests a priority ordering for the future revisions of the model, in that those sections having the largest average error (e.g., MARINESED, MINZONE, and so on) are the ones that could benefit most by further fine-tuning. By comparing these values in Table 3 with the analogous averages in a table corresponding to a revised model, one can measure quantitatively the extent to which the revisions achieved the objectives that motivated them. In point of fact, the fine-tuning of the RWSSU model to its current status was based on just such a feedback process.

2.3 Sensitivity Analysis

The user's certainty about inputs provided to PROSPECTOR are expressed on −5 to 5 scale (as opposed to simply 'yes' or 'no', for example). Hence PROSPECTOR's conclusions depend on the degree of certainty of its inputs. To measure the sensitivity of conclusions to perturbations in the certainties of the inputs, we make two additional executions of PROSPECTOR for each set of input data. In one case we change each of the user's input certainties by one unit closer to zero, so that, for example, a 4 becomes 3 and a −3 becomes −2. In the other

Table 3 – Differences between target value and PROSPECTOR score (RWSSU model, 8 regions, 14 sections of the model).

Test regions:	A	B	C	D	E	F	G	H	Avg.
RWSSU	−0.83	0.44	0.53	0.82	0.61	−0.00	−0.67	−1.62	0.69
FTRC	−0.85	−0.04	0.16	0.16	0.33	−1.14	1.00	−1.60	0.53
TECTON	0.10	0.00	0.00	0.00	0.10	0.00	−0.01	0.04	0.03
AHR	−1.65	−0.45	0.05	−0.20	−0.05	−0.64	0.94	0.12	0.51
FAVHOST	−1.50	−0.50	−0.20	−0.50	0.00	−0.50	0.90	−2.00	0.76
SEDTECT	−1.10	−0.26	−0.08	−0.28	−0.06	−0.35	−0.50	−1.57	0.52
FAVSED	−0.88	0.12	0.22	0.08	−0.05	−0.84	−3.67	−2.23	1.01
FLUVSED	−0.88	0.12	0.22	0.08	−0.05	−0.84	−3.59	−2.23	1.00
MARINESED	0.76	0.49	−1.43	−0.08	−0.75	−2.83	−0.17	2.10	1.07
AEOLSED	0.15	0.23	−0.40	−0.61	0.26	−1.29	−0.64	0.44	0.50
FMA	−0.85	0.86	0.54	1.91	0.20	0.05	−0.23	0.11	0.59
RBZONE	−0.84	−0.29	0.40	0.11	−0.31	−0.05	0.69	−3.01	0.71
AIZONE	−0.78	0.66	−0.77	1.94	−0.24	−0.23	0.36	1.50	0.81
MINZONE	−2.00	−0.40	0.00	−0.10	−0.50	−0.40	−3.20	−1.73	1.04
Average:	0.94	0.35	0.36	0.49	0.25	0.58	1.18	1.45	0.70

(Note: averages are averages of absolute values).

Key to test regions:

A = Black Hills, B = Crooks Gap, C = Gas Hills,
D = Shirley Basin, E = Ambrosia Lake, F = Powder River,
G = Oil Mountain, H = Fox Hills

case, each input certainty is changed by one unit toward 5 or −5, e.g., −3 becomes −4. These 'less certain' and 'more certain' variants of the original input data set are then run through PROSPECTOR, so that the resulting conclusions can then be compared with those obtained from the original 'standard' data set, as illustrated in Table 4. In Table 4 'maximum difference' denotes the larger of the two differences between standard and less certain conclusions on the one hand, and standard and more certain conclusions on the other.

Table 4 – Standard, 'more certain', and 'less certain' conclusions (RWSSU model, overall conclusions, 8 test regions).

	Less certain	Standard	More certain	Maximum difference
Black Hills	4.20	4.33	4.37	0.13
Crooks Gap	4.12	4.26	4.25	0.14
Gas Hills	4.30	4.37	4.40	0.07
Shirley Basin	3.99	4.13	4.13	0.14
Ambrosia Lake	4.32	4.39	4.40	0.07
Southern Powder River	4.36	4.40	4.42	0.04
Fox Hills	1.78	2.17	2.23	0.39
Oil Mountain	1.90	3.32	3.49	1.42

Average: 0.12

Of the eight test regions compared, the data in Table 4 indicate that Oil Mountain is the most sensitive to more certain or less certain changes in input certainties. The other seven cases are very stable in this respect. As indicated at the bottom of Table 4, there is an average 1.2% change in conclusions in response to a 10% change in input certainties (i.e., one unit of certainty over a 10-point scale).

