ERROR EFFECT SIMULATION IN FIRE HAZARD MODELLING

De Vliegher B.M.* Brackman P.* and Basigos P.S.**

*Department of Geography, University Gent - Krijgslaan 281 (S8-A1)
B-9000 Gent, Belgium. tel. +32.9.264 47 09, fax. +32.9.264 49 85

** **Dieuthisi Dason Messinias, Ydras 5 - GR-24100 Kalamata, Greece.
tel. +30.721.20 820, fax. +30.721.93 723

ABSTRACT

The combination of remote sensing techniques and GIS is a very useful tool for hazard modelling. Nevertheless, it is widely known that geographical information is subject to error and uncertainty. Therefore, the success and the usefulness of such hazard models depend upon the correctness of the input data. These errors are due to the quality of the original analogue data, the digitising of the data, the field measurements, and the digital treatment of the data (interpolation, modelling).

This paper discusses the results of a study with the aim to investigate the effect of random errors in input data on hazard models. This is done by means of a theoretical approach, based on the Monte Carlo simulation. The methodology consists of the generation of known errors in a "supposed" error-free data set. The size and the statistical distribution of the errors are controlled individually for each thematic map and are changed during different simulations.

A first set of results is obtained for the error sensitivity of different parameters. It is concluded that this sensitivity was remarkable different according to the parameter and in relation to the degree of error.
A second set of results is obtained about the holistic influence of error in the computed hazard model for wild fires . The difference between each of the resulting hazard maps and the error-free hazard map is made. The spatial distribution of the anomalies allows one to note the areas where the hazard model is doubtful, based only on error in input data.

INTRODUCTION

The power of geographic information systems (GIS's) for capturing, integrating, manipulating and modelling map data and the explosive growth of the applications during the last decades give rise to the question about the reliability of the results. It is widely known that GIS is subject to error. However, errors and uncertainties in map data and in GIS are often undervalued or even ignored as it concerns environmental modelling.

"What could be cuter
Than to feed a computer
With wrong information
But naive expectation
To obtain with precision
A Napoleonic decision?"
(Eco, 1991)

Yet, the analysis of spatial data with unknown accuracy will result in output products with low confidence limits and with restricted use in the decision making process (Lunetta *et al.*, 1991). Also, there are no standard procedures in commercial GIS for error handling (Brunsdon & Openshaw, 1993).

Errors and uncertainties in GIS's can have many different causes. Three main groups of factors governing errors that are commonly associated with geographic information processing have been identified by Burrough (1986):

1. Obvious sources of errors, including age of data, areal coverage, map scale and density of observations.
2. Errors associated with the natural variability or from original measurements, including positional accuracy (e.g. errors due to digitising), qualitative and quantitative accuracy, and sources of variations in data (e.g. measurement errors, observer bias).
3. Errors arising through processing, including numerical errors in the computer because of the limitations of computer representation of

numbers, faults associated with topological analysis, and classification and generalisation problems. This group also includes errors in interpolation.

Another factor having an important contribution to the success of the hazard models is the resolution of the raster data set. This problem of scale is not new as it concerns the use of cartographic documents. The scale or resolution of a map influences the type of data which can be extracted from the map (analogue or digital). According to the spatial resolution of the digital elevation model strong errors may occur in derived information, such as aspect and slope. This topic of error effect is, however, not discussed in this paper.

The study of error propagation can be split into five tasks (Openshaw *et al.*, 1991):

1. Modelling the distribution of errors to show the uncertainty characteristics of digital map databases.
2. Developing procedures to investigate what will be the effect of errors (in input data) on GIS procedures. One solution is a Monte Carlo-simulation.
3. Models and techniques must be applied to some case studies to derive empirical estimations of error in output.
4. Development of techniques to utilise output data uncertainty estimates.
5. Incorporation of the technology as standard GIS procedures.

AIM OF THE STUDY

The purpose of the study was to answer the following questions:

1. What will be the effect of errors occurring in a DEM on the calculation of other variables (for example slope degree)?

2. What will be the variance of the error in for example slope orientation, if the DEM has a root mean square error (RMSE) of 1 m?

3. Will there be a difference in the amount of error in for example slope orientation, if the RMSE of the DEM is increased from 2 m to 3 m, as from 1 m to 2 m?

4. What will be the distribution of errors of the derived data, if initially a random distribution is considered?

5. Can it be concluded that the reliability of the derived data is spatially determined?

These questions need to be answered and are of primary interest when additional information is derived from the input data throughout the GIS process.

In general, the accuracy of a calculated variable is checked by means of random sampling, comparing the values of the calculated variable with the values obtained through a more accurate method of measuring. This procedure, however, cannot be applied if one does not have access to a more accurate set of data.

To try to answer the above questions and to have an idea about the variance of error in the variable with increasing degree of error in the initial data, an error propagation was made, based upon the principle of a Monte Carlo simulation (Eastman *et al.*, 1993; Fisher, 1991a).

The presented study also aimed to determine the holistic influence of error in a computed hazard model for fire spreading. The resulting hazard maps and the spatial distribution of errors in the fire spreading hazard allowed the reliability of the model to be determined.

