

Uncertain spatial reasoning of environmental risks in GIS using genetic learning algorithms

Rouzbeh Shad · Arefeh Shad

Received: 17 August 2011 / Accepted: 14 October 2011
© Springer Science+Business Media B.V. 2011

Abstract Modeling the impact of air pollution is one of the most important approaches for managing damages to the ecosystem. This problem can be solved by sensing and modeling uncertain spatial behaviors, defining topological rules, and using inference and learning capabilities in a spatial reasoning system. Reasoning, which is the main component of such complex systems, requires that proper rules be defined through expert judgments in the knowledge-based part. Use of genetic fuzzy capabilities enables the algorithm to learn and be tuned to proper rules in a flexible manner and increases the preciseness and robustness of operations. The main objective of this paper was to design and evaluate a spatial genetic fuzzy system, with the goal of assessing environmental risks of air pollution due to oil well fires during the Persian Gulf War. Dynamic areas were extracted and monitored through images from NOAA, and the data were stored in an efficient spatial database. Initial spatial knowledge

was determined by expert consideration of the application characteristics, and the inference engine was performed with genetic learning (GL) algorithms. Finally, GL (0.7 and 0.03), GL (0.7 and 0.08), GL (0.98 and 0.03), GL (0.98 and 0.08), and Cordon learning methods were evaluated with test and training data related to samples extracted from Landsat thematic mapper satellite images. Results of the implementation showed that GL (0.98, 0.03) was more precise than the other methods for learning and tuning rules in the concerned application.

Keywords Genetic · Fuzzy · Learning · Reasoning · GIS

Introduction

Among the various damages to the ecosystem, those that originate from air pollution comprise the main elements of environmental pollution. Air pollution produces environmental risk and affects the environment in many ways. Moreover, it is difficult to determine the full impact of this phenomenon in risk assessments. Several aspects of this problem, such as the growth change of vegetation covers, can be considered as an urgent application for making decisions in geo-spatial information systems (GISs). The advanced analytical functions and decision tools available through GISs enable specialists to include

R. Shad · A. Shad
Civil Department, Faculty of Engineering,
Ferdowsi University,
Mashad, Iran

A. Shad
e-mail: Arefeh_Shad@yahoo.com

R. Shad (✉) · A. Shad
Industrial Faculty, Amirkabir University of Technology,
Tehran, Iran
e-mail: Rouzbeh_Shad@yahoo.com

spatial knowledge and relationships in the solution of environmental pollution.

Given the uncertain nature and indeterminate boundaries of air pollution, uncertain modeling and spatial knowledge must be used to monitor dynamic regions and to assess environmental risks. Impact assessments generally are carried out with a minimum of data, high uncertainty, and poor conceptual understanding (Bojórquez-Tapia et al. 2002). Consequently, soft computing, as a combination of computational methods that rectify human mental capabilities to inference and learn in an uncertain environment (Kortenkamp et al. 1998; Konar 2005), can be very effective.

In this paper, remote sensing data were used as the primary data source for monitoring air pollution, and fuzzy modeling was developed by genetic learning (GL) for reasoning and learning in GIS. Our system evaluates different GL methods to determine proper rules for reasoning. When a desirable domain of knowledge is provided, the system can make inferences on a set of realities similar to how a human would find a solution to a problem. The required topological knowledge used for inference and learning is specified in descriptive and predictive modes. In the predictive mode, the aim is to obtain models for the simulation of complicated spatial behaviors. In descriptive mode, not all of the relationships between rules and variables are considered, and the knowledge is used only for some data (Lavrac et al. 2004). For each snapshot, inference rules are learned and tuned with consideration of the uncertain conditions. It is essential to determine an appropriate GL algorithm to use in the final inference process for obtaining risk areas.

This paper is organized as follows. In “[Related works](#),” related works and practical experiences are described relative to the application framework, system design, and algorithm properties. In “[Application and case study](#),” the specific problem, case study, and related spatial data are explained on the basis of the physical characteristics of the environmental area and input data properties. “[Genetic fuzzy system concepts and methods](#)” defines the fundamentals and concepts of fuzzy inference systems and GL. This section describes fuzzy logic formulas to create inputs and to allow inferences, the knowledge base, and rules in a spatial fuzzy controller system, and discusses different combination methods of genetic algorithms with a

fuzzy system. In “[Implementation and evaluation of system](#),” air pollution is monitored on the basis of the input spatial database, the topological knowledge base of environmental risk is updated, and the GL of the database is implemented with a fuzzy controller background. Analytical results of the various GL methods are discussed and evaluated based on crossover and mutation probabilities. Finally, the “[Conclusions](#)” discuss the most robust GL algorithm for estimating environmental risks and describes the preciseness of this system.

Related works

Decision makers of forest policies require scientific information on environmental risk to justify their decisions, so there is a need to evaluate all potential risks of environmental damage (Darbra et al. 2008). The impacts of air pollution on the environment, agriculture, and health in areas surrounding thermo-electric power plants and industrial sites have received increasing attention, with the goal of preventing or minimizing adverse effects or restoring already compromised ecosystems (Triolo et al. 2008). For environmental risk assessments, change detection systems can use various remote sensing and geographic information analytical tools with expert knowledge to delineate the risk areas (Brink and Eva 2009; Kaiser 2009; Makkeasorn et al. 2009). Disadvantages of using these systems include their lack of real-time capability for estimating risk areas, dependency on the knowledge of experts, use of complex models in uncertain situations, and difficulty of integrating with GIS (Tømmervik et al. 1998).

Fuzzy air pollution monitoring (Bojórquez-Tapia et al. 2002) can overcome some of the difficulties of crisp change detection by modeling uncertainty and using expert judgments for assessing environmental risks. Fuzzy systems are able to express the ambiguity of human thinking and translate environmental expert knowledge into computable numerical data. However, these systems for estimating environmental risks have two limitations (Ferson 2003): strong reliance on subjective inputs, and failure to capture the ranges of values in complex data sets and the correlations among them.

Because it is difficult for experts to examine all of the input and output data from a complex system to

find proper rules (Lin and Xu 2006), researchers have proposed hybrid methods, such as mixing fuzzy-set concepts with other methods (Kentel and Aral 2005; Li et al. 2008) to solve this problem. One of the most popular mixing approaches is the use of genetic algorithms (GAs) that lead to fuzzy systems.

For example, Cordon et al. employed fuzzy logic to create a knowledge base and to learn automatically from a database using GA (Cordon et al. 2000). They demonstrated that the use of GA can equip fuzzy controller systems by learning and adopting capabilities for generating and tuning rules (Cordon et al. 2001a). Other researchers have designed different methods to develop membership functions and rule sets by using GA capabilities (Karr and Gentry 1993; Homaifar and McCormick 1995; Lin 2004). Numerous applications of GA have been suggested, including fuzzy knowledge base design, selection of fuzzy rules, and intelligent multiobjective decision making (Wang et al. 2005; Sánchez and Couso 2005; Cordon et al. 2004; Freitas 2002; Setzkorn and Paton 2005). The results obtained from these studies indicate that GL in the complicated space of rules increases robustness capabilities. However, most of these studies did not compare or evaluate various GL methods (Sánchez et al. 2001; Mikut et al. 2005; Ishibuchi and Yamamoto 2004; Cano et al. 2006).

In general, there is a great demand for hybrid systems that can achieve advantages of tuning and learning a knowledge base. However, most developed hybrid systems have a major problem when choosing optimal GL. In this paper, a spatial genetic fuzzy system is proposed to achieve trends that make it possible to extract, store, and optimize topological knowledge by performing optimal inference and learning in a GIS, to assess environmental risks of air pollution plumes.

Application and case study

Several significant events have occurred in the Persian Gulf that influenced the surrounding environment for many subsequent years (Price 1993). In January 1991, millions of barrels of oil entered the Persian Gulf through tankers and oil terminals off Kuwaiti shores. Within less than 1 month, nearly 700 Kuwaiti wells, tankers, and refineries caught fire, which led to the burning of 900 million barrels of oil.

Plumes of smoke and soot covered hundreds of miles for 9 months (Fig. 1). Wartime conditions impeded the rapid reconstruction of oil wells. Consequently, massive quantities of smoke, sulfur dioxide, and unburned hydrocarbons were released into the upper atmosphere. Weather patterns were changed, solar radiation was blocked, and the light available for the photosynthetic growth of plants was limited throughout the Persian Gulf and surrounding region.