Besides the overall conclusions reported in Table 4, quite detailed information was collected for each individual test region. Inspection of detailed sensitivity conclusions for various sections of the model reveals the source of the sensitivity reflected in the overall conclusions represented in Table 4.

We shall present one example in detail below, for the case of Oil Mountain, for which sensitivity about overall conclusions in relatively large. For the cited case, Table 5 compares PROSPECTOR's standard, 'more certain,' and 'less certain' conclusions for each of the 14 major sections of the RWSSU model that were detailed in Table 2 and 3. (See the key following Table 2 to identify the geological assertations corresponding to the symbolic names of these sections.)

The data in Table 5 concerning Oil Mountain reveal that the sensitivity of the overall conclusion (maximum difference of 1.42) is due to the various sensitivities of the FTRC, FAVSED, and FMA sections (maximum difference 62, 1.08,

Table 5 – Detailed standard, "more certain", and "less certain" runs (RWSSU model, Oil Mountain region, 14 sections of the model).

	Less certain	Standard	More certain	Maximum difference
RWSSU	1.90	3.32	3.49	1.42
FTRC	3.23	3.85	3.66	0.62
TECTON	3.05	3.81	3.81	0.76
AHR	1.04	1.28	0.96	0.32
FAVHOST	3.99	5.00	5.00	1.01
SEDTECT	4.32	4.57	4.58	0.25
FAVSED	3.15	4.23	4.42	1.08
FLUVSED	3.15	4.23	4.42	1.08
MARINESED	−0.27	−3.10	−4.22	2.83
AEOLSED	−3.28	−4.44	−3.76	1.16
FMA	0.78	1.64	1.93	0.86
RBZONE	3.98	4.51	4.73	0.53
AIZONE	−1.59	0.00	0.08	1.59
MINZONE	1.45	3.23	3.82	1.78

Average: 1.09

and 0.86, respectively). To illustrate a deeper level of analysis, within the FMA section the data in Table 5 indicate that the mineralized zone section displays highest sensitivity (1.78), followed by the altered interior zone section (1.59), and finally by remote barren zone (which is rather stable in this case – maximum difference = 0.53). Hence, this sort of analysis can pinpoint the section(s) of a model most sensitive to uniform changes in certainty of input data. For most of the sections, there is little change from standard to 'more certain' data, since a large fraction of the standard inputs already expressed near or complete certainty. Note also that for several sections of the model (FTRC, AHR, MARINESED) the 'more certain' score in Table 5 is actually less than the corresponding standard score. These cases reflect a negative effect produced by LS or LN values (discussed in Fig. 1) that is larger than the positive effect of those values greater than 1.

2.4 Comparison of the WSSU, RWSSU, and EDSU Models

Two other sandstone uranium models were also subjected to performance evaluations analogous to those described above for the RWSSU model. Details of these experiments with the prospect-scale 'Western States' sandstone uranium model (WSSU) and the epigenetic carbonaceous sandstone uranium model (ECSU) are given in Gaschnig 1980a. Below we summarize some of the results concerning the RWSSU, WSSU, and ECSU models.

First we shall compare the sizes of these three models. As described previously, a PROSPECTOR model is represented as a network of assertions and the rules of geological inference that connect them. Hence, the size or complexity of a model can be measured in a simple way by the number of assertions and rules it contains. Two types of assertions can be distinguished: those concerning directly observable field evidence, and others concerning higher-level conclusions that can be inferred from such field evidence. Table 6 summarizes these statistics for the WSSU, RWSSU, and ECSU models.

Table 6 – Size statistics for the PROSPECTOR uranium models.