METHODOLOGY

The Monte Carlo simulation

A Monte Carlo simulation is based upon a randomiser algorithm for generating errors in the original data set. It is supposed to have an error-free (theoretical) data layer (e.g. DEM) where the values of the grid cells represent the true values (in the following, indicated by MAP_0).

In the frame of this research, data layers (MAP_X) are produced showing anomalies with the MAP_0. It is supposed that these anomalies are randomly distributed. This means that spatial autocorrelation is not taken into account. Moreover, errors are distributed normally around a mean $\mu=0$ with a standard deviation s corresponding to the degree of error.

The procedure for error simulation is given in figure 1.

The uncertainties are included in the original input data set by combining the MAP_0 with the error maps (ERR_X) to create the other MAP_X. MAP_0 and MAP_X (e.g. DEM_0 and DEM_X) are then used as basic documents to derive other information: e.g. slope (SLOPE_0 and SLOPE_X).

If no systematic error occurs, the difference between MAP_X and MAP_0 is normally distributed. In the case of this study, this distribution is user defined, with $\mu=0$ and s. The size of s is a measure for the "quality" of the input map. The larger s is, the less precise are the data of the input map. Figure 1 represents the histogram of the difference map DEM_X-DEM_0 with a normal distribution around $\mu=0$ and with s = 5 m.

In the next step, an algorithm is applied to the input data (DEM_0 and DEM_X). The difference between the results (SLOPE_X-SLOPE_0) is also calculated and the distribution is an indicator of the errors that might be expected when the initial information layer is characterised by an inaccuracy of s (e.g. 5 m). The distribution of the difference between the derived variables does not have to be normal.

Errors are brought in the digital elevation model and the distance maps.

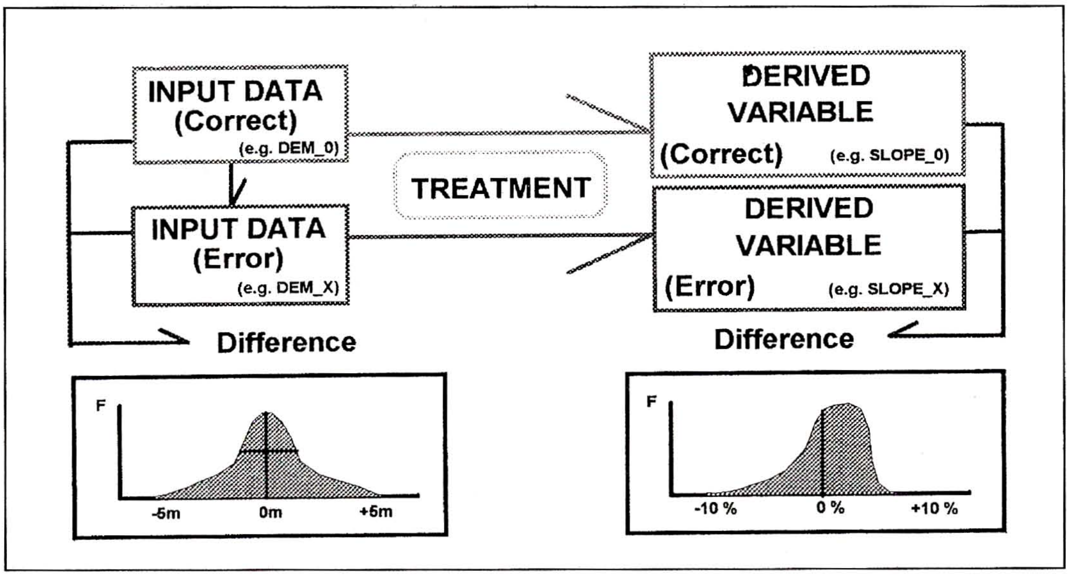

Figure 1 - Procedure followed for error simulation

Errors in the DEM

Seven different error maps (ERR_X) are created with $\mu=0$ and a standard deviation (s) of 1, 2, 3, 4, 5, 10 and 20 m. The distribution of these errors for the different s-values is given in figure 2. These ERR_X are used to produce the "error"DEM's (e.g. DEM_1 to DEM_20).

Errors in the distance maps

Distances to roads, coast and waste dumps are calculated using the standard procedure in the ILWIS-software (Integrated Land and Watershed Information System, Enschede, The Netherlands) (ILWIS, 1993). Initially, each non-target pixels gets a maximum distance to the source pixels. For each pixel, the distance to its neighbouring pixels is calculated using a 3*3 filter with values 7,5,7 - 5,0,5 - 7,5,7 (7/5 is a good approximation of Ã2, that is the distance between two diagonally connected pixels when the raster cell size is equal to 1). Using this 7/5 approximation, a maximum error of 5 % is obtained for straight lines of 0, 45 and 90° (5 % too small) and for lines of 22.5 and 67.5° (5 % too large). Distances for all other lines are calculated correctly (ILWIS, 1993).

