

01 Jan 1990

## An Expert System to Convert Knowledge-Based Geological Engineering Systems into Fortran

Ralph W. Wilkerson

Missouri University of Science and Technology, [ralphw@mst.edu](mailto:ralphw@mst.edu)

Jill J. Cress

Follow this and additional works at: [https://scholarsmine.mst.edu/comsci\\_facwork](https://scholarsmine.mst.edu/comsci_facwork)



Part of the [Computer Sciences Commons](#)

---

### Recommended Citation

R. W. Wilkerson and J. J. Cress, "An Expert System to Convert Knowledge-Based Geological Engineering Systems into Fortran," *Proceedings of the Fourteenth Annual International Computer Software and Applications Conference, 1990*, Institute of Electrical and Electronics Engineers (IEEE), Jan 1990.

The definitive version is available at <https://doi.org/10.1109/CMPSAC.1990.139349>

This Article - Conference proceedings is brought to you for free and open access by Scholars' Mine. It has been accepted for inclusion in Computer Science Faculty Research & Creative Works by an authorized administrator of Scholars' Mine. This work is protected by U. S. Copyright Law. Unauthorized use including reproduction for redistribution requires the permission of the copyright holder. For more information, please contact [scholarsmine@mst.edu](mailto:scholarsmine@mst.edu).

# An Expert System to Convert Knowledge-Based Geological Engineering Systems into Fortran

Jill J. Cress and Ralph W. Wilkerson  
Department of Computer Science, University of Missouri-Rolla,  
Rolla, MO 65401 ralphw@cs.umr.edu

## Abstract

A knowledge-based geographic information system (KBGIS) for geological engineering map (GEM) production was developed in GoldWorks, an expert system development shell. Using this shell, the geological engineer is able to develop a rule base for a particular application that results in a valid GEM. However, this implementation failed as a practical production system due to the excessive execution time required to produce a GEM. To solve this problem, a Conversion Expert System was developed which accepted, as input, a KBGIS and produced, as output, the equivalent Fortran code. Two major objectives are accomplished as a result of this system: GEM production time is dramatically reduced and the versatility of the KBGIS development environment is retained.

## Introduction

Expert system development is rapidly becoming an integral part of almost every field. This trend can be traced to the growing need to add "knowledge" to processes and the proven success of using expert systems to fulfill this need. In fact, the need for a more knowledgeable production process recently led to the introduction of expert systems into yet another area, that of map production in geological engineering. This introduction resulted in the development of a knowledge-based geographic information system (KBGIS) for geological engineering map (GEM) production.

GEMs are maps that represent the "geological engineering conditions" of an area; they "show surficial and bedrock geologic patterns classified according to engineering suitability for urban development..." [6],[7]. The specific information which is represented on any single GEM is determined by the chosen classification scheme, which, in turn, is based on the intended use of the map.

Classification schemes refer to the methods used to

classify the data of a GEM. A classification scheme is developed by determining exactly what features are to be represented on the map and then defining each class in terms of the chosen features. A key is also included on the map to show the definition of each class. Features can refer to properties such as soil, geology, and topography, as well as specific attributes of these properties, i.e. flooding, permeability, karst, plasticity index, and slope.

One of the primary uses of GEMs is as a tool in determining the best site for a particular type of development. Site selection GEMs would show the geological characteristics of an area classified according to their suitability for the specified type of development. For example, if a possible landfill site was being investigated, characteristics such as flooding frequency and soil permeability would obviously be important. Therefore, in response to this need, a GEM would be produced to classify the area under investigation in terms of these factors.

## GEM Production Method

The GEM production cycle begins by identifying an area which needs to be mapped. Once this area has been chosen, a series of six processing steps are performed. These steps, which are performed each time a GEM is produced, are usually done by a geological engineer. In fact, the first three processing steps rely heavily on input and interpretation from an engineer skilled in both geological principles and mapping concepts. Most, if not all, of these production steps are performed manually.

Obtaining the area maps necessary to generate the GEM is the first production step. This involves determining what features are going to be used and then gathering the maps of the area which contain information about these features. Once all of the necessary maps have been obtained, the next production step involves establishment of the specific criteria. The en-

gineer determines which characteristics are the most important for the specific map being produced. After the desired criteria has been established, all of the maps that were obtained in the first step, both published and on site investigations, are analyzed. The fourth step of the production process is printing the map. Determining acceptance of the map is the fifth step in the GEM production system. This involves not only confirming that the map meets all quality guidelines, but also gaining the approval of its users.