Model	Designer	Number of evidence assertions	Total number of assertions	Number of rules
WSSU	R. Rackley	105	200	148
RWSSU	R. Rackley	107	205	152
ECSU	S. Adams	109	197	153
Total:		321	602	453

As enumerated in Table 6, these three models are of comparable size. Other recently developed PROSPECTOR models are of the same approximate size as the present models (see Table 1 in Duda *et al.* 1979).

Table 7 compares the performance of three uranium models.

The numbers in Table 7 are derived as follows. The total number of PROSPECTOR runs per model (item 2) is three times the number of test sites or regions (item 1), because a 'standard,' 'more certain,' and 'less certain' run were performed for each completed questionnaire. The number given in item 3a is derived from Table 1 for the RWSSU model, and from analogous data for the WSSU and ECSU models. The numbers given in item 3b are derived from Table 3 for the RWSSU model, and from analogous data for the WSSU and ECSU models. Similarly, the numbers in item 4 are derived from Table 4 in the case of the RWSSU model. The number given in item 6 is the product of the number of test sites or regions (item 1) and the number of detailed sections of the model for which target values were obtained from the model designer (item 5): $8 \times 12 = 112$ for the WSSU and RWSSU models; $9 \times 19 = 171$ for the ECSU model. This product is the number of data points averaged to obtain the numbers in item 3b. Note that the values in items 3 and 4 in Table 7 are expressed as percentages of our 10-point certainty scale used to express PROSPECTOR conclusions. The fourth column in Table 7 gives the total of the other three columns for items 1, 2, 5, and 6; it gives the average of the other three columns for items 3 and 4.

Inspection of the data in Table 7 reveals that the three models have excellent average performance both in overall conclusions and in detailed subconclusions.

Table 7 – Comparative performance of the WSSU, RWSSU, and ECSU models.

	WSSU	RWSSU	ECSU	Total/ average
1. Number of test sites or regions:	8	8	9	25
2. Total number of PROSPECTOR runs:	24	24	27	75
3. Average difference between PROSPECTOR score and model designer's target value ...				
(a) for overall conclusions:	6.6%	6.9%	8.0%	7.2%
(b) for detailed conclusions:	7.2%	7.0%	7.8%	7.3%
4. Average sensitivity of overall conclusions to unit change of certainty in input data:	2.2%	1.2%	4.0%	2.5%
5. Number of sections of model represented by detailed conclusions:	14	14	19	47
6. Total number of comparisons for detailed conclusions:	112	112	171	395

Specifically, the difference between PROSPECTOR score and corresponding model designer's target value averages to a 7.3% difference on a 10-point scale in the case of overall conclusions, and to 7.4% difference in the case of detailed conclusions. Each of the three models has low average sensitivity to fluctuations in the certainty ascribed to the field observations on which the tests are based, averaging 2.5% difference from the 'standard' case for overall conclusions. Finally, the differences in performance levels of the three models are small. One should note, however, that the performances of the WSSU and RWSSU models are slightly better than that of the ECSU model, reflecting the fact that the former models have been subjected to somewhat more tuning and testing than the latter. Additional tuning of the ECSU model could be expected to result in improved performance.

3. A PILOT STUDY FOR DOE's NATIONAL URANIUM RESOURCE EVALUATION

3.1. Overall Results

Having established the credibility of the RWSSU model by the test results just discussed, we then undertook an evaluation of five test regions selected by the Department of Energy. For this purpose USGS and DoE geologists completed questionnaires for this model. As a sensitivity test, several geologists independently completed questionnaires for each test region. For comparison, the model

Table 8 – Overall conclusions about five test regions in NURE pilot study.

Geologist	A	B	C	D	USGS team	Rackley data	Average	Range
Monument Hill	4.17	3.32	3.97			4.40	3.97	1.08
Pumpkin Buttes	4.20	3.30	4.19			4.40	4.02	1.10
Moorcroft			3.92	3.88, 4.00		4.00	3.95	0.12
Northwest Gillette			3.64	0.10		3.42	2.39	3.54
White River					0.13	0.01	0.07	0.12

designer, R. Rackley, also completed questionnaires for five test regions. The overall results are reported in Table 8†.

The results in Table 8 indicate that the Monument Hill, Pumpkin Buttes, and Moorecroft regions are very favourable, and about equally favourable, for occurrence of 'Western States' sandstone uranium deposits. Northwest Gillete is scored as moderately favourable, whereas White River is neutral (balanced positive and negative indicators).