To investigate the influence of errors in distance maps, a population of random errors is created that are distributed normally around a mean μ with a standard deviation s, corresponding to the degree of error. Five different error maps (ERR_X) are created with $\mu=0$ and a standard deviation (s) of 1, 2, 3, 4, and 5%. The ERR_X are combined with the different distance maps to create ROAD_X, COAST_X and DUMP_X.

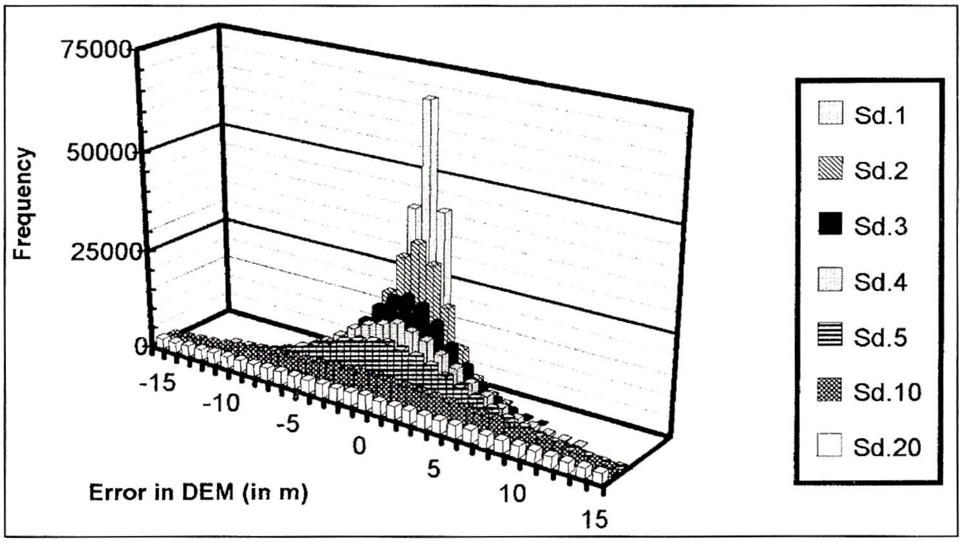

Figure 2 - Histogram of errors generated in the input data

ERROR SENSITIVITY OF THE DERIVED DATA

Slope gradient

The effect of error in slope gradient is illustrated in figure 3. It shows that the average calculated slope gradient will increase with increasing degree of random distributed error. In other words: slopes become steeper with increasing random error in a DEM.

Moreover, error in slope gradient will become larger with increasing error in the DEM. This relationship is quasi-linear. This means that if the accuracy of the height determination is higher (thus with lower s), the accuracy of the derived data (slope gradient) is also higher (s-value is lower).

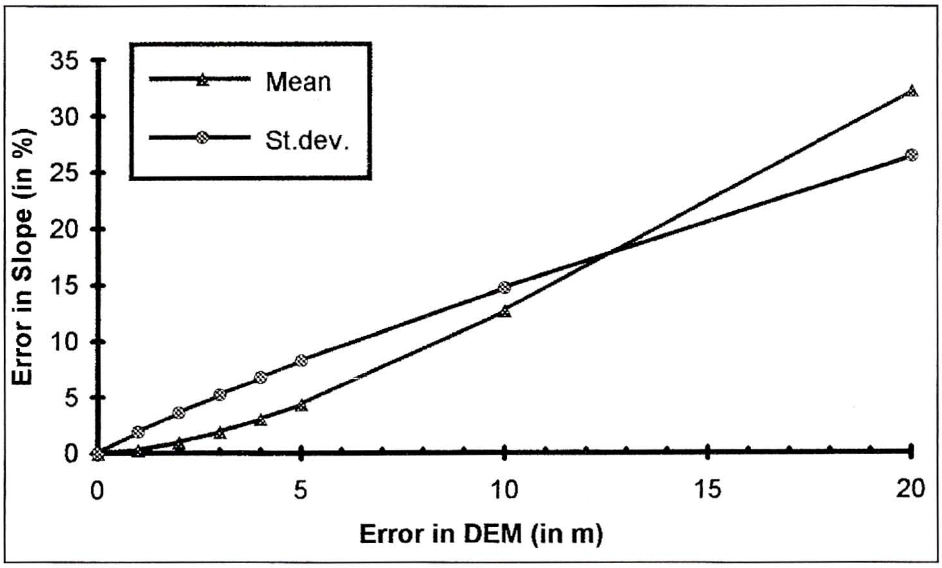

Figure 3 - Influence of errors generated in the DEM on the calculation of slope gradient