Sulfur dioxide and nitrogen oxide injected into the atmosphere can fall back to earth as acid rain. In consequence of the aforementioned events in the Persian Gulf, the production rate of sulfur dioxide was reported to be twice that of nitrogen oxide, and rainfall stations in the western part of Iran reported black rain during the period from 25 to 28 March 1991 (Price 1993). Acid rain can seriously damage soil biology and chemistry. Some microbes are unable to tolerate changes to low pH levels and are killed. The leaching of base cations, such as calcium and magnesium, by acid rain also affects sensitive species. In southwestern Iran, the reaction of the soil to acidification was very poor, and the acidic rains substantially affected vegetation growth and animal life.

The objective in this paper is to infer the spatial relationship between forest areas in southwestern Iran and smoke pollution from Kuwaiti oil wells. Uncertain topological relations will be determined and entered into a knowledge base. The desired inferences and learning will be performed on these relationships in the genetic fuzzy system.

The coast line of Iran along the Persian Gulf has a total length of 2,440 km, and the maritime claims of Iran include a contiguous zone of 24 nm (nautical miles) in width. The concerned study area (Fig. 2) is located at 32° N and 53° E in the Middle East. Resources influenced by the pollution include forests, water resources, uncultivated lands, agricultural lands, meadows, etc. The percentage of natural resources existing in the area include 8% uncultivated lands, 27% meadows, 11% forests, and 54% other resources.

Data used in the inference process include spatial layers, such as forest areas, smoky plumes in NOAA images, persistency of smoke, dominant wind direction, acidic rains, soil type, samples of negative plant growth, and amount of dissipated smoke. Table 1 presents the resources used to gather data.

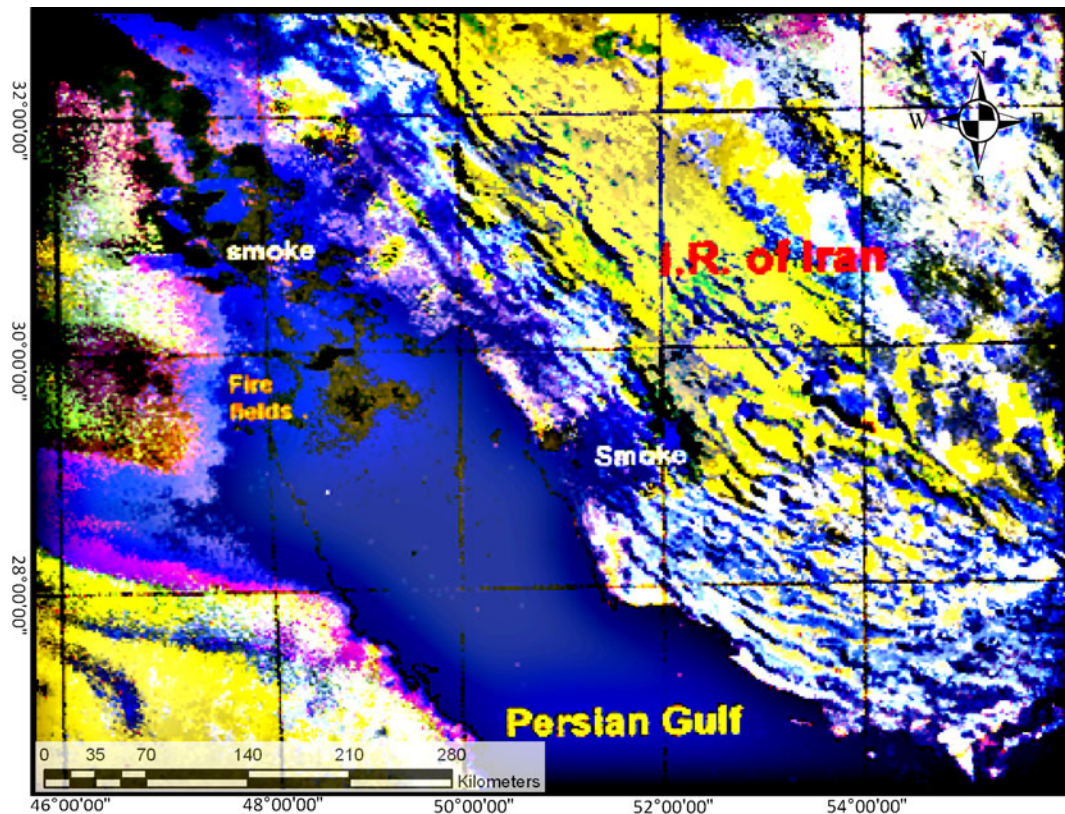


Fig. 1 Movement of smoke plumes toward Iran territory

This paper utilized 1,267 NOAA-AVHRR (1.1 km at nadir) images (stored in high-density Exabyte tapes) that were converted to Erdas software format. The AVHRR sensors are located on four satellites (NOAA 9, 10, 11, and 12), and they provide at least four images every day. Because there were no fully cloudy images and the information quality of bands 1 and 2 was weak, the estimation and prediction of smoke are major problems that should be considered during preprocessing. Band 3 of NOAA-AVHRR is very suitable for fire detection and is used to detect burning oil wells. The transportation of smoke was determined by processing the available images and recorded wind direction vectors. By estimating smokes changes over time, the monthly smoke persistency was calculated. Landsat thematic mapper (TM) monthly mosaic images, with 30-m spatial resolution for bands 1 to 5 and 7, was used to determine the sensitivity of the forest cover area, based on NDVI profiles at several sample points. These samples were collected and analyzed from 1991 to 1993.

Genetic fuzzy system concepts and methods

A fuzzy system is comprised of IF–THEN rules that define relationships between inputs and outputs. Fuzzy systems that fulfill learning or adopting by genetic algorithms are called genetic fuzzy. Genetic algorithms can simultaneously evaluate many points in the search space and adjust and learn the parameters of designed fuzzy systems. The general performance of a genetic fuzzy system for adaptation and learning rules is shown in Fig. 3.

This system has two parts, fuzzy rule base and GL, which are described briefly in the following sections.

Fuzzy rule base

Fuzzy systems can model uncertainty of the input and output parameters by defining fuzzy numbers and setting them as linguistic variables. Fuzzy rule-based systems are performed based on IF–THEN verbal rules that are overlapped throughout the space to handle nonlinear relationships. IF–THEN rules, which

