Application of a Bayesian Network in a GIS based Decision Making System

A. Stassopoulou, M. Petrou & J. Kittler

Dept. of Electronic and Electrical Engineering
University of Surrey
Guildford GU2 5XH, U.K.

Abstract

In this paper we show how a Pearl Bayes network of inference can be used with a GIS in order to combine information from different sources of data for the purpose of classification. Data may include satellite images, topographic maps, geological maps etc, each one with its own resolution and accuracy. We show how this uncertainty in the input data is incorporated in the network and develop also a method to construct the conditional probability matrices used by the network. We demonstrate our approach within the framework of the problem of assessing the risk of desertification of some burned forests in the Mediterranean region.

1 Introduction

In Geographic Information Systems most of the time the data represented by the various layers are of diverse origins with varied degrees of accuracy. These data usually have to be combined for the inference of some conclusions expressed in the form of labellings. The straightforward combination of information coming from the various layers often used for convenience (e.g. see Grunblatt *et al.* 1992, Chuvieco and Congalton 1989, Sader *et al.* 1995), not only fails to take into consideration the reliability of each source of information, it also ignores the fact that the rules of combining the information themselves may be unreliable. Thus, inspite of the uncertain

rules and the uncertain data, the output of such a GIS is a hard labeling with no indication given as to how reliable this classification is.

This is because most GIS perform reasoning with simplistic inference mechanisms in the form of IF-THEN rules. In rule-based reasoning, one has to know for certain that an assertion is true or false in order to draw a conclusion. Although the truth of certain premises may be suggestive of the truth of a conclusion, it may not imply it conclusively. It is evident that performing inference in any real world domain always requires some simplifications to be made. It is necessary therefore to bear in mind that conclusions drawn even from absolutely correct data may not always be correct.

In this paper we present a Pearl Bayes network for probabilistic reasoning with a GIS for the purpose of assessing the risk of desertification after a forest fire. The reason we chose this particular type of network, is because it allows the modelling of both types of uncertainty in the reasoning process, namely the uncertainty in the reasoning rules (modelled with the help of the conditional probability matrices involved in the network) and the uncertainty in the data (modelled by the probability with which a certain input parameter belongs to a certain class).

A Bayesian network (Pearl 1988, Pearl 1986, Pearl 1993, Charniak 1991, Henrion 1989, Goldman *et al.* 1993) is a powerful tool which can be used for the fusion of information coming from disparate sources with varying degrees of reliability. In these networks each node represents a variable the label of which is either given or has to be assessed and the relationships between the nodes are expressed in terms of conditional probabilities. Each variable can be labelled with a label from a corresponding finite set with a certain degree of confidence. The input data may concern any of the nodes and once they are fed into the right node, a mechanism is provided for the propagation of their information to all other nodes of the system.

So far Bayesian networks have been very widely applied in medical systems to perform medical diagnosis (Olesen *et al.* 1993, Hamilton *et al.* 1994). Medicine is one area where categorical decisions are not sufficient but instead one has to reason under various sources of uncertainty. Other applications involve forecasting (Abramson 1994, Gu *et al.* 1994), computer vision

(Gong and Buxton 1993) etc. Bayesian networks are particularly applicable to cause-effect problems i.e. problems that can easily be represented by cause-effect relationships between the variables of interest, as is the case in most medical problems and in the particular application presented in this paper. Their ability to perform bidirectional reasoning has also made them useful for troubleshooting system failures, including software and hardware problems as well as mechanical failures of cars and jet engines (Heckerman *et al.* 1995). The use of Bayesian networks in Geography, however, has been very limited with a few notable exceptions (e.g. Ducksbury 1993, Hass 1991).

In section 2 we shall present a brief literature survey of the various methods that have been used so far for reasoning with a geographic information. In section 3 we shall present a very brief overview of Bayesian networks so that we can establish our terminology. In section 4 we shall present the particular problem we have to solve with the help of which we shall exemplify the various aspects of using a Bayesian network with a GIS. We shall also describe the Bayesian network constructed for solving this problem. In section 5 we shall discuss how to model the uncertainty in the data assuming a certain type of error in the corresponding measurements and how to incorporate the uncertainty of the data into the network. In section 6 we shall discuss ways of deriving the conditional probability matrices needed by the Bayesian network. In section 7 we present some results of applying the theory developed in the previous sections to a set of real data concerning an area of study in Greece for which the expert's classification is available from ground inspection. We also compare our method with the IF-THEN rule based inference method. We conclude in section 8.

2 Literature Survey

We classify the various methods used for fusion of information to two major categories, namely distributed versus centralised approaches. In the distributed approaches, the data from each sensor are processed separately and the outcomes from all the sources are combined. The data from each sensor are processed anyway whatever approach one uses, but in the case we are talking about, the data from each source are processed fully so that each pixel acquires a label

from the list of labels to be used in the final classification one tries to achieve. Subsequently, each source is given a different weight of significance and the results from each separate classification are combined into a final labeling scheme.

More often is preferred to combine data rather than conclusions drawn from individual sources. The combined data could be raw (as for example in Benediktson and Swain 1990, Benediktson *et al.* 1990,) or partly processed (as for example in Peddle and Franklin 1992, Richards *et al.* 1982). Such systems are centralised systems where there is a central inference mechanism employed to do the classification on the basis of all the data received.

Starting with distributed systems, Solberg *et al.* (1993) used satellite images and GIS ground cover data for land-use classification. The basic fusion model performed a linear combination of the probabilities of a pixel belonging to a particular class given the measurements from different sources of information. This linear function is the result of an assumption that the measurements from each source are independent.

Mason *et al.* (1992) developed the MuSIP (MultiSensor Image Processing) system for the development of data fusion techniques for the exploitation of multisensor and multitemporal images. The approach used was region based and the fusion of information happened at the high level of region classification. As the regions of each individual image were assigned labels from the final list of labels to be used, with varying probabilities, this system is a distributed system. The system was applied to forestry and one of the objectives was to detect changes occurring in forests, particularly those concerned with deforestation and afforestation. The input to the high level data fusion are two PUDs (Picture Understanding Databases) designed for storing 2D vector information, like regions, lines, points etc. For each region, information is stored concerning its attributes (e.g. size and shape), its bounding frame and adjacent regions. The output of the fusion model is another PUD, containing the matching regions and an error PUD containing regions that could not be matched. The matching of regions was based on a number of matching tests, namely attribute matching, label, relational and boundary matching. Each of these tests produced a confidence value supporting a specific region. These values were then combined using the Dempster-Shafer approach(Shafer 1976).

Some centralised systems that combine data instead of conclusions are reviewed next. Grunblatt *et al.* (1992) used a GIS together with simple ecosystem models for desertification assessment and mapping in Kenya. In this study five desertification indicators were selected for investigation: water erosion, wind erosion, vegetation degradation, range utilisation and human settlement. After deriving the status i.e. state or condition of these indicators by assigning values, the desertification hazard was calculated as a summation of the various indicator class values.

A forest fire hazard mapping was developed by Chuvieco and Congalton (1989). In this study digitally processed Thematic Mapper data were integrated with other layers of geographic information to derive a forest fire hazard map. The variables in this model included the basic factors that affect forest fires. For the integration of the five layers of information mentioned above, it was assumed that some of those layers have a higher influence on fire hazard than others and were ranked accordingly. The approach included several steps; first each data layer was weighted according to its impact on increasing the fire hazard. Second, each data layer was then divided into different levels which were assigned coefficients of 0, 1 and 2 based on the ranking of high, medium and low fire hazard respectively. The final formula index was a linear combination of the results of the different layers.

Sader *et al.* (1995) used a GIS rule-based model to identify forest wetlands in Maine. An additive model was developed for this study. The model consisted of four hierarchical layers: each layer incorporated the GIS data and was arbitrarily assigned a weight based on its presumed contribution to identify wetlands. Each pixel was analysed at each layer in the model. The weights assigned to themes and attributes in the GIS model produced a numerical value that indicated the potential for a forest wetland to exist at any location. A lower weight was given to a GIS layer which was thought to be correlated with other variables in the database. As we will see in the description of our model later on, dependencies and independencies among the variables are presented in the structure of the Bayesian network and the propagation of information assures that these dependencies or independencies are not violated.

Binaghi and Rampini (1993) used fuzzy reasoning to combine land-use observables and ancillary observables to deduce the best fire risk judgement for each pixel in the studied area.

Land-use observables were classified into five classes of land cover using Landsat Thematic Mapper image. Ancillary observables included elevation, slope and other factors that affect forest fires. Linguistic labels *low*, *medium* and *high* were then introduced to describe these observables and were quantified with standard membership functions. The result was a degree of membership in the *low*, *medium* and *high* fire risk classes.

Middelkoop and Janssen (1991) developed a method of knowledge-based classification based on temporal relationships between classes. In this study spectral image information, information stored in a geographic information system and knowledge about crop rotation by means of state transition matrices were combined in a Bayesian maximum likelihood classification. The study involved acquisition of knowledge about the temporal relationships between classes from the available data and experts and representation of the knowledge.