Note that each respondent has had different exposure to the target regions, in terms of both first-hand, on-site experience and familiarity with field data reported in the literature. These difficulties in experience are reflected in their answers on the questionnaires. Since different inputs yield different conclusions, one would expect a spread in the certainties about each region, reflecting the differences in input data provided by the various geologists. Inspection of Table 8 reveals that the scores derived from different geologists' input data about the same region agree rather closely for each region except Northwest Gillette (see the column labelled "Range"). These generally close agreements reflect the capability of PROSPECTOR models to synthesize many diverse factors, mechanically ascertaining general commonalities without being unduly distracted by occasional disparities.

In cases such as Northwest Gillette in which a large difference in conclusions occurs, it is easy to trace the source of the disagreement by comparing the individual conclusions for different sections of the model (representing different geological subconclusions), as in Table 9.

Table 9 – Comparison of detailed conclusions about Northwest Gillette region.

Geologist	C	D	Rackley	Average
RWSSU	0.10	3.66	3.42	3.56
FTRC	4.67	3.80	4.63	4.37
TECTON	4.90	4.50	4.50	4.63
AHR	4.95	1.03	4.94	3.64
FAVHOST	5.00	5.00	5.00	5.00
SEDTECT	4.98	4.33	4.78	4.69
FAVSED	0.04	3.92	4.79	2.92
FLUVSED	0.04	3.92	4.79	2.92
MARINESED	−4.60	3.34	0.02	−0.41
AEOLSED	−4.99	−2.10	−3.23	−3.44
FMA	0.27	2.45	1.33	2.18
RBZONE	4.10	4.83	4.73	4.55
AIZONE	−3.29	2.40	0.00	−0.30
MINZONE	0.41	2.82	2.59	1.94

† Note that geologist 'D' divided the Moorcroft region into two subregions and completed a questionnaire for each, resulting in two conclusions (3.88 and 4.00 in Table 8).

Inspection of Table 9 reveals that the conclusions agree fairly closely for the FTRC section of the model, and less closely for the FAVSED and FMA sections. Tracing the differences deeper, one sees that of the three factors on which FMA depends, there is fairly good agreement about RBZONE, but larger differences in the cases of the AIZONE and MINZONE sections. In some cases, such a detailed analysis can isolate the source of overall disagreement to a few key questions about which the respondents disagreed. These can then be resolved by the respondents without the need to be concerned with other disagreements in their questionnaire inputs that did not significantly affect the overall conclusions.

3.2. Sensitivity Analysis

Table 10 lists PROSPECTOR's overall conclusions about the five test regions, comparing standard, 'more certain,' and 'less certain' interpretations of the questionnaire input data, analogous to those described earlier in validating the RWSSU model. (The standard results are the same as those in Table 8.) The column labelled 'Maximum difference' in Table 10 is the maximum of the difference between the standard and 'less certain' scores on the one hand, and between the 'more certain' and standard scores on the other. Hence this column gives the maximum sensitivity for each case.

Table 10 – Standard, 'more certain', and 'less certain' runs (RWSSU model, overall conclusions, five test regions).

Site	Geologist	Less certain	Standard	More certain	Maximum difference
Monument Hill:	A	4.02	4.17	4.20	0.15
	B	3.70	3.32	3.34	0.38
	C	4.21	3.97	3.88	0.24
	Rackley	4.32	4.40	4.41	0.08
Pumpkin Buttes:	A	4.09	4.20	4.22	0.11
	B	3.69	3.30	3.31	0.39
	C	4.13	4.19	4.22	0.06
	Rackley	4.34	4.40	4.41	0.06
Moorcroft:	C	3.72	3.92	3.97	0.20
	D	3.81	3.88	3.87	0.07
	D	3.94	4.00	4.01	0.06
	Rackley	3.71	4.00	4.24	0.29
Northwest Gillette:	C	3.26	3.64	3.94	0.38
	D	0.91	0.54	0.89	0.37
	Rackley	2.99	3.42	3.63	0.43
White River:	USGS team	0.34	0.13	0.07	0.21
	Rackley	0.13	0.01	0.00	0.12

Average: 0.21

The data in Table 10 indicate very stable performance in each case tested, averaging 2.1% difference on the ten-point certainty scale.