The map is not considered complete until it has been field checked; the last step in the production process. Field checking means confirming that the new map is correct by actually going into the field to check it. Once the GEM has been successfully field tested, the production process is complete and a published map is produced.

Several weaknesses can be observed in this GEM production system. One of the more serious of these is the non-standardization of the system. Since there are no set policies or guidelines for the development of classifications, it is entirely possible to discover two maps which define the same specific classification in different ways. In fact, studies of existing maps show that a definite variance exists among classification definitions. These definitions are obvious victims of the lack of standardization.

Another weakness inherent in the GEM production process is the demands it places, in terms of both time and effort, on a skilled geological engineer. Few of the steps can be performed without extensive interaction with the engineer, which makes the system extremely inefficient. The system also uses a "back to the drawing board" approach for all of its resource gathering and data manipulation. This means that not only does the engineer have to spend a large amount of time developing each GEM, he will also be performing the same basic processing steps over and over again.

In summary, although the GEM production system fulfills its purpose by producing a GEM, it does so in neither an efficient nor reliable manner. In fact, the weaknesses inherent in this system – its lack of standardization and extreme reliance on interaction with a skilled engineer – make it a prime target for modernization.

### **Development of the KBGIS**

It was felt that the GEM production system could be significantly improved by automation. Preliminary in-

vestigations determined that the best way to automate would be to "...utilize expert knowledge in GIS (geographic information system) processing" [6]. In other words, create a rule based expert system with the ability to interact with existing GISs, i.e., a knowledge-based GIS (KBGIS) [5], [6].

GIS refers to an information system which allows for the input, manipulation, and output of geographic databases. Geographic databases refer to files which contain digitized geographic data. There are a variety of different types of geographic databases; including digital line graphs (DLGs), which are created by digitizing an existing map.

During the initial investigation of existing maps it was discovered that the variety of GEM classifications varied dramatically from region to region. It was clear that the development of a single set of valid classifications would be virtually impossible.

In an effort to determine the minimal factors, i.e., properties, necessary for the creation of a GEM, a sample of GEMs were chosen for detailed investigation. This involved not only studying the maps but also contacting their creators in an effort to establish exactly what thought processes were used in their development. The result of this investigation was a list of each output classification and exactly what factors made up the classification.

Experimentation with the 1st Class expert system shell [4] resulted in the discovery that by reordering the factors on the table, different decision trees could be created. The table was then repeatedly reordered until a decision tree was developed which contained the minimum number of factors needed to create a GEM. These factors were identified as slope, karst, plasticity index, and flooding.

An appropriate classification scheme had to be chosen before the rules could be developed. In other words, a standardized classification scheme had to be established. This meant, that in addition to resolving any variations between classification definitions, a determination had to be made on what kind of end product was desired.

It was decided that a general GEM, instead of a specific site selection GEM, was a desirable end product. This general end product increased the efficiency of the production system by removing the need to repeat the entire production process each time a specific site selection GEM was needed. Instead of repeating

the process, the general GEM could be used to derive any of the specific end products. In other words, a one step translation from a general GEM to a specific GEM could easily be performed.

Development of the rule base was essentially completed once the minimum factors and a classification scheme were developed and chosen. All that remained was simply converting the correct decision tree into if-then structured rules.

The first KBGIS design used the LISP environment to implement the rule-base. A LISP program was created that interacted with both the rule base and the input files (overlays) in order to generate a file of the GIS operations needed to produce a GEM. Production of the GEM was then simply a matter of using a GIS to perform these operations. The Earth Resources Data Analysis System (ERDAS) software package, which is "a raster-based GIS and image processing system", was chosen as the GIS for this implementation [6].

### Current KBGIS Implementation

"The final system design implements GIS processing within the expert system" [6]. In other words, the final system design required that an expert system be developed that would actually produce the GEM. The GoldWorks expert system shell package from Gold Hill was used to develop this expert system [1].

The KBGIS uses conceptual values, that represent the symbolic values of the various overlays, for processing. For instance, when the karst input is being processed by the expert system, it will have one of the following conceptual values; "no", "yes", or "water". These conceptual values are then used by the rule base to determine a conceptual GEM classification value, like "class-Ia". In other words, since both the input and output files consist of numerical pixels, the expert system must have the ability to perform the following mappings: "pixel to conceptual" and "conceptual to pixel".