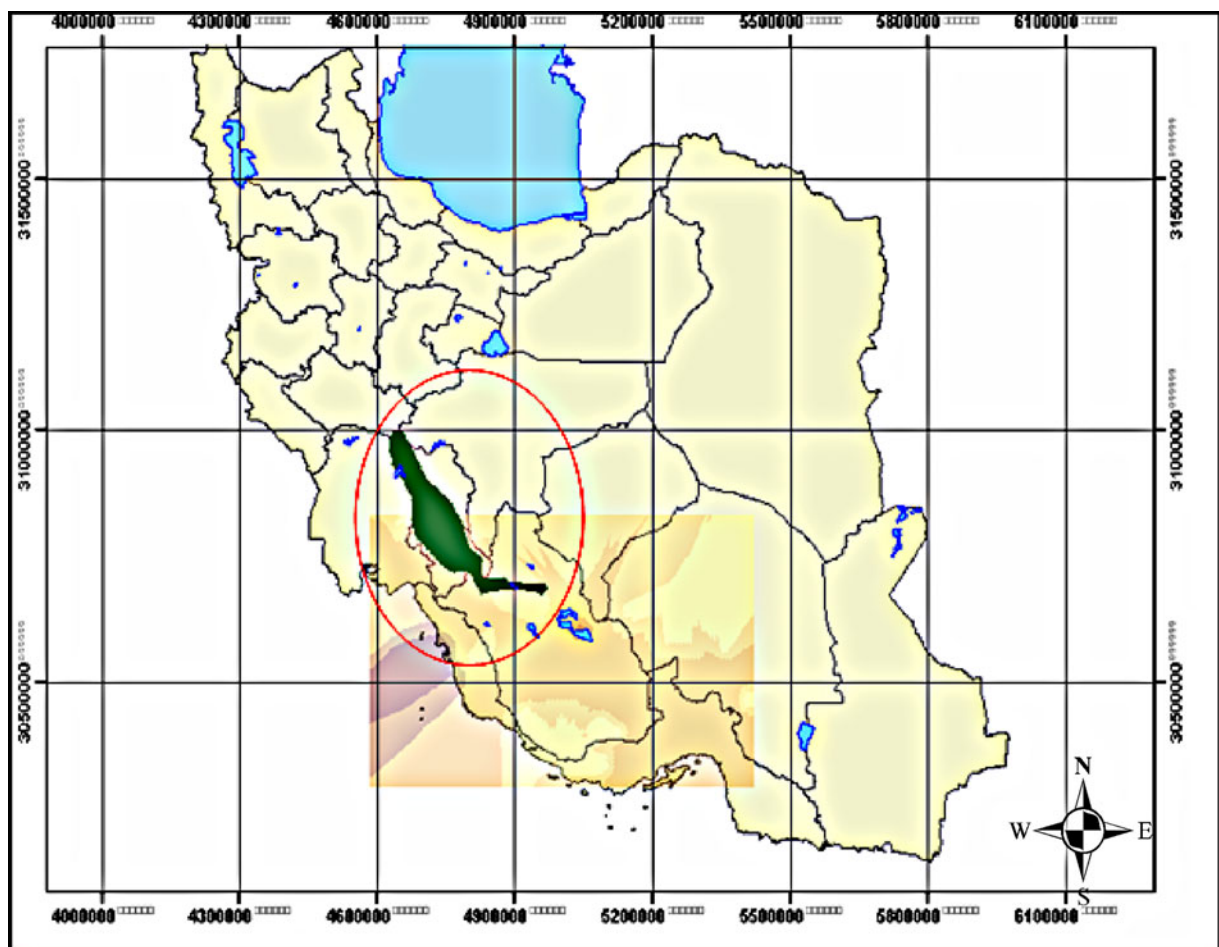


Fig. 2 Study area

are conditional statements, are comprised of a set of conditions to be satisfied (IF) and a set of consequents that can be inferred (THEN). The schematic diagram of the fuzzy rule-based system is shown in Fig. 4. The system is composed of five operational layers,

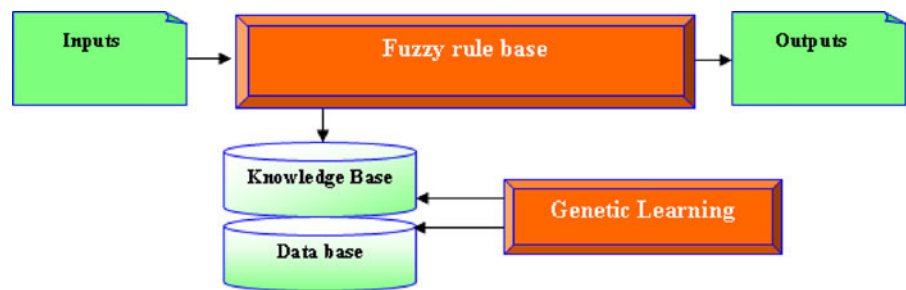
including input linguistic nodes, input term nodes, rule nodes, output term nodes, and output linguistic nodes.

A fuzzy linguistic node represents a fuzzy variable, term node indicates the mapping degree

Table 1 Data used and their collection resources

Data	Source	Scale
Forest area	National Cartographic Center and Environment Organization	1:250,000
NOAA Satellite images	Soil conservation and watershed management research center	—
Smoke persistency	Soil conservation and watershed management research center and oceanography corporation	1:1,000,000
Monthly wind direction	Meteorological organization and oceanography corporation	1:250,000
Monthly acidic rainfall	Meteorological organization	1:250,000
Monthly negative growth of vegetations	Soil conservation and watershed management research center	1:250,000
Soil type	Geological survey	1:250,000
Monthly dissipated smoke	Soil conservation and watershed management research center	1:1,000,000

Fig. 3 General performance of the genetic fuzzy system



of a fuzzy variable, and rule node demonstrates a rule that decides the severity of firing during inference. The first and second layers are called premise layers and the last two layers are called consequential layers. To layers are further described as follows:

Layer 1 In this layer, nodes can directly transfer input values to layer 2. If the input vector for the t th mode is $I_t = (I_{t1}, I_{t2}, \dots, I_{tm})$, with I_{tm} as the input value of the m th fuzzy variable for the t th mode, then the t th output vector of this layer is:

$$\mu : I \rightarrow [0, 1]$$

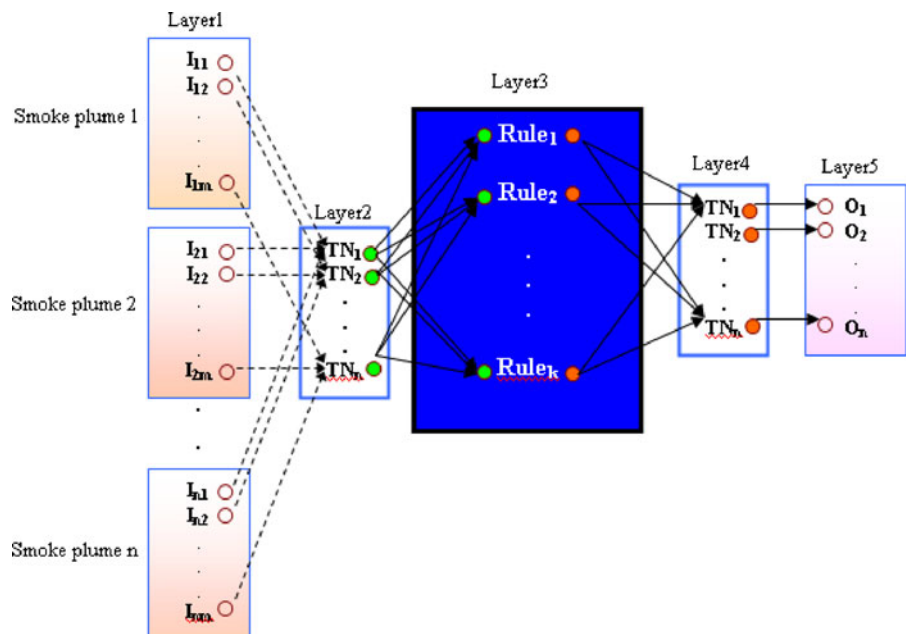
$$\mu_{\text{Layer1}}^t = \left((I_{t11}; I_{t12}; \dots; I_{t1n_1}); (I_{t21}; I_{t22}; \dots; I_{t2n_2}); \dots; (I_{tm1}; I_{tm2}; \dots; I_{tmn_m}) \right), \quad (1)$$

where I_{tmk} is the input value of the k th linguistic term in the m th fuzzy variable for the t th mode, and a fuzzy linguistic term is represented by membership function μ . The great advantage of these linguistic terms is that

they can model the expert's thoughts about the application.

Layer 2 Here, outputs of fuzzy variables are linked to the nodes of the third layer

Fig. 4 Fuzzy rule-based system



to create lawful conditions, by considering some specified rules. These outputs should be obtained from linguistic nodes that are specified in fuzzy rules. Consequently, this layer performs the first step

$$\mu_{\text{Layer2}}^t = ((M_{t11}; M_{t12}; \dots; M_{t1n_1}); (M_{t21}; M_{t22}; \dots; M_{t2n_2}); \dots; (M_{tmn_1}; M_{tmn_2}; \dots; M_{tmn_m}))$$

$$M_{\text{Triangular}}(x) = \begin{cases} 0 & x < x_1 \\ \frac{x-x_1}{x_2-x_1} & x_1 \leq x < x_2 \\ \frac{x_3-x}{x_3-x_2} & x_2 \leq x < x_3 \\ 1 & x_3 \leq x \end{cases} \quad x_1 < x_2 < x_3, \quad (2)$$

where M_{tmk}^2 is the membership degree of the k th linguistic term in the m th fuzzy variable for the t th mode. Membership functions can classify an input variable into varying degrees of different labels, rather than the “0” or “1” used in binary logic. In this paper, trapezoidal functions are used for membership functions of linguistic terms. For example, fuzzy variables of smoke persistency and wind direction were designed and introduced into the knowledge-base system as shown in Fig. 5.

Layer 3 In this layer, each rule node represents a fuzzy IF–THEN rule. These rules are experimentally introduced to the knowledge

base as initial preconditions by using expert questionnaires. The following provides a calculation of the matching degrees of preconditions $\mu_{\text{Layer3}}^{t-\text{mode}1}$ and the weight of the output link $W_{\text{Layer3}}^{t-\text{mode}1}$ for rule node 1 in the t th mode:

$$\mu_{\text{Layer3}}^{t-\text{mode}1} = \min(M_{t11} \dots M_{tmn_1})$$

$$W_{\text{Layer3}}^{t-\text{mode}1} = \prod_{o=1}^m M_{to1}, \quad (3)$$

where \min is a fuzzy minimum operator and \prod multiplies membership degrees M .