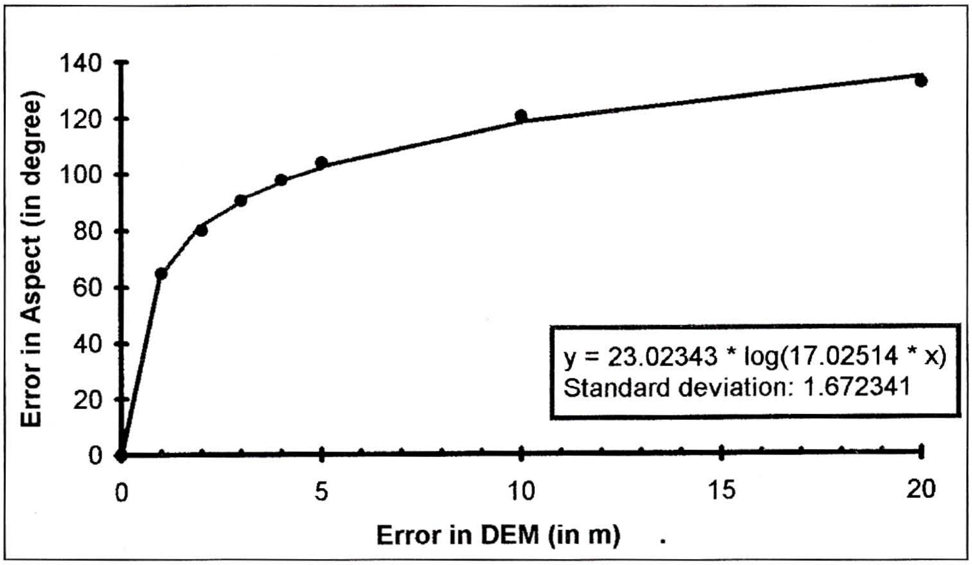

Figure 4 - Influence of errors generated in the DEM on the calculation of aspect

The effect of error in aspect is illustrated in figure 4. It shows that the errors of aspect (expressed as s-value in degree on Y-axis) increase with increasing error in the DEM (expressed as s-value in m on X-axis). A relative small error in the DEM (s=1) result in important errors for the aspect. The increase of error becomes less with increasing s-value of errors in the DEM.

The relationship between the errors in the aspect map and the errors in the DEM follow a logarithmic curve. Because the difference in aspect between the theoretical "error-free" aspect and the aspect calculated using the "error" DEM cannot exceed 180°, the error in aspect (thus the s-value on the Y-axis) is limited to a maximum.

EFFECT OF ERRORS IN THE FIRE HAZARD MODEL

The model

In previous studies (De Vliegher & Basigos, 1994), research with respect to fire hazard modelling has been made, using remote sensing techniques and geographical information systems. These studies were undertaken in the district (Eparchy) of Pylias (SW-Peloponnese, Greece). Statistical data for the period January 1978 to August 1993 include records of 118 wild fires in the district of Pylias covering an area of 4454 ha of forested land and 7015 ha in total, including agricultural land.

The fire hazard was investigated based upon the relationship between the fire history and the physical and human characteristics of the study region. The influencing factors were chosen with respect to the results of multi-temporal field work. These are: 1) altitude, 2) slope degree, 3) aspect (slope orientation), 4) vegetation, 5) accessibility of the area, 6) distance to waste dumps and 7) influence of tourism.

Two important steps can be distinguished in the hazard modelling: 1) determination of the thematic hazard and 2) calculation of the hazard for fire spreading.

Determination of the thematic hazard

The procedure to determine the relationship between the environmental problem (i.e., the occurrence of wild fires) and the different parameters (influencing factors) comprises four steps, as shown in figure 5.

1. Dividing the influencing factor into thematic classes (e.g., slope 0-2 %). The frequency (%) of each thematic class is defined. The fire location map (period 1978-1992) is combined with the thematic map and the area affected by wild fires within each thematic class is determined and expressed in %.
2. The obtained results are represented by column-line histograms, where the X-axis refers to the thematic class and the Y-axis represents the proportional frequency of the thematic class (lines refer to fire frequency, columns refer to the frequency of the thematic class).
3. Calculation of ratio "Fire frequency per thematic class (%)/Frequency thematic class (%)"
4. Definition of hazard function for each of the influencing factors.

Using this function, thematic maps are converted into thematic hazard maps.

a.

Slope Class (%)	Slope Frequency (%) =A	Fire Frequency (%) =B
0..2	6.5	3.6
3..6	16.9	14.7
7..12	23.1	19.9
13..18	19.3	17.8
19..25	13.1	13.9
26..35	10.6	13.2
36..50	6.6	9.8
>50	3.8	7.0

c.

Slope Class (%)	Ratio B/A
0..2	0.6
3..6	0.9
7..12	0.9
13..18	0.9
19..25	1.1
26..35	1.2
36..50	1.5
>50	1.8

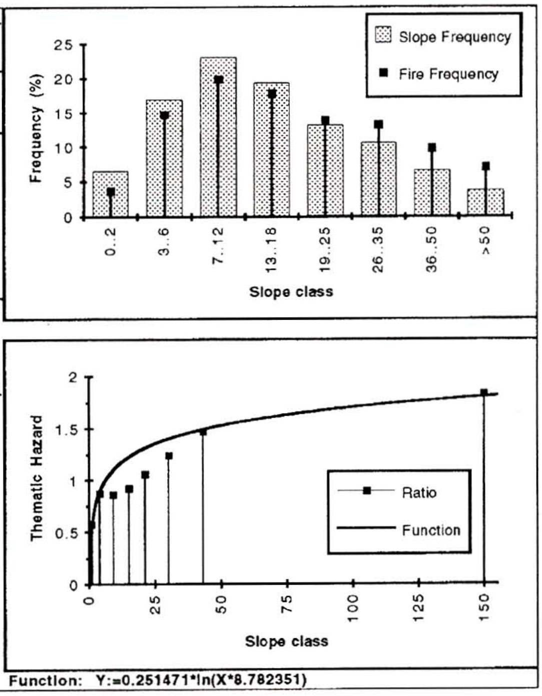

Figure 5 - Procedure of thematic hazard determination

Calculation of hazard for fire spreading

The hazard for fire spreading is obtained by multiplying the different thematic hazard map values :