These mappings are performed by implementing the GIS recode operation in the KBGIS. The recode operation is more complex in the KBGIS environment than it was in the GIS, since it requires more than a "pixel to pixel" mapping. Referring back to the recoding example for the slopes, in the KBGIS the final values would be the conceptual ones, "low", "medium", and "high".

Implementation of the KBGIS recode operation was

done by making use of the frame-based data structures. In order to perform the KBGIS recode operation, first a frame structure had to be created for each of the GIS data files. The set of slots for each frame represented all of its possible conceptual values. For example a frame called karst would have "yes", "no", and "water" as its slots. In order to perform the recode operation, each time the KBGIS is executed an instance for each of the frames is created and the slot values are set equal to the appropriate pixel values. The KBGIS recode can also map multiple pixel values into a single conceptual value since GoldWorks allows multi-valued slots. The pixel values that are placed in the specific slots are user-supplied for the input GIS files and system-supplied for the output GEM GIS file.

Currently, the KBGIS can produce a GEM given one of two different types of input. The first type of input requires four separate GIS overlay files; flooding, slope, plasticity, and karst. In addition to supplying these input files, the user must also supply the information necessary to map the input pixel value to a conceptual value for each of the files. This information is then used by the system to convert the input pixel from each of these files into a conceptual value, which is then placed in a pixel instance. The rule base then performs the matrix operation and sets the GEM slot of the pixel instance. This conceptual value is then recoded into a pixel value and output to the GIS formatted GEM file. Once all of the pixels have been processed, a GIS trailer file is created which contains a legend to aid in interpreting the GEM pixels.

The second type of input that the KBGIS can accept is that of a GIS soils file. A soils file is simply a composite of the flooding, slope, plasticity, and karst overlays. Therefore, it can also be used to produce a GEM file. In order to use a soils file, however, the user must first of all describe to the system each of the possible soil pixels in terms of the flooding, slope, plasticity, and karst overlays. In other words, a specific pixel instance is created for each of the possible soils pixels and the appropriate conceptual values are filled in. The system then performs a matrix operation to set the GEM slot for each of these pixels. To produce a GEM file now simply requires a KBGIS recode of the soils file and the creation of a trailer file.

### KBGIS Evaluation

The final KBGIS was tested by first selecting an area that already had a manually produced GEM and then generating a KBGIS GEM for it. These two maps were then compared and the results of the comparison

were analyzed. The Creve Coeur, Missouri, area was chosen for this test because a manually produced GEM already existed for it [3],[6].

The first generation step required that the needed input files be in ERDAS GIS format. It was decided that the soils file would be used for the GEM generation. Therefore, the published map of the soils overlay was digitized and then converted into ERDAS format. Once this input was in the correct format, the KBGIS was used to produce a GEM. This GEM was then displayed on a graphics system, as was the digitized manually produced GEM, and side by side comparisons were performed.

It was concluded that a "good" correlation existed between the two maps, taking into account land use changes that may have occurred in the 10 years since the manual GEM was produced [6]. The major differences that were observed between the two maps were caused by differences in the classification methods, rather than problems with the system [6]. In fact, the KBGIS produced a more detailed GEM than the one produced manually. It was, therefore, concluded that a KBGIS could be used to produce a GEM.

Unfortunately, the system did not prove to be very practical. The amount of execution time required by the KBGIS severely hampered its practicality. Generation of the Creve Coeur GEM took in excess of 76 hours. In fact, the system was not allowed to run to completion. Instead, processing was stopped and the soils file was recoded by ERDAS by utilizing the KBGIS generated table. This meant that, although the KBGIS automated the GEM production process, it was not at all efficient in terms of time.

### The Conversion Expert System

Results of the testing of the final KBGIS supported the theory that an automated approach could be used to produce GEM maps. In fact, automation strengthened the production process by standardizing the classification definitions, reducing the time commitment required of an engineer, and removing the need to start from scratch each time a specific site selection GEM was required. Unfortunately, the automated process still required a significant amount of production time.

In response to this problem, an investigation was performed to determine how the developed knowledge base could be more efficiently implemented as a production system. This resulted in development of an expert system which converts any existing KBGIS into

Fortran code, which can then be compiled and used efficiently in production.

One of the primary causes of the slow KBGIS execution time is the large size of the input and output files, since a typical GIS overlay file could easily contain as much as 250,000 bytes of data. The overhead created by the processing of such large files is significant. Expert system inference algorithms, while supporting extremely powerful reasoning abilities, also, by their very nature, require a larger execution time for the processing of larger data files.

