Doline probability map using logistic regression and GIS technology in the central Ebro Basin (Spain)

Lamelas, M.T.[a,b,*], Marinoni, O.[a], Hoppe, A.[a], de la Riva, J.[b]

[a] *Institut für Angewandte Geowissenschaften, Technische Universität Darmstadt, Schnittspahnstrasse, 9, D-64287 Darmstadt, Germany*

[b] *Dep. Geografía y Ordenación del Territorio, Facultad de Filosofía y Letras, Universidad de Zaragoza, Pedro Cerbuna 12, 50009, Zaragoza, Spain*

*Corresponding autor: María Teresa Lamelas. Adress: Dept. Geografía y Ordenación del Territorio, Facultad de Filosofía y Letras, Universidad de Zaragoza, Pedro Cerbuna 12, 50009, Zaragoza, Spain.

E-mail: tlamelas@unizar.es, tlamelas@geo.tu-darmstadt.de.

Tel.: +34976761000 ext.3909, +496151163205.

Fax: +34976761506, +496151166539.

Abstract. In the surroundings of Zaragoza, karstification processes are especially intense in covered karst areas where fluvial terraces lie directly on Tertiary evaporites. Since the beginning of Quaternary, these processes have lead to the development of collapse and subsidence dolines with a wide range of sizes, which have significant economic impacts. To reduce economic impact and increase safety a regional analysis of this phenomenon is needed for spatial management. Therefore, a probability map of dolines was developed using logistic regression and Geographic Information System (GIS) techniques. This paper covers the selection of input data, manipulation of data using GIS technology, and the use of logistic regression to generate a doline probability map. The primary variable in the doline development in this area is geomorphology, represented by the location of endorheic areas and different terrace levels. Secondary variables are the presence of irrigation and the water table gradient.

Keywords: dolines; geo-hazards; GIS; logistic regression; Ebro Basin

Introduction

Zaragoza city is located in the central part of the Ebro Basin, in the north-east of the Iberian Peninsula. This triangular-shaped basin is bounded in the north by the Pyrenees, in the south-west by the Iberian Range and in the south-east by the

1

Catalan Coastal Range. The continental sedimentary infill of this basin is composed of conglomerates and sandstones grading into clays at the margins, marls, evaporites and carbonate facies towards the depocentre of the basin (Benito et al. 1998). In the central part of the basin, the playa-lake deposits, from the Zaragoza Formation, form the largest gypsum outcrop in the area (Quirantes 1978). In this area, the different terrace levels and glacis lie directly on these Tertiary evaporites (Figure 1).

The rate of gypsum dissolution in water is approximately between 30-70 (Klimchouk et al. 1996) to 100-150 (Martínez et al. 1998) times faster than that of limestone. Thus, karstification processes are especially intense in this covered evaporite karst areas although the supply of water by annual precipitation in this semiarid environment is less than 350 mm.

Four conditions must be met for evaporite karstification to be considered active. These include (Johnson 2005; Lamont-Blanck et al. 2002):

1.- An evaporite deposit at the surface.

2.- Water that is unsatured with respect to the evaporite mineral.

3.- An outlet for the escape of solvent water.

4.- Energy to cause water to flow through the system.

These conditions are met in the study area. Karstification is an active process and has formed dolines, which results in collapse (sinkholes) or subsidence dolines with a wide range of sizes.

In the last decades, subsidence and doline development in this area has led to enormous economic losses caused by rupture of irrigation infrastructures and water supply networks as well as the destruction of transport infrastructures and buildings. The total economic losses due to all these factors are estimated to be in the order of tens of millions of dollars (Soriano and Simón 1995).

Most of the studies performed in relation with doline development in this area are aimed at solving selective problems, but many experts agree on the necessity of a regional analysis of this phenomenon for urban planning (Benito et al. 2000; Guerrero et al. 2004; Gutiérrez-Elorza and Gutiérrez-Santolalla 1998; Simón et al. 1998a; Simón and Soriano 2002; Simón et al. 1998b). This is especially necessary in complex areas characterized by a dynamic industrial and urban development and an intensive agricultural use as in the case of Zaragoza area.

According to Paukštys et al. (1999), the most cost-effective way of planning in these areas lies in avoiding existing dolines and most subsidence prone areas. The application of this preventive philosophy requires the recognition of the areas affected by subsidence and the production of hazard maps (Guerrero et al. 2004). In the last decade, several countries have built Karst databases integrated in a GIS environment where new interpretations of the relationship between the causes of the data can be readily made (Cooper et al. 2001; Gao et al. 2001, 2005; Green et al. 2002; Lei et al. 2001).

Our main objective is the development of a doline probability map in a GIS environment that can help to differentiate between high and low hazardous areas as an aid for regional land use management in the Zaragoza area. We aim to improve the existing knowledge about the factors that best determine doline distribution in this dynamic area.

In order to fulfil this main objective several secondary objectives need to be satisfied as well:

- Revision of the state of the art in the study area for the recognition of the main factors involved in this phenomenon.

- Revision of possible methodologies to be used in the hazard assessment.

- Mapping of the dolines present in the study area.

- Development of models for the different geological, environmental and human factors that can play a role in the hazard development.

- Multivariate analysis of the information for producing a doline probability map with the use of logistic regression.

Logistic regression relates predictor variables to the occurrence or non occurrence of dolines and uses the relationship to produce a map showing the probability of dolines. The main advantage of this technique, in comparison with multiple linear regression, is based on the possibility it provides of analysing a qualitative variable (as the occurrence or not of dolines) as a function of several qualitative and quantitative explanatory variables.

The study area

General settings

Zaragoza is located in the Ebro corridor, a highly dynamic economic area within the Iberian Peninsula. Its strategic location in the middle of three of the most important developed areas inside the Iberian Peninsula (Madrid, Barcelona and Valencia) facilitated a fast and great economic and urban development in the last decades. In 1964, Zaragoza was declared Focus of Industrial Development resulting in a rapid increase in population to 500 000 inhabitants in the 70s. Nowadays, the city has a total population of about 700 000 inhabitants (more than 60% of the population in the Region of Aragón) and is expected to increase even more in the next decade. Zaragoza will be the venue of the International Fair in 2008 with the title "Water and sustainable development", thus implying the creation of many new infrastructures and positions that will be very attractive for new inhabitants to come. Figure 2 shows the extension of new industrial and urban areas as the logistic platform (PLAZA) or the recycling technologic station (PTR) as well as the assigned location of the EXPO 2008.