$$\text{Hazard} := V * A * S * O * R * D * C$$

with V = vegetation hazard; A = altitude hazard
S = slope hazard; O = aspect (orientation) hazard
R = accessibility hazard; C = distance to coast hazard
D = waste dump hazard

Four hazard classes are distinguished:

Hazard class	Hazard value
No to slight	1
Moderate	2
Severe	3-10
Very severe	> 10

HOLISTIC INFLUENCE OF ERRORS IN THE HAZARD MODEL

In order to compute the holistic influence of errors, the obtained thematic maps with various degrees of error are used for the fire hazard modelling, according to the procedure described above. Seven different hazard maps are produced, named HAZ_1 to HAZ_20. These maps are made as a combination of the respective thematic "error"-maps (Table 1).

Table 1 - Degree of error in the input data to create the "error" hazard maps

Degree of error in input data	Hazard		Thematic hazard map for						
			DEM (A)	Slope (S)	Aspect (O)	NDVI (N)	Dump (D)	Road (R)	Coast (C)
s= 1	Haz_1	*	A_1	S_1	O_1	N	D_1	R_1	C_1
s= 2	Haz_2	*	A_2	S_2	O_2	N	D_2	R_2	C_2
s= 3	Haz_3	*	A_3	S_3	O_3	N	D_3	R_3	C_3
s= 4	Haz_4	*	A_4	S_4	O_4	N	D_4	R_4	C_4
s= 5	Haz_5	*	A_5	S_5	O_5	N	D_5	R_5	C_5
s= 10 (5)	Haz_10	*	A_10	S_10	O_10	N	D_5	R_5	C_5
s= 20 (5)	Haz_20	*	A_20	S_20	O_20	N	D_5	R_5	C_5