A detailed study of the GEM expert system has shown that the knowledge base can essentially be split into two separate parts; "an application-specific knowledge base", and a GIS knowledge base [6]. The GIS part consists of the knowledge needed to perform the recode and matrix operations that enable the system to convert the input data into a GEM. This represents the stable part of the knowledge base; regardless of the specific application, this part of the system remains unchanged. The application knowledge base, however, is not stable. This is the part of the knowledge base consisting of the specific rules that were developed to generate the GEM from the minimal factors, both of which were established for the Midwest region only. This means that in order to create a production system for any other region, the development process would have to be repeated to establish a valid application knowledge base for the new region.

The application knowledge base consists of two parts; data structures for the minimal factors and the rules which determine the classifications. The rules themselves follow a fairly standard if-then format. The if part consists of a combination of the minimal factors and associated conceptual values; the then part sets the correct classification value. These rules are mutually exclusive, and will always be mutually exclusive since classification schemes demand that unique factor combinations be established for each classification definition. The number of rules will, therefore, be equal to the number of possible classifications, which will vary region to region. The data structure of the minimal factors will also remain basically stable region to region. This structure associates with each of the minimal factors all of its possible input values.

Although the GoldWorks expert system is obviously vital to the development of the application knowledge base, it is inefficient when used to implement the resultant knowledge base in a production environment. It was also observed that once the rule base is estab-

lished, its characteristics did not require implementation in an expert system environment. This discovery led to the conclusion that, once the final rule base was established, it could be converted into another, more efficient, programming language for production purposes. However, it was also observed that this conversion from a GoldWorks based expert system into another language would require access to an experienced computer programmer. This conversion process would also have to be performed multiple times and possibly over a prolonged time period, since any changes to a region's established rule base must also be reflected in the production system.

It was, therefore, decided that an intermediate expert system should be developed that could accept as input any GEM knowledge base and convert it into another computer language. This new program could then be compiled and used for the production system. Development of an expert system to perform the conversion process would solve both the need to have access to a programmer, and the need to perform the conversion process repetitively.

### **Establishing Validity of the Conversion System**

The first step in developing a conversion expert system was to choose the language to convert the KBGIS into, and then confirm that significant time savings would be obtained. Fortran was chosen as the language because of its inherently fast execution time for data intense files, which can be seen in Fortran based ERDAS processing. It was also chosen because of its availability, since the majority of the computer systems currently available in the field support Fortran.

After Fortran was chosen, the existing KBGIS for the Midwestern region was manually converted into equivalent Fortran code. This hand-coded program was compiled and an executable program developed. A sample data set was then used to develop a GEM, using both the Fortran program and the GEM expert system. Comparisons were made between the two GEMs to confirm that the Fortran code was producing identical results. A byte by byte comparison was made between the two GEM files and no discrepancies were found. Therefore, it was concluded that the Fortran program was producing valid results.

A comparison was also done between the execution times of the two GEM production systems, which were both executed on the same hardware in order to eliminate any hardware related execution factors. The For-

tran program was able to produce the GEM in 5 minutes and 49 seconds whereas the KBGIS had still not completed production after 3 hours. It was, therefore, concluded that the Fortran production system was able to produce a valid GEM with a significant time savings as compared to the GEM expert system.

The first task of this expert system was to convert the GoldWorks framed-based data structures into appropriate Fortran data structures. Fortran arrays were chosen, since they allowed a specific number of elements to be associated with a single data structure. This was a vital requirement, since each of the frames in the KBGIS had associated slots. Conversion of a frame into an equivalent Fortran array was done by creating an instance of the specific frame, and then giving the name of the equivalent Fortran array element as a value for each of the slots. The name of the array is used as the instance name, and a key list containing the frame name and associated array name was constructed.

An equivalent Fortran array data structure was created for each of the frames representing one of the minimal factors. Determination of these frames was performed by searching the rule base and developing a list of all frames used in the antecedent part of the rules which set the GEM classification. This list was then used as a key for which frames needed to be converted. Once all of the appropriate frames were converted, it was time to convert the rules.

Conversion of the GoldWorks rules into Fortran if-then rules required extensive use of recursion. Only the rules which deduce a GEM classification needed to be converted. Therefore, the first step was to identify those rules. This was done by retrieving each rule's consequence and then searching it for the name which represented the classification slot of the pixel frame. GEM is the slot name in the existing KBGIS. If a match was found, then the specific classification value which the rule sets was checked for validity. Confirmation of validity was done by making sure that the classification value specified in the rule matched one of the slots for the GEM frame in the KBGIS.