Frank (1988) combined Landsat Thematic Mapper data with topographic and topoclimatic variables to map dominant vegetation communities in the Colorado Rocky Mountain Front Range. The thematic spectral bands were transformed into five band ratios and normalised difference variables to characterise the spectral patterns of vegetation community cover types. Slope, elevation, aspect and relief measures were also obtained to examine topographic effects on vegetation distributions. A topoclimatic index was created from a Digital Elevation Model to distinguish between favourable habitats for windblown and snow-covered communities. Slope-Aspect index was used to characterise prevailing wind effects on soil, moisture and subsequent vegetation distributions. All these variables obtained were combined linearly to provide a discriminant score for an observation, for each dominant community. Then, using discriminant scores each observation was assigned to the dominant community using the posterior probability: the probability that an observation with a discriminant score of D belonged to dominant vegetation community group G was estimated by the conditional probability and the observation was assigned to the group which produced the largest conditional probability.

Srinivasan and Richards (1990) classified remotely sensed data into urban, woody vegetation, grass land and soil. The reasoning was based on IF-THEN rules which fired and gave a classification of the reason for a cover type (label) into three categories: prima facie reason, supportive

or criterion, each of which indicated a certain strength of the conclusion given the premises. After all the rules fired, the class labels were given endorsements based on the quantity of reasons for and against a class label. The endorsement was mainly biased by the strong reasons that confirm or deny a label.

Srinivasan and Richards (1990) also developed another system for land-use classification where the central inference mechanism used was based on the Dempster-Shafer theory (Shafer 1976). This evidential system was used to classify the pixels in the image of Sydney Harbor, into three primary features: water, cultural (man-made) and vegetation. Four Landsat Multispectral Scanner (MSS) images have been used from the bands 4, 5, 6 and 7. The system incorporated explicit knowledge in the form of several knowledge bases, each relating to a different source. These contain facts and rules specific to that source. Based on the data available to the source, the appropriate rules fired giving rise to the basic probability assignments (degrees of belief showing to what extend the data support the various class labels) to various nodes in a predefined hierarchy of possible class labels. The control mechanism forward chains through each rule base going from data to possible labels. Dempster-Shafer theory is then used to provide a consistent set of beliefs over all available data sources.

Ducksbury (1993) used a Pearl Bayes network to classify urban regions versus non-urban in aerial photographs. He did not use multisource data, but rather multiprocessor data. In fact all his data came from an aerial image which, however, was processed in various ways so that different measurements concerning the same region were deduced. The different measurements (statistics) were then assessed and combined using a Pearl Bayes network, into a belief of the region to be urban.

3 Bayesian Networks

Belief or Bayesian networks are directed acyclic graphs in which the nodes represent multi-valued variables, comprising a collection of mutually exclusive and exhaustive hypotheses. The arcs signify direct dependencies between the linked variables and the strength of these depend-

encies are quantified by conditional probabilities.

Consider a typical node X having m children $Y_1, Y_2, ..., Y_m$ and n parents $U_1, U_2, ..., U_n$ as shown in figure 1 (The node variables will be denoted by capital letters and the parent and children will also have a subscript to distinguish among them). What follows is a description of the propagation algorithm developed by Pearl (Pearl 1988, Pearl 1986).
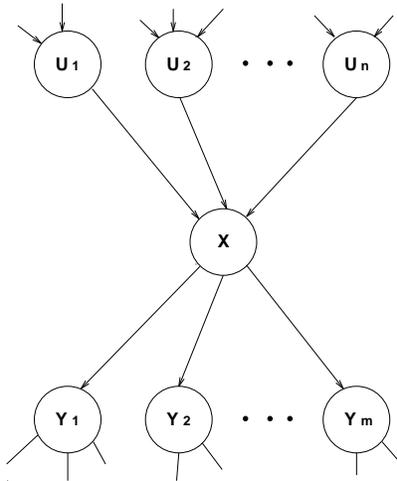

Figure 1: The parents and children of a typical node X in a polytree

The purpose of the Bayesian network is to give a belief in each possible labeling for each node after some evidence arrives. That is, it gives the conditional probability of a node being in each of its states given the evidence observed. In order to estimate the belief of a node we need the information sent by the parents (causal), the information sent by the children (diagnostic) and the conditional probability matrices. The messages that communicate this information obtained by the parents are denoted by π and the messages that carry information from the children are denoted by λ. We denote the belief in each node X by $\text{BEL}(X)$ which is a vector with elements $\text{BEL}(x_i)$ where the small letter indicates a particular value of a state i indicated by the subscript.

Before any propagation commences we initialise the network. Every node is assigned a vector λ, a vector π and a vector BEL. These vectors have as many elements as there are possible states for this node. To initialise the network, we set all elements of all λ vectors to 1. We also set all elements of the π vectors of the root nodes equal to the prior probabilities for the corresponding states. All the values of the remaining variables of the network can be computed now

form the above initialised quantities, and the elements of the conditional probability matrices. An element of the conditional probability matrix looks like $P(x_i|u_{1j_1}, ..., u_{nj_n})$ and gives the probability of state i for node x conditioned on the states of its parent nodes. In addition to the above, each node, except the leaf nodes has a $\pi_{(send)}$ vector that can be communicated to its children. This $\pi_{(send)}$ vector carries as a subscript the name of the child it goes to. For example $\pi_X(u_{ij})$ indicates the jth element of the π vector of the U_i parent sent to its child X. Also, each node except the root nodes carries a $\lambda_{(send)}$ vector which can be communicated to its parents. This $\lambda_{(send)}$ vector carries as a subscript the name of the child it comes from. It has as many elements as the states of the parent node it goes to. For example $\lambda_{Y_j}(x_i)$ means the contribution of the Y_jth child to the ith state of parent X.

Having established our terminology, we can now summarize below the steps of the algorithm:

- **Step 1: Belief updating**

  When node X is activated it inspects both the $\pi_X(u_{ij})$ messages communicated by its parents and the $\lambda_{Y_j}(x)$, $j = 1, ..., m$ messages communicated by each of its children (In case of two subscripts, the first indicates which variable's value and the second indicates the corresponding state i.e. $\pi_X(u_{ij})$ indicates the π value of the jth state of the U_ith parent).

  By using these it updates its belief to

  $$BEL(x_i) = \alpha\lambda(x_i)\pi(x_i) \tag{1}$$

  where

  $$\pi(x_i) = \sum_{j_1}\cdots\sum_{j_n}[P(x_i|u_{1j_1}, ..., u_{nj_n})\prod_m \pi_X(u_{mj_m})] \tag{2}$$

  where each summation ranges over all possible values of each parent,

  $$\lambda(x_i) = \prod_j \lambda_{Y_j}(x_i) \tag{3}$$

and α is a normalising constant rendering $\sum_x BEL(x) = 1$

- **Step 2: Bottom-up propagation**

  Using these messages, X computes a new λ message to be sent to its parents U_i:

  $$\lambda_X(u_{ij}) = \beta \sum_m$$

  $$\left[\lambda(x_m) \sum_{j_1} \cdots \sum_{j_{i-1}} \sum_{j_{i+1}} \cdots \sum_{j_n} [P(x_m | u_{1j_1}, ..., u_{(i-1)j_{i-1}}, u_{ij}, u_{(i+1)j_{i+1}}, ..., u_{nj_n}) \prod_{k \neq i} \pi_X(u_{kj_k})] \right]$$

  (4)

  where $\lambda(x_m)$ is given by 3 and the summation is over all possible values of all parents except parent U_i, for which we compute the λ message for its u_{ij} value.

- **Step 3: Top-down propagation**

  Each node computes the new π messages to be sent to each of its children. The message that X sends to its jth child Y_j is given by:

  $$\pi_{Y_j}(x_i) = \alpha \prod_{k \neq j} \lambda_{Y_k}(x_i) \sum_{j_1} \cdots \sum_{j_n} [P(x_i | u_{1j_1}, ..., u_{nj_n}) \prod_i \pi_X(u_{ij_i})] \qquad (5)$$

  or equivalently

  $$\pi_{Y_j}(x_i) = \alpha BEL(x_i) / \lambda_{Y_j}(x_i)$$

However, in order to apply the formulae and propagation algorithm above the network has to be singly connected i.e. only one path should exist between any two nodes. Otherwise the network is considered to have loops and various methods exist in the literature for handling them (Suermondt and Cooper 1991, Pearl 1988).

4 Problem Formulation

The purpose of this work was to assess the risk of desertification of certain burned forest areas in Greece, by combining data from various sources together with expert knowledge about the desertification processes.

Desertification is land degradation in arid, semi-arid and dry sub-humid areas resulting from various factors, including climatic variations and human activities. The Mediterranean region is characterised by extensive aridity, forest fires, overgrazing and improper land use. This, combined with irregular but intensive precipitation can lead to land degradation.

