TECHNICAL UNIVERISITY OF CIVIL ENGINEERING BUCHAREST

DOCTORAL SCHOOL

SCIENTIFIC RESEARCH REPORT NO. 1

Scientific coordinator:
Univ. Prof. Dr. Eng. Iohan NEUNER

PhD Student:
Eng. Codruț-George BENEDIC

2016 - 2017 school year

THE DOCTORAL THESIS THEME:

THE PREDICTIVE MODELLING AND THE SPATIAL DATA VISUALIZATION IN THE RISK MANAGEMENT

THE SCIENTIFIC RESEARCH REPORT THEME:

THE CURRENT STAGE OF THE PREDICTIVE MODELLING PROBLEM

Contents

List of figures

Abstract

In this research report we have presented some aspects from different fields of activity, from witch it can be seen the evolution manner of the predictive modelling, both on international plan, and well in Romania.

Thus, the predictive modelling of one phenomenon presupposes above all, the influence factors analysis of that phenomenon and then it must create a mathematical model for the evolution of that phenomenon. For this, it is considering the relations between the variables involved in the model.

In the forestry field, the predictive modelling can be used both for the conservation of the wooded areas, and well for the restoration of the degraded forests. For this, it is estimated the evolution of the different species of trees. The environmental variables which can be considered are the sun exposure of the wooded areas, the minimum or maximum temperatures or the amount of rainfall.

The predictive modelling of forest fires follows among other the weather conditions modelling in case of fire, the prediction of fire behavior, the fire intervention plan, which implies the necessary resources estimation and the places where there are at one time, as well as the prediction of a possible fire extension. For this, it requires a high-performance geographic information systems, with an updated database, from which it is possible to extract at any time the information which is needed.

In the archeology field, the predictive modelling aims at the identifying of the areas where archaeological sites are located. For this, is done an analysis of the soil structure and the sediment deposition along the time. The relation between the location of an archaeological site and the geomorphological evolution is influenced by the ratio of solid and liquid flows from a watercourse. This leads to the accumulation of a different amount of sediment from time to time.

The modelling of a water flow from a river basin involves the basin dividing into distinct areas, depending of the relief, climatic conditions, soil structure and vegetation. These areas are analyzed separately, mathematical models are generated for each area and then these patterns unite in a model which shows the prediction of water flow for the entire river bazin.

In our country, the predictive modelling using the geographic information systems is at an early stage of development. This is due to the fact that the geographic information systems are not very developed and still the difficult access to the data needed to build the databases. On the risk management side, flood risk maps are made, based on the past information, with regard to the areas in which its occurred, but the predictive modelling of the risk also presupposes the estimation of the evolution of the floods, using the mathematical statistics.

1. Introduction to the predictive modelling terminology

The predictive modelling it is the given name to a collection of mathematical techniques, which have as purpose the discovering of a mathematical relation between a target variable, an answer or a “dependent” variabile and different estimates of that variable or between “independent” variables. According to [5], by this it follows the measuring of the future values of these estimates and the introducing of them into a mathematical relation in order to predict the future values of the target variable.

These mathematical relation can never be perfectly put into practice. Therefore, it is desirable to give some measures of uncertainty for estimates. Ussualy, it is used a prediction interval, who it has a 95% assumed confidence interval.

1.1. The regression - method of correlation between variables

1.1.1. The simple linear regression

In the case of the simple linear regression, the measurements errors are independent. There are not errors in the measuring of the dependent variable values. The relational model, on which the regression depends, it must be correctly identified. The values variation of the Y variable it is the same for all values of the X variable. The values of the Y variable follow a normal distribution, as it can be seen within [2].

The resulting variable is the Y variable, and the predictor variable is the X variable.

An example in this sence it is the transformation in Fahrenheit degrees, when it is known the value in Celsius degrees, or inversely.

$$F = 32 + 1.8\,^{o}\,C \quad (1.1)$$

This expression describe a perfect line.

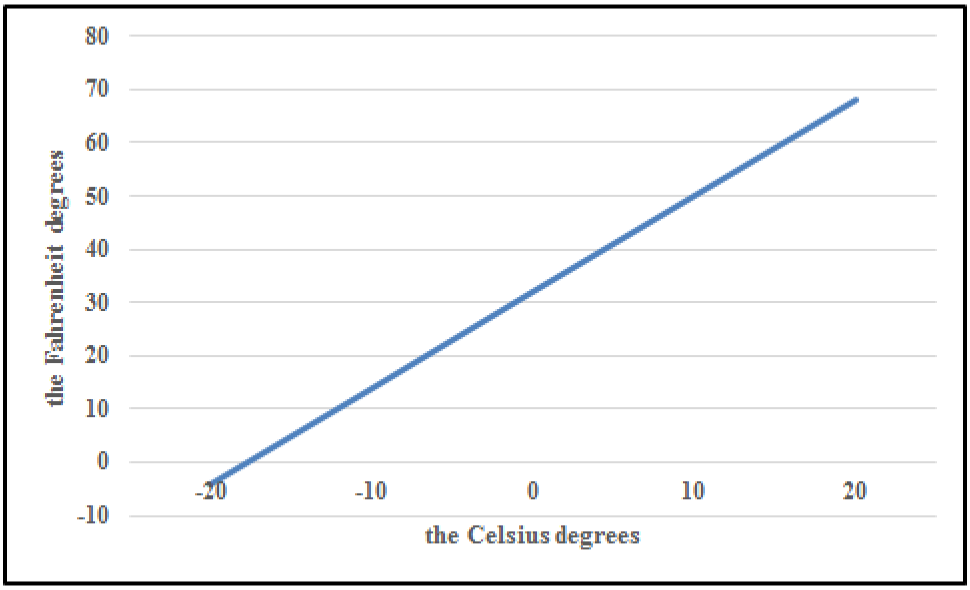


Figure no. 1. The straight line equation in the case of the simple linear regression for expressing the correspondence between the Celsius degrees and the Fahrenheit degrees

The general expression of the regression straight line it is:

$$Y = a + b \cdot X \quad (1.2)$$

The predict equation is:

$$\tilde{Y} = a + b \cdot X \quad (1.3)\text{, where:}$$

a - represents *the intercept*;

b - represents *the coefficient* (*the slope*) *of the straight line*;

X - represents *the predictor* (*the estimated variable*).

In the case of the linear regression equation, a and b are constant, and the values which are modified it is those of the X and Y variables, as it is shown within [2].

The "b" slope represents the quantity with which the values of the "Y" dependent variable there are modified, in the case in which the value of the "X" independent variable it is modified with one unit.

The "a" intercept represents the value of the "Y" dependent variable, when the value of the "X" independent variable it is equal with zero.

$$a = \bar{Y} - b \cdot \bar{X} \quad (1.4)$$

1.1.2. The regression with multiple inputs

When there are two or more estimates for one variable, it can appear additional problems. This thing it is known as multicollinearity.

For the exemplification of the regression with multiple inputs, it is presented the mode in which the shops managers from one commercial chain they allocate the advertising costs on radio and television spots.

Y represents the sales, X_1 represents the radio advertising, and X_2 represents the T.V. advertising.

The sales in comparation with the money allocation for T.V. advertising, for radio advertising and for both on radio and on T.V. advertising.