Besides the overall conclusions reported in Table 10, quite detailed information was collected for each individual run. We shall present one example in detail below, for the case of geologist A's data about Monument Hill. For the cited case, Table 11 compares PROSPECTOR's standard, 'more certain,' and 'less certain' conclusions for each of the 14 major sections of the RWSSU model that were detailed in previous tables.

Table 11 – Detailed standard, 'more certain', and 'less certain' runs (RWSSU model, Monument Hill region, Geologist A).

Section of model	Less certain	Standard	More certain	Maximum difference
RWSSU	4.02	4.17	4.20	0.15
FTRC	4.02	4.43	4.57	0.41
TECTON	3.60	4.50	4.50	0.90
AHR	4.89	4.93	4.95	0.04
FAVHOST	1.97	2.98	3.99	1.01
SEDTECT	4.36	4.79	4.94	0.43
FAVSED	4.22	4.56	4.66	0.34
FLUVSED	4.22	4.56	4.66	0.34
MARINESED	0.02	−2.39	−3.65	2.37
AEOLSED	0.38	0.14	0.12	0.24
FMA	3.30	3.42	3.41	0.12
RBZONE	4.57	4.74	4.84	0.17
AIZONE	3.28	3.41	3.40	0.13
MINZONE	4.66	4.84	4.86	0.18

Average: 0.49.

The 'maximum difference' column in Table 11 identifies three sections of the model exhibiting significantly greater sensitivity than the other sections for Monument Hill, namely, TECTON, FAVHOST, and MARINESED. The sensitivity of MARINESED is irrelevant in this case, since the sedimentology (FAVSED) is clearly established as fluvial (FLUVSED) rather than marine. The sensitivities of TECTON and FAVHOST are reflected in the somewhat smaller sensitivity of FTRC, which, in turn, contributes to a small sensitivity in the overall conclusion. Hence, senstivities propagate upward through the model, but the impact of a single sensitive section of the model is usually diluted when combined with other more stable factors. This is a salient consequence of the hierarchical structure of PROSPECTOR models. In any case, this kind of analysis can pinpoint the section(s) of a model most sensitive to uniform changes in certainty of input data.

3.3. Conclusions

This pilot study for DoE's NURE project has attempted to demonstrate the methodological features of the PROSPECTOR approach for problems of resource assessment. The numerical data tabulated in the preceding subsection, along with complete analogue sets of tables in (Gaschnig 1980a), provided extensive evidence addressing a variety of questions:

- Measuring the faithfulness of the RWSSU model to its designer's reasoning across several regions and across several major sections of the model.
- Ranking the test regions.
- Determining the range in certainties for each region, reflecting different geologists' input data.
- Determining whether agreement about a region extends from overall conclusions to detailed subconclusions as well.
- Pinpointing the source of disagreement about input data that resulted in any disagreements in overall conclusions about a region.
- Measuring the sensitivity of conclusions to 'more certain' or 'less certain' variations in each individual's input data.

Table 12 highlights many of these results.

Table 12 – Summary of results of the DONE NURE pilot study.

	Results/Remarks
1. Average difference between PROSPECTOR score and model designer's target value ...	
(a) for overall conclusions:	7.5% over 5 regions
(b) for detailed conclusions:	9.2% over 70 combinations of 5 regions and 14 sections of model
2. Ranking test regions: (Table 8)	Very favourable – Monument Hill, Pumpkin Buttes, Moorcroft Moderately favourable – Northwest Gillette No match – White River
3. Variability of conclusions about each region, reflecting different geologists' input data: (Table 8; see also Table 9)	Negligible – Moorcroft, White River Small – Monument Hill, Pumpkin Buttes Large – Northwest Gillette
4. Average sensitivity to unit change in certainty of input data: (Table 10; see also Table 11)	2.1% over 17 questionnaires about 5 regions

In sum, we have performed a precise, step-by-step evaluation of five target regions, beginning with an independent detailed assessment of the accuracy of the RWSSU model in predicting its designer's reasoning; we then scored each region in accordance with the input data provided by several geologists; we further examined the numerical results in detail to determine their sensitivity to a variety of factors; finally we demonstrated how the PROSPECTOR approach allows one routinely to identify and resolve disagreements in conclusions resulting from differences in input data provided by different geologists.