Two major factors that directly influence the degree of risk of desertification of a burned forest are its regeneration potential and the danger of soil erosion. Attica is an area of arid and semi-arid climate that suffers from repeated forest fires. The dominant forest species is Aleppo pine and a number of bushy species. Aleppo pine regenerates naturally after a forest fire due to the presence of mineral elements in the ash and the lack of competition for nutrients and water from other plant species, provided that the pine trees that were burned had left on the ground enough seeds of good quality. Maquis-type plants regenerate naturally after a forest fire by re-sprouting. The extent and rate of natural regeneration of a burned forest area, under Mediterranean conditions, depend on precipitation, surface geology, surface structure etc. This area of study in Greece is not thought of containing microclimates in it. We are actually interested in the *relative* ranking of burned forests within the same relatively small area of study, for the purpose of prioritising the reforestation resources of the country. Thus, over the whole area one can easily assume constant climatic conditions. Therefore, from all the factors that are known to influence the forest regeneration and soil erosion only those which are expected to vary from one site to the other are considered, namely soil depth, ground slope, rock permeability, aspect and animal grazing.

What follows is a description of the network constructed to tackle this problem.

The constructed network is shown in figure 2 alongside the labels (states) each node can take. Due to the nature of the problem we have designed the network so that the variables are discrete.

The main reason for this is that some of the variables in the network cannot be represented by real values (i.e. rock type) and due to the restrictions involved in having a network with both discrete and continuous variables, we have decided to have only discrete valued nodes. The number of possible classes of each variable is also shown in figure 2.

The arcs between the nodes show causal dependencies and are quantified by conditional probability matrices which reflect the rules the experts use when making similar decisions. The exact values of their elements are chosen using training data. In order to specify a Bayesian network we need to assign prior probabilities to all root nodes and conditional probabilities for all labels of all non-root nodes given all possible combinations of labels of their parents i.e. direct predecessors. When no prior knowledge is available, equal probabilities are assigned to all possible states of the root nodes.
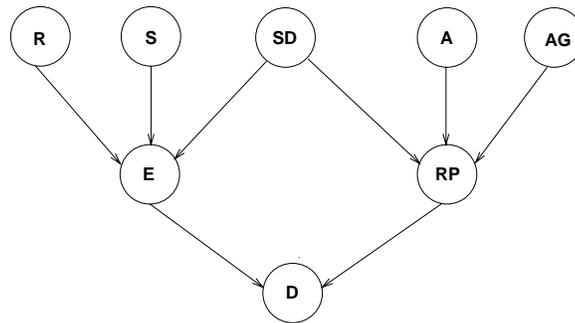
R : Rock Type (permeable, semi-permeable, impermeable)
S : Slope   (gentle, middle, steep)
SD : Soil Depth (bare, shallow, deep)
A : Aspect   (south, west/east, north)
AG : Animal Grazing (slightly, moderately, heavily grazed)
E : Erosion   (low, medium, high)
RP : Regeneration Potential (low, medium, high)
D : Desertification (no/slight, low, medium, high,very high)

Figure 2: The Bayesian network constructed

The network we have is not singly connected, because there exists more than one path between two nodes (SD and D). Before we discuss how we deal with this connectivity, we shall give the basic steps by which the beliefs for each of nodes E, RP and D are updated, ignoring the fact that SD influences both E and RP directly. In particular we may for the moment split node

SD into two, each part influencing one only child node as shown in figure 3.

We now demonstrate the propagation equations of the previous section with an example input to the network of figure 3:

At the initial state, before observing any evidence, all λ's are unit vectors since no variable has any observed descendant and therefore no evidence exists to favour any particular state. The root nodes have a π equal to their prior probability i.e. $(1/3, 1/3, 1/3)$ except $aspect$ which has a prior probability $(1/4, 1/2, 1/4)$ since the second state covers two possible classes (west and east) and therefore does not have equal prior probabilities.

Suppose that R receives evidence indicating $permeable$ rocks. The actual evidence that R receives is represented by a dummy child sending a λ message which is of the form $(1, 0, 0)$ with 1 at the corresponding position indicated by the evidence. This λ message is multiplied element by element with the λ message received from the genuine child node E, which is $(1,1,1,)$. Node R will therefore update its belief using the equation:

$$BEL(R_{perm}) \quad = \quad \alpha\lambda(R_{perm})\pi(R_{perm}) = \alpha \times 1 \times 1/3 = \frac{\alpha}{3}$$

$$BEL(R_{semi}) \quad = \quad \alpha\lambda(R_{semi})\pi(R_{semi}) = \alpha \times 0 \times 1/3 = 0$$

$$BEL(R_{imperm}) \quad = \quad \alpha\lambda(R_{imperm})\pi(R_{imperm}) = \alpha \times 0 \times 1/3 = 0 \tag{6}$$
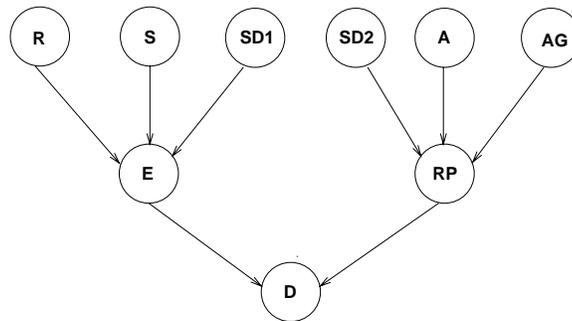
Obviously α must be 3.

Figure 3: Bayesian network without loop

After updating its belief, R sends a π message to its child E which in this case is equal to its belief i.e. (1,0,0).

Upon receiving this π message, E will update its belief using:

$$
\begin{aligned}
BEL(E_i) &= \alpha\lambda(E_i)\pi(E_i) = \alpha\pi(E_i) \\
&= \alpha\pi(E_i)
\end{aligned}
\tag{7}
$$

since $\lambda(E_i) = \lambda_D(E_i) = 1$, for all i, i.e the λ message that D sends to E is a unit vector since D was not instantiated and the $\pi(E_i)$ is given by:

$$
\pi(E_i) = \sum_{k,l,m} [P(E_i|R_k, S_l, SD1_m)\pi_E(R_k)\pi_E(S_l)\pi_E(SD1_m)]
\tag{8}
$$

where the summation ranges over all possible values of each parent R, S and $SD1$. Since we have not yet quantified the matrix $P(E_i|R_k, S_l, SD1_m)$, for all i, k, l, m, we will omit writing down equation (8) explicitly. However, the vectors $\pi_E(R)$, $\pi_E(S)$ and $\pi_E(SD1)$ are equal to $(1,0,0)$, $(1/3, 1/3, 1/3)$ and $(1/3, 1/3, 1/3)$ respectively (R sends its current belief to E whereas S and SD_1 send as π their prior probability). The components of these three vectors are the ones, whose possible product combinations will be used in equation (8). However, the final sum will only include all the product combinations *given* the *permeable* state of rocks, since this is the only non-zero element in the vector $\pi_E(R)$.

Since S and $SD1$ represent data nodes and are therefore clamped, their beliefs are not updated and therefore E does not send any λ message to these parents. However, it will send a π message to its child D given by:

$$
\pi_D(E_i) = \alpha BEL(E_i)/\lambda_D(E_i) = \alpha BEL(E_i)
\tag{9}
$$

i.e. the π message that E will send to D will be its belief.

Upon receiving this information D will update its belief using:

$$
BEL(D_i) = \alpha\lambda(D_i)\pi(D_i) = \alpha\pi(D_i)
\tag{10}
$$

since D did not receive diagnostic information (i.e. λ), the $\lambda(D_i) = 1$, and $\pi(D_i)$ is given by:

$$\pi(D_i) = \sum_{k,m}[P(D_i|E_k, RP_m)\pi_D(E_k)\pi_D(RP_m)] \tag{11}$$

where the summation ranges over all possible values of each parent E and RP. Equation (11) is identical to equation (8) where again we omit the calculation due to the yet undefined matrix involved.

Node D will then send a λ to all its parents except the one it received the message from i.e. E. So the message that it will send to RP is given by:

$$\begin{aligned}\lambda_D(RP_i) &= \beta \sum_m[\lambda(D_m)\sum_k[P(D_m|E_k, RP_i)\pi_D(E_k)] \\ &= \beta \sum_m[\sum_k[P(D_m|E_k, RP_i)\pi_D(E_k)] \\ &= 1\end{aligned} \tag{12}$$

$\lambda(D_m)$ is 1, for all m and the summation is over all possible values of all parents except parent RP, for which we compute the λ message for its ith value. Since D did not receive any diagnostic message and was triggered by a π message, the λ vector send by D to RP is a vector with all its elements 1. Therefore RP will receive a vector which will not alter its belief. This is based on the fact that evidence gathered at a particular node (E in our case) does not influence any of its spouses (RP) until their common child (i.e. node D) gathers diagnostic support. This is sensible since evidence on rock type should not alter our belief in regeneration potential.