Once it was established that a rule needed to be converted, a list of the rule's antecedent was retrieved. This list can consist of several layers, depending on the complexity of the antecedent. Each member of the list was searched until the minimal sublist was found. This sublist is a list which actually contains the pattern used to search the fact base. This list is then converted into Fortran code and, if necessary, the

appropriate logical connective is also converted into Fortran. This process was recursively repeated until each of the minimal sublists were converted. After the rules antecedent was processed, the consequence was retrieved and converted. This conversion was much simpler, since the list of the consequence contains only atoms, that can be easily converted.

The only part of the KBGIS left to convert was its user interface. This conversion was relatively simple, since the only anticipated changes would be the names and numbers of the input files since these are based on the minimal factors used by the system. The format of both the input and output files was constant since it was based on the use of GIS files; and the creation of a trailer file was also constant. Essentially, all that was required for this part of the conversion was to create the user input interface. This required using the list of the minimal factors (developed earlier in the conversion) to generate the Fortran code needed for input/output processing by using formatted stream output statements.

### Conclusions

Once the generated Fortran code is converted into an executable program, it can then be used to produce a GEM. This program is now equivalent to the KBGIS system in terms of input and output requirements, i.e., it requires one of two types of input and produces two outputs (GEM and trailer). The only differences in the systems are their methods of processing; the Fortran program performs pixel to pixel recodes instead of pixel to conceptual to pixel recodes like the KBGIS. In addition, the Fortran code is limited to one-to-one mappings, unlike the KBGIS which supports multiple values to a single value mapping. This is, however, the only additional constraint added by the Fortran production system. And, if a multiple to single mapping is required of the system, it can easily be performed in ERDAS as a pre-processing step.

The Conversion Expert System was tested by converting the Midwestern KBGIS expert system. After the KBGIS was converted into Fortran code by the expert system, an executable Fortran program was created. In other words, the Fortran code which was output by the expert system was compiled and linked. Next, the executable Fortran program was used to produce a GEM and then compared to a GEM produced by the KBGIS from the same input. Comparisons resulted in no differences being found between the GEMs.

The Fortran production system was able to process 179 bytes per second (.0056 sec/byte) as compared to the KBGIS which processed only .77 bytes per second (1.30 sec/byte). And since this comparison was based on a data set of only 62500 bytes and it had already been determined that the larger the data set the worse the KBGIS system performs (in fact the time required to process a byte grows exponentially) it was obvious that the Fortran production system performs in a significantly more efficient manner.

This system significantly improved the practicality of the automated production of GEMs by drastically reducing the production time. The Fortran approach was also able to solve the possibly inhibiting effect of the cost of the system. The KBGIS could then be developed and converted at the central office, and only the resulting executable program would need to be sent out to the field offices.

### References

- [1] Gold Hill. GoldWorks Expert System Development and Delivery. Cambridge, Massachusetts: Gold Hill Computers, INC., 1987.
- [2] E. E. Lutzen and J. D. Rockaway, Jr. "Engineering Geology of St. Louis County, Missouri," Engineering Geology Series No. 4 , 1971, 23.
- [3] E. E. Lutzen and J. D. Rockaway, Jr. "Engineering Geology of the Creve Coeur Quadrangle, St. Louis County, Missouri," Engineering Geology Series No. 2 , 1970, 19.
- [4] Programs-In-Motion. 1st Class User's Guide , Wayland, Massachusetts: Programs-In-Motion, Inc., 1987.
- [5] T. R. Smith and M. Pazner. "Knowledge-Based Control of Search and Learning in a Large-Scale GIS," Proceedings International Symposium on Spatial Data Handling , 1984, 498-519.
- [6] E. L. Usery, Phyllis Altheide, Robin Deister, and David Barr. "Knowledge-Based GIS Techniques Applied to Geological Engineering," Photogrammetric Engineering and Remote Sensing , Vol. 54 (November 1988), 1623-1628.
- [7] E. L. Usery, Robin Deister, and David Barr. "A Geological Engineering Application of a Knowledge-Based Geographic Information System," Proceedings of the American Congress on Surveying and Mapping/American Society for Photogrammetry and Remote Sensing , Vol. 2 (1988), 176-185.