This evidence demonstrates clearly and extensively the usefulness of PROSPECTOR not only for evaluating regional mineral potential, but also for actually quantifying the credibility and stability of its conclusions. Given the variabilities and uncertainties inherent in the task of resource assessment, the PROSPECTOR methodology introduces a powerful new tool by which to obtain assessments significantly more objective, repeatable, uniform, self-calibrating, detailed, and open to public inspection (hence defendable), than those presently available using other methods.

4. A LAND USE STUDY IN ALASKA

PROSPECTOR has also been applied to several other tasks, which we mention heré briefly. In one case, PROSPECTOR was used to evaluate several regions on the Alaskan Peninsula for uranium potential (Cox, Detra, and Detterman 1980). In this case the practical issue was to evaluate the mineral potential of Federal lands for the purpose of deciding their ultimate disposition (e.g., wilderness status versus commercial exploitation). As in the NURE study, geologists familiar with the locales completed questionnaires corresponding to two PROSPECTOR models representing different types of uranium deposits; the questionnaire data were processed by PROSPECTOR, resulting in evaluations that included several levels of geological detail. Table 13 shows the results. (See the key following Table 2 to identify the geological assertions corresponding to the symbolic names of the various sections of the model listed at the left in Table 13.)

Inspection of Table 13 shows that the three formations exhibit mild to moderate favourability for this type of deposit. Looking into the three major categories of evidence underlying these evaluations, we see that all three formations are moderately to very favourable with respect to two of the factors (namely, FTRC – tectonic and regional conditions; and FAVSED – sedimentology). However, none of the three formations offers more than a very weak match against the third major category, namely FMA – favourable mineralization and alteration. By inspecting the three factors underlying the latter (namely RBZONE, AIZONE, and MINZONE), it turns out that the three formations offer essentially no match with respect to the most dominant of the three factors, namely MINZONE – a favourable mineralized zone. In sum, Table 13 shows the three regions tested have environments generally favourable for the type of deposit being assessed, but do not satisfy the key requirement of having a good mineralized zone.

Table 13 – Uranium favourability of three areas is Alaska (using the RWSSU model).

Section of model	Chignick formation	Tolstoi formation	Bear Lake formation
RWSSU	2.91	2.28	1.06
FTRC	3.59	4.01	4.32
TECTON	3.81	3.81	3.81
AHR	1.09	1.66	4.53
FAVHOST	4.73	5.00	3.99
SEDTECT	4.04	4.51	3.57
FAVSED	4.81	4.68	3.64
FLUVSED	4.81	4.68	3.64
MARINESED	3.74	0.04	−1.08
AEOLSED	2.27	0.05	−4.90
FMA	1.08	0.58	0.19
RBZONE	4.62	4.47	1.99
AIZONE	3.19	1.64	0.00
MINZONE	0.41	0.41	0.25

To provide evidence concerning the sensitivity of these results to the uncertainties and omissions of the field observations on which the results are based, this study also included 'more certain' and 'less certain' runs analogous to those described earlier. In doing so we demonstrated that the results are rather insensitive to unit perturbations in the certainties of the questionnaire input data. In short, small perturbations of the inputs produced smaller perturbations of the outputs.

One interesting facet of this study was that one of the USGS geologists charged with evaluating these regions was not himself an expert about uranium deposits. By using PROSPECTOR, these geologists, in effect, augmented their own experience with the judgment of a noted authority on the specilized types of deposits being considered. Hence, this is another case in which specialized geological expertise has been disseminated to where it was needed.

5. AN EXPERIMENT MEASURING THE VALUE OF A MAP

A somewhat different application of PROSPECTOR was concerned with measuring quantitively the economic value of a geological map (Shapiro & Watson 1979). The USGS expends great time, effort, and money in creating maps detailing the geological characteristics of various geographical districts. The question naturally arises as to whether the benefits obtained from using such maps justify their cost. In an attempt to answer the question quantitatively, the USGS conducted an experiment using a porphyry copper model of the PROSPECTOR system.