In our network there are two paths between SD and D and in order to handle the loop we used the method of conditioning(reasoning by assumptions) which is based on the ability to change the connectivity of a network and render it singly connected by instantiating a selected group of variables (Suermondt and Cooper 1991, Pearl 1988).

5 Coping with Uncertain Data

If there was no uncertainty in the data, the input data concerning a node should be in the form of hard labeling (Pearl 1988, Olesen at al. 1989). For example the input to a node S of our example network could be of the form (1,0,0) implying that S is labelled *gentle* with probability 1 and

both other labels are rejected (they have probability zero). In practice, however, data are never certain, even when they are given to us as hard classifications. We shall examine below the various forms of input data and how uncertainty in them can be handled. Then we shall discuss how the parameters of such process may be assessed in a GIS.

5.1 Data expressed as a continuous-valued variable

Quite often a variable in a problem is a real valued function that can be measured directly. This is the case for example of soil depth or aspect. Such data are stored in raster form in the corresponding GIS layer. In the case of continuous variables the way to incorporate uncertainty is relatively straight forward. In this section, we shall show how, if we have a model of the error distribution in the measuring process, this error can be taken into consideration when we define the uncertainty in the data. For simplicity we shall demonstrate our approach assuming Gaussian error distribution.

Assume that we are given a measurement for one of the input nodes of the network of figure 2. If we assume that the input measurement, μ, has an error which is normally distributed with mean, μ, and variance σ^2, then the probability of the input variable belonging to class i i.e. C_i, given the measurement μ is given by:

$$\mathrm{P}(C_i|\mu) \;=\; \frac{K_i}{K} \tag{13}$$

where

$$K_i \;=\; \frac{1}{\sqrt{2\pi}\sigma} \int_{l_i}^{u_i} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \tag{14}$$

and

$$K \;=\; \frac{1}{\sqrt{2\pi}\sigma} \int_{min}^{max} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \tag{15}$$

The l_i and u_i denote the lower and upper limit of class i respectively and min and max denote the minimum and maximum values that the measured variable can take.

After some manipulation we get:

$$K_i \;=\; \frac{1}{2} \left\{ \mathrm{erf}\left(\frac{u_i - \mu}{\sqrt{2}\sigma} \right) - \mathrm{erf}\left(\frac{l_i - \mu}{\sqrt{2}\sigma} \right) \right\} \tag{16}$$

and

$$K = \frac{1}{2}\left\{\text{erf}\left(\frac{max - \mu}{\sqrt{2}\sigma}\right) - \text{erf}\left(\frac{min - \mu}{\sqrt{2}\sigma}\right)\right\} \tag{17}$$

where $erf(x)$ is the error function defined by:

$$erf(x) \equiv \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt \tag{18}$$

5.2 Data expressed as class labels

The formula derived above is appropriate if we are given the value of the measured variable directly. This, however is not usually the case. Quite often the measured variables are quantised grossly and all we are given is a class label of the object stored in vector or raster format. Even when the input data are of discrete nature, there is always some uncertainty associated with them because often this discretization is achieved by measuring a variable that can take continuous values and grossly discretizing it into a few ranges that define the corresponding classes, without taking into consideration the error of the measurement.

In the absence of any information, all we can assume is that the measured variable could equally likely have been any of the values in the range of values that characterise the particular class. In other words, in terms of the notation introduced above, all we can say is that μ has a uniform probability density function in the range (l_i, u_i). Then we can calculate the probability of the object to belong to any class C_j as follows:

$$P(C_j|l_i < \mu < u_i) = \int_{l_i}^{u_i} P(C_j|\mu)P(\mu)d\mu \tag{19}$$

The above formula can be easily derived as follows:

$$P(C_j|l_i < \mu < u_i) = \frac{P(l_i < \mu < u_i|C_j)P(C_j)}{P(l_i < \mu < u_i)}$$

$$= \frac{P(C_j) \int_{l_i}^{u_i} P(\mu|C_j)d\mu}{\int_{l_i}^{u_i} P(\mu)d\mu}$$

Assuming that μ is uniformly distributed over the range $[l_i, u_i]$, the integral in the denominator is equal to 1 and the equation becomes:

$$P(C_j|l_i < \mu < u_i) = P(C_j) \int_{l_i}^{u_i} P(C_j|\mu)P(\mu)/P(C_j)d\mu$$

$$= \int_{l_i}^{u_i} P(C_j|\mu)P(\mu)d\mu$$

which gives equation 19. Now substituting equation 13 into this equation and assuming a uniformly distributed μ as before we can obtain the probability of the variable to be in class j, i.e. C_j, given that the measuring process indicated class i, i.e. C_i', by:

$$P(C_j|C_i') = \frac{1}{u_i - l_i} \int_{l_i}^{u_i} \frac{K_j}{K}d\mu \qquad (20)$$

Thus, the probability of a variable to belong to class i instead of being taken as equal to 1 (after the evidence), is set to:

$$P(C_i|C_i') = \frac{1}{u_i - l_i} \int_{l_i}^{u_i} \frac{K_i}{K}d\mu \qquad (21)$$

An exact solution to the above integral is difficult to be obtained. Therefore, standard numerical integration techniques have been employed in order to obtain a satisfactory approximate solution.

5.3   Data not obviously related to a continuous-valued variable

Some of the data pertaining to a problem may not be easily associated with a measurable continuous variable. Such data are for example, rock type and human influence and they are stored in vector format. Although not obvious, both these data, in spite of the fact that they are expressed by linguistic descriptors, they can still be quantified. For example, there are models now being developed that attempt to quantify human influence by the number of cattle heads in

the region, or by the distance from the nearest road or town, the population density, etc. If such a model is available, the uncertainty in this variable can be expressed by one of the methods described in section 5.1 and 5.2.

Rock type on the other hand cannot really be mapped on the real axis. However, the *property* of rock type that is relevant to a particular problem may be. For example, if what enters into the reasoning process is not directly the rock type, but the property of rocks in terms of their water permeability, then this is a measurable quantity that can be mapped on the real axis. Rock permeability ranges from 10^{-1} m/sec for pebble beds down to 10^{-12} m/sec for granites. The three classes of permeability are roughly characterised by the following ranges of this value:

- impermeable: 10^{-12}m/s-10^{-8}m/s

- semi-permeable: 10^{-8}m/s-10^{-5}m/s

- permeable: 10^{-5}m/s-10^{-2}m/s

One can envisage that now we have mapped this property of the rocks on the real axis, we can apply the method of section 5.1 to define the uncertainty associated with it. In this particular case however, the range of values is so large that uncertainty in the measurement is probably of secondary relevance. The most significant uncertainty probably stems from the intrinsic variability of the properties of the particular rocks, rather than the measuring process. For example, hard limestones could be characterised by permeabilities anywhere between 10^{-4} and 10^{-10} depending on the amount of solution features they contain. This uncertainty should be the one we model in this case rather than the uncertainty in the measuring process.

5.4   Application to the constructed network

Most of the networks considered deal with either continuous or discrete variables. This is because of the restriction that continuous parents must have continuous children. In the case of our network, we would like to have as continuous variables only the input (data) nodes but not the intermediary ones, since some of these variables cannot take continuous values. Therefore,

we formed a network with only discrete variables. To accommodate for input variables of continuous nature, we can imagine that our network has an extra layer of input nodes each linked to one of the existing root nodes. These new parent nodes will now form the new set of roots in the modified Bayesian network. These parents will represent the input as obtained using the GIS input data with a possible error. So the new set of root variables will consist of nodes R', S', SD', A' and AG' were the $(')$ represents the *given* input.
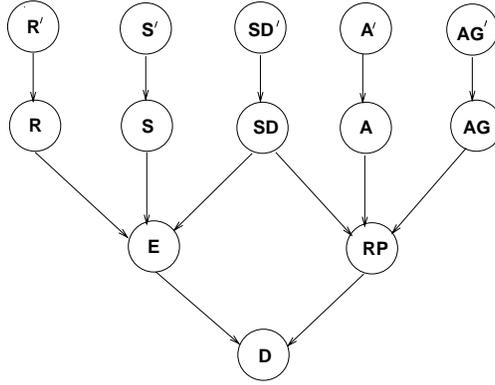

Figure 4: Bayesian network constructed

Each one of these roots will be linked with the corresponding variable which will represent the *actual* variable i.e. the variable after the incorporation of uncertainty or error in the measuring process. These variables consist of the nodes R, S, SD, A and AG. As it can be seen from figure 4 we added 5 extra nodes and 5 more arcs. We therefore need to construct 5 more conditional probability matrices, one for each arc. Each of these matrices will be of dimension 3×3. An example of an entry in the second matrix, which relates the ideal slope with the one given, would be: $P(S = steep|S' = steep)$ which is the probability of slope being $steep$ given that the measuring process indicated $steep$.