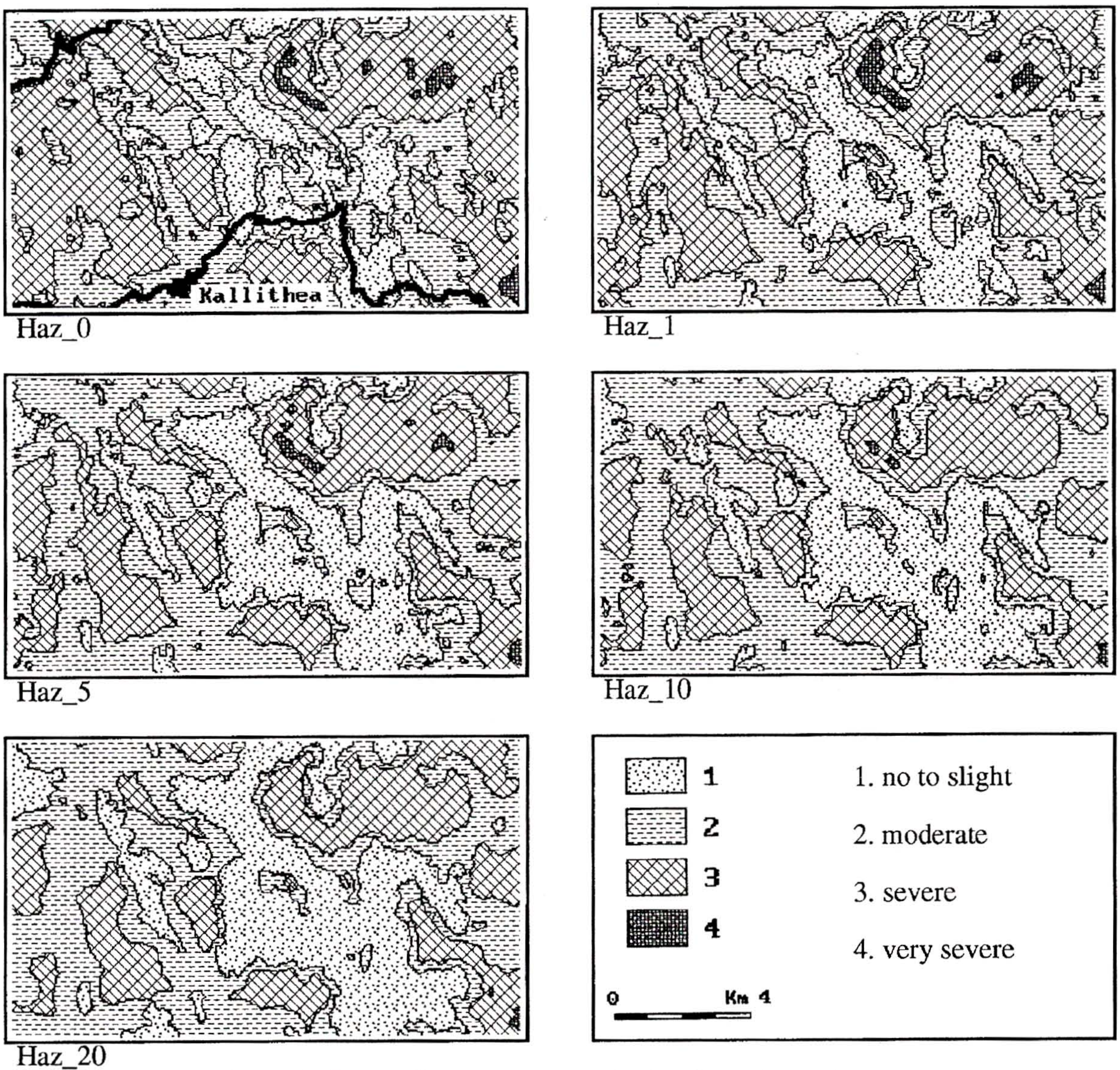

Figure 6 - Extract of hazard maps created with different degree of error

Figure 6 includes extracts of hazard maps created with different degrees of error in the input data. This figure shows clearly that the fire hazard map becomes more uniform with increasing degree of error in the input data. Classes with high hazard disappear in favour of the lower classes. Compared to the hazard map HAZ_0, nearly no zones of very severe hazard remain on the HAZ_20. Also, the very fragmented area of "No to slight" hazard in the centre of HAZ_0 has grown to a large and uniform area on the HAZ_20.

The increase in areas belonging to the "no to slight" hazard class is shown in figure 7. It is coupled to a decline of the high hazard classes "severe" and "very severe". It can thus be concluded that the degree of hazard will be less pronounced with increasing random error in the initial data set.