The climate in this area is semi-arid with mean annual precipitation of about 350 mm and a mean annual temperature of about 15°C. The continental Mediterranean climate of Zaragoza is also characterized irregular distribution of precipitation, with annual peak values during spring and autumn, which also varies between different years. Other characteristics are the existence of high thermal differences between summer (average monthly maximum temperature of 24-25°C in July) and winter (monthly minimum temperature of 5-6°C in January) and the intensity and frequency of the wind called "Cierzo" (NW wind).

The combination of high anthropogenic impact and semiarid climate, means that the remnants of scarce natural vegetation corresponds to sclerophyllous scrub, gypsum steppes and riparian forest and thickets.

The Subsidence phenomenon

Hazard description

Alluvial dolines are generated when surficial deposits (terraces and glacis) overlying Neogene evaporites fall into conduits enlarged by dissolution, causing

depressions at the surface. Their generation follows two mechanisms (Soriano and Simón 1995):

a)   Mobilization of material by ground water, which may take place by either dissolution of soluble substrate or piping of the detrital cover.

b)   Cave-in of the alluvial cover, developed by either sudden collapse or slow subsidence.

According to Soriano and Simón (2002), the forms of the present-day alluvial dolines northwest of Zaragoza correspond to those described in classical papers (Cvijic 1981; Palmquist 1979): pan, funnel and well-shaped types have been identified. Doline size varies according to the diameter and ranges between several meters to 100 m, with depths varying between 1 and 20 m. There are also larger shallow depressions up to 1100 m long and 600 m wide. Most of these were identified as uvalas. Soriano and Simón (2002) observed in this sector, that dolines density is higher in the second terrace levels than in the flood plain. In addition, they monitored the subsidence of different dolines during a period of about 4 years and estimated subsidence rates ranging between 21 to 92 mm/year. In the sector southeast of Zaragoza, most of the dolines are shallow closed depressions, commonly less than 1.5 m deep, with non-scarped edges. They show highly variable geometries and dimensions, reaching up to 0.35 km^2 in area and 2 km in length. A small number of scarp edged collapse sinkholes (up to 7 m in length and 2.5 m in depth) have also been recognised in the floodplain (Gutiérrez-Santolalla et al. 2005b).

Controlling factors

Gutiérrez-Elorza and Gutiérrez-Santolalla (1998) classified the factors that play a role in the dissolution process based on geological and environmental factors. On the other hand, Soriano and Simón (1995) distinguished between physical, hydrological and human factors. Despite this difference in classification their studies in the area agree on the following factors that control the subsidence process:

1. Geological factors:
    a. Lithological: characteristics of the substrate and Quaternary cover (texture, porosity and permeability).

b. Stratigraphical: thickness of gypsum formations and Quaternary cover, intercalation of insoluble bodies, existence of halite beds.

c. Structural: structure of the gypsum formations, discontinuity planes, joints, faults.

d. Geomorphological: relief configuration.

e. Hydrogeological: flow velocity and regime, water infiltration, hydraulic gradient, depth and seasonal variations of the water table, thickness of the saturated Quaternary deposits, chemical composition of the groundwater.

2. Environmental factors:

a. Climate: existence of stormy events and high temperatures.

3. Anthropogenic factors: irrigation, pumping, and constructions.

Guerrero et al. (2004) reported that recent studies based on borehole data and interpretations of the paleokarst features highlight the relevance of salt and sodium sulphate karstification in the subsidence phenomena. They confirm the presence of halite close to the surface in some locations along the Huerva Valley like Cadrete village (Figure 2). It is possible that previously existing halite beds have been removed by dissolution from the outcropping evaporite sequence. The existence of halite and Na-sulphates in the bedrock is a crucial factor for the development of dissolution-induced subsidence phenomena due to their high solubility. Whereas the solubility of gypsum at 25º C is 2.4 g/l, halite, glauberite and thenardite solubility reach 360, 118 and 519 g/l respectively (Ford and Williams 1989).

Guerrero et al. (2004) also observed that most of the subsidence problems do not occur inside the city but in the peripheral areas of Zaragoza (Figure 6) where there is a relatively thin alluvial mantle. They attribute this to the fact that the majority of the city is built on thickened and slightly cemented alluvial deposits. This explanation matches the results obtained by Simón et al. (1998b); Soriano and Simón (1995) and Simón et al. (1998b). However, a recent study (Gutiérrez-Santolalla et al. 2005b) reveals that downstream of the Ebro River flood plain the alluvium thickness does not have a significant influence on the formation of dolines.

Two main sets of vertical joints with prevalent N–S and NW–SE directions have been recognized in the study area (Arlegui and Simón 2000). Several authors

agree in the high morphogenetic control of the axis NW-SE in the study area. Maldonado et al. (2000) studied the dolines developed in La Puebla de Alfindén village (10 km NE Zaragoza). They discovered that many dolines developed along this NW-SE axis. Soriano (1992) compared the direction of the axis of dolines with the direction of the fractures and observed a high correlation.

Maldonado et al. (2000) also monitored the dolines' history and discovered that the subsidence activity was directly related to storm events. Furthermore, Soriano and Simón (2002) reported that subsidence rates correlate with rainfall but they stated that the changes are very subtle, suggesting that the influence of rainfall on doline evolution is not evident.

Another possible factor is the location of areas where the Tertiary aquifer discharges into the alluvial aquifer following the model suggested by Jiménez-Torrecilla et al. (2004).

There is a reciprocal interaction between anthropic activities and doline generation. Many infrastructures are affected by subsidence but at the same time certain activities favour the generation of dolines. Benito and Gutiérrez-Elorza (1988) observed that dolines preferentially form near unlined canals. Guerrero et al. (2004) observed that the generation of dolines is frequent along the Imperial Canal (Figure 2 and 4), particularly on the downgrading side.

Other human-induced triggering factor also cause variations in the water table since these are mainly due to the supply of water by irrigation (Soriano and Simón 1995). As the water level declines it causes a loss of buoyant support to the ground, increases the flow gradient and velocity, facilitates the aquifer recharge and reduces the geomechanical strength of the alluvial cover (Gutiérrez-Elorza and Gutiérrez-Santolalla 1998).

The geological hazard

The presence of dolines causes many problems both in farming activities and civil engineering. Farmers lose arable land and attempt to replace it by filling the depressions each year. In the upstream area of Zaragoza problems in civil engineering have increased concern since the 1970s when a large number of factories were built. Many buildings were built on depressions, which had been filled several years before and no special inspections for construction were taken at that time (Soriano and Simón 1995). Nowadays, much damage is evident in the

form of fractures which have developed in walls and floors of some factories (Figure 3).

Due to the filling of voids by the farmers, the surface of the area occupied by dolines in present air photographs has been reduced with respect to photographs taken in the 50s. Local farmers fill the depressions with the intention of transforming the land into arable production. However, according to some landowners, this is an ineffective practice, as the filled and levelled dolines often remain unproductive (Gutiérrez-Santolalla et al. 2005b).