Ten sites were selected, five of which were known to contain a deposit matching the characteristics of the specified model; the remaining five sites were known to be barren. Three different maps were available for each site: one at a scale of 1:24,000; another at 1:250,000; and a third at a scale of 1:1,000,000. Two geologists were selected to provide input data to PROSPECTOR about each test site, based only on information contained on the corresponding maps they were provided. For each site, the designated geologist completed three copies of the PROSPECTOR questionnaire for the copper model, each copy corresponding to a different scale map. Using the completed questionnaire as input, PROSPECTOR assigned three sets of scores to each test site, one for each scale map.

These results were used to determine how well PROSPECTOR distinguished the sites containing deposits from the barren sites, as measured by the Spearman Rank Correlation Coefficient. In the case of the 1:24,000 maps, PROSPECTOR's predictions obtained a correlation of 0.73 (on a −1 to 1 scale), which was statistically significant at the 5% confidence level (even though the sample size was small). Statistical significance was not obtained in the case of the smaller scale maps, indicating either that these maps contain insufficient information to evaluate the test regions, or alternatively, that the sample size was too small to obtain significance in these cases. To distinguish the latter two possibilities, an extension of this study to an additional 20 sites is now in progress.

6. OTHER APPLICATIONS IN PROGRESS

Since the completion of the NURE pilot study, the U.S. Geological Survey and the Department of Energy have funded additional evaluations similar to the NURE and Alaska tests. One project will extend the NURE pilot study to additonal regions selected by DoE. Another will score the relative merits of a dozen areas in the San Jaun Basin, New Mexico, for possible occurrence of epigenetic carbonaceous sandstone uranium deposits, for which a PROSPECTOR model (called ECSU) is available.

PROSPECTOR has also made its first prediction about the location of an as yet undiscovered ore body. For this purpose we used a drilling site selection model developed for porphyry molybdenum deposits. Whereas the prospect-scale and regional-scale models discussed previously are intended to provide an overall evaluation of a prospect-sized property or a larger region, PROSPECTOR's drilling site selection models are intended to pinpoint exact spots where an ore body will be found. Toward this end, an area is overlaid with a grid of cells (typically 128 × 128), so that each cell represents an area about 30 meters square. Using graphic input data obtained by digitizing features from a geological map, PROSPECTOR evaluates each cell in the grid, using an efficient network compiler to speed the evaluation process (Duda *et al.* 1978, Konolige 1979). Each of the 16 384 resulting cell scores is colour-coded, and then the entire map is displayed on a colour graphic display, so that the brightest areas are most favourable for drilling. To test their accuracy, the drilling site selection models have been subjected to a number of *a posteriori* tests, using areas where ore deposits have already been discovered and mined, but using maps from an early stage of exploration for the purpose of the test. In these tests of a porphyry copper drilling site selection model (Duda *et al.* 1978), PROSPECTOR's predictions agreed very closely with the outline of the known orebody. Since mining companies sometimes drill dozens or hundreds of holes for every commercially viable deposit discovered, the potential of accurate 'first-hole' predictions is very great. Accordingly, we were encouraged to develop another drilling-site selection model, for porphyry molybdenum deposits (Duda, 1980). This model has been used to predict the possible occurrence of a deposit in a relatively unexplored area of the Mount Tolman region in Washington State. These results which were published in Duda 1980, are currently being evaluated as further exploration data are becoming available.

Another type of evaluation of PROSPECTOR is now being planned, namely a 'peer review' workshop. Each of the experts who designed PROSPECTOR models will present his model to a group of geologists who are also knowledgeable about that type of deposit. The object is to elicit comments and criticisms (and perhaps a consensus) among the geologists present about the model under discussion. While it is rather common for geologists to disagree among each other to certain degrees, an advantage of the PROSPECTOR approach is that the geologists attending the upcoming workshop will be focussing on very precisely stated issues (e.g., this factor in the model or that rule strength value). Our hope is that the PROSPECTOR methodology will contribute to the increased codification of economic geology.