Assume for example, that we are given by the source that slope is $steep$. We then input at node S' a vector $(0, 0, 1)$ with 1 at the position indicated by the data. The belief that slope is indeed $steep$, i.e. $BEL(S = steep)$, will be the probability of slope being $steep$, given that the measuring process indicated steep i.e. $P(S = steep|S' = steep)$, i.e. the corresponding entry of the conditional probability matrix relating S to S'. This is the actual value that will get propagated downwards. This value is given by equation 21.

This trick is needed because in a classical Bayesian network with hard classification input, confidence in the input data can be altered by the evidence propagation procedure. In our case we want to avoid that. As long as we realize it, we can input our data directly to the unprimed parent nodes in the form of soft classification which is not to be altered during the reasoning process. So figure 4 is only a conceptual structure, and in practice instead of changing the network structure, we modify the beliefs of the input nodes of the original network, so that the belief will be set according to the error estimation.

Another general issue is the way we choose the uncertainty level. For the case of continuous valued variables we tried to refer everything to one number. For example, we choose the σ of the variable coming from the most reliable source in such a way that the confidences in each class were almost hard. Then we chose the standard deviations of the error of the less reliable variables to be multiples of the previous value appropriately scaled to take into account the fact that the variables are measured by different units in different scales. These multiplication factors were chosen according to linguistic expressions of opinion by experts, in the form "I say that soil depth is twice as unreliable as aspect" and also by tuning the network to agree with the experts' opinion in the assessment of some training sites.

For each region we have the DEM data with 20m$\times$20m resolution. From these we calculate the aspect and slope. Aspect is expressed as the angle between the normal to the ridge and the north direction. This is expressed in positive degrees form 0 to 360 measured clockwise from the north. Aspect is therefore represented by a circle starting clockwise from the north and completing 360 degrees, as:

- North: $0°$-$45°$

- West: $45°$-$135°$

- South: $135°$-$225°$

- East: $225°$-$315°$

- North: $315°$-$360°$

Due to the nature of these boundaries (non continuous), we created for our network, the following classes together with the limits which apply only to half the circle as follows:

- South: 180°-225°

- West/East: 225°-315°

- North: 315°-360°

Any measurements, μ, that fall below 180 degrees, are taken as 360-μ.

Slope is calculated by the tangent of the given slope angle times 100 i.e. $tan\theta \times 100$ where θ is the slope angle calculated from the DEM data. The three classes of slope are:

- Gentle: 0%-20%

- Middle: 20%-40%

- Steep: 40%-∞

The GIS data regarding both variables are available in the form of measurements. For each measurement, μ, of each of the variables we have associated with it a standard deviation σ. Since the two variables (namely *slope* and *aspect*) were obtained from the most accurate source, we assign a σ which will indicate high confidence in the data. Therefore we set σ=2 for slope and set a σ for aspect which is equivalent to that of slope, appropriately scaled to correspond to the range of values of aspect. Therefore σ of aspect is estimated to be 9. Since the measurement is given, we apply equation 13. This will give the probability of each class of each variable given the measurement, i.e. $P(C_i|\mu)$ which is effectively the belief in that class, i.e. $BEL(C_i)$. This is shown diagrammatically by figures 5a to 5c.

Figure 5a represents the belief in the variable, *slope* say, before the uncertainty i.e. $BEL(S')$. It is represented by the belief vector $(0, 1, 0)$ with 1 at the position being indicated by the measurement. The τ_1 and τ_2 are the bounds of the class indicated. Now, we assume that this input measurement has an error which is normally distributed with mean μ=measurement and

variance σ^2 as shown in figure 5b. This is the Gaussian which we integrate in equations 14 and 15 to obtain equation 13. Equation 13 will give the belief in each class of the *actual* variable, i.e. $BEL(S)$, which is shown diagrammatically by figure 5c.

(a)

(b)

(c)

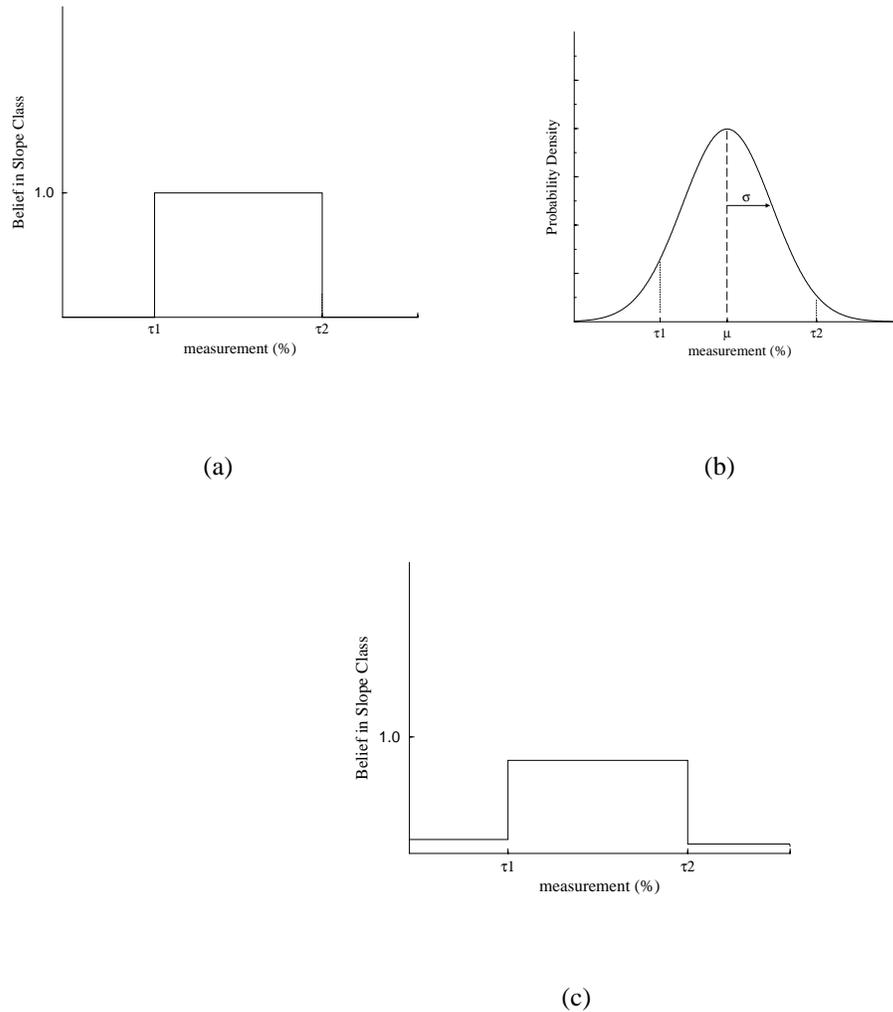Figure 5: Belief distribution of slope assuming (a) no uncertainty and (c) uncertainty according to the probability density function (b)

For the other two variables soil depth (SD) and rock type (R), the GIS data indicated the particular class to which each variable belonged.

Soil depth, measured in centimeters (cm), is divided into the following three classes according to experts:

- Bare: 0cm-5cm

- Shallow: 5cm-30cm

- Deep: 30cm-∞

The GIS data indicate the class that each pixel belongs to and therefore we apply equation 20 to calculate the belief in SD after the incorporation of uncertainty. We choose a $\sigma = 2.5$ for this variable which is twice the sigma taken for slope and aspect, scaled accordingly. This is due to the fact that soil depth is less accurate than the slope and aspect which are obtained from DEM data. The fact that we set a sigma twice as that of slope and aspect is justified a posteriori by the training data (i.e. the sites whose desertification risk is known) which are used to tune the parameters of the network.

The possible rock types for our problem were:

- Hard Limestones

- Mica Schists

- Metamorphic

- Calcareous tertiary deposits

- Siliceous tertiary deposits

- Colluvium

Out of these types, mica schists and metamorphic rocks, are definitely considered as impermeable, whereas hard limestones, colluvium and tertiary deposits are on the boarder between permeable and semi-permeable. In the first case we set the belief of rock permeability to (0,0,1) with 1 at the $impermeable$ state and in the second case we set the belief of rock permeability to $(0.50, 0.50, 0)$ with equal probability in the given rock being permeable and semi-permeable.

6   Coping with Uncertainty in the Rules

The propagation of information in the network relies on the conditional probabilities concerning the nodes that are related to each other by explicit causal relations. In this paper we also present a novel methodology for constructing the conditional probability matrices for use by the Bayesian network to perform the inference. A conditional probability matrix such as $P(A|B, C)$ can be represented by a table with one entry for each possible combination of the states of the variables A, B and C. However, when the number of parents is large or/and when the number of possible states of the participating variables is large, then direct elicitation of such a table is difficult because the required data grows exponentially with the number of variables involved. There have been nevertheless proposed some approximation techniques of this conditional probability $P(A|B, C)$ from pairwise relations such as $P(A|B)$ and $P(A|C)$ (Kim and Pearl 1987, Kim and Pearl 1983). The most common model of this type is disjunctive interaction (the "noisy OR-gate") (Pearl 1988). The basic theory behind this is that any member of a set of conditions is likely to cause a certain event and this likelihood does not diminish when several of these conditions exist simultaneously. However disjunctive interaction is based on some assumptions which are only true in some cases. Other methods of estimating the conditional probability matrices are ad hoc based on expert rules (Henrion 1989, Haas 1991).