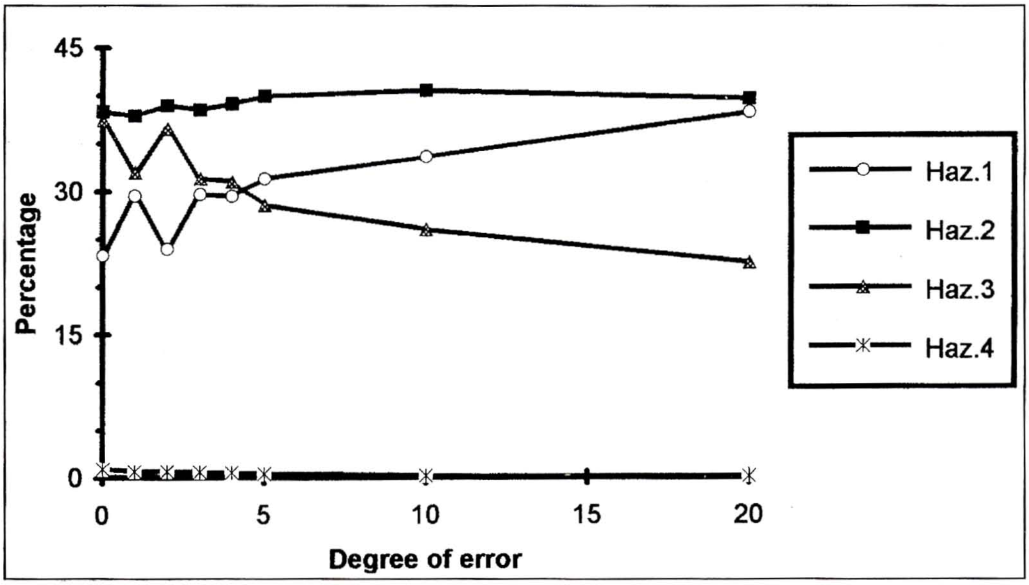

Figure 7 - Proportion of hazard classes (%) as a function of error

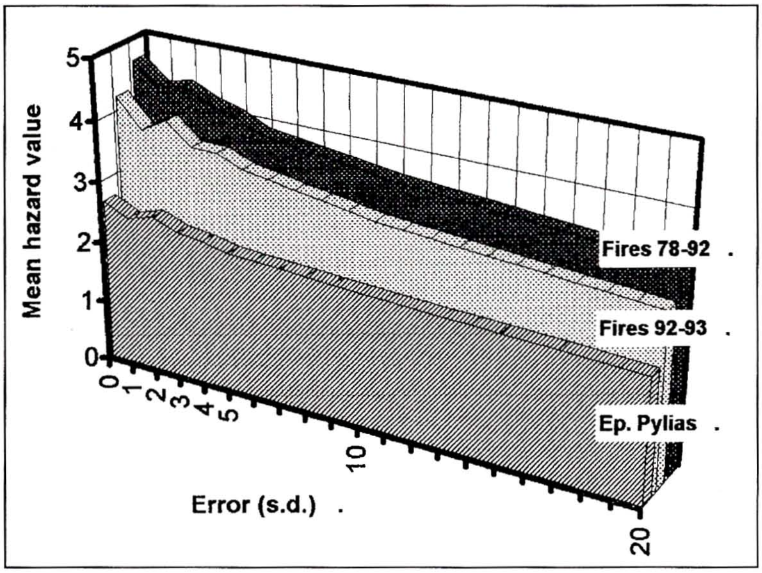

Figure 8 - Mean value of fire hazard in function of error in input data

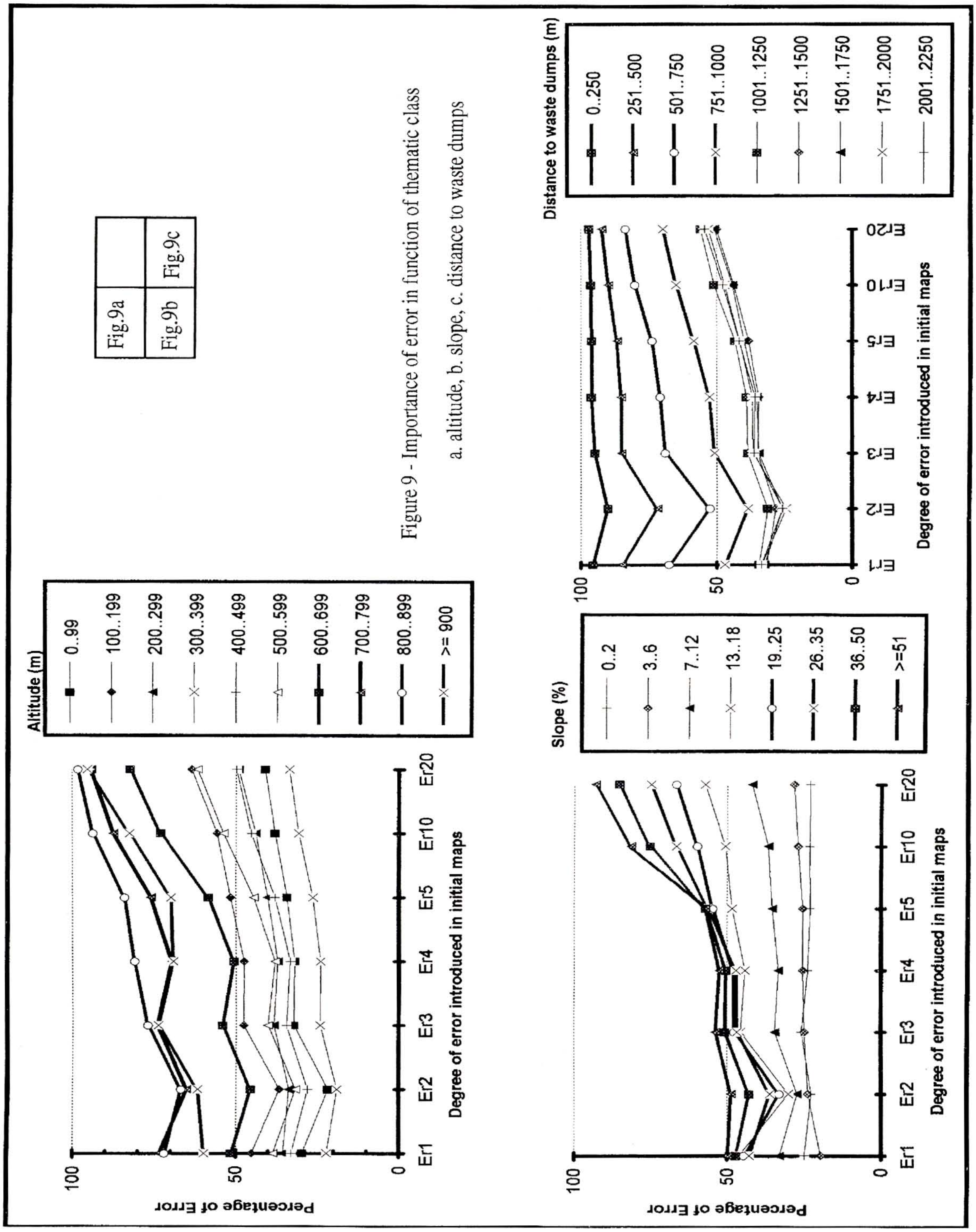

Figure 9 - Importance of errors in function of thematic class

This trend of decreasing hazard with increasing error in input data can also be deduced from the mean hazard value of the different hazard maps. Figure 8 gives the mean hazard value for the entire hazard map of the Eparchy of Pylias, as well as the mean value of the areas burnt during the period 1978-1992 and during the period 1992-1993.

The decrease in mean hazard value is most pronounced in the case of relatively smaller errors: The difference between the mean value for HAZ_0-HAZ_5 and for HAZ_5-HAZ_20 is 0.36 and 0.26, respectively. The same trend occurs when only the burnt areas are considered.