In the downstream area of Zaragoza damage to buildings has been reported in La Puebla de Alfindén village (Figure 2), although in some cases it is difficult to elucidate whether it is due to dissolution subsidence or the hydrocollapse of the gypsiferous silts. The Madrid-Barcelona motorway (A-2) crosses a few shallow closed subsidence depressions (Gutiérrez-Santolalla et al. 2005b).

As to the water management infrastructures, an old section of the Canal Imperial de Aragón, the biggest canal in this region, had to be abandoned and rebuilt several meters to the north of the old section because the construction itself intercepted several collapses (Figure 4).

Another case of doline problem was observed in La Cartuja village (Figure 2). In 2003 a doline developed underneath the municipal swimming pool. Approximately 15 years ago, the swimming pool was broken in the middle and rebuilt without major protective investigations.

Many examples can be collected from the previous studies, newspapers and personal meetings with the local residents in the area reasserting the importance of this phenomenon and the necessity of a better knowledge of its spatial distribution for a proper land use management.

Methodological background

Many examples concerning subsidence hazard mapping can be found in the literature, especially in European and North-American developed countries. Soriano and Simón (1995) observed that groundwater sulphate content, grain size of the detrital cover, topography of the Tertiary/Quaternary boundary, annual variation of the water table and the thickness of detrital cover were the main factors controlling doline development. Taking into account these variables a theoretical spatial hazard model was elaborated and expressed as a mathematical

equation. They also produced a 1: 50 000 scale doline hazard map of their study area located upstream of Zaragoza.

In subsequent studies Simón et al. (1998b) stated that this model cannot be applied at a larger, 1:25 000 scale and used a qualitative assessment. They introduced three main factors in the subsidence hazard, by order of importance:

1. Topography of the Tertiary-Quaternary contact.
2. Thickness of Quaternary sediments.
3. Lutitic percentage in Quaternary deposits.

They also introduced the gradient in groundwater sulphate content as a secondary factor. A potential hazard map was developed based on these factors. The potential hazard map overlaps with a cartography of present hazard developed by air photograph analysis and cartography of dolines. The resulting map presents 7 categories of hazard from Very high real hazard to Low potential hazard.

Kaufmann and Quinif (2002) also used a GIS-based qualitative assessment in the Tournaisis area. Hydrogeological data and geological mapping information were used to create a geo-hazard map. Records of former collapses were also investigated. These records were of great interest since doline distribution was obviously clustered in the area. Their study showed that zones of high doline occurrence coincide with zones of significant drop of piezometric heads. Combining the density of former collapses with the dewatering of the limestone enabled them to delineate zones of low, moderate and high collapse hazard.

Green et al. (2002) introduced a karst unit delineation using GIS technology as well. Many different overlays of the karst features were combined to better understand the landscape dynamics.

Sinkhole probability maps have also been constructed for south-eastern Minnesota. These maps were constructed as paper maps with boundaries drawn by the maps' authors using subjective criteria. As part of the transition of this mapping effort into a digital GIS environment, Gao and Alexander (2003) developed a mathematical decision tree model for the construction of maps of relative sinkhole risk based on the distribution of distances to the nearest sinkhole and the sinkhole density.

Whitman and Gubbels (1999) used GIS to investigate the spatial relationships between hydrogeologic factors and sinkhole formation near Orlando, Florida. Landsat TM imagery, digital topography, and well data were used to construct a

model of head difference between a discontinuous set of surficial aquifers and the Florida aquifer, a regionally extensive confined aquifer.

Lei et al. (2001) studied sinkhole distributions influenced by karstification of bedrock, soil types and properties, groundwater condition, human activities, land use, infrastructures, and economic development. They developed a sinkhole hazard map using a qualitative weighting method within GIS. The three factors they introduced were: water level in dry and rainy seasons, thickness of soil and finally the karstification of bedrock together with the fault distance.

Qualitative methods are very flexible and permit a complete inclusion of expert knowledge. The main pitfall is that they involve a great level of subjectivity, so that the maps produced by different researchers can be very different. Although a fully objective procedure does not exist, quantitative methods as logistic regression assure that the same results can be achieved provided the same basic assumptions apply (Beguería and Lorente 2003).

Specific methodology

Logistic regression

Commonly applied linear regression techniques are appropriate only when the dependent variable and the explanatory variables are quantitative and continuous. To analyze a binary qualitative variable (0 or 1) as a function of a number of explanatory variables, special techniques must be used if the analysis is to be performed adequately. One of the most common regression models that accomplish this is the logit or logistic regression model (Bledsoe and Watson 2001).

The logistic regression technique has already been used for many environmental studies, in many cases with more success than multiple linear regression. Battaglin and Goolsby (1997) compared the results of both techniques to identify natural and anthropogenic variables of drainage basins that have strong relations to agricultural chemical concentrations and mass transport measured in rivers. They concluded that logistic regression was somewhat more successful than multiple linear regression.

This technique has also been applied successfully to predict thresholds of channel pattern and instability (Bledsoe and Watson 2001) and to forecast short term hail

risk (Sánchez et al. 1998). But the most common use of this technique in geo-hazards has been the development of landslides hazard maps (Beguería and Lorente 2003; Lee and Min 2001; Ohlmacher and Davis 2003).

Logistic regression establishes a functional relationship between the binary coded hazard locations (existence or not of dolines) and different factors which are recognised to play a role in the hazard development. It states that the natural logarithm of odd (logit) is linearly related to the independent variables (Beguería and Lorente 2003):

$$\text{Logit } (P) = \text{Ln } (P/\ 1\text{-}P) = B_0 + B_1 X_1 + \ldots + B_n X_n \qquad (1)$$

Where P is the probability of occurrence, X_n is the set of n independent variables, and B_n is the set of n+1 parameters. Solving P from expression 1:

$$P = \exp (B_0 + B_1 X_1 + \ldots + B_n X_n)/\ (1 + \exp (B_0 + B_1 X_1 + \ldots + B_n X_n)) \qquad (2)$$

Slope coefficients for the logistic equation are fit into the categorical data using a maximum likelihood method that optimises the probability that the observed data be estimated from the set of slope coefficients (Battaglin and Goolsby 1997).

In ordinary regression analysis, the coefficient of determination (r^2) is frequently used as a measure of model performance. In logistic regression, it is common to be more concerned with whether the predictions are correct or incorrect than with how close the predicted values are to the observed (0 or 1) values of the dependent variables. Therefore, r^2 has little meaning in logistic regression analysis (Bledsoe and Watson 2001).