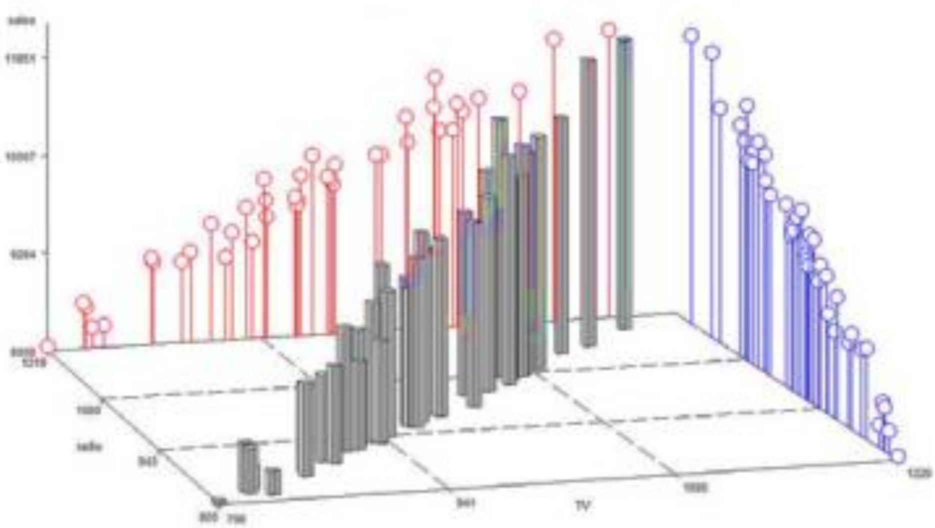


Figure no. 2. The sales in comparation with the money allocation for advertising (fictitious data) (*Source:* Dickey, Alan David, 2012)

Together with the 3D graph, in **the figure no. 2** there are shown two 2D graphs, too, representing: the sales in comparation with the money allocation for T.V. advertising, in the back part, respectively the sales in comparation with the money allocation for radio advertising, in the right part, as it shown in [5].

It can observe that the 3D prisms have the bases placed nearly on the line in which the costs for the T.V. advertising are equal of those for the radio advertising. This it means that the shops managers, that receive different total allocations for advertising, they tend to divide them equally between the radio advertising and the T.V. advertising. In other words, X_1 is approximately equal with X_2, where X_1 and X_2 represent the spent sums for the radio and T.V. advertising. From here it results that if $X_1 = X_2$ and the sales are equal with $\beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2$, we can replace $X_1 = X_2$ and to obtain the sales $\beta_0 + 0 \cdot X_1 + (\beta_1 + \beta_2) \cdot X_2$ sau $\beta_0 + (\beta_1 + \beta_2) \cdot X_1 + 0 \cdot X_2$ or $\beta_0 +$

$5 \cdot X_1 + (\beta_1 + \beta_2 - 5) \cdot X_2$ or any of the infinite number of good models of equations for sales, having the form $\beta_0 + C \cdot X_1 + (\beta_1 + \beta_2 - C) \cdot X_2$ and we can choose any value for the "C" constant (for example, $C = \beta_1 + \beta_2$ or $C = 5$).

1.1.3. The regression with variable indicators

In **the figure no. 3** it is shown on graph with the number of accidents with deers, in which the cars from North of Carolina there are involved.

The evolution of the accidents with deers

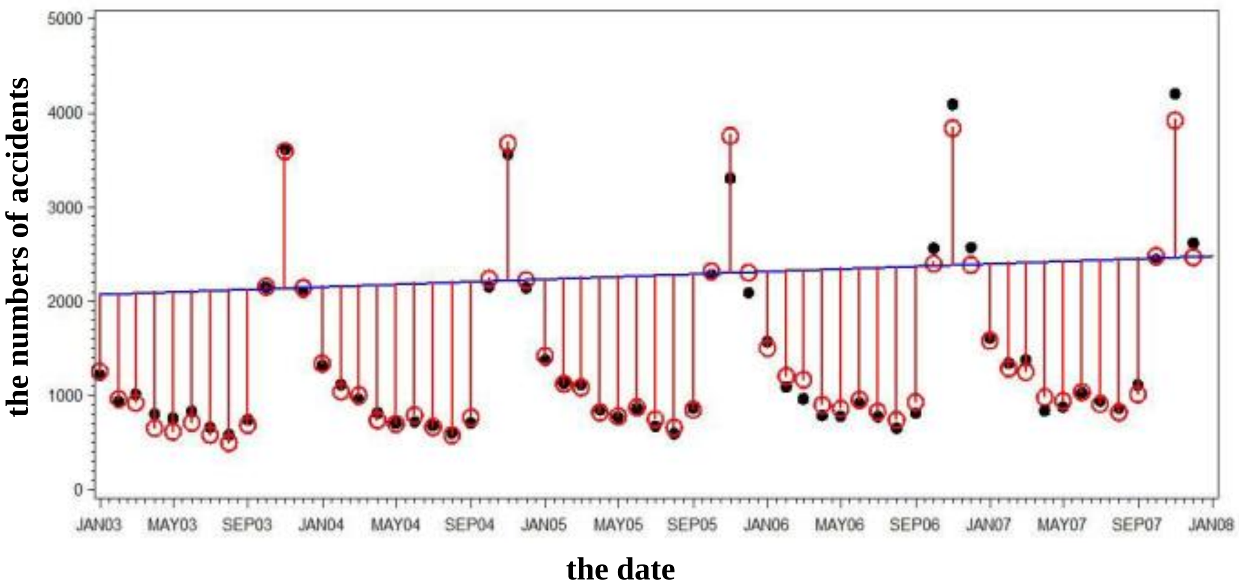


Figure no. 3. The monthly car accidents in North Carolina, involving deers.
The numbers are points and the predictions are circles.

(*Source:* Dickey, Alan David, 2012)

The data have been taken over the time and it can be observed a small tendency of increase of the accidents number with deers. This thing can be because of the increased speed of movement of the cars, of the increasing of the deers number or both. The data are monthly and it can be observed a model which is repeated each month.

From the graph shown in **the figure no. 3** it can be observed high values of the accidents numbers with deers in the October month, when it is the period of the deers mating.

According to [5], this effect can be modelated using the "fictitious" indicators or variables. For example, it can be done a NOV variable, which takes the 1 value in the November month and the 0 value in the other months of the year.

A model having the form $Accidente = 100 + 20 \cdot NOV$ it estimates that it could occur 120 accidents for the November month and 100 accidents for all the other months of the year.

Another model, $Accidente = 100 + 3 \cdot t + 20 \cdot NOV$, it estimates that the accidents numbers increase linearly in the "t" period with the November values, which are situated with 20 units over a line, considered as basis line. For each month the indicator variables can be included in model, allowing of the each values to deviate from the line. If it is used 12 indicator variables, one for each month, the basis line will have an arbitrary height. In the case of the using of 12 variables, the coefficients are not uniquely determined, what it is a problem. A solution it is the elimination of a monthly indicator, for example the one from the December month, such that the line predicts the accidents in the December month, and the others 11 indicators have the coefficients which show the deviations from the basis line of the December month in comparation with the other months.

In **the figure no. 3** it is shown a graph, in which the data are represented by points. The predictions are circles, and the line passes through the predicted values for the accidents in the December month. The high values in each year its are sunt the ones from the November month (the period mating), from which the data in the December month are easy to identify. It can be observed, also, that the deviations in the October month, in comparation with the basis line for the December month, there are close to zero for the data taken in consideration in this study.

2. The spatial predictive models use for the territorial planning of the forested areas

The territorial planning of the forests areas it tries to resolve the different problems, like the existing forests conservation and the restoration of the areas with degraded forests. For an accurate planning it is necessarly to know the forest potential of the territory. This it is cartographically expressed with the help of the suitable models. Such models can be generated through metods of spatial analysis, with reference to a set of variables which influences the environment. This variables indicate the presence or the absence of a certain type of forest in a certain area, as it is presented in [8].

In an area such as the Iberian Peninsula where the forests was been în mod constantly cut down, over the centuries, two of the main objectives of the any plan of forest zoning there are the decreasing of the forests fragmentation and the conservation of their biodiversity.

For the making of these objectives there are necessarly the information from land of a good quality, which help at the knowing of the territory, as well as actions of orientation through a three-dimensional view of the study area.

The suitable metods are rarely used in the forest planning and their utility it is demonstrated in the folowing.

A suitable model for the forest planning it is a map in which each pixel has associated a specific value of the reflectance, for a taken usage. The suitable models can be generated through various techniques, such as the logistic regression or the multiple linear regression.

According to [17], through the logistic regression "it is modeled the relation between a lot of independent variables and a dependent variable".

All the techniques it requires a map with the vegetation (dependent variable) and a set of variables of the environmental (climatic, topographical, litological, and so on), which have potential of influence on the vegetation distribution. The metods assume the determination of the relations between the environmental variables and the spatial distribution of the vegetation. In current mode, each type of vegetation will correspond to a different model. This is a consequence of the different requirements of the vegetation types on the environmental.

The "suitable" area in a model it is the area which must to reflect the potential area for the considered vegetation type, as it is specificed in [8]. Usually, the areas with current distribution there are fewer than in the areas with potential distribution. This is due to the fact that the forests was been identified starting from the areas in which there were in the past.

Knowing the potential distribution area, we can to obtain valuable data for the actions of reconstitution, as well as to delimit the areas in which such actions are a priority.

2.1. The study area, the materials and the methods used

2.1.1. The study area

Extremadura is one of the 17 autonomous provinces in Spain (**the figure no. 4**). Having a surface of 41.680 km^2, it have a temperate mediterranean climate because of the sea proximity and of the penetration of the ocean fronts from the west. In this region of Spain the 3 species of trees predominate, namely: the rock oak (on a surface of 19.600 km^2), the cork oak (on a surface of 3.140 km^2) and the Pyrenees - Eiche oak (on a surface of 2.160 km^2).

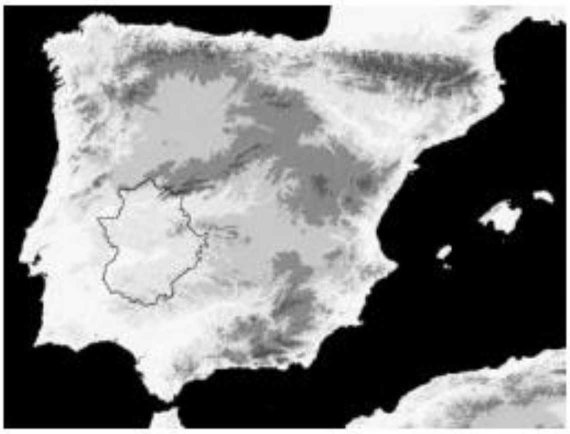


Figure no. 4. The geographical localization of the Extremadura province in the Iberian Peninsula (*Source:* Felicísimo, M. Ángel, 2003)

2.1.2. The dependent variable: the current distribution of the Pyrenees - Eiche oaks forests

The map with the current distribution of the Pyrenees - Eiche oak, shown in **the figure no. 5**, it was been generated having as its basis "The forests map from Spain". This map of the forests in Spain it is made by the Nature Conservation General Direction, belonging to the Ministry of the Environment.

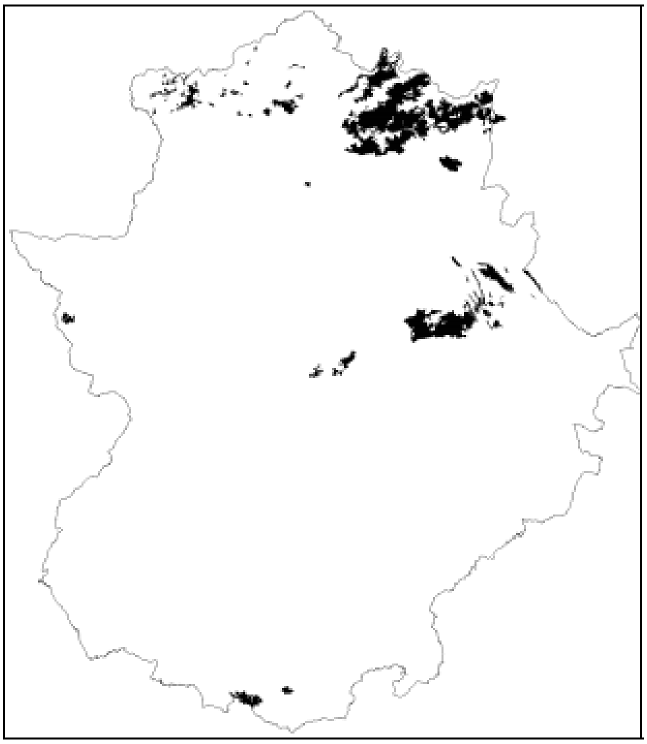


Figure no. 5. The current distribution of the Pyrenean - Eiche oak in the Extramadura province, according to "The forests map from Spain" (the areas with black, approximately 2.160 km^2)

(*Source:* Felicísimo, M. Ángel, 2003)

2.1.3. The independent variables: the digital terrain models

According with [8], the environmental variables, with potential of influence on the areas with Pyrenean - Eiche oak forests, there were the following:

- ✓ the elevation: the digital elevation model (DEM) was been making using the Delaunay triangulation algorithm, through the digitization of the hispographic curves, followed by a transformation in a grid with regular structure, with the cell size of 100 m;

According with [4], the hipsography represents "a branch of the geomorfology, which deals with the description of the relief by altitude".

- ✓ the potential exposure at sun: the models were been done by simulation, starting from a digital elevation model (DEM), analizând the land shading like a function of the sun trajectory for standard time periods. The result is a estimator of the time period in which each point on the land surface it receives direct solar radiation, with a temporal resolution of 20 minutes and a spatial resolution of 100 m;

The temporal resolution, as it is specified in [18], it represents "the time interval in the limits of which a remote sensing system was recorded pictures. This represents the time moment, well specified (the year, the month, the day, the hour), at which was registered a picture".

According to the same source, the spatial resolution is defined as being "the linear dimension of the smallest object in land, represented in a picture. It can been considered, also, as the width of the line which separates two small dimensions neighboring objects in a picture, such as a car and a building. At the digital pictures, the spatial resolution corresponds with the dimension in land of the side of a pixel, the smallest element which forms the respective picture".

- ✓ the maximum and minimum averages of the temperatures: interpolated from a total of 140 weather stations, through spline functions, represented by flat, thin curves, with a spatial resolution of 500 m;
- ✓ the total quantity of rainfall over a period of 3 months: interpolated from a total of 276 weather stations, through spline functions, represented by flat, thin curves, with a spatial resolution of 500 m.

Note: The last two data sets were been obtained from the National Institute of Meteorology in Spain.

2.1.4. The methods: the multiple logistic regression (MLR)

The multiple logistic regression was used like a method of prediction, for to generate the probability models in different domains, like: epidemiology, geological prospecting, silviculture or the conservation of the wild animals. The multiple logistic regression (MLR) is suitable because the dependent variable is dichotomous (present / absent) and the model accepts the non-Gaussian independent variables. In the end, it results values which varies from 0 to 1. From this cause it is recommended to generate a suitable probabilistic model.

The introduction of a spatial component inside of the multiple logistic regression for generating cartographic models, it started to be done recently. Usually, the multiple logistic regression method is integrated inside of the geographical information systems. An example in this sense it is the generation of a model of the distribution of the "*Carex curvula*" species of plants in Swiss Alps, using the "ArcGIS" application, developed by the "ESRI Inc." company. A similar study was made on the aquatic vegetatiton, using the "GRASS GIS" application, developed by the "US Army Construction Engineering Research Laboratory".

In the end it were been generated the logistic models for the Cantabria region in Spain, using the "ArcGIS" application. The procedure and the statistical bases are detailed here. The logistic model shows the suitable character, namely the taking in consideration of the values of the "n" explanatory variables, which verify the next relation:

$$P(i) = \frac{1}{1 + e^{-\left(b(0)+b(1)\cdot x(1)+\cdots+b(n)\cdot x(n)\right)}} \quad (2.1)$$

where $P(i)$ represents the suitable (good) value, $x(1) \div x(n)$ the values of the environmental variables, and $b(1) \div b(n)$ the coefficients of the environmental variables. The results for each pixel vary between 0 (incompatible) and 1 (ideal).

The regression coefficients were been determinated by a stratified sampling in the presence / absence areas of the Pyrenean - Eiche oaks forests, as it is specificed within [8]. The procedure for this is the following:

1. Its were made two randomly samplings on each areas with or without forests. The first sampling was been used for generating the model (forming sample) and the second for testing its performance (testing sample). Both samples have an equal number of positive and negative cases.

2. The logistic regression was made using a commercial software for statistics, through the "*stepwise*" method, through the maximization of the probability.
3. The the suitability degree of the model it was estimated through the comparison of the results of the forming sample with those of the testing sample, for some limit values, measuring the area from under the "ROC" curve (the characteristic curve of the receiver operating).

According to [19], the "Receiver Operating Characteristic" ("ROC"), it is a bidimensional curve, in which on the Y axis we have the sensitivity and on the X axis we have the specificity. This curve we help us to measure the efficiency of a model. When the area from under curve it is bigger (the maximum is 1), the model is better.

For the study of the suitability degree of a model, it must be followed the relation between prediction and answer.

The sensitivity represents the number of the cases with positive predictions divided at the number of cases with negative answers.

The specificity represents the number of cases with negative predictions divided at the number of cases with negative answers.

By classification it is obtained the probability of belonging to a certain class, according to [14]. In the logistic regression, the predicted values are the probabilities of belonging to a certain class. The discrimination limit it is the limit imposed of the estimated probabilities of assigning of the observations for every class.

In the end it was applied the equation of the logistic regression for entire the territory, for generating the probabilistic model. The results was compared with the existing data on the vegetation maps.

2.2. The results of the regression method

2.2.1. The evaluation of the logistics model

The capacity of the predictive model (its accuracy) it can be evaluated as the percent of the cases of correct classification, both for the presence, and for the absence of the forests from the study area. The usual method is to consider the area from under the "ROC" curve („AUC" - "Area Under the ROC curve"), calculated starting from the sensitive or specific values. "AUC" takes a maximum value equal with 1 for a model with a perfect suitability. In our case, the value of the "AUC" indicator it includes an area of 0,973, what it represents an accuracy very good of the model.

2.2.2. The suitability of the model for the Pyrenees - Eiche oak

The last stage in order to make the model, it is the generation of a map, by applying the logistic model for the entire territory. The results are shown in **the figure no. 6**. The values was been classified in 5 suitable classes, in order to facilitate the interpretation.

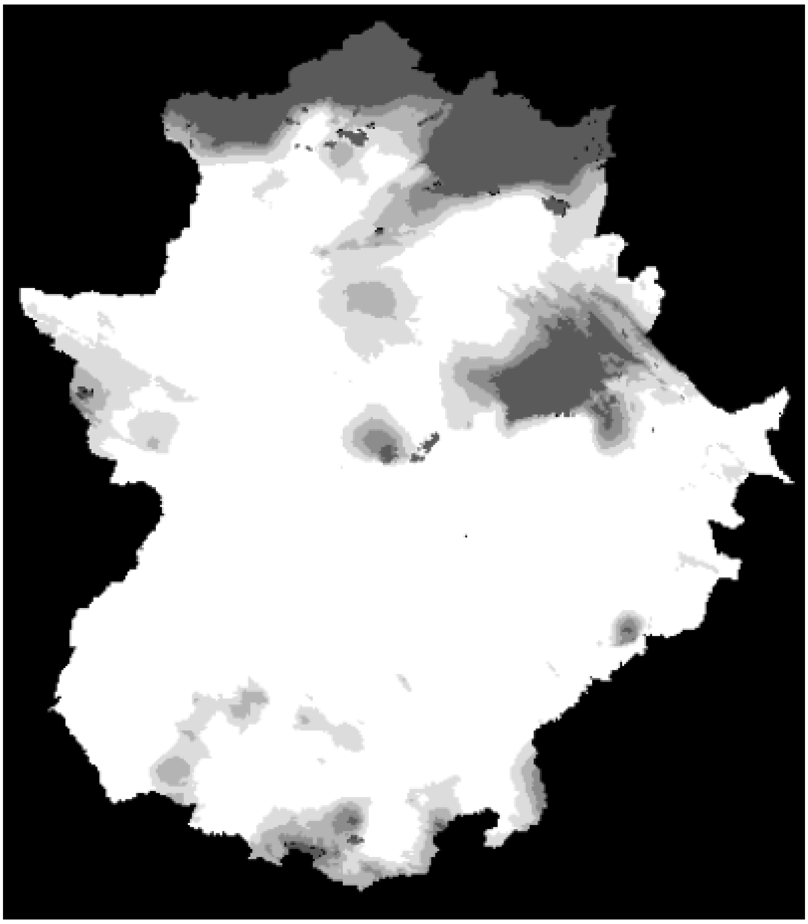


Figure no. 6. The application of the logistic model for the study area. The values from legenda are the follows: white: 0,00 ÷ 0,10 (incompatible); light gray: 0,11 ÷ 0,25; medium gray: 0,26 ÷ 0,50; dark gray: 0,51 ÷ 0,75; very dark gray: 0,76 ÷ 1,00. The area from outside of the study area it is shown with black.

(*Source:* Felicísimo, M. Ángel, 2003)

3. The application of the GIS technology in the forests fires management from Canada

The management agencies against the fires from Canada forests it were became more dependent of the decision taking systems, based on planning and the decision taking in real time, as it is specificed within [9]. Especially, the integration of the Geographical Information Systems it were became more and more important for the realistic modelling of the fires danger of the forests and for the prediction of the evolution mode of their. Some operational systems were been implement with successfully and efficiently. The "Intelligent Fire Management Information System" ("IFMIS"), a personal computer, based on a support of taking decisions, developed by the "Forestry Canada", for the operational management of the forest fires, it is an exception. Following the "IFMIS" design and principles, "Forestry Canada" was designed the next generation of the "Fire Management Information System" ("FMIS"). By the GIS technology usage, commercial available, the prototype of the "FMIS" system was been ready for testing in the 1993 year.

3.1. The integration of the vector and raster data into the fires management system

The operational management of the forest fires it is a spatial problem. The used data for the fires management are both vector data (point, line and polygon), and raster data (structures, networks) in nature. According to [9], the spatial techniques which are used in the management of the forest fires there are the continuous interpolation of the surface, the temporal sequencing, the Euclidean distance euclidiană and the analysis of the damages on the destroyed surfaces of forests fires.

The models of forest fires, developed over the years by the researchers from "Forestry Canada", there was incorporated in the prototype which uses specific functions of the GIS systems and application programming interfaces ("API").

This applications uses a basic map, topographic data, data concerning the forest inventory, data concerning the forest fuels and meteorological data in real time, integrated.

The implemented scenarios within the management system of the fires there taken in consideration a series of requirements concerning the training planning for the forest fires, which include: the weather conditions modelling in the case of the fires, the combustible type modelling, the prediction of the fire behavior, the forest fires evaluation, as well as the resources deployment for the fires extinguishing.

3.1.1. The modelling of the weather conditions in case of the fires

The "Canadian Forest Fire Weather System" ("FWI"), developed by the "Forestry Canada", it is used in Canada, Alaska and other countries from the world, for modelling and estimating the weather conditions in the case of the fires. The "FWI" system estimates the humidity conditionsof the forest fuels. The weather conditions include the temperature, the relative humidity, the wind speed of over 10 $m/_{s}$ and the rainfall.

The entrance data are obtained by a network of the weather stations for the forest fires, using either manual fire stations, or automatic fire stations. This continuous monitoring (following) of the weather conditions in the time of the fires perriod, it leads at a high volume of data which must to be spatially extrapolate for the wooded area, at daily or hourly intervals.

The weather stations of fire represent test points (vectors), on a continuous surface, with many layers. Usually, the weather stations are placed at 50 kilometers away. Using the function of modelling of the GIS systems, the entrance data of the test points there are extrapolated over a regular grid. For this thing it is used the most suitable tehniques of interpolation. After there were made the layers for the entrance data, using the GIS application, it can be determinated the indices of combustible humidity and of fire behavior. For this it is applied the "FWI" models for every cell. As it is specified within [9], the modelling is finished through "the overlaping" of the each layer which contains the entrance data, resulting 6 new layers, which have as attribute the danger situation due of the forest fires.

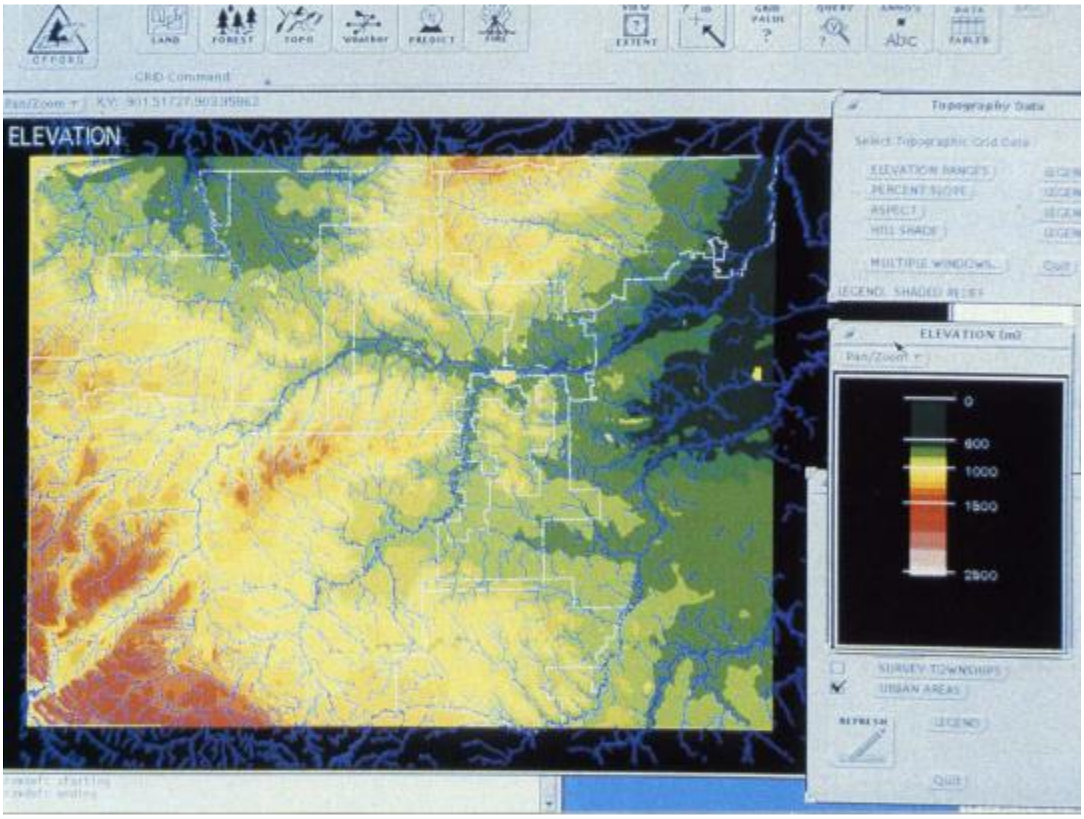


Figure no. 7. The land information, which are compiled using a digital elevation model (DEM) and there are used as inputs to a fire behavior prediction model

(*Source:* Lee, Bryan S. și David J. Buckley, iunie 1992)

3.1.2. The behavior prediction in case of fire

The fire behavior prediction is quantitatively, spatially and temporary. According with [9], using the information about the weather conditions, combustibles and those related to the land as entrances, its can make quatitative estimations of the fire evolution, the combustible consumption or related to the potential of fire expansion.

The prototype incorporates empirically developed equations, from the “Canadian Forest Fire Behavior Prediction System” (“FBP”). The “FBP” system estimates the “rate of spread” (“ROS”) of the fire for different combustibles, using entrance data concerning the weather conditions, the type of combustible and data related to the land. The additional output data of the “FBP” system, there include the “head fire intensity” (“HFI”), the “crown fraction burned” (“CFB”) and the “total fuel consumed” (“TFC”).

The cartographic model shown in **the figure no. 8** demonstrates how every layer of data from the component of the Geographical Informational System (GIS) it is processed for making secondary and tertiary maps.

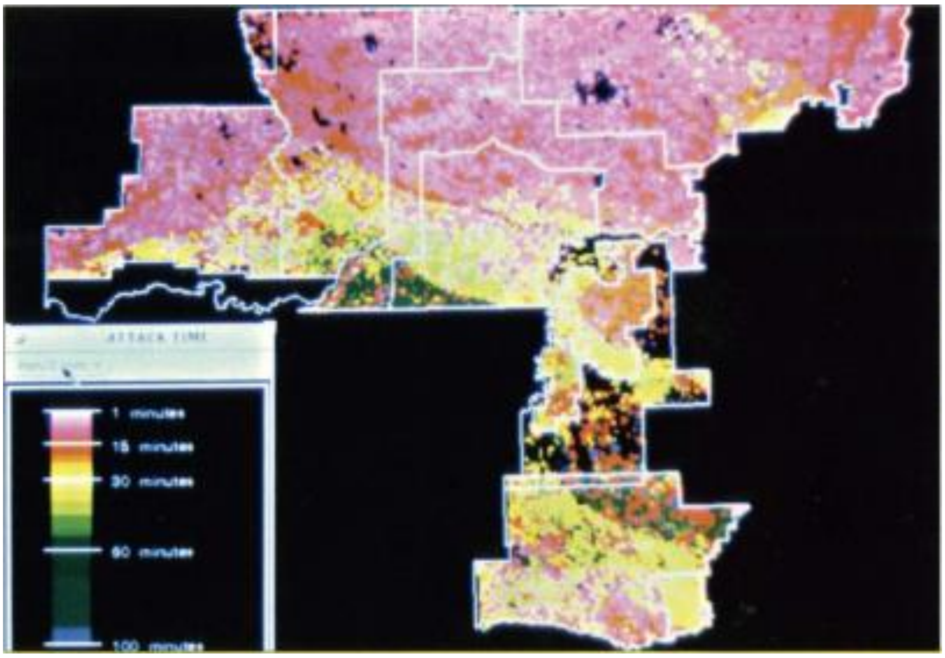


Figure no. 8. The initial firefighting times in the “Whitecourt” forest
(These times are based of the expansion rate calculated for fires and the extent of the fires.)

(*Source:* Lee, Bryan S. și David J. Buckley, iunie 1992)

Using the “FBP” system together with the database concerning the environmental conditions within the forests and the interpolation of the “FWI” values (those regarding the weather conditions), through the “IFMIS” system it can be obtained the maps of the potential behavior at fire, such as “ROS”, “HFI”, “CFB” or “TFC”.

Observation: The significance of the abbreviatings it was presented above.

3.1.3. The training plan in case of fire

In the west part of Canada, some management agencies of the forest fires its were adopted the training plans in the case of the forest fires. This thing it was made with the purpose to determine the daily requirements of extinguishment of the fires. The training plan in case of fire it is the assurance process with suitable resources for the fires extinguishment, for to deal of the daily anticipated fires. The training plan in case of fire it is based on the philosophy of the early detection of the forest fires and on the fast action in the case of them starting. For to be realized this objective, for all the fires it must act from the initial phase, before to reach in a critical situation.

Using a such strategy, the prototype can evaluate the efficiency of the previously located resources in a certain region of a forest. This evaluation can be done daily or from hour to hour.

For every cell, the prototype calculates the necessary time of reaction in the case of a possible fire. This criterion of the elapsed time, named reaction time, it can be displayed on the map. Having this information, the forest fires manager can determinate how many resources can reach to a cell at a certain time, on predeterminated bases. The elapsed time includes both the time return, and the displacement time towards a cell. The usage of a representation of raster type, instead of the vectorial representation, it allows the GIS application to calculate, constantly, the necessary time for to reach from basis at every cell. The number of necessarly resources and their optimal location in forest for the fast intervention in the case of to a possible fire, it is determinated through the integration of the spatial database of the Geographical Information System (GIS), with linear programming algorithms elaborated by the “Forestry Canada”. As it is specified within [9], using the “IFMIS” system and this approach, it is estimated that in the Alberta province there managed to save 4 milions dollars yearly, the necessary costs for the of the extinguishing forest fires.

3.1.4. The fire expansion modelling

Once the fire was detected, the GIS system can play a very important role. Using current and forecasted weather data, there can make estimations concerning the increasing of the fire intensity. This thing it is possible only through the usage of a raster data structure, based on which there can obtained a potential surface of cost, “ROS”. By using the temporal sequencing, it can estimated the increasing of the intensity and the expansion of the fire. This information can be used by the fire manager, for to determinate the optimal strategies for extinguishing of the fires. According to [9], this strategies consider the following: the number of crews required to be sent, the number of tanks with air required at the place of the fire, the places where it is necessarly to make ditchs for the prevention of the fire expansion, as well as the areas which are threatened by fire and there must protected.

4. The predictive modelling using in the archeology field

In the archeology field, the predictive modelling it is a technique which tries to predict the places where are located the archaeological sites. The predictive modelling in this domain there starts from the hypothesis that the localization of the archaeological remains it is not accidentally, but it is related to certain characteristics of the natural environmental, as it is specified within [13].

The reason for which there are wishing the making of a predictive model for archeology there are one of practical nature. When the time and the money not allow a complete archaeological research on an area, a predictive model can serve as a tool for the choosing of the areas where it is the probablest to be the archaeological phenomena of interest. The research will be focused, then, on this areas. Thus, there will be obtained a maximum profitability of the investment.

This situation is frequently met in the cultural resources management, when the archaeologists are constrained to decide what to investigate, taking in consideration the limited budgets and the short available time.

Also, the designation of the important areas by archaeological point of view, using the predictive modelling, there can be used for to convince the politicians and the developers to choose the areas with the lowest “archaeological risk” for their plans.

4.1. The predictive modelling of the archeological sites in the Tricastin - Valdaine region (the middle valley of the Rhône, France)

4.1.1. The localization of the study area

The middle valley of the Rhône river it is located at the border of the Mediterranean, Central European and Alps mountains climatic zones. In the area of the middle valley of the Rhône it is the Tricastin - Valdaine region. This region have a surface of 1086 km^2 and it is located in the eastern part of the Rhône river, as it can be seen in **the figure no. 9**.

Figure no. 9. The localization of the study area in France

(*Source:* Verhagen, Jacobus Wilhelmus Hermanus Philippus, 2007)

As it is shown within [13], the area is composed of two distinct regions. In the northern part it is the Valdaine basin. This includes the valleys of the Roubion and Jabron rivers, surrounded by the prealpine hills, dating from the Tertiary. The most important city in this region it is Montélimar. To the south, the basin of the Tricastin river forms a large transition area between the bed of the Rhône river, the lateral alluvial fissures which date from Holocene and the prealpine hills which date from the Tertiary. The most important cities in this region there are Pierrelatte, St-Paul-Trois-Châteaux and Bollène.

Because of its location at a climatic and geological border zone, both the vegetation growth, and the geomorphological processes in the middle valley of the Rhône river there are very sensitive at the climate change and at the changes produced by man. In this area there was occured erosion and sedimentation phenomena since the beginning of the Holocene. Because of this change of the landscape, many archaeological remains are buried under the current surface, especially in the alluvial zone of the Rhône river.

The purpose of this study it is to show that the usage of the results of the traditional field research, there will not detect the buried archaeological sites. Also, the usage of this results for the relation analisys between the localization of an archaeological site and the landscape characteristics, there can lead both at an incorrect representation of the sites distribution in comparison with the landscape units, and of the volume of sediments existing in an archaeological site. For this there was made the qualitative predictive maps of the area and there was made a quantitative extrapolation of the sites density where there are buried the most sediments.

4.1.2. The predictive model

For the most studies of predictive modelling, the relation between the sites location and one or more landscape characteristics there are deduced through the overlaping in a GIS application of the known locations of the archaeological sites on the available cartographic background. After this overlaping, there are making a quantitative analisys of the observed distribution model, known as the name of inductive modelling. According with [13], in the most cases, this analisys are made assuming that the known sit sample is representatively for the whole population.

By this research method there are determining how many archaeological sites will be uncovered. It is evidently that the buried sites will not be uncovered through by the going on the land. The bored and digging ditchs are pretty expensive forms of research, which are not available, ussualy, for amateurs. Even the professional archaeologists will not use this forms of study than if is a clear necessity. In practice, this thing means that, in majority of the archaeological databases the number of buried sites will be underestimated.

Moreover, when there are not knowing the dimension of the researched area, there are not knowing where there are archaeological sites. This thing is important for the statistical analisys.

Also, the researched surface, so the archaeological site sample, there can be not representative for the entire study area. This can be a consequence of the difficult access in the land, for example because of the steep slopes or because of the research only a certain aspects.

For the Tricastin - Valdaine region there are detailed information about the buried sites, as well as a map of the researched areas. Nevertheless, the study can not take in consideration the differentiated visibility of the archaeological sites, existing under the different land use forms.

For to have information about of the sedimentation and erosion of the area, the landscape must be interpreted, taking in consideration its geomorphological and pedogenetical history.

According with [4], ***the geomorphology*** represents a science "which are studying the genesis (the form mode) and the landforms characteristics, their grouping and spreading mode on the Earth surface".

According to the same source, ***the pedogenesis*** is “part of the pedology which deals with the forming and evolution study of the soils. The pedology represents the science which deals with the genesis study, of the soils evolution and distribution, as well as with the knowing of their physical, chemical and biological composition, for the determination their fertility degree”.

For to obtain the used map as taxonomic basic layer, the different geological and sol maps there was been digitized and combined into a GIS.

According with [4], ***the taxonomy*** represents “the science of the classification laws”. The ***taxonomy*** term have, also, the meaning of “study of a groups of animals or plants, by the point of view of the species classification and description (taxonomic = with classification character)”.

The only basic maps available for the entire region there was been the France geological maps, at the 1:50.000 scale.

This maps was been edited where there was been necessary. The portions of the geological map was been assigned to one of those 18 distinct taxonomic category.

As it is specified within [13], the basic geological information was been updated with other available information, concerning the geological and pedological conditions from the area. This information came from 3 sources, namely:

- ✓ a classified picture of the Tricastin region, obtained from satellites;
- ✓ a delimitation of the main pedological and sedimentary units, obtained during the work in the land, in the Valdaine and Tricastin regions;
- ✓ the existing pedological maps of the area, at the 1:25.000 scale.

The obtained picture from the satellite sensors there are offering the detalied information about the location of the old beds of the river, the alluvial platforms, the terraces and basins in the Tricastin region. The data on the land there are providing additional information about the location of the colluvium and alluvial deposits in the Valdaine region (of the colluvium, of the river deposits and of the cuvettes). The pedological maps were been updated for to be able to observe the real extent of the colluvium and alluvial deposits on the Roubion and Jabron valleys, for to find the location of the stable Pleistocene terraces and the alluvial platforms in the Valdaine region and for to be able to know the distribution of the colluvium and the cuvettes in the Tricastin region.

According with [4], ***the terrace*** represents “a landform with appearance of step, coming from the Earth bark lifting or as result of the waters erosion, along a river”.

The cuvette represents “a closed depression, like a basin, permanently or temporary covered with water”, according with [4].

According with [4], ***the colluvium*** represents "a detritic material (detritic = which comes from another rock, through an action of granulation due to the external agents), accumulated at the slopes base and coming from the rocks disintegration and the displacement of their fragments from the mountains slopes, under the gravitation action".

The additional information were been used for to update the reclassified geological maps. So, the final taxonomic map is a mosaic of several maps, with varied scale and precision.

The sample from the archaeological site there are consisting of data which come from several sources. The collected data were been obtained by digging of the ditchs and by archaeological researchs and excavations for the high speed railway connection which connects Lyon with Marseille, made between the 1995 and 1998 years. The second method of working in the land there was having as result the detection of the buried archaeological sites. This data form the most important part of the archaeological database. The data for other archaeological sites there were been obtained from literature, by reading of the archaeological map and through the data came from the regional archaeological service for the Rhône - Alps region, as it is specified within [13].

The relation between the dynamics of the site location and the geomorphological evolution it is understood in terms of river systems. In the hydrological Rhône system, the river regime is changed according to the variations in the report of the liquid and solid flows. The branched or sinuous channels systems are not conducive for the sediments deposition, with the exception of the seasonal periods (in the period after the Paleolithic, in the middle period of the early Neolithic and in the first part of the Iron Age). The flat or convex meadows of the sinuous rivers courses there are favorable for the sediments deposition, with the exception of the periods when there are occuring floods (in the late Neolithic period and in the middle period of the Iron Age). The most favorable situation for the sediments deposition it is a sinuous river system, together with the rooting of the river bed (in the late Mesolithic period, at the middle of the Neolithic and in the Bronze Age). This thing leads to the stabilization of the meadow for a long time. It was observed that there are a cyclical alternation between the water courses stability and the expansion of the floodable areas occupation, on the one hand, and water courses instability and the depopulation of the same areas, on the other hand.

Using the mentioned sources, there was made a comprehensive archaeological base. This database contains the location of 510 prehistorical archaeological sites, buried in their natural environment. For every site there was researched if the site was uncovered at surface or if it was covered with sediments. The number of necessary sites for a reliable analysis of the archaeological site location in relation to the taxonomic map it is about 40. For the early Neolithic period this requirement are not accomplished. According with [13], in the other periods there are enough sites for to be able to make the analysis of the archaeological sites location.

For to see if the non-random sampling influence the analysis of the site location from the study area, the available sample of archaeological sites there was divided in a visible sample and a complete sample, which includes, also, the sites which are not visible, too.

In the case of the visible sites, there was been taken in consideration only the areas in which the sediments was been deposited on the ground. In the case of the complete sample there was been included both the areas in which the sediments was covered, and the areas in which the sediments was been deposited on the ground.

Due to the small number of involved locations, there was not been possible the making of a separated analysis only for the areas in which the sediments was been covered. However, the chances to discover a site in an area in which the sediments was been covered there are bigger than in the areas in which the sediments was been carried on the ground. Because the area in which the sediments was been covered there are representing a relatively small part of the total investigated area, the real importance of the covered archaeological sites can be bigger than that suggested by the predictive modelling.

As it is specified within [13], for the analysis of the relations between the archaeological site location and the taxonomic map, there was made 3 separate analysis.

4.1.2.1. The χ^2 test (The Pearson test)

For starting, there can use the χ^2 test, with the purpose to be seen if there are siqnificant differences between the location of a archaeological site and the units on the map. However, the χ^2 test not provides information related to the relative importance of the units from the map for the location of a archaeological site. The application of that test as the only statistical tool of predictive modelling there was criticized on several occasions.

Due to the limits of the χ^2 test, respectively the requirement to have 5 sites on a unit on the map, requirement that depends, in turn, of the sample dimension, the taxonomic map was been reclassified in 9 categories. Even so, in some cases the statistical requirements could not be fulfilled. In this cases there was been applied the "***Yates***" correction for to calculate the value for χ^2. However, it must take into account the fact which in the case of less than 40 observations, the results of the χ^2 test, even through the application of the "***Yates***" correction, which have a low confidence degree.

According with [16], the "***Yates***" correction, also known as the "**continuity corrected chi-square**", it is applied at the tables with 2 lines and 2 columns, when there are cells with less than 5 elements.

4.1.2.2. The ratio between the size of the archaeological site and of the researched area

The ratio between the size of the archaeological site (p_s) and of the researched area (p_a) it is a simple and direct mode for to examine the importance of certain categories of maps for the site location. For example, this ratio was used in Netherlands for making "The orientative map of the archaeological values". However, the value of this ratio do not provide information regarding to the relative weight of the maps categories depending on the size. This problem is the best showed through the taking in consideration of the zero case: a big unit, without archaeological sites, will be less important for the site location, than a small unit, without sites. In other words, a small unit, without sites, it is more significant by statistical point of view. However, by calculation of the ratio p_s/p_a it will be obtained the 0 value for both units. This thing means that for both units it will be attributed an equal importance.

For to show this effect, it is calculated a relative weight factor for every map unit. In the case of a 3 map units: "α", "β" and "γ", it is calculated the following weights:

$$A = a' \cdot b \cdot c/(a' \cdot b \cdot c + a \cdot b' \cdot c + a \cdot b \cdot c') \quad (4.1)$$

$$B = a \cdot b' \cdot c/(a' \cdot b \cdot c + a \cdot b' \cdot c + a \cdot b \cdot c') \quad (4.2)$$

$$C = a \cdot b \cdot c'/(a' \cdot b \cdot c + a \cdot b' \cdot c + a \cdot b \cdot c') \quad (4.3)$$

where:

„a", "b" and "c" represent the sizes of the researched areas on the "α", "β" and "γ" map units;

„a'", "b'" and "c'" represent the sizes of the sites on the "α", "β" and "γ" map units.

This thing is aritmethical equivalent with the division of every p_s/p_a value at the sume of the all p_s/p_a values, what it means that the obtained values for weights are normalized values. So, there do not resolve the problem of the relative weights.

Besides this, the calculation of the p_s/p_a ratios values there do not offer information about the statistical meaning of the observed model. The meaning of the model can be tested by the comparison of the weights values with those obtained through the simulation models for the archaeological sites localization. According with [13], this analysis assumes the making of maps with the random distribution of the sites, against with to be tested the real model.

4.1.2.3. The K_j parameter method

A more complex method for the evaluation of the categories of maps for the archaeological sites location it is the use of the K_j parameter. This parameter is thus calculated:

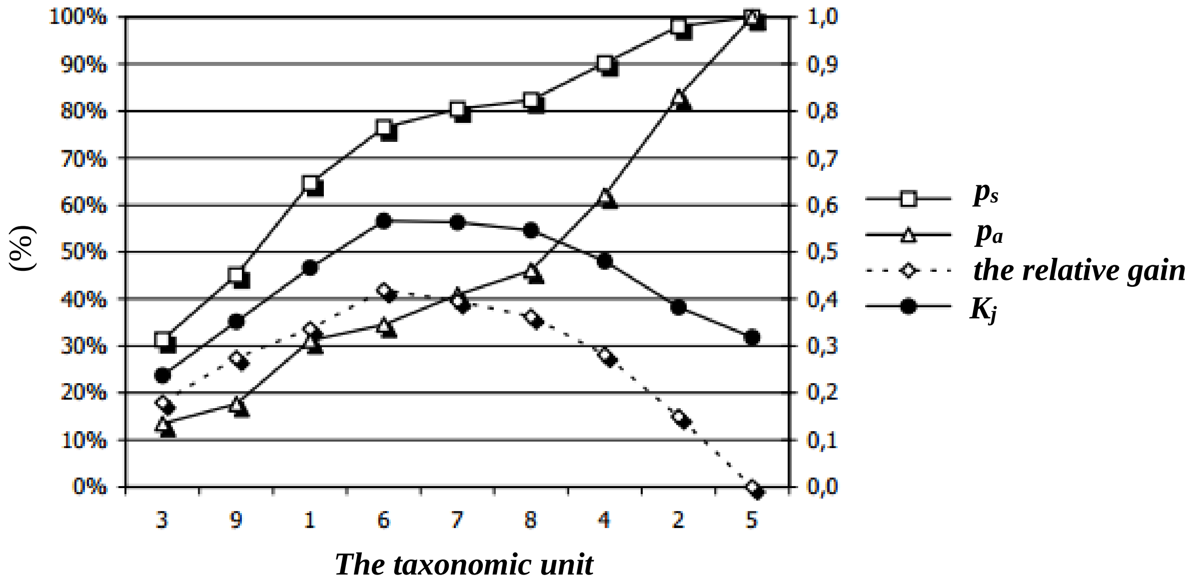
$$K_j = \sqrt{[p_s \cdot (p_s - p_a)]} \quad (4.4)$$

In the original equation, the $p_s - p_a$ difference is divided at p_w. The p_w value represents the proportion of the areas without the archaeological sites. This change is made only when it is used the estimated surfaces of the site, which it is not the case for this model. The K_j parameter is calculated for every map category. The category for which it is obtained the highest value of the K_j parameter it is the most successful. By the method of the successive iterations, K_j is recalculated, rezulting the following category in the model. As it is specified within [13], this process is repeating until the all categories are resulting.

The utility of the calculation method of the K_j parameter value it is that it takes in consideration că the relative importance of the observed sites density. Every time when it is calculated the K_j parameter, it is determined the relative gain, $p_s - p_a$, for to evaluate the model performance. A model with high predictive power will have high values of the gain. The model performance increases with the increase of the K_j value, after every executed interation. However, in certain situations the gain decreases, while the value for K_j continues to increase. This thing happens because the equation assigns a higher weight for the categories of maps which contain a higher number of archaeological sites. A gain of 40% can be obtained through a model which contains 50% of sites on a surface of 10%, but and through a model which contains 80% of the archaeological sites on a surface of 40%. In the second case, the model have better results, although the gains are the same. However, in order of managing of archaeologica resources, a model based on the values of the gains it is more useful when the total surface which must taken in consideration it is smaller.

In the analysed cases wihtin this study, some categories of maps contribute greatly at the increase of the gain and at the K_j parameter value, and other units have only a limited effect on the total gain, existing, also, units which lead to the great decrease of the model gain. In other words, it can be said which these groups have a positive (high), neutral (intermediate) or negative (low) predictive power. These groups can be easily identified by graphical representation of the model. An example is presented in **the figure no. 10**.

The surface and buried archaeological sites from Paleolithic / Mesolithic,
used for development of the predictive model, based on the K_j parameter values



The figure no. 10. The making of a model for the K_j parameter determination, for all the discovered and buried archaeological sites, from the Paleolithic and Mesolithic periods (The 3, 9 and 1 units has very big predictive power, the 7 and 8 units has intermediary predictive power, and the 4, 2 and 5 units has very low predictive power)

(*Source:* Verhagen, Jacobus Wilhelmus Hermanus Philippus, 2007)

5. The prediction of the water flow on the ungauged river courses

Currently there are not theories or equations for to estimate the flow at the scale of a hydrographic basin. The water flow equations are available at laboratory scale. The prediction supposes the transition from the laboratory equations, to those for the basin scale. A way of approach of this problem it is the division of the basin in small enough elements, for to can be applied the equations available at laboratory scale. As it is specified within [1], these equations are united at the end, in order to form a model for the entire hydrographic basin, for to be done the required flow predictions.

5.1. The fragmentation in hydrology

For to be done flow predictions in the ungauged basins, it was developed methods which are based on the notion of similarity. There is not a standard method of flow prediction in such basins. On the contrary, the existing flow models differ depending on the model structure, the involved parameters, as well as depending on the entrance data. Also, the models differ depending on the processes which its represent. Depending on the natural environments, the role of the different phenomena, such as the snow, the water flow or the transpiration process at plants, there may be different. Likewise, the factors which control these processes there can be different. Some differences between models are directly related to the different climate and the characteristics of the hydrographic basin.

The land surface is organized in hydrographic basins, with a small water exchange between them. A hydrographic basin can be studied successfully, not influencing the studying of the other hydrographic basins. In the hydrology field it is studied different basins, each of these having different answer characteristics. This is a fundamental difference, which appears in the hydrology field.

All this factors contribute to the fragmentation of the hydrology at different levels.

The processes: Many a time, the processes in hydrology are treated separately. The annual yield of water it is studied, usually, independently of the flows level from a interest basin. Also, the floods are often studying independently of the knowing of the seasonal flow models from the hydrographic basins. There is asked the question if it is a deep relation between these processes and if it is needed of a simultaneous treatment of them, at different time scales.

The places: A certain group of researchers in the hydrology field there has the tendency to analyze only certain basins. Because of this there was understand the flow processes only for the restricted areas. As a result, it is difficult the assigining of the another researched area for that group. Also, the water flow models are most often adapted at a certain hydrographic basin, which have a certain structure of the model and parameters, different of those of other hydrographic basins.

The scales: The research was made at different scales. Because of this, the reunion of the research results at the same scale it was made with difficulty. When it is verified the infiltration equations at laboratory scale for the basin scale, there must estimate the quantity of rainfall in the area of the respective hydrographic basin, during the analyzed period.

According with [1], for the making of a water flow model, it is needed of engineers, geologists, soil scientists or meteorologists. Each of them can have different points of view regarding to the involved processes in the hydrologic model.

5.2. The generalization made by the comparative analysis

A way for flow models making it is the usage of the Newtonian mechanics concepts, which to show the component processes in a certain hydrographic basin. These models can be used for to make simulations over several years or decades, for to see if its look like the observed models in the natural hydrographic basins. Also, it can be done the models for to simulate the hydrological processes on shorter time scales, for the purpose of estimating the effect of the rainfall in the ungauged hydrographic basins. These models are based on the causality of the component processes. This causality can be represented in a deterministic and detailed way. However, it is difficult to represent "*the feedback*" beetween the processes which act over a period of time. This aspect of the complex systems it leads to their limited predictability. The modelling of the interactions and the reactions between the different processes which act within the hydrographic basins there can be improved by the understanding of the common effects of the climate, the vegetation, the landforms and the soil types on the functioning of the hydrographic basin.

According with [4], by the notion of ***feedback*** it is understand "the inverse action which manifests itself at the level of different systems (biological, technical, aso.), with the purpose of the maintaining of their stability and balance against of the external actions".

As it is specified within [1], instead of the detailed studying only one a hydrographic basin, an alternative approach is the observation of several hydrographic basins and the studying by comparison of the emerging models (of the resulting models). The purpose is that of the generalization development, by studying of the differences between the hydrographic basins. There is a high potential for this type of comparative analysis.