COMPARISON OF HAZARD_ERROR MAPS WITH THE "ERROR-FREE" HAZARD MAP

The ambiguity of the hazard model due to errors in the input data is investigated by comparing each of the resulting hazard maps (HAZ_X) with the "error-free" hazard map (HAZ_0). The cartographic representation of the spatial distribution of "hazard"-errors allows one to note the areas where the hazard model is doubtful, based only on errors in the input data.

A comparison with the different thematic maps is performed. An illustration is given for altitude (Figure 9a), slope (Figure 9b) and distance to waste dumps (Figure 9c). It can be said that:

1. the hazard model for fire spreading in the Eparchy of Pylias becomes less reliable with increasing degree of error for areas located at higher altitudes (> 600 m). For each of these thematic classes, more than half of the area has a different hazard value than the one obtained when "error-free" maps are used as input data.
2. Two groups of slope classes can be distinguished. The limit between these groups is 12°. Only slopes steeper than 12° will cause problems when errors greater than Er10 are present in the input data. For those steeper than 18%, significant problems occur beyond Er5.
3. A clear distinction can be made between areas close to a waste dump (< 1000 m) and areas at greater distances: the higher the degree of error in the input data set, the less reliable will be the fire spread hazard for areas within a distance of 1000 m from a waste dump. This is particularly significant for distances beyond 500 m at errors greater than Er4.

CONCLUSION

The advantage of GIS applications is that information can be extracted automatically from input data saving time and sparing costs. The input data can be combined with the derived variables to set up hazard models which can be used in decision making and planning. It would be nearly impossible to arrive at this point if this information were to be calculated in a non-automatic way.

However, GIS is very prone to errors and uncertainties. This point should be of primary interest if (hazard) models are used. Even relatively small errors in the input data can result in important errors for the derived data and can disturb the outcome of the model influencing as such the decision of the end user. It is essential to know what is the sensitivity of the different parameters used in the model to error and uncertainty in the input data.

As illustrated in this paper, error sensitivity can be studied by means of a theoretical approach that is based on the Monte Carlo simulation. The results of this research illustrate the error sensitivity of different parameters. The sensitivity to error is strongly different according to the parameter and in relation to the degree of error. It is found that a relative small error in the input DEM may result in important inaccuracies as is the case with aspect. The errors of aspect increase with increasing error in the DEM.

The holistic influence of error in a computed fire hazard model is also discussed. It is found that the fire hazard map becomes more uniform with increasing degree of error in the input data. Classes with high hazard occur less frequent in favour of the lower classes. This means that the degree of hazard will be less pronounced with increasing random error in the initial data set.

In conclusion, it can be formulated that it is important to know, prior to modelling, what variables are highly sensitive to errors. This will allow to bypass these parameters with a "chaotic" behaviour and to obtain better results for the model or to define these areas where the hazard model is doubtful, based only on errors in the input data.

REFERENCES

Brunsdon C. & Openshaw S. (1993). Simulating the effects of error in GIS. In: Mather P.M. (ed.) *Geographical Information Handling - Research and Applications.* Chichester: J.Wiley & Sons, pp.47-61.

Burrough P.A. (1986). *Principles of Geographical Information Systems for Land Resources Assessment.* Oxford: Clarendon Press, 193 p. (Monographs on soil and resources survey n°12).

De Vliegher B.M. & Basigos P.S. (1994). Fire hazard modelling using remote sensing and GIS. A case study of the Eparchy of Pylias (Messinia, Greece). In: *Proc. International Workshop "Satellite Technology and GIS for Mediterranean Forest Mapping and Fire Management"*, 4-6 Nov. 1993, EUR 15861 EN, pp. 107-118, Thessaloniki (Greece).

Eastman J.R., Kyem P.A.K., Toledano J. & Jin W. (1993). *GIS and Decision Making.* Geneve: United Nations Institute for Training and Research Explorations in GIS technology. UNITAR European Office, 112 p.

Eco U. (1988). *Foucault's Pendulum.* London: Picador Pan Books, 641 p.

Fisher P.F. (1991a). Spatial data sources and data problems. In: Maguire D.J., Goodchild M.F. & Rhind D.W. (eds.) *Geographical Information Systems - Principles and Applications*, Vol.1. Harlow: Longman Scientific and Technical, pp.175-189.

Fisher P.F. (1991b). First Experiments in Viewshed Uncertainty: The Accuracy of the Viewshed Area. *Photogrammetric Engineering & Remote Sensing*, **57(10)**, 1321-1327.

ILWIS (1993). *Integrated Land and Water Information System. User manual V1.3.* Enschede (The Netherlands): International Training Centre.

Lunetta R.S., Congalton R.G., Fenstermaker L.K., Jensen J.R., McGwire K.C. & Tinney L.R. (1991). Remote Sensing and Geographic Information System Data Integration: Error Sources and Research Issues. *Photogrammetric Engineering & Remote Sensing*, **57(6)**, 677-687.

Openshaw S. (1991). Developing appropriate spatial analysis methods for GIS. In: Maguire D.J., Goodchild M.F. & Rhind D.W. (eds.) *Geographical Information Systems - Principles and Applications*, Vol.1. Harlow: Longman Scientific and Technical, pp.389-402.