Goodness-of-fit tests may aid in the interpretation of the results of logistic regression. The likelihood L_0 for the null model, where all slope parameters are zero, may be directly compared with the likelihood L_1 of the fitted model. Specifically, one can compute the X^2 statistic for this comparison as:

$$X^2 = -2 (\log (L_0) - \log(L_1)) \qquad (3)$$

The degree of freedom for this X^2 value is equal to the number of independent variables in the logistic regression. If the P-level associated with this X^2 is significant, the estimated model yields a significantly better fit to the data than the null model and the regression parameters are statistically significant.

Dependent and independent variables mapping

In this section, we aim to explain the development of the dependent variable (doline/no doline) and the independent variables that will be introduced in the stepwise logistic regression procedure. Due to the lack of information about some factors such as location of salty layers and fractures in the Tertiary substrate, these factors were not included in the model. This fact may limit the accuracy of the doline probability map.

Digital Terrain Models (DTM) of the different factors that can be regionalised were created. These models have a spatial resolution of 20 m and are restricted to the Ebro alluvial due to the availability of geological and hydrogeological information.

Several methodologies were used for the models development: simply conversion of existing maps into a common reference system or change on data structure (rasterization); in other cases it was necessary to analyze and digitalize air photographs; and finally the interpolation of punctual data gathered from reports or field analysis (i.e. precipitation from meteorological stations, boreholes data, etc.) and information modelling.

ArcGIS 9.1 (ESRI 2005) and Gocad (Earth Decision Sciences 2005) were used to implement the model. Figure 5 shows the different steps of the data preparation. The information about the dependent and independent variables at every pixel in the DTM was stored in a database and then analysed with a statistical software package, SPSS 11 (SPSS 2001). In the database, every row corresponds to a case, every location, and every columns represents a different variable.

ERDAS 8.7 (Leica 2003) was used to create the database. This software has the capability of importing grids into a same Layer stack. This Layer stack can then be exported as ASCII files which in turn are easy to be imported into SPSS. SPSS performs the logistic regression analysis and returns the required logistic regression model parameters in terms of the regression coefficients (see eq. 1). The last step is the implementation of the model which was carried out using the raster calculator in ArcGIS.

The dependent variable

For the development of the dependent variable, doline location and cartography of previous studies in the upstream area of Zaragoza were georeferenced and

digitized (Simón et al. 1998a; Simón et al. 1998b). For areas with no previous studies available air photographs from 1984 and 1997 were analyzed. Figure 6 shows that doline development is a severe problem in the region along the Ebro valley. The doline vector map was then converted into raster format and reclassified into two different values, existence (1) and no existence (0) of dolines.

Geological variables

In the case of the lithological and stratigraphical variables, a 3D model of the Quaternary alluvial deposits of the study area was created. The information from more than 900 boreholes was obtained from the Water Points Inventory (IPA, Inventario Puntos de Agua) of the Ebro Basin Authority (CHE, Confederación Hidrográfica del Ebro). This information was completed with several boreholes collected from different private enterprises (Control-7, Entecsa, Z-amaltea, CTA, ESHYG) and from some previous studies carried out for the construction of several roads (M.O.P. 1967, 1970, 1973, 1994, 2000, 2003).

Since doline development is mainly limited to the Ebro valley where the major concentration of infrastructures exists, boreholes outside the valley were not considered for the model. Due to the lacking, fragmentary or imprecise information on the lithology, it was necessary to omit a number of additional boreholes. In the end, about 400 boreholes could be used for modelling the Quaternary-Tertiary boundary. Together with the digital elevation model (DEM), this boundary represents the outer limit of a 3D model body which was filled with lithologic information. The spatial construction of the model was carried out with Gocad.

Due to the lack of precise information about the permeability of the Quaternary valley deposits, the permeability of the Quaternary cover was qualitatively assessed using lithological information. The approach uses the mean grain-size of the encountered layers. Clay, silt, sand and gravel represent grain-size classes with sedimentologically well defined upper and lower grain-size limits which allow the determination of a mean grain size. For layers which consisted of more than one grain-size class, the mean layer grain-size was determined as a function of the proportion of each grain-size class, which in turn was estimated from the lithological layer description. The mean grain-size was then spatially interpolated within the previously modelled 3D body, returning a 3D model of the grain-size.

In this model, areas having a mean grain-size spectrum belonging to the clay-silt fraction were categorized as low permeable whereas the rest was classified as permeable.

For the geomorphological factors, the geological maps scale 1:50 000 from the National Geological Institute (ITGE 1995; ITGE 1998) were used to locate the different terrace levels. We wanted to assess if doline distribution is more concentrated in particular terrace levels. This information was originally in vector format from ArcInfo and required conversion to raster format.

Several hydrogeological variables were also considered. The information about water table level and hydrochemistry was also obtained from the IPA. Since this inventory contains all hydrogeological information from different studies the information is not consistent with respect to time and space.

In the case of the water table models, only the points with more than 10 measurements were selected. Figure 7 shows the distribution of the points in the study area. In order to obtain the water table surfaces, the information from the points in the IPA needs interpolation. Interpolation is a complex issue, which consists of developing a continuous surface from punctual data. Interpolation procedures can be simple deterministic mathematical models (inverse distance weighting, trend surface analysis, Thiessen polygons, etc.), or more complex models (geostatistical methods, such as kriging and thin plate splines) (Ninyerola et al., 2000). The mean, maximum and minimum water table level values were queried from the database and interpolated with ordinary Kriging. This method belongs to geostatistical methods. Geostatistics constitute a theory about the statistical behaviour of natural phenomena with a spatial variability. Its fundamental concept is the regionalized variable, which can always be represented by a continuous surface over the map plane.

For the creation of the groundwater depth models, a subtraction of the groundwater level from the Digital elevation model (DEM and groundwater level model with 20 m pixel size) was conducted (expression 4).

$$X = A-B \qquad\qquad\qquad (4)$$

Where X is the groundwater depth, A is the groundwater elevation above sea level and B is the Digital elevation model of the surface.

Besides, a subtraction between the maximum and minimum water table level models was also made for the development of the water table variations between summer and winter.

The water table gradient was created using a focal analysis filter. The standard deviation of a 5x5 pixel kernel was used to model the areas with more water table variations in space, thus implying greater flow velocity.

With respect to the hydrochemistry, the number of points with several measurements was scarce, so the information from all boreholes was used. To avoid erroneous data the error balance was calculated based on the method by Custodio and Llamas (1983). All the points with an inadmissible error were rejected. A simple kriging interpolation technique and a cross validation of the model were conducted. A Leave-one-out cross validation of the model was selected due to the lack of enough information to divide the sample into training sets and testing sets. This involves using a single observation from the original sample as the validation data, and the remaining observations as the training data. This is repeated such that each observation in the sample is used once as the validation data.