For the network of the size presented here these conditional probabilities are expressed by a $3 \times 3 \times 3 \times 3$ matrix, the elements of which have to be specified. For example, one element of the conditional probability matrix relating $erosion$ to its parent variables i.e. $rock\ permeability$, $slope$ and $soil\ depth$ would give the conditional probability of erosion to be in a particular state given the states of each of the three influencing variables. There is no mention in the literature how the elements of this matrix can be specified in practice.

There are some geophysical phenomena, however, for which either an analytic model exists in the form of an equation that has been deduced from careful consideration of the underlying physical processes, or an equation has been defined in an empirical way to express quantitatively observations and expertise built up over the years. We shall demonstrate here how the elements

of the conditional probability matrix can be defined when such an equation is available. We shall also discuss what can be done when no such equation is available.

6.1   When an analytic equation relating causes to effects is available

We will demonstrate here a way to specify the independent elements of the conditional probability matrix of $erosion$ which will take into account not only the relationship between a child node and its parents but also the importance that each of the parent variables bares on the child. This is possible because an analytic formula exists that expresses quantitatively the relationship between erosion and the factors that affect it. The formula used here is the one proposed by Stehlik (Morgan 1986) and is very similar to the Universal Soil Loss Equation (USLE) and is as follows:

$$\text{mean annual soil loss} = D \cdot R \cdot P \cdot S \cdot L \cdot O \tag{22}$$

where D is the climatic factor which in our case is assumed to be constant over the whole study area, R is the petrological factor and assesses the rock type according to the permeability of its weathered debris and P expresses the erodibility of the soil. P is related to soil depth which is inversely related to erosion. Deep soils, for example, due to their larger water storage capacity, are less sensitive to erosion than shallow soils. The way we incorporate soil depth into this erodibility factor is to assume that for a constant type of soil in the study area, soil depth and soil erodibility are related using a linear function. We can assume this linear function since we are only interested in a specific, small range of soil depth in which a more or less linear relationship between the two may be assumed. (As long as the range of the independent variable is small, any non linear function can be approximated by a linear one.). In this way we derive a function that will give us P given the depth of the soil. S and L are the slope steepness and length factor respectively. The length factor may be assumed constant over all the sites. S is given by:

$$S = 0.24 + 0.106s + 0.0028s^2$$

where s is the slope in per cent.

O in equation 22 is the vegetation cover which is dependent upon the percentage cover. In our case O is constant since the area of interest is burned and therefore there is no vegetation cover. Thus for our problem this formula is simplified as follows:

$$\text{mean annual soil loss} = f(S, R, P) = \alpha \cdot R \cdot S \cdot P$$

where α is a constant and P represents erodibility of the soil which is a function of soil depth. In order to estimate the constant α we will need to estimate D, L and O from equation 22. We know that D is expressed in terms of the mean annual precipitation, r, using the equation:

$$D = 0.0014r - 0.38$$

In our study area we know r to be constant to 437mm and therefore we estimate $D = 0.23$. O is constant equal to 4 since this corresponds to 0 percentage cover and L is constant equal to 1 which corresponds to 20m length of slope. We therefore derive $\alpha = 0.92$.

Each one of these variables f, R, S and P has three possible states or classes corresponding to the three classes of each one of the variables E, R, S and SD in the network. Assume that we denote the lower and upper boundaries of a certain class of a node by subscripts 1 and 2.

Now assume that we want to estimate the conditional probability of erosion belonging to a certain class given the classes of the variables that it depends on. This conditional probability clearly can be estimated by evaluating the following formula:

$$P(e|r, s, sd) = \frac{\int_{R_1}^{R_2} \int_{S_1}^{S_2} \int_{P_1}^{P_2} u(f(S, R, P) - E_1) * u(E_2 - f(S, R, P)) \, dS \, dR \, dP}{(R_2 - R_1) * (S_2 - S_1) * (P_2 - P_1)} \quad (23)$$

where small letters for the variables mean their actual assigned class labels and u is a heaviside function which is equal to 1 when its argument is positive and 0 otherwise.

The above formula seems pretty straight forward and in principle one can calculate the integrals analytically and derive an expression which for given values of the class boundaries will yield the value of the conditional probability. However, in practice the analytic calculation of the triple integral is prohibitively complicated for the general case due to the great number of

possible cases one has to consider. Thus, we estimated the above probabilities by the Monte Carlo method: For each possible combination of classes of the arguments of the conditional probability we drew 40000 samples uniformly distributed over the volume $[R_1, R_2]$, $[S_1, S_2]$ and $[P_1, P_2]$. The number of samples that make the integrand in the numerator in equation 23 non-zero divided by the total number of samples drawn is an estimate of the conditional probability. Table 1 gives the entries of this matrix calculated as described above. Each entry gives the conditional probability of erosion being in one of the states low, $medium$ and $high$ (L, M, H respectively) given the states of the three variables as shown in the first three columns.

We follow the same procedure for all possible combinations of states of E and its parents in order to complete the whole conditional probability matrix.

6.2 When no analytic formula is available between causes and effects

When no analytic formula exists between causes and effects, one has to deduce the elements of the conditional probability matrix using training data and expert rules expressed in vague linguistic terms. This is the case for assessing the regeneration potential of a forest and its risk of desertification.

No quantitative formula exists that relates the regeneration potential to its contributing factors. One can easily reason and derive one such formula but to our knowledge there are no detailed studies done which would allow us to calculate, for example, the constants that will appear in it. This situation is similar regarding the dependence of the factor of desertification on the regeneration potential and erosion. Thus, for the calculation of the elements of these matrices appropriate numbers had to be chosen to reflect the rules followed by the experts, namely:

1. The more soil, the more seeds and more nutrition and therefore the higher the potential of a forest to regenerate naturally.

2. The way the land surface area is oriented towards the sun affects the amount of radiation it receives. South facing aspects due to the higher amounts of heat they receive from the sun, tend to be drier. On such slopes natural vegetation is usually sparser compared to

vegetation on north-facing aspects.

3. The more animals that graze the land, the lower the potential to regenerate.

4. The higher the erosion, the higher the desertification risk.

5. The lower the potential of an area to regenerate, the higher the risk of desertification. We are also given that this factor influences desertification more than erosion.

The above general rules were used to give some initial values to the elements of the conditional probability matrix, interpreting that $high$ means 80% confidence, low means 20% confidence and so on. Then these values were tuned using training sites for which the experts' (hard) classification was available. It must be noted that ideally this tuning should be done using *historical* data where the state of a number of sites is compared with the state predicted by he experts some years earlier and statistics are performed. However, in view of the lack of such data one can only tune the parameters in such a way that the systems' assessment agrees with that of the expert in as many sites as possible.

7   Experimental Results

In this section we will present the results of the network when applied on 53 sites using GIS data. We have implemented Pearl's propagation algorithm in C. We then compare our method with the results obtained using the GIS data with reasoning in the form of IF-THEN rules i.e. ignoring the uncertainty. In each case we compare the results with the expert's classification which was based on field data.

7.1   Reasoning using the Bayesian network

The inference mechanism described above could be applied either at pixel or, preferably, at region level. The latter would obviously make the algorithm faster. At whatever level it is applied, it is necessary for the entity that has to be classified to have uniform attributes. For example, our method could not be applied to a region consisting partly from south looking slopes and partly from west looking slopes. In that case fuzzy logic would have been more appropriate, as fuzzy logic deals with partial membership to various classes. Thus, to apply our inference mechanism to the region level, some preprocessing is needed: As each layer of

			Erosion		
Rock Type	Slope	Soil Depth	L	M	H
permeable	gentle	bare	0.28	0.53	0.19
		shallow	0.39	0.57	0.04
		deep	0.58	0.42	0.00
	middle	bare	0.00	0.00	1.00
		shallow	0.00	0.07	0.93
		deep	0.00	0.34	0.66
	steep	bare	0.00	0.00	1.00
		shallow	0.00	0.00	1.00
		deep	0.00	0.00	1.00
semi-permeable	gentle gentle	bare	0.21	0.43	0.36
		shallow	0.29	0.54	0.17
		deep	0.45	0.55	0.00
	middle	bare	0.00	0.00	1.00
		shallow	0.00	0.01	0.99
		deep	0.00	0.14	0.86
	steep	bare	0.00	0.00	1.00
		shallow	0.00	0.00	1.00
		deep	0.00	0.00	1.00
impermeable	gentle	bare	0.15	0.33	0.52
		shallow	0.21	0.42	0.37
		deep	0.33	0.59	0.08
	middle	bare	0.00	0.00	1.00
		shallow	0.00	0.00	1.00
		deep	0.00	0.02	0.98
	steep	bare	0.00	0.00	1.00
		shallow	0.00	0.00	1.00
		deep	0.00	0.00	1.00

Table 1: Conditional probability matrix for $erosion$.

the GIS has different tessellation of the land into polygons, a composite tessellation must first be created by considering a union of all polygons in all layers. This way each small polygon created will have a unique set of attribute classes or measurements which can be fed into our system to assess its risk of desertification. In our case, part of the data, i.e. aspect and slope, were given in raster format. Thus, the smallest "polygon" we had was a single pixel and that is why we had to apply our system at a pixel level.