Fig. 5 Use of trapezoidal membership functions for smoke persistency and wind direction



Layer 4 The output fuzzy variable of the fuzzy rule base is a risk value ranging between zero and 1. As Fig. 6, this output variable is specified with

five terms: “very high,” “high,” “medium,” “low,” or “very low”. Then, the output of the fourth layer is calculated as follows:

$$\mu_{\text{Layer4}}^t = \underbrace{(C(\max\{\mu_{\text{Layer3}}^{\text{tPVL3}}\} \times \text{verylow}), C(\max\{\mu_{\text{Layer3}}^{\text{tPL}}\} \times \text{low}), C(\max\{\mu_{\text{Layer3}}^{\text{tPM}}\} \times \text{medium}))}_{\substack{S_{t1} \\ \text{PVL} \in \text{Verylow} \quad \text{PL} \in \text{low} \quad \text{PM} \in \text{Medium}}} , \underbrace{(C(\max\{\mu_{\text{Layer3}}^{\text{tPH}}\} \times \text{high}), C(\max\{\mu_{\text{Layer3}}^{\text{tPVH}}\} \times \text{veryhigh}))}_{\substack{S_{t4} \\ \text{PH} \in \text{High} \quad \text{PVH} \in \text{Veryhigh}}} \quad (4)$$

where C (centroid) is the center of gravity function that is used for defuzzification, and \max is the fuzzy maximum operator. In our case, the fuzzy reasoning method used was the classical Mamdani method, which played the roles of the implication and conjunctive operators, and the center of gravity was weighted by the matching strategy, which acted as the defuzzification operator (Cordón et al. 1997).

Layer 5 In this layer, by considering Eq. 4, y_i is calculated with Eq. 5:

$$y_i = \frac{\sum_{h=1}^5 w_{th} \times s_{th}}{\sum_{h=1}^5 w_{th}} \quad (5)$$

where W_{th} is the aggregate weight of rule nodes for each output linguistic term.

Genetic learning

Genetic algorithm is a revolutionary search algorithm that provides robust search capabilities in complex spaces, such as the space of rule sets (Cordón et al.

2004). To solve a problem by GA, a population representing the candidate solutions (randomly generated chromosomes) is used through a process of competition and controlled variation. Chromosomes are rated for their adaptation as solutions during each generation (successive iteration). Therefore, the population moves towards better chromosomes based on the evaluation, by using the fitness function. In this process, new populations are formed through the use of selection mechanisms and the application of genetic operators, such as crossover and mutation.

Genetic algorithms offer advantages for learning in fuzzy systems by searching in their complex space. They can learn and tune different components of the fuzzy rule-based system by optimizing the parameters (Adoption) and generating rules (Learning) in a genetic fuzzy framework. The knowledge-base part of the fuzzy rule-based system consists of databases (in which fuzzy linguistic variables can be defined) and rules (fuzzy IF–THEN relations). These two parts create a searching space in which to apply genetic adoption and learning in a genetic fuzzy system. The following methods are used for learning in genetic fuzzy systems.

Genetic learning of rules To learn rules in a knowledge base, it is necessary to have a predefined database. A typical solution for defining such database is as follows: (1) select an odd number of linguistic terms (between 3 and 9) for each linguistic variable and (2) set the values of the system parameters by using the integrated distribution of linguistic terms in the variable universe of discourse (Fig. 6).

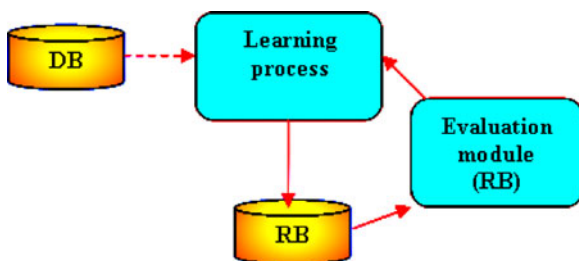


Fig. 6 Genetic learning of rules

Genetic tuning With the goal of increasing the efficacy of the fuzzy rule base system, genetic tuning first optimizes the predefined database and then extracts the rules. At this point, a posterior tuning process considers the whole knowledge base and adjusts the parameters of the membership function. In this mode, the shape of the membership functions changes and the number of linguistic terms in each fuzzy section remains fixed (Fig. 7).

Genetic learning of knowledge base In this method, both components of the knowledge base are trained simultaneously. Genetic learning of the knowledge base makes it possible to develop better definitions and concepts in the inference process. However, enlargement of the search space will render the learning process more difficult and slower (Fig. 8).

Genetic learning of database In this method, which is used in this paper, the knowledge base learning operation is performed to derive the components of rules (RB) and database (DB). For this purpose, DB development covers the RB learning process, based on the condition that “whereas DB is obtained by the execution of the definition process, the RB generation method is applied for the extraction of rules.” Error calculation should be used to validate the whole knowledge base. In this method, an important aim is the optimal partitioning of the learning and tuning operations in the knowledge base to decrease the complication of the searching space (Fig. 9).

Implementation and evaluation of system

The designed genetic fuzzy system (Fig. 10) includes data extractor, topological simulator, fuzzy inference, and GL parts.

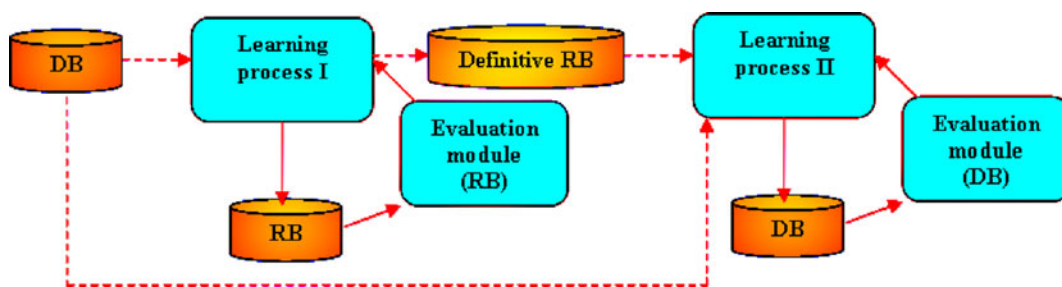


Fig. 7 Genetic tuning

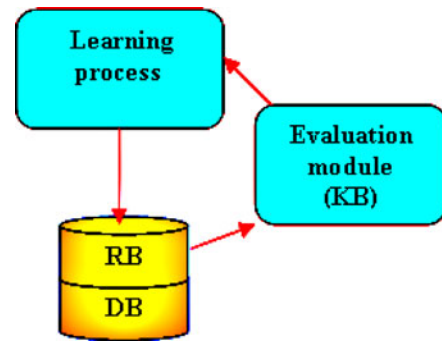
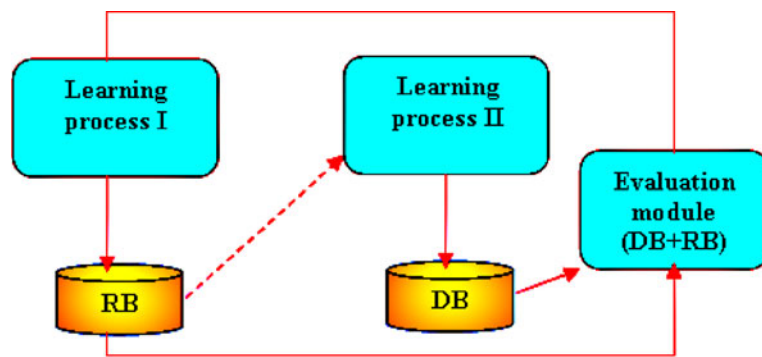


Fig. 8 Genetic learning of the knowledge base

This system is able to determine risk areas in the problem of “the impact of smoke plumes resulting from Kuwaiti oil well fires in 1991 on forest areas in Southwestern Iran”. To apply this problem, when a polluted air plume moves towards a forest area in southwestern Iran, the data extractor obtains the fuzzy limits of the smoke and forest area through NOAA satellite images and sends them to the topological simulator. In the topological simulator, the fuzzy spatial relation between the smoke and forest areas is obtained, and the resulting decision variables are simulated and entered into the knowledge base of the fuzzy inference part. Then, fuzzy inference together with GL determines the risk values in the forest area (monthly as a snapshot) by using a designed inference engine, and updates the existing knowledge in the database. After the completion of a computational period (1 year), the total risk area is specified and risk limits are determined.