Environmental variables

In the study area, the main environmental triggering factor for doline development is the occurrence of storm events. However, the regression analysis does not include this variable because rain storms are relatively uniformly distributed throughout the study area.

Anthropogenic variables

The cartography of irrigated areas in 2000, which is available online (CHE 2005), was used for locating irrigated land, since the irrigated land in the alluvial plain has not changed since 2000. This information was converted from vector format to raster format.

In addition to this, the irrigation amount was also approximated. A database with information about the availability of water resources for every agrarian administrative division and irrigation system was available. More details about the origin of this information can be found in Cruz et al. (1997). The irrigation area map was intersected with the agrarian administrative division map, which is also

available through the internet (CHE 2005). The water availability values were assigned to every system in every agrarian division. This only applies to the real water supply by irrigation since the information represents the common water availability under regular conditions and as a consequence the amount of water commonly irrigated. However, it does not take into account dry periods when the real availability is reduced. Therefore, in case of selection of this factor by the model, the doline susceptibility would be exaggerated in dry periods when the amount of irrigation is lower than the one introduce in the model.

The impermeability of some infrastructures (roads and train rails) causes a water concentration at their edges, which is believed to be an important factor in the doline development. In addition, the breaks in the irrigation network (canals) represents a grate support of water. In many previous studies, it was asserted that dolines locate in the proximity of irrigation canals. Thus, the infrastructure information was obtained from the digital 1:25 000 scale topographic map from the National Geographical Institute (IGN). The information had to be previously updated by digitising new roads and train rails (constructed after the IGN map creation) by using air photographs. Models representing the distance to these infrastructures were performed to assay whether doline development probability increases proportionally in response to reduced distances of this infrastructures. For the development of these models a distance operation into ArcGIS was performed.

Results

In order to better understand the factors controlling doline distribution, models of different factors were visually analysed together with the doline map. In addition, some of the variables were classified into different categories taking into account previous studies and the visual analysis. Both variables, the continuous and categorical, were introduced in the analysis. Table 1 shows the different variables introduced in the model and their continuous or categorical condition.

A subset of 60% of the total information, selected randomly, was introduced in a stepwise logistic regression procedure; stepwise selection method with variables introduction based on the significance of the score statistic, and removal based on the probability of a likelihood-ratio statistic based on conditional parameter estimates.

First, the fitness of the logistic regression equation was tested using this 60%
sample data. For the overall model fit the goodness-of-fit test called "chi-square"
in the SPSS was used. This test is simply the chi-square difference between the
null model (i.e., with the constant only) and the model containing one or more
predictors. This is one use of the likelihood ratio test between two nested models.
It is an assessment of the improvement of fit between the predicted and observed
values by adding the predictor(s). A classification table of observed dolines (0 and
1) and adjusted probability classes (determining a dividing point by matching the
quantity of the number of 1s in the observed values of the dependent variable) to
obtain the overall percentage of correct classifications was performed.

In the final model, a total of 11 variables were introduced (or used) where eight
variables were categorical and three of them were continuous. Table 2 shows the
variables introduced in the model. β is the coefficient of the logistic equation. The
Wald test is the usual test for the significance of a single predictor. It tests the
hypothesis that the predictor is useful in predicting the outcome. In this case, all
the variables are significant ($\alpha < 0.05$) since the method used is a stepwise
procedure. And finally, the exp of β is the change in odds for a unit increase in the
independent variable. This last statistic shows the relative importance of the
variables.

According to the model results, the most significant variables are geomorphology,
represented, in order of importance, by the location of endorheic areas and
different terrace levels, T4, T3, T2 and T1. The secondary variables are the
presence of irrigation and the water table gradient.

The next variable in importance is the percentage of low permeable layers in the
Quaternary cover. This only plays a role when classified into two categories.
Thus, the existence of low percentage, less than 10%, of low permeable
(impermeable) layers increases doline probability, according to the positive
coefficient .

The factors with negative coefficients are of less importance in this case due to
their low exp of β. The results suggest that high ground water sulphate contents
reduce doline probability. Doline probability increases in areas with lower
summer water table, which imply a larger supply by irrigation. Finally, doline
probability is significantly reduced in areas where the Quaternary sediment is
more than 30 m thick. The variables selected by the model represent the doline

distribution in the study area and do not imply that they are the only ones that take part in the hazard development. This is exemplified by the water table variations. Despite its importance, this variable is not introduced in the model because it correlates high with water table gradient.

The determined model was validated with the 40 % of information not introduced in the logistic regression by comparing the predicted outcome (probability value) with the reality (dolines occurrence or not). Due to the difference in percentage between the two groups (existence or not of dolines) and the difficulty in understanding the validation results all the doline cases and 12000 randomly selected cases lacking dolines were selected making a total validation sample of 24059 cases. The resulting confusion (error) matrix (Table 3) yields an overall map accuracy of more than 68 % and a Kappa index of agreement of 0.36, which means a classification of 36% better agreement than that obtained by chance. The confusion matrix is commonly used in classificatory approaches as a way to test the model performance (Beguería and Lorente 2003; Martínez-Casanovas et al. 2004).

Logistic models are frequently used in a classification approach. This implies selecting given values of the response variable and classifying all the cells in one of the groups formed according to these. The method more commonly adopted in literature is to divide the probability map histogram into different categories based on expert opinions (Lee and Min 2001; Dai and Lee 2002; Ohlmacher and Davis 2003). In the case of division into two groups, the threshold value is normally 0.5 probability, since the two sample groups are usually similar in size. For the case where the two groups are very dissimilar, the proportion of ones in the sample (proportion of dolines, 0.06 in our study area) can be used instead of the 0.5 value (Beguería and Lorente 2003; Martínez-Casanovas et al. 2004). According to Ayalew and Yamagishi (2005), this type of changing continuous data into two or more categories does not take into account the relative position of a case within the probability map and is neither fully automated nor statistically tested. They suggested classification systems that use quantiles, natural breaks, equal intervals and standard deviation to choose the one that best suits the information and the scale of investigation.

Figure 8 shows the doline probability map and, superimposed, the location of dolines since they need to be considered as areas with a higher doline probability.

In this case the map has been classified into four categories using the mean and standard deviation values. Note that the division in categories has been made for visualization purposes, which does not imply a categorization into safe and unsafe areas. In fact, all the study area presents a certain probability of doline development.

Discussion and conclusions

One of the major advantages of logistic regression is the possibility to introduce categorical variables in the model. Our results corroborate this affirmation as many of the most important variables in the doline distribution were categorical i.e. geology, represented by the different terrace levels. In addition, it allowed us to introduce anthropogenic factors that are very important in the doline probability, namely location of irrigated land.