7. DISCUSSION

We have measured PROSPECTOR's expertise explicitly and found that its detailed conclusions match those of the expert who designed it, to within about 7% on a 10-point certainty scale used as a basis for comparison. Having so validated the models, we presented results of several applications to practical tasks for the USGS and DoE. In so doing, we demonstrated in particular how the PROSPECTOR approach deals effectively with the variabilities and uncertainties inherent in the task of resource assessment (i.e., by sensitivity analysis).

This work illustrates that expert systems intended for actual practical use must accommodate the special characteristics of the domain of expertise. In the case of economic geology, it is not rare for field geologists to disagree to some extent about their field observations at a given site. Accordingly, the use of various sorts of sensitivity analysis is stressed in PROSPECTOR to bound the impact of such disagreements and to isolate their sources. In so doing, we provide geologists with new quantitative techniques by which to address an important issue, thus adding to the attractiveness of PROSPECTOR as a working tool.

Other domains of expertise will have their own peculiarities, which must be accommodated by designers of expert systems for those domains. A more mundane, but nevertheless practically important, example concerns the use of a questionnaire as a medium for obtaining input data to PROSPECTOR from geologists. Most geologists have little or no experience with computers; furthermore, acess to a central computer from a remote site may be problematic in practice. On the other hand, geologists seem to be quite comfortable with questionnaires. Our point is simply that issues ancillary to artificial intelligence usually have to be addressed to ensure the practical success of knowledge-based systems.

REFERENCES

Cox, D. P., Detra, D. E., & Detterman, R. L., (1980). Mineral Resources of the Chignik and Sutwick Island Quadrangles, Alaska, U.S. Geological Survey Map MF-1053K, (in review).

Duda, R. O., Hart, P. E., Barrett, P., Gaschnig, J., Konolige, K., Reboh, R., and Slocum, J., (1978). Development of the PROSPECTOR consultation system for mineral exploration, *Final Report, SRI Projects 5821 and 6415*, Menlo Park: Artificial Intelligence Center, SRI International.

Duda, R. O., Gaschnig, J. G., and Hart, P. E., (1979). Model design in the PROSPECTOR consultant system for mineral exploration, *Expert Systems in the Microelectronic Age*, pp. 153–168 (ed. Michie, D.). Edinburgh: Edinburgh University Press.

Duda, R. O., Hart, P. E., Konolige, K., & Reboh, R., (1979). A computer-based consultant for mineral exploration, *Final Report, SRI Project, 6415*, Menlo Park: Artificial Intelligence Center, SRI International.

Duda, R. O., (1980). The PROSPECTOR system for mineral exploration, *Final Report, SRI Project 8172*, Menlo Park: Artificial Intelligence Center, SRI International.

Gaschnig, J. G., (1979). Preliminary performance analysis of the PROSPECTOR consultant system for mineral exploration, *Proc. Sixth International Joint Conference on Artificial Intelligence*, (Tokyo), pp. 308–310.

Gaschnig, J. G., (1980a). Development of uranium exploration models for the PROSPECTOR consultant system, *SRI Project 7856*, Menlo Park: Artificial Intelligence Center, SRI International.

Gaschnig, J. G., (1980b). An application of the PROSPECTOR system to DoE's national uranium resource evaluation, *Proceedings of the First Annual Conference of the American Association for Artificial Intelligence*, pp. 295–297.

Konolige, K., (1979). An inference net compiler for the PROSPECTOR rule-based consultation system, *Proc. Sixth Intl. Joint Conference on Artificial Intelligence.* (Tokyo), pp. 487–489.

National Uranium Resource Evaluation, (1979). Interim Report, *Report GJO-111(79)*, Grand Junction, Colorado: U.S. Department of Energy, June.

Roach, C. H., (1979). Overview of NURE Progress Fiscal Year 1979, *Preprint of Proceedings of the Uranium Industry Seminar*, Grand Junction, Colorado: U.S. Department of Energy October 16–17.

Shapiro, C., & Watson, W. (1979). An Interim Report on the Value of Geological Maps, Preliminary Draft Report, Director's Office, U.S. Geological Survey, Reston, Virginia.

Yu, V. L., *et al.*, (1978). Evaluating the performance of a computer-based consultant, *Memo HPP-78-17*, Stanford: Heuristic Progamming Project, Dept. of Computer Science, Stanford University.