Various components are implemented and connected to each other with the help of VB, Matlab, and Arcobjects programming languages. After initial processing and quality control, the obtained data are entered into a spatial database. To increase the data retrieval speed and capabilities of the ASCII format for this purpose, all data were stored in ASCII. Some samples of the stored data are shown in Fig. 11.

Fig. 9 Genetic learning of the database



By using the designed pointers in the object-oriented programming, the data extractor is connected to the spatial database and starts sensing the prepared information, which is stored in ASCII format, for each period of smoke plumes change. This part of system is able to process the portal-received information in a snapshot. In the concerned application, a snapshot means monthly tracked information that demonstrates the spatial behavior variation of smoke plumes, considering the forest areas. Because there are many changes in plant growth and smoke dissipation during February, March, May, June, July, August, and September of 1991, seven snapshots were considered as behavioral change periods (Snapshots). In each period, the connection of the extractor with the database includes the following three operational phases:

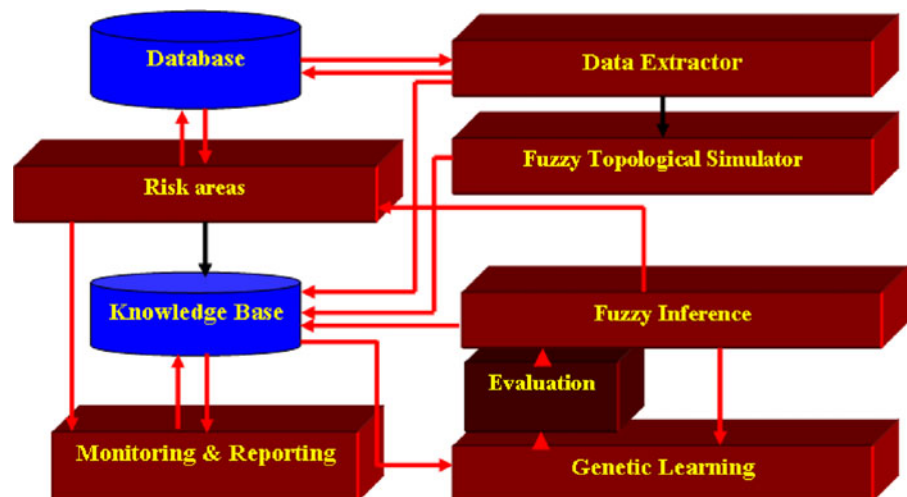
1. Extracting smoke plumes and forest areas in each snapshot in a fuzzy manner, and sending them to the topological simulator.

2. Tracking the variation of smoke persistency, dominant wind direction, and amount of smoke for each snapshot, and sending them to the knowledge base of the inference part of system to fuzzify them.
3. Sensing the received acidic rains, soil type variables, and annual computational risk, and sending them to the knowledge base of the inference part for inference and learning.

The topological simulator, which has the most sensitive function of sensing the topological behavioral variations, receives data from the data extractor continuously. The topological simulator part is comprised of the following three main functions:

1. Simulating fuzzy topological relations between fuzzy regions of smoke plumes and fuzzy areas of forests in each snapshot, and extracting decision variables to send to the knowledge base.

Fig. 10 Architecture of the proposed system



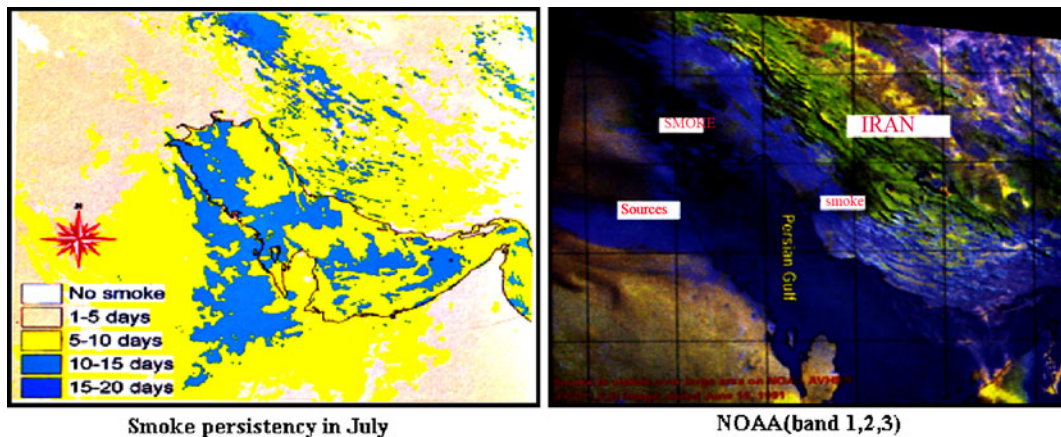


Fig. 11 Samples of prepared spatial data for the application

2. Calculating the approximate distance of smoke plumes to fuzzy forest areas, and fuzzifying them in the knowledge base.
3. Justifying the movement direction of plumes through the forest areas to identify the effective wind direction.

In the inference part, 22 rules are experimentally introduced as its initial preconditions through the completion of a questionnaire by experts. This part is not limited to these rules, and the user is able to increase or decrease rules considering his/her concerned application (Fig. 12).

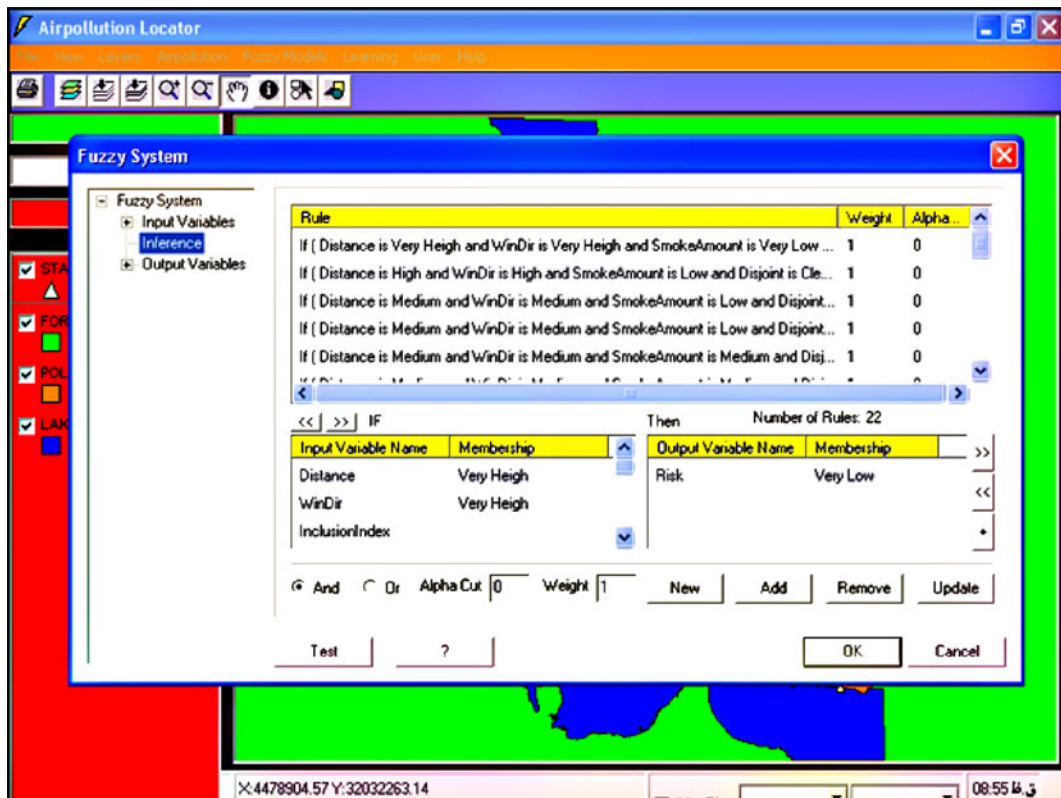


Fig. 12 Twenty-two defined rules

For each condition and to sense the behavior of each smoke plume, the GL section immediately connects to the database, extracts the data required for training, and sends the data to the inference section. The required training data are spatial layers related to samples of the monthly negative plant growth, which are obtained by analyzing Landsat TM satellite images and drawing monthly NDVI profiles for sample points. Figure 13 presents the results obtained from inference and learning in the designed genetic fuzzy system for

several snapshots, in which high-risk areas are shown as flash points.