In some cases, it is necessary to categorize continuous variables as the changes produced from one unit to another are insignificant or are not constant along the range of values of the variable. This implies that the relation between the dependent and independent variable does not follow a lineal regression. This is the case of the thickness of Quaternary deposits or the percentage of permeable layers. Our results suggest that there is a threshold of 30 m thickness of Quaternary cover. Thicker covers reduce doline probability while, in the range between 0 and 30 m thickness, dolines develop independently from the sediment cover thickness. Similar results can be found in previous studies (Johnson 2005; Simón et al. 1998b; Soriano and Simón 1995).

Several authors have demonstrated, in different mantled karst areas, that the formation of one doline (the ''mother'') promotes subsurface conditions that favour the formation of additional dolines (the ''daughters'') in the adjacent area (Cooper and Saunders 1999; Drake and Ford 1972; Gutiérrez-Santolalla et al. 2005a; Kaufmann and Quinif 2002; Palmquist 1979).

In the downstream area of Zaragoza Gutiérrez-Santolalla et al. (2005a) discovered that the dolines show a clear tendency to form clusters. They calculated the spatial distribution index given by the Clark and Evans index (Clark and Evans 1954). This index quantifies the clustering or dispersion of elements in a particular area. In the La Puebla de Alfindén sinkhole field, the Clark and Evans index has a value close to 0 (R=0.00089) indicating a clustered distribution. Gao et al. (2005), after

applying the Clark and Evans index to active kast areas in Minnesota, also concluded that the sinkholes tend to be clustered. However, they found that the pattern change from clustered to random to regular as the scale of the analysis decreases.

To this respect, we agree with Gutiérrez-Santolalla et al. (2005b) that the mapping of areas affected by subsidence is an effective approach to avoid future subsidence for land-use planning. Consequently these areas have to be considered as areas with high probability of doline development. However, as they pointed out these maps have some limitations related to the difficulty of mapping all the subsidence areas as they may be masked by the morphosedimentary activity of the fluvial system and anthropogenic fillings. Therefore, the study of the relationship between doline distribution and different environmental factors is very important. Such studies lead to a better recognition of hazardous areas which would be considered as non hazardous .

As a result, the doline probability map created with GIS and logistic regression technique provides valuable information for land use planning at a regional scale. However, there are some limiting factors such as the impossibility of mapping some factor playing a role in the doline development. This is the case of the mentioned salty layers or faults in the Tertiary evaporites. This factor could not be introduced in the logistic regression due to the lack of information. Thus more research is still needed to characterised the geology of this area since the introduction of these factors may improve the result of the doline susceptibility map.

An excellent example of management in subsidence prone areas can be found in some countries like Slovenia and China where governments have founded karst research institutes (Veni 1999). A better management approach in the area surrounding Zaragoza city would be the creation of a Karst research institute or organization in the Central Ebro Basin. Its main objective should be the creation of a karst database. This database must integrate the karst features mapped in previous studies carried out separately in different areas around the city of Zaragoza by different research groups. Future studies need to integrate information related to the geological, environmental and human factors and to create new data in areas having limited data and information such as the Ebro

alluvial downstream of Zaragoza. And finally, but not the least important is the active cooperation between different research groups and land use managers.

References

Arlegui L, Simón JL (2000) Geometry and distribution of regional joint sets in a non-homogeneous stress field: case study in the Ebro Basin (Spain). J Struct Geol 23:297-313

Ayalew L, Yamagishi H (2005) The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. Geomorphology 65:15-31

Battaglin WA, Goolsby DA (1997) Statistical modeling of agricultural chemical occurrence in midwestern rivers. Journal of Hydrology 196:1-25

Beguería S, Lorente A (2003) Landslide hazard mapping by multivariante statictics: comparison of methods and case study in the spanish pyrenees. Instituto Pirenaico de Ecología, Contract No EVG1 - CT-1999-00007: http://damocles.irpi.pg.cnr.it/docs/reports/df_modelling.pdf

Benito G, Gutiérrez-Elorza M (1988) Karst in gypsum and its environmental impact on the Middle Ebro Basin Spain. Environmental Geology Water Sciences 12(2):107–111

Benito G, Gutiérrez-Santolalla F, Pérez-González A, Machado MJ (2000) Geomorphological and sedimentological features in Quaternary fluvial systems affected by solution-induced subsidence (Ebro Basin, NE-Spain). Geomorphology 33:209-224

Benito G, Pérez-Gonzalez A, Gutiérrez-Santolalla F, Machado MJ (1998) River response to Quaternary subsidence due to evaporite solution (Gállego River, Ebro Basin, Spain). Geomophology 22:243-263

Bledsoe BP, Watson CC (2001) Logistic analysis of channel pattern thresholds: meandering, braiding, and incising. Geomorphology 38:281-300

CHE (2005) Contenido Cartográfico: Regadíos: http://oph.chebro.es/ContenidoCartoRegadios.htm

Clark PJ, Evans FC (1954) Distance to nearest neighbour as measure of spatial relationships in populations. Ecology 35:445–453

Cooper AH, Farrant AR, Adlam KAM, Walsby JC (2001) The development of a national Geographic Information System (GIS) for the British karst geohazard and risk assessment. In: Beck BF, Herring JG (eds) Geotechnical and Environmental Applications of Karst and Hydrology. Balkema, Rotterdam, pp 125-130

Cooper AH, Saunders JM (1999) Road and bridge construction across gypsum karst in England. In: Beck BF, Pettit AJ, Herring JG (eds) Hydrogeology and Engineering Geology of Sinkholes and Karst. Balkema, Rotterdam, pp 407-412

Cruz JM, Reiné R, R. de Rivera F (1997) Metodología utilizada en el estudio "Dotaciones de riego" borrador del Plan Hidrológico del Ebro. Riegos y drenajes XXI 93:45-54

Custodio E, Llamas MR (1983) Hidrogeología subterránea. Omega, Barcelona.

Cvijic J (1981) The dolines: translation of geography. In: Sweeting MM (ed) Karst geomorphology. Hutchinson, Pennsylvania, pp 225–276

Dai FC, Lee CF (2002) Landslide characteristics and slope instability modeling using GIS Lantau Island, Hong Kong. Geomorphology 42:213– 238

Drake JJ, Ford DC (1972) The analysis of growth patterns of two-generation populations: the examples of karst sinkholes. Can Geograph 16:381–384

Earth Decision Sciences (2005) GOCAD®. Earth Decision Sciences, Nancy, France.

ESRI Inc. (2005) ArcGIS 9.1. ESRI Inc., Redlands, CA.

Ford D, Williams P (1989) Karst geomorphology and hydrology. London.