The final classification of a site was performed by averaging the corresponding probabilities of the site pixels.

There are 5 possible desertification classes:

- Class 1 : Sites without any risk of desertification after a forest fire

- Class 2 : Sites with a low risk of desertification

- Class 3 : Sites with a moderate risk of desertification

- Class 4 : Sites with a high risk of desertification

- Class 5 : Sites with a very high risk of desertification

All sites have been classified by an expert using ground data.

Out of the 53 available sites, 39 sites were used for training and 14 sites were used for testing the final system. Table 2 shows the results of the 39 sites using GIS data with uncertainty. Table 3 shows the results of the 14 test sites using GIS data with uncertainty. Using GIS data the system agreed with the expert on 28 out of the 39 training sites, with the majority of the misclassified sites falling into the adjacent classes. Out of the 14 test sites, 8 agreed with the expert. The sites that were correctly classified are indicated by a "$\sqrt{}$" in the last column of the tables. We have also tested the system on the 53 sites using GIS data with no uncertainty. In this case, for slope and aspcet we calculated the mean slope and aspect of each site and classified it by inputting 100% probability to the class it fell in. For the rest of the variables we gave again 100% probability to the class indicated. When tested with these certain input, the system agreed with the expert on 28 out of the 53 sites.

7.2 Rule-based reasoning

We compare our results with those obtained when reasoning without uncertainty, but simply performing inference using IF-THEN rules.

An example of such a rule is:

IF   slope is steep

AND   rock-type is permeable

AND   soil-depth is shallow

THEN   risk-of-erosion is high

So for each pixel in each site we test to see if the premises of a rule are satisfied in order to draw a conclusion. We then classify the site according to the classification of the majority of the pixels.

For the raster data available for slope and aspect we simply classify them according to which class range the measurement falls. The rest of the data were already in classes. Out of the total of 53 sites 18 (13/39+5/14) agreed with the expert's classification. Results on the 39 sites are shown again in table 2. The classification of the 14 sites is shown in table 3. In both tables agreement with the experts is indicated with a "•" in the last column.

The lower accuracy observed when reasoning with IF-THEN rules lies in the fact that hard labellings were used thus ignoring the degree of uncertainty in the GIS data. This should not be the case since the degree of belief in the input is significant for determining the output. Although the *classification* of the input may suggest the class to which the output belongs, when we have degrees of beliefs attached to these input variables we may end up with a different output classification. This is why probabilistic reasoning with a Bayesian network gives better results. By including uncertainty only in the inference (i.e. using Bayesian network with certain data) still we have better results than by ignoring any sort of uncertainty and applying rule-based inference. However the best results are obtained by incorporating uncertainty in the data as

well as in the inference.

Another notable advantage of the Bayesian network over the traditional IF-THEN inference is that we can include in the structure of the network variables which latter on can leave uninstantiated in case of lack of information. For example, in the network created for this particular application we have included the variable *animal grazing* which although affects the potential of an area to regenerate, there were no data available for input from the GIS. So when performing the propagation one can leave this variable uninstantiated, as in our case. If we later on receive some information regarding this variable we can simply input it in the network. In contrast, in IF-THEN reasoning any variables on which we do not have information are omitted from the rules. Therefore any future available information on this variable will require altering the rule base and constructing new rules to account for the new variable.

8   Discussion and Conclusions

In this paper we have shown how a Pearl Bayes network was constructed for the purpose of fusing information in a GIS systems. The information includes data of different resolution and accuracy. It was therefore necessary for the system to be able to reason with uncertainty.

We have presented techniques to handle data input as raw measured values or as class memberships. We also discussed ways of dealing with data in vector or raster form.

The designed system was tested on 53 sites using the GIS data assuming uncertainty with promising results. The majority of the misclassified sites were not given a completely wrong classification but rather a classification into a neighbouring class. This is expected due to the discrepancy between field and GIS data. Results also show that reasoning with uncertainty using a Pearl Bayes network performed significantly better than rule-based reasoning of the IF-THEN form. Results of the Baysian network with certain data also performs better than rule-based reasoning since it still applies some uncertainty in the inference. However, results showed that the incorporation of uncertainty in the data is also essential in order to get satisfactory results. This proves the fact that the confidence with which a label is assigned is significant in deriving

the correct conclusion. The incorporation of uncertainty also provides the additional information on how probable the alternative labels are.

To summarise, we believe that the major contributions of this work are:

- It showed how artificial intelligence techniques can successfully be applied to geographical issues. More specifically, how a Bayesian network can be used as a tool for performing probabilistic reasoning in a GIS.

- It presented a technique for introducing uncertainty not only in the knowledge provided but in the input data as well, which in most cases are obtained from various sources with varied degrees of reliability.

- It presented a novel methodology for obtaining the parameters to be used by the system in case a formula exists relating the connected variables. We believe that this is particularly useful since the quantification of the Bayesian network, still remains a largely unexplored area.

Acknowledgements

References

ABRAMSON, B., "The design of belief network-based systems for price forecasting," *Comp. Electr. Eng.*, **20**, 1994, pp. 163–180.

BENEDIKTSSON, J. A. AND SWAIN, P. H., "A method of statistical multisource classification with a mechanism to weight the influence of the data sources," in *International Geoscience and Remote Sensing Symposium(IGARSS'89)*, Vancouver, Canada, 1989, pp. 517–520.

BENEDIKTSSON, J. A., SWAIN, P. H. AND ERSOY, O. K., "Neural network approaches versus statistical methods in classification of multisource remote sensing data," *IEEE Transactions on Geoscience and Remote Sensing*, **28**(4), July 1990, pp. 540–551.

BINAGHI, E. AND RAMPINI, A., "Fuzzy decision-making in classification of multisource remote sensing data," *Optical Engineering*, June 1993.

CHARNIAK, E., "Bayesian networks without tears," *AI Magazine*, **12**, 1991, pp. 50–63.

CHUVIECO, E. AND CONGALTON, R. G., "Application of remote sensing and geographic information systems to forest fire hazard mapping," *Remote Sensing of the Environment*, **29**, June 1989, pp. 147–159.

DUCKSBURY, P. G., "Parallel texture region segmentation using a Pearl Bayes Network," in *Proceedings of the 4th British Machine Vision Conference*, Surrey, UK, September 1993, pp. 187–195.

FRANK, T. D., "Mapping dominant vegetation communities in the Colorado Rocky Mountain Front Range with Landsat Thematic Mapper and digital terrain data," *Photogrametric Engineering and Remote Sensing*, **54**(12), December 1988, pp. 1727–1734.

GOLDMAN, R. P. AND CHERNIAK, E., "A language construction of Belief Networks," *Pattern Analysis and Machine Intelligence*, **15**(3), March 1993, pp. 196–208.

GONG, S. AND BUXTON, H., "Bayesian nets for mapping contextual knowledge to computational constraints in motion segmentation and tracking," in *British Machine Vision Conference*, Guildford, Surrey, 1993.

GRUNBLATT, J., OTTICHILO, W. K. AND SINANGE, R. K., "A GIS approach to desertification assessment and mapping," *Journal of Arid Environments*, **23**, 1992, pp. 81–102.

GU, Y., PEIRIS, D., CRAWFORD, J., NCNICOL, J., MARSHALL, B. AND JEFFERIES, R., "An application of belief networks to future crop production," in *Proceedings of the 10th Conference on Artificial Intelligence for Applications*, San Antonio, Texas, 1994, pp. 305–309.

HAAS, T. C., "A bayesian belief network advisory system for aspen regeneration," *Forest Science*, **37**(2), 1991, pp. 627–654.

HAMILTON, P. W., ANDERSON, N., BARTELS, P. H. AND THOMPSON, D., "Expert system support using bayesian belief networks in the diagnosis of fine needle aspiration biopsy specimens of the breast," *Journal of Clinical Pathology*, **47**, 1994, pp. 329–336.

HECKERMAN, D., BREESE, J. S. AND ROMMELSE, K., "Decision- Theoretic Troubleshooting," *Communications of the ACM*, **38**(3), 1995, pp. 49–56.

HENRION, M., "Some practical issues in constructing Belief Networks," in Kanal, L. N., Levitt, T. S. and Lemmer, J. F. (eds.), *Uncertainty in Artificial Intelligence (3)*, Elsevier Science Publisher, 1989, pp. 161–173.