In the GL algorithm, two encoded components, including membership functions of fuzzy variables and linguistic modifiers (or fuzzy rules), are considered for defining each chromosome. Each chromosome includes two Ch_1 and Ch_2 parts, as shown in Fig. 14, where I_{t1} , I_{t2} , and I_{tm} are input fuzzy variables and O_t is an output fuzzy variable for the t th mode.

To encode these variables, the following equations are used:

$$\begin{cases} I_{t1} = \{[M_{a-t1-Li1}, M_{b-t1-Li1}, M_{c-t1-Li1}, M_{d-t1-Li1}], \dots, [M_{a-t1-LiN1}, M_{b-t1-LiN1}, M_{c-t1-LiN1}, M_{d-t1-LiN1}]\}, \\ \cdot \\ \cdot \\ \cdot \\ I_{tm} = \{[M_{a-tm-Li1}, M_{b-tm-Li1}, M_{c-tm-Li1}, M_{d-tm-Li1}], \dots, [M_{a-tm-LiNM}, M_{b-tm-LiNM}, M_{c-tm-LiNM}, M_{d-tm-LiNM}]\}, \\ O_t = \{[M_{a-t1-Lo1}, M_{b-t1-Lo1}, M_{c-t1-Lo1}, M_{d-t1-Lo1}], \dots, [M_{a-t1-LoN'}, M_{b-t1-LoN'}, M_{c-t1-LoN'}, M_{d-t1-LoN'}]\}, \end{cases} \quad (6)$$

Figure 15 indicates a fuzzy variable, I_{t1} , with three linguistic terms $[M_{a-t1-Li1}, M_{b-t1-Li1}, M_{c-t1-Li1}, M_{d-t1-Li1}]$, $[M_{a-t1-Li2}, M_{b-t1-Li2}, M_{c-t1-Li2}, M_{d-t1-Li2}]$, and $[M_{a-t1-Li3}, M_{b-t1-Li3}, M_{c-t1-Li3}, M_{d-t1-Li3}]$.

To improve the performance and efficacy of GL and to protect meaningful fuzzy variables, encoding limitations for Fig. 15 can be defined as follows:

$$\begin{aligned} M_{a-t1-Li1} &\leq M_{b-t1-Li1} \leq M_{c-t1-Li1} \leq M_{a-t1-Li2} \\ &\leq M_{b-t1-Li2} \leq M_{c-t1-Li2} \leq M_{a-t1-Li3} \\ &\leq M_{d-t1-Li2} M_{b-t1-Li3} \leq M_{c-t1-Li3} \\ &\leq M_{d-t1-Li3} \end{aligned} \quad (7)$$

To encode the rules, a linguistic modifier is used as a function of δ , which enables GL to modify membership functions related to the concerned linguistic term. For example, two main types of linguistic modifiers include very ($M(x)^\delta$, $\delta=2$) and more-or-less ($M(x)^\delta$, $\delta=0.5$). Therefore, to define Ch_2 , the following relations can be used:

$$\{[\delta_{t11}^1, \delta_{t12}^1, \dots, \delta_{t1m}^1, \delta_{t10}^1], \dots, [\delta_{t11}^K, \delta_{t12}^K, \dots, \delta_{t1m}^K, \delta_{t10}^K]\} \quad (8)$$

where δ_{t1m}^k and δ_{t10}^k are linguistic modifiers of the m th input and output fuzzy variables, respectively, for the k th fuzzy inference rule.

The initial population for forming the gene pool includes four groups, as follows:

- First group: including original Ch_1 s and Ch_2 s (with $\delta=1$ modifier);
- Second group: including original Ch_1 s and random Ch_2 s;
- Third group: including random Ch_1 s and Ch_2 s (with $\delta=1$ modifier); and
- Fourth Group: Including random Ch_1 s and random Ch_2 s.

By encoding and considering limitations, the GL part specifies the initial population and stores it as the current population. In the current population, chromosomes are evaluated by a fitness function. If the concerned evaluation does not satisfy the fitness function, then GL will process the existing generation and use an elitism mechanism. At the beginning of selection, two optimized chromosomes are selected from the current population to enter into the new population without interference by crossover or mutation operators. The remaining chromosomes are selected by the Roulette Wheel Selection Mechanism for entering into the new population. When the number of population elements becomes equal in the

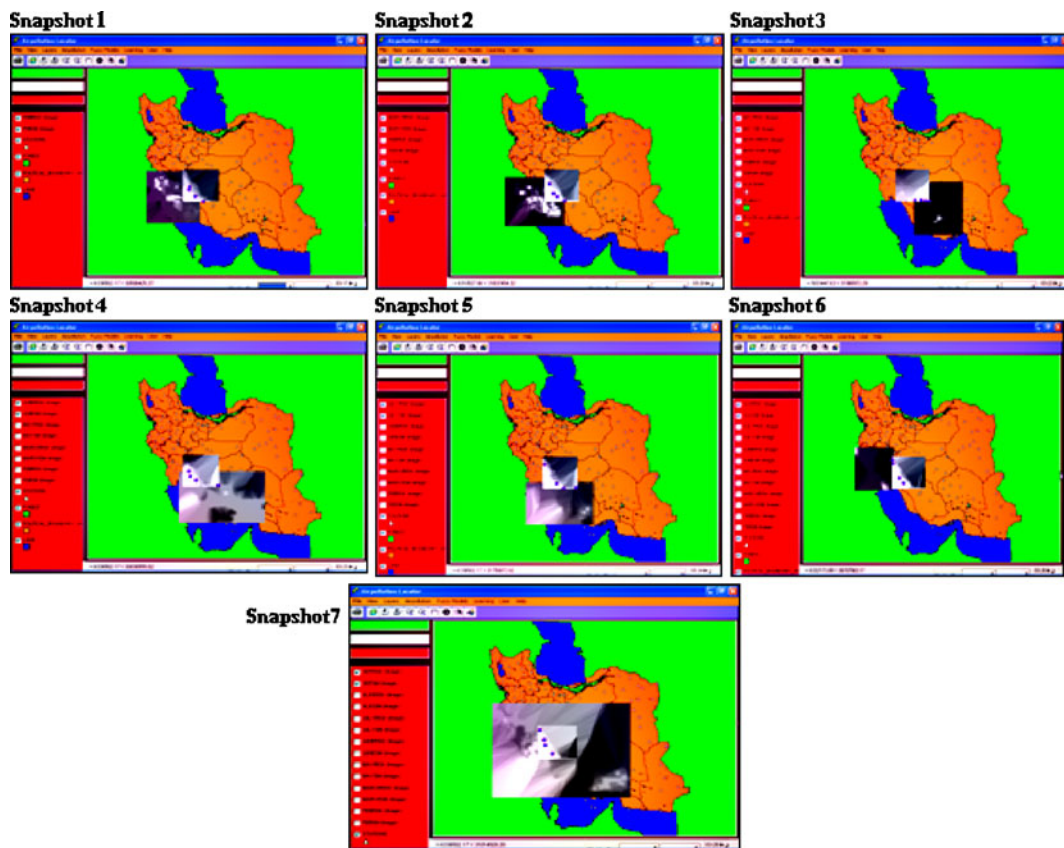


Fig. 13 Inference and learning results for seven snapshots

current and new situations, then the new population is replaced with the current population.

After selection by elitism mechanism, two crossover and mutation operators are implemented. In this mode, a single-point crossover method is fitted and one crossover point is defined randomly. For formation of offspring, some parts of two adjacent chromosomes are exchanged with the right side of the cut-off point. In this mode, a performance rate of 0.7 to 1 is used as the crossover probability. After execution of the crossover operator, the mutation process is carried out separately for each offspring. This operator randomly modifies each gene with a probability of <0.1.