Gao Y, Alexander EC (2003) A mathematical model for a sinkhole probability map in Fillmore County, Minnesota. In: Beck BF (ed) Proceedings of 9th multidisciplinary conference on sinkholes and the engineering and environmental impacts of karsts. ASCE Geotechnical Special Publication, Huntsville, Alabama, pp 439-449

Gao Y, Alexander EC, Tipping RG (2001) Application of GIS technology to study karst features of southern Minnesota. In: Beck BF, Herring JG (eds) Geotechnical and Environmental Applications of Karst Geology and Hydrology. Swets & Zeitliger, Lisse, pp 83-88

Gao Y, Alexander EC, Tipping RG (2005) Karstdatabase development in Minnesota: design and data assembly. Environmental geology 47:1072-1082

Gao Y, Alexander EC, Barnes RJ (2005b) Karst database implementation in Minnesota: analysis of sinkhole distribution. Environmental geology 47: 1083-1098

Green JA, Marken WJ, Alexander EC, Alexander SC (2002) Karst unit mapping using geographic information system technology, Mower County, Minnesota, USA. Environmental Geology 42:457-461

Guerrero J, Gutiérrez-Santolalla F, Lucha P (2004) Paleosubsidence and active subsidence due to evaporite dissolution in Zaragoza area (Huerva River valley, NE Spain): processes, spatial distribution and protection measures for transport routes. Engineering Geology 72(3-4):309-329

Gutiérrez-Elorza M, Gutiérrez-Santolalla F (1998) Geomorphology of the Tertiary gypsum formations in Ebro Depression (Spain). Geoderma 87:1-29

Gutiérrez-Santolalla F, Gutiérrez-Elorza M, Marín C, Desir G, Maldonado C (2005a) Spatial distribution, morphometry and activity of La Puebla de Alfindén sinkhole field in the Ebro river valley (NE Spain): applied aspects for hazard zonation. Environmental Geology 48:360-369

Gutiérrez-Santolalla F, Gutiérrez-Elorza M, Marín C, Maldonado C, Younger PL (2005b) Subsidence hazard avoidance based on geomorphological mapping in the Ebro River valley mantled evaporite karst terrain (NE Spain). Environmental Geology 48:370-383

I.T.G.E. (1995) Mapa Geológico de España a escala 1:50.000: hoja 354-Alagón (27-14). Ministerio de Industria y Energía, Madrid

I.T.G.E. (1998) Mapa Geológico de España a escala 1:50.000: hoja 353 Pedrola (26-14), 355 Leciñena (28-14), 383 Zaragoza (27-15), 384 Fuentes de Ebro (28-15). Ministerio de Medio Ambiente, Madrid

Jiménez-Torrecilla N, Galve JP, Asta MP, Sánchez-Navarro JA (2004) Estudio hidrogeológico de la subsidencia y los humedales salinos en el entorno de Zaragoza. Ayuntamiento de Zaragoza, Universidad de Zaragoza, Zaragoza. Unpublished.

Johnson KS (2005) Subsidence hazard due to evaporite dissolution in the United States. Environmental geology 48:395-409

Kaufmann O, Quinif Y (2002) Geohazard map of cover-collapse sinkholes in the "Tournaisis" area, southern Belgium. Engineering Geology 65:117-124

Klimchouk A (2005) Subsidence hazards in different types of karst: evolutionary and speleogenetic approach. Environmental Geology 48:287-295

Klimchouk K, Cucchi J, Calaforra J, Calaforra M, Askem S, Finocchiaro F, Forti P (1996) Dissolution of gypsum from field observations. International Journal of Speleology 25(3-4):37– 48

Lamont-Blanck J, Younger PL, Forth RA, Cooper AH, Bonniface JP (2002) A decision-logic framework for investigating subsidence problems potentially attributable to gypsum karstification. Engineering Geology 65:205-215

Lee S, Min K (2001) Staistical analzsis of landslide susceptibility at Yongin, Korea. Environmental geology 40:1095-1113

Lei M, Jiang X, Li Y (2001) New advances of Karst collapse research in China. In: Beck BF, Herring JG (eds) Geotechnical and Environmental Applications of Karst Geology and Hydrology. Swets & Zeitlinger, Lisse, pp 145-151

Leica (2003) Leica Geosystems software ERDAS IMAGINE 8.7. Leica Geosystems, Atlanta, GA.

M.O.P. (1967) Proyecto Red arterial de Zaragoza: Autopista Zaragoza -Alfajarín. Tramo II. Ministerio de Obras publicas, Zaragoza. Unpublished.

M.O.P. (1970) Red arterial de Zaragoza: N-II. Autopista ronda norte. Zaragoza. Ministerio de Obras Públicas, Zaragoza. Unpublished.

M.O.P. (1973) Proyecto de Trazado de la red arterial de Zaragoza: 2º cinturón de la red arterial de Zaragoza, conexión de la N-II con la N-123. Ministerio de Obras Públicas, Zaragoza. Unpublished.

M.O.P. (1994) Proyecto de construcción: N-232 de Vinaroz a Santander. P.K. 223,6 al P.K. 234,8. Acceso a población (duplicación de calzada. Tramo: El Burgo de Ebro-Zaragoza. Ministerio de Obras Públicas, Zaragoza. Unpublished.

M.O.P. (2000) Proyecto de construcción. Conexión de la carretera N-232 de Vinaroz a Santander con la Autopista A-68 de Bilbao a Zaragoza. Tramo Casetas. Ministerio de Obras Públicas, Zaragoza. Unpublished.

M.O.P. (2003) Proyecto de construcción. Nueva carretera autopista. Ronda sur. Cuarto cinturón de Zaragoza. De la N-II a ña N-232 (Vinaroz). Ministerio de Obras Públicas, Zaragoza. Unpublished.

Maldonado C, Gutiérrez-Santolalla F, Gutiérrez-Elorza M, Desir G (2000) Distribución espacial , morfometría y actividad de la subsidencia por disolución de evaporitas en un campo de dolinas de colapso (Valle del Ebro, Zaragoza). Cuaternario y Geomorfología 14(3-4):9-24

Martínez JD, Johnson KS, Neal JT (1998) Sinkholes in evaporite rocks. American Scientist 86:39–52

Martínez-Casasnovas JA, Ramos MC, Poesen J (2004) Assessment of sidewall erosion in large gullies using multi-temporal DEMs and logistic regression analysis. Geomorphology 58:305–321

Ninyerola, M, Pons X, Roure JM (2000). A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques. International Journal of Climatology, 20: 1823-1841

Ohlmacher GC, Davis JC (2003) Using multiple logistic regression and GIS technology to predict landslides hazards in northeast Kansas, U.S.A. Engineering Geology 69:331-343

Palmquist R (1979) Geologic controls on doline characteristics in mantled karst. Z Geomorph Suppl Bd 32:90–106

Paukštys B (1999) Is human activity accelerating karst evolution? In: Beck BF, Pettit AJ, Herring JG (eds) Hydrogeology and Engineering Geology of Sinkholes and Karst. Balkema, Rotterdam, pp 103-124

Paukštys B, Cooper AH, Arustiene J (1999) Planing for gypsum geohazard in Lithuania and England. Engineering Geology 52:93-103

Quirantes J (1978) Estudio sedimentológico y estratigráfico del Tercio Continental de los Monegros. Instituto "Fernando el Católico", Zaragoza.