KIM, J. AND PEARL, J., "A computational model for causal and diagnostic reasoning in inference systems," in *Proceedings of the 8th International Joint Conference on Artificial Intelligence*, 1983, pp. 190–193.

KIM, J. H. AND PEARL, J., "Convince: A conversational inference consolidation engine," *IEEE Transactions on Systems, Man and Cybernetics*, **17**(2), April 1987, pp. 120–132.

MASON, D. C., HINDLEY, N., JOHNSON, D. G., JONES-PARRY, I. H., ODDY, C. J., PIKE, T. K., PLASSARD, T., RYE, A. J., DE SALABERT, A., SAWYER, G., SERPICO, B. AND WIELOGORSKI, A., "The MuSIP multisensor image processing system," *ITC Journal*, **1**, 1992, pp. 73–86.

MIDDELKOOP, H. AND JANSSEN, L. L. F., "Implementation of temporal relationships in knowledge based classification of satellite images," *Photogrametric Engineering and Remote Sensing*, **57**(7), July 1991, pp. 937–945.

MORGAN, R. P. C., *Soil Erosion and Conservation*, Edited by D.A. Davidson Harlow:Longman Scientific and Technical, 1986.

OLESEN, K. G., KJAERULFF, U., JENSEN, F., JENSEN, F. V., FALCK, B., ANDREASSEN, S. AND ANDERSEN, S. K., "A munin network for the median nerve-a case study on loops," *Applied Artificial Intelligence*, **3**, 1989, pp. 385–403.

PEARL, J., "Fusion, propagation, and structuring in Belief Networks," *Artificial Intelligence*, **29**, 1986, pp. 241–288.

PEARL, J., *Probabilistic reasoning in intelligent systems:Networks of plausible inference*, Morgan Kaufmann Publishers Inc., 1988.

PEARL, J., "Belief networks revisited," *Artificial Intelligence*, **59**, 1993, pp. 49–56.

PEDDLE, D. R. AND FRANKLIN, S. E., "Multisource evidential classification of surface cover and frozen ground," *International Journal of Remote Sensing*, **13**(17), February 1992, pp. 3375–3380.

RICHARDS, J. A., LANDGREBE, D. A. AND SWAIN, P. H., "A means for utilizing ancillary information in multispectral classification," *Remote Sensing Environment*, **12**, 1982, pp. 463–477.

SADER, S. A., AHL, D. AND LIOU, W., "Accuracy of Landsat-TM and GIS Rule-Based methods for forest wetland classification in Maine," *Remote Sensing of the Environment*, **53**, March 1995, pp. 133–144.

SCHISTAD SOLBERG, A. H., JAIN, A. K. AND TAXT, T., "Fusion of multitemporal satellite images and GIS data for land-use classification," in *8th Scandinavian Conference on Image Analysis(SCIA)*, Tromso, Norway, May 1993, pp. 691–697.

SHAFER, G. A., *Mathematical Theory of Evidence*, Princeton University Press, Princeton, N.J., 1976.

SRINIVASAN, A. AND RICHARDS, J. A., "Knowledge-based techniques for multi-source classification," *International Journal of Remote Sensing*, **11**(3), April 1989, pp. 505–525.

SUERMONDT, H. J. AND COOPER, G. F., "Initialization for the method of conditioning in bayesian belief networks," *Artificial Intelligence*, **50**(1), 1991, pp. 83–94.

Site	Rule-based Output D class	Bayesian Network Output BEL(D)	Expert Output D class		
lavrio1	3	(0.02, 0.45, 0.40, 0.07, 0.06)	3	●	
lavrio2	3	(0.02, 0.39, 0.40, 0.09, 0.10)	3	●	√
lavrio3	3	(0.01, 0.38, 0.43, 0.07, 0.11)	3	●	√
lavrio4	5	(0.02, 0.45, 0.39, 0.07, 0.07)	2		√
lavrio5	2	(0.03, 0.54, 0.36, 0.05, 0.02)	2	●	√
lavrio6	3	(0.02, 0.47, 0.42, 0.05, 0.04)	2		√
pateras1	1	(0.01, 0.33, 0.43, 0.09, 0.14)	3		√
pateras2	1	(0.01, 0.34, 0.42, 0.09, 0.14)	3		√
pateras3	1	(0.00, 0.25, 0.40, 0.13, 0.22)	3		√
pateras4	1	(0.06, 0.60, 0.28, 0.05, 0.01)	3		
pateras5	1	(0.00, 0.34, 0.47, 0.07, 0.12)	2		
pateras6	1	(0.02, 0.43, 0.43, 0.10, 0.02)	3		√
pateras7	2	(0.03, 0.46, 0.38, 0.07, 0.06)	3		
pateras8	1	(0.05, 0.55, 0.32, 0.05, 0.03)	3		
pateras9	5	(0.01, 0.31, 0.38, 0.13, 0.17)	3		√
pateras10	2	(0.02, 0.49, 0.43, 0.03, 0.03)	2	●	√
pateras11	2	(0.03, 0.44, 0.36, 0.11, 0.06)	2	●	√
pateras12	2	(0.03, 0.48, 0.35, 0.11, 0.03)	2	●	√
pateras13	1	(0.02, 0.44, 0.44, 0.10, 0.00)	2		√
pateras14	1	(0.00, 0.27, 0.41, 0.12, 0.20)	3		√
pateras15	1	(0.03, 0.47, 0.37, 0.07, 0.06)	3		
pateras16	1	(0.03, 0.48, 0.37, 0.06, 0.06)	3		
pendeli1-1	3	(0.02, 0.42, 0.42, 0.06, 0.08)	2		√
pendeli1-2	5	(0.01, 0.35, 0.43, 0.07, 0.14)	3		√
pendeli1-3	3	(0.01, 0.34, 0.45, 0.07, 0.13)	3	●	√
pendeli1-4	3	(0.02, 0.40, 0.43, 0.06, 0.09)	3	●	√
pendeli1-5	3	(0.01, 0.34, 0.40, 0.10, 0.15)	2		
pendeli1-6	3	(0.02, 0.39, 0.39, 0.09, 0.11)	3	●	√
pendeli1-7	3	(0.01, 0.39, 0.43, 0.07, 0.10)	3	●	√
pendeli2-1	5	(0.01, 0.38, 0.44, 0.07, 0.10)	3		√
pendeli2-2	5	(0.00, 0.21, 0.34, 0.12, 0.33)	3		√
pendeli2-3	5	(0.01, 0.34, 0.44, 0.07, 0.14)	3		√
pendeli2-4	5	(0.00, 0.31, 0.45, 0.07, 0.17)	3		√
barnavas1	3	(0.01, 0.32, 0.42, 0.09, 0.16)	4		
barnavas2	3	(0.01, 0.39, 0.49, 0.04, 0.07)	3	●	√
barnavas3	3	(0.02, 0.41, 0.43, 0.06, 0.08)	3	●	√
barnavas4	5	(0.01, 0.33, 0.49, 0.05, 0.12)	3		√
barnavas5	3	(0.01, 0.39, 0.45, 0.06, 0.09)	3	●	√
barnavas6	3	(0.01, 0.39, 0.43, 0.07, 0.10)	2		

Table 2: Results of the risk of desertification for the 39 training sites using Rule-based reasoning and Bayesian network inference with GIS data

	Rule-based Output	Bayesian Network Output	Expert Output		
Site	D class	BEL(D)	D class		
Tlavrio1	3	(0.02, 0.43, 0.45, 0.04, 0.06)	3	•	√
Tlavrio2	3	(0.01, 0.41, 0.43, 0.06, 0.09)	3	•	√
Tpateras1	5	(0.02, 0.39, 0.43, 0.06, 0.10)	2		
Tpateras2	5	(0.01, 0.43, 0.42, 0.06, 0.08)	2		√
Tpateras3	3	(0.01, 0.40, 0.47, 0.04, 0.08)	3	•	√
Tpateras4	1	(0.01, 0.33, 0.47, 0.06, 0.13)	2		
Tpateras5	1	(0.03, 0.44, 0.37, 0.08, 0.08)	3		
Tpateras6	1	(0.04, 0.50, 0.35, 0.06, 0.05)	3		
Tpendeli1-1	3	(0.02, 0.43, 0.44, 0.05, 0.06)	3	•	√
Tpendeli1-2	3	(0.00, 0.37, 0.52, 0.03, 0.08)	4		
Tpendeli1-3	3	(0.00, 0.36, 0.50, 0.04, 0.10)	4		
Tpendeli2-1	5	(0.01, 0.33, 0.44, 0.07, 0.15)	3		√
Tbarnavas1	3	(0.01, 0.41, 0.46, 0.05, 0.07)	3	•	√
Tbarnavas2	3	(0.01, 0.46, 0.43, 0.04, 0.06)	2		√

Table 3: Results of the risk of desertification for the 14 test sites using Rule-based reasoning and Bayesian network inference with GIS data