The introduction of crossover and mutation probabilities depends on an objective function. The objective function is a measurement mechanism for

the evaluation of chromosomes, and is defined as follows:

$$\text{MSE} = \frac{1}{2 \times N} \sum_{m=1}^N (O_{tm} - O_{tmd})^2 \quad (9)$$

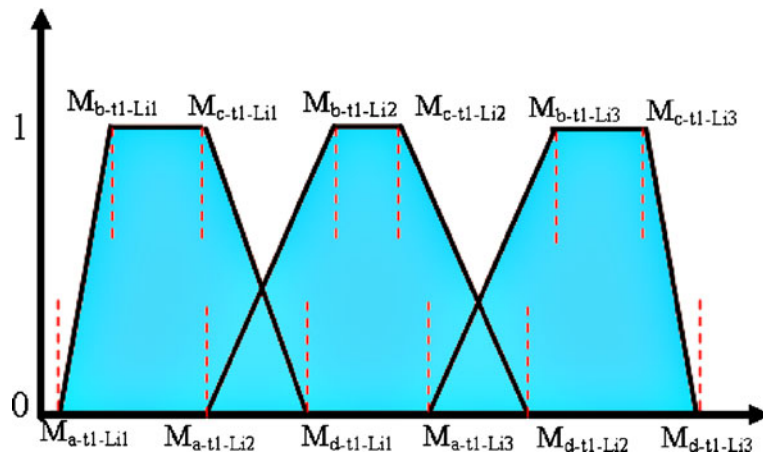
where N is the number of training data for the t th mode, O_t is an output of inference for the m th training data, and O_{tmd} is a desired output. The computational value in each snapshot is introduced as an input fuzzy variable in the next snapshot. Computational results in various periods are dependent on each other, because results obtained from inference are entered into the knowledge base.

Different algorithms, including Cordón et al.'s (2001b), GL (0.7 and 0.03), GL (0.7 and 0.08), GL (0.98 and 0.03), and GL (0.98 and 0.08), were tested.

Fig. 14 Ch_1 chromosome for the t th mode



Fig. 15 I_{tl} with three linguistic terms



The optimized algorithm was determined for use in the learning process of the system. For this purpose, a report section was designed in the user interface for presenting reports of the performances of various genetic algorithms and checking their convergence conditions. Figure 16 shows the convergences of various GAs in the fourth snapshot (June).

In Fig. 16, the MSE diagram of various GAs is drawn for 55 generations with the test data. In most of these algorithms, convergence occurred after 10 to 20 generations. The MSEs obtained in Table 2 indicate that the performances of the GL algorithms were better than that of Cordon, with GL (0.98 and 0.03) having the best performance of all algorithms. The

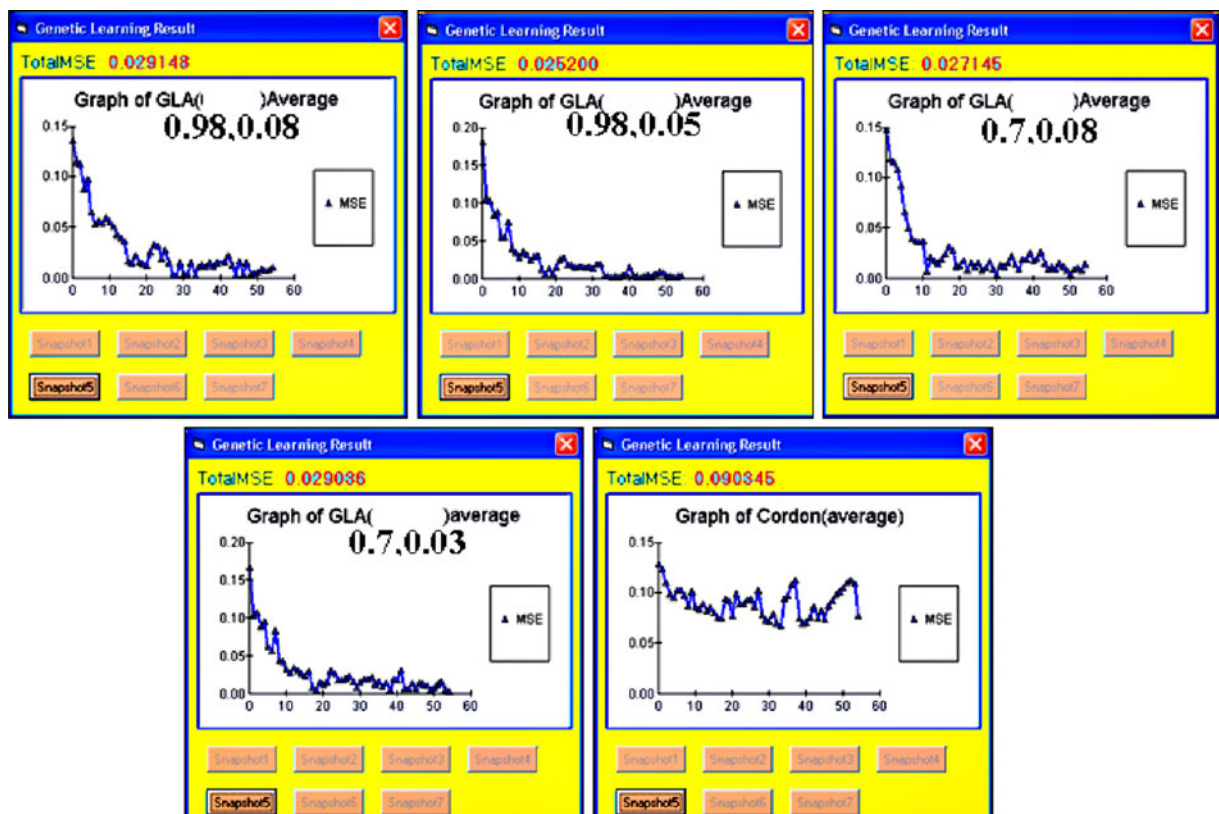


Fig. 16 Genetic algorithm convergence for snapshot 4

Table 2 Final MSE for seven snapshots

Learning algorithm	Total MSE
GL (0.98–0.08)	0.031696
GL (0.98–0.03)	0.028174
GL (0.7–0.08)	0.030425
GL (0.7–0.03)	0.030630
Cordon	0.095162

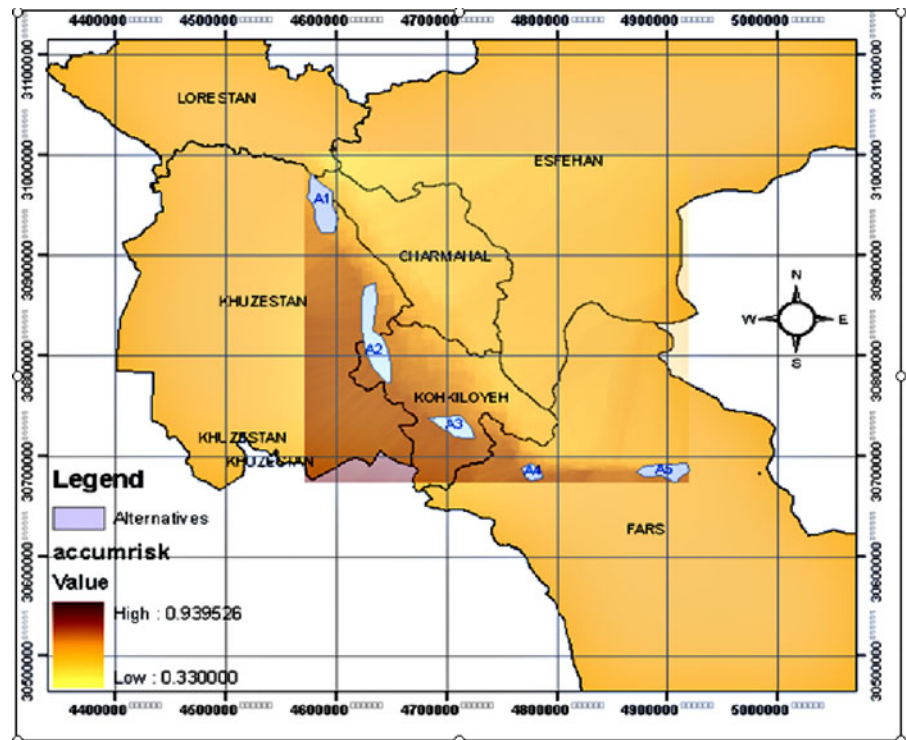
final risk map and high risk areas (A_1 , A_2 , A_3 , A_4 , and A_5) obtained after 1 year are shown in Fig. 17.

Conclusions

In dynamic and complex applications, such as assessing the environmental risks of air pollution plumes, designing, and implementing a robust spatial system that is able to process uncertain data and analyze them logically (like humankind) is vital for environmental protection agencies to control environmental impacts. As shown in this study, the use of a remote sensing database integrated with topological analysis using GIS and a spatial fuzzy reasoning

system provides the spatial tools necessary to make quick decisions about the effects of air pollution. Given the uncertain nature and indeterminate boundaries of input data for reasoning in this application, uncertain modeling capabilities must be combined with expert judgments for defining variables and rules. However, it is difficult for experts to examine all of the data from this complex system to find proper rules. Adopting and generating rules with GL methods will increase the preciseness of the system and will improve robustness.

In this paper, a spatial genetic fuzzy system with GL database architecture was proposed through the evaluation of various GL methods (Cordon, GL (0.7 and 0.03), GL (0.7 and 0.08), GL (0.98 and 0.03), and GL (0.98 and 0.08)] for estimating and determining risk areas. This system successfully determined risk areas in the problem of “the impact of smoke plumes resulting from Kuwaiti oil well fires in 1991 on forest areas in Southwestern Iran,” and RB and DB for decreasing the complication of the searching space. To apply this problem, the data extractor obtained the fuzzy limits of smoke and forest area through NOAA satellite images and sent them to the topological simulator. The topological simulator determined the

Fig. 17 Final risk areas


fuzzy spatial relations and decision variables to send to the knowledge base of the fuzzy inference part. Fuzzy inference, together with GL, updated the existing knowledge in the database and calculated the total risk area. In the GL algorithm, two encoded components—membership functions of fuzzy variables and linguistic modifiers (or fuzzy rules)—were considered for defining each chromosome. For this purpose, the required training data were obtained through the analysis of Landsat TM satellite images and NDVI profiles for some sample points. Different genetic algorithms, such as Cordon, GL (0.7 and 0.03), GL (0.7 and 0.08), GL (0.98 and 0.03), and GL (0.98 and 0.08), were tested based on sample points.

The results from this analysis revealed that the performance of GL (0.98 and 0.03) was more appropriate for generating and tuning rules in the concerned application. The algorithm showed the highest robustness and precision for learning and tuning the 22 rules of the designed genetic fuzzy system. Substantial time was spent on designing the spatial genetic fuzzy system, which revealed interesting information regarding intelligent solutions for predicting and simulating spatial solutions based on topological concepts.

References

- Bojórquez-Tapia, L. A., Juárez, L., & Cruz-Bello, G. (2002). Integrating fuzzy logic, optimization, and GIS for ecological impact assessments. *Environmental Management*, 30(3), 418–433.
- Brink, A. B., & Eva, H. D. (2009). Monitoring 25 years of land cover change dynamics in Africa: a sample based remote sensing approach. *Applied Geography*, 29, 501–512.
- Cano, J. R., Herrera, F., & Lozano, M. (2006). On the combination of evolutionary algorithms and stratified strategies for training set selection in data mining. *Applied Soft Computing*, 6, 323–332.
- Cordón, O., Herrera, F., & Peregrín, A. (1997). Applicability of the fuzzy operators in the design of fuzzy logic controllers. *Fuzzy Sets and Systems*, 86(1), 15–41.
- Cordón, O., Herrera, F., Villar, P., Hoffmann, F., & Magdalena, L. (2000). *Genetic fuzzy systems—evolutionary tuning and learning of fuzzy knowledge bases*. Singapore: World Scientific Co.
- Cordón, O., Herrera, F., Hoffmann, F., & Magdalena, L. (2001). *Genetic fuzzy systems. Evolutionary tuning and learning of fuzzy knowledge bases*. River Edge: World Scientific.
- Cordón, O., Herrera, F., & Villar, P. (2001). Generating the knowledge base of a fuzzy rule-based system by the genetic learning of the data base. *IEEE Transaction of Fuzzy Systems*, 9(4), 667–674.
- Cordón, O., Gomide, F., Herrera, F., Hoffmann, F., & Magdalena, L. (2004). Ten years of genetic fuzzy systems: current framework and new trends. *Fuzzy Sets and Systems*, 41, 5–31.
- Darbra, R. M., Eljarat, E., & Bracelo, D. (2008). How to measure uncertainties in environmental risk assessment. *Trends in Analytical Chemistry*, 27, 377–385.
- Ferson, S. (2003). *RAMAS Risk Calc 4.0 Software: risk assessment with uncertain numbers*. Boca Raton: Lewis Publishers.
- Freitas, A. A. (2002). *Data mining and knowledge discovery with evolutionary algorithms*. Berlin: Springer.
- Homaifar, A., & McCormick, E. (1995). Simultaneous design of membership functions and rule sets for fuzzy controllers using genetic algorithms. *IEEE Transaction of Fuzzy Systems*, 3(9), 129–139.
- Ishibuchi, H., & Yamamoto, T. (2004). Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining. *Fuzzy Sets and Systems*, 141, 59–88.
- Kaiser, M. F. (2009). Environmental changes, remote sensing, and infrastructure development: the case of Egypt's East Port Said harbour. *Applied Geography*, 29, 280–288.
- Karr, C. L., & Gentry, E. J. (1993). Fuzzy control of pH using genetic algorithms. *IEEE Transaction of Fuzzy Systems*, 1, 46–53.
- Kentel, E., & Aral, M. M. (2005). 2D Monte Carlo versus 2D fuzzy Monte Carlo health risk assessment. *Stochastic Environmental Research and Risk Assessment*, 19(1), 86–96.
- Konar, A. (2005). *Computational intelligence: principles, techniques and applications*. Berlin: Springer.
- Kortenkamp, D., Bonasso, R. P., & Murphy, R. (1998). *Artificial intelligence and mobile robots: case studies of successful robot systems*. Cambridge: AAAI Press.
- Lavrac, N., Cestnik, B., Gamberger, D., & Flach, P. (2004). Decision support through subgroup discovery: three case studies and the lessons learned. *Machine Learning*, 57(1–2), 115–143.
- Li, H. L., Huang, G. H., & Zou, Y. (2008). An integrated fuzzy-stochastic modeling approach for assessing health-impact risk from air pollution. *Stochastic Environmental Research and Risk Assessment*, 22(6), 789–803.
- Lin, C. J. (2004). A GA-based neural fuzzy system for temperature control. *Fuzzy Sets and Systems*, 143, 311–333.
- Lin, C. J., & Xu, Y.-J. (2006). A novel genetic reinforcement learning for nonlinear fuzzy control problems. *Neuro-computing*, 69, 2078–2089.
- Makkeasorn, A., Chang, N.-B., & Li, J. (2009). Seasonal change detection of riparian zones with remote sensing images and genetic programming in a semi-arid watershed. *Journal of Environmental Management*, 90, 1069–1080.
- Mikut, R., Jäkel, J., & Gröll, L. (2005). Interpretability issues in data-based learning of fuzzy systems. *Fuzzy Sets and Systems*, 150, 179–197.
- Price, A. R. G. (1993). The Gulf: human impacts and management initiatives. *Marine Pollution Bulletin*, 27, 17–27.

- Sánchez, L., Couso, I. (2005). Advocating the use of imprecisely observed data in genetic fuzzy systems. In *Proc. of the First Int. Workshop on Genetic Fuzzy Systems (GFS 2005), Granada, Spain*. pp. 124–129.
- Sánchez, L., Casillas, J., Cordon, O., & del Jesus, M. J. (2001). Some relationships between fuzzy and random classifiers and models. *International Journal of Approximate Reasoning*, 29, 175–213.
- Setzkorn, C., & Paton, R. C. (2005). On the use of multi-objective evolutionary algorithms for the induction of fuzzy classification rule systems. *Bio Systems*, 81, 101–112.
- Tømmervik, H., Johansen, M. E., Pedersen, J. P., & Guneriusen, T. (1998). Integration of remote sensed and IN-SITU data in an analysis of the air pollution effects on terrestrial ecosystems in the border areas between Norway and Russia. *Environmental Monitoring and Assessment*, 49, 51–85.
- Triolo, L., Binazzi, A., Cagnetti, P., Carconi, P., Correnti, A., De Luca, E., et al. (2008). Air pollution impact assessment on agroecosystem and human health characterisation in the area surrounding the industrial settlement of Milazzo (Italy): a multidisciplinary approach. *Environmental Monitoring and Assessment*, 140, 191–209.
- Wang, H., Kwong, S., Jin, Y., Wei, W., & Man, K. F. (2005). Multiobjective hierarchical genetic algorithm for interpretable fuzzy rule-based knowledge extraction. *Fuzzy Sets and Systems*, 149, 149–186.