Sánchez JL, Marcos JL, de la Fuente MT, Castro A (1998) A Logistic Regression Model applied to short term forecast of hail risk. Phys. Chem. Earth 23(5-6):645-648

Simón JL, Martínez-Gil J, Soriano MA, Arlegui L, Caballero J (1998a) Estudio de riesgos naturales en los terrenos de la orla sudoeste del suelo urbanizable. Ayuntamiento de Zaragoza and Departamento de Geología de la Universidad de Zaragoza, Zaragoza. Unpublished.

Simón JL, Soriano MA (2002) Actual and potencial doline subsidence hazard mapping: Case study in the Ebro basin (Spain). In: Bobrowsky PT (ed) Environmental mapping : methods, theory and practice. Balkema, Rotterdam, pp 649-666

Simón JL, Soriano MA, Arlegui L, Caballero J (1998b) Estudio de riesgos de hundimientos Kársticos en el corredor de la carretera de Logroño. Ayuntamiento de Zaragoza and Departamento de Geología de la Universidad de Zaragoza, Zaragoza. Unpublished.

Soriano MA (1992) Characteristics of the alluvial dolines developed due to de dissolution of gypsum materials in the cetral Ebro Basin (Spain). Geomorphology 85:59-72

Soriano MA, Simón JL (1995) Alluvial dolines in the central Ebro basin, Spain: a spatial and developmental hazard analysis. Geomorphology 11:295-309

Soriano MA, Simón JL (2001) Subsidence rates of alluvial dolines in the central Ebro Basin, Northeastern Spain. In: Beck BF, Herring JG (eds) Geotechnical and Environmental Applications of Karst and Hydrology. Balkema, Rotterdam, pp 47-52

Soriano MA, Simón JL (2002) Subsidence rates and urban damages in alluvial dolines of the Central Ebro basin (NE Spain). Environmental geology 42:476-484

SPSS Inc. (2001) SPSS 11 for Windows. SPSS Inc., Chicago, IL.

Veni G (1999) A geomorphological strategy for conducting environmental impact assessments in Karst areas. Geomophology 31:151-180

Whitman D, Gubbels T (1999) Application of GIS technology to the triggering phenomena of sinkholes in Central Florida. In: Beck BF, Pettit AJ, Herring JG (eds) Hydrogeology and Engineering geology of sinkholes and karst. Balkema, Rotterdam, pp 67-73

Illustrations

Figure 1: Geomorphology of the study area.

Figure 2: Infrastructure development in the surrounding areas of Zaragoza.

Figure 3: Fractures developed in the parking lot of a factory in the northwest of Zaragoza.

Figure 4: Doline developed in an old section of the Canal Imperial de Aragón.

Figure 5: Scheme of the logistic regression variables development.

Figure 6: Dolines distributed in the surrounding areas of Zaragoza.

Figure 7: Location of boreholes for the groundwater table models.

Figure 8: Dolines probability map created with logistic regression technique. Please note, that the map does not allow a categorization into safe and unsafe areas (see text).

Tables

Table 1: Variables introduced in the logistic regression.

Table 2: Variables introduced by the model.

Table 3: Confusion matrix.

Table 1: Variables introduced in the logistic regression.

Factors	Name	Type
Lithological		
Percentage of impermeable layers in Quaternary	PERMEABI	Continuous
Percentage of impermeable layers in Quaternary 0-10 %	PERME_10(1)	Categorical
Percentage of impermeable layers in Quaternary >10 %	PERME_100(1)	Categorical
Stratigraphical		
Thickness of Quaternary deposits 0-15 m	DEPQ_15(1)	Categorical
Thickness of Quaternary deposits 15-30 m	DEPQ_30(1)	Categorical
Thickness of Quaternary deposits >30 m	DEPQ_80(1)	Categorical
Thickness of Quaternary deposits	Q_DEP	Continuous
Elevation of Quaternary deposits	COTA_Q	Continuous
Geomorphological		
Geomorphology. Terrace level T4	T4(1)	Categorical
Geomorphology. Terrace level T3	T3(1)	Categorical
Geomorphology. Terrace level T2	T2(1)	Categorical
Geomorphology. Terrace level T1	T1(1)	Categorical
Geomorphology. Endorheic area	GEOENDOR(1)	Categorical
Hydrogeological		
Groundwater table mean depth	WT_MEAN	Continuous
Groundwater table minimum depth	HIGWTDEP	Continuous
Groundwater table maximum depth	LOWTDEP	Continuous
Watertable mean level	WT_MEAN	Continuous
Watertable low level	WT_LOW	Continuous
Watertable high level	WT_HIGH	Continuous
Watertable gradient	GRAWTMEA	Continuous
Annual groundwater table variations	WT_VARIA	Continuous
Hydraulic conductivity	CONDU	Continuous
Groundwater content of sulphates	SULPHATE	Continuous
Groundwater content of residuum	RESID	Continuous
Anthropogenic		
Irrigation location	IRRIG_CL(1)	Categorical
Distance to canals	DIS_CAN	Continuous
Distance to roads	DIS_ROAD	Continuous
Distance to trains	DIS_TRAIN	Continuous
Location of canals	CANAL(1)	Categorical
Location of trains	TRAIN(1)	Categorical
Location of roads	ROAD(1)	Categorical

Table 2: Variables introduced by the model.

Factors	β	Wald	Exp(β)
HIGWTDEP	-0.09	1682.69	0.92
SULPHATE	-0.02	204.05	0.98
T4(1)	2.35	1714.70	10.50
T3(1)	2.04	1425.65	7.68
T2(1)	1.71	1342.72	5.53
T1(1)	0.57	151.92	1.77
GEOENDOR(1)	2.33	1372.01	10.26
DEPQ_80(1)	-0.64	309.93	0.53
IRRIG_CL(1)	0.48	269.61	1.61
PERME_10(1)	0.14	66.31	1.15
GRAWTMEA	0.33	167.38	1.39
Constant	-4.08	5579.57	0.02

Table 3: Confusion matrix.

		Predicted			Correct %
		0	1	total	
Observed	0	8823	3177	12000	73.5
	1	4477	7582	12059	62.9
	Total	13300	10759	24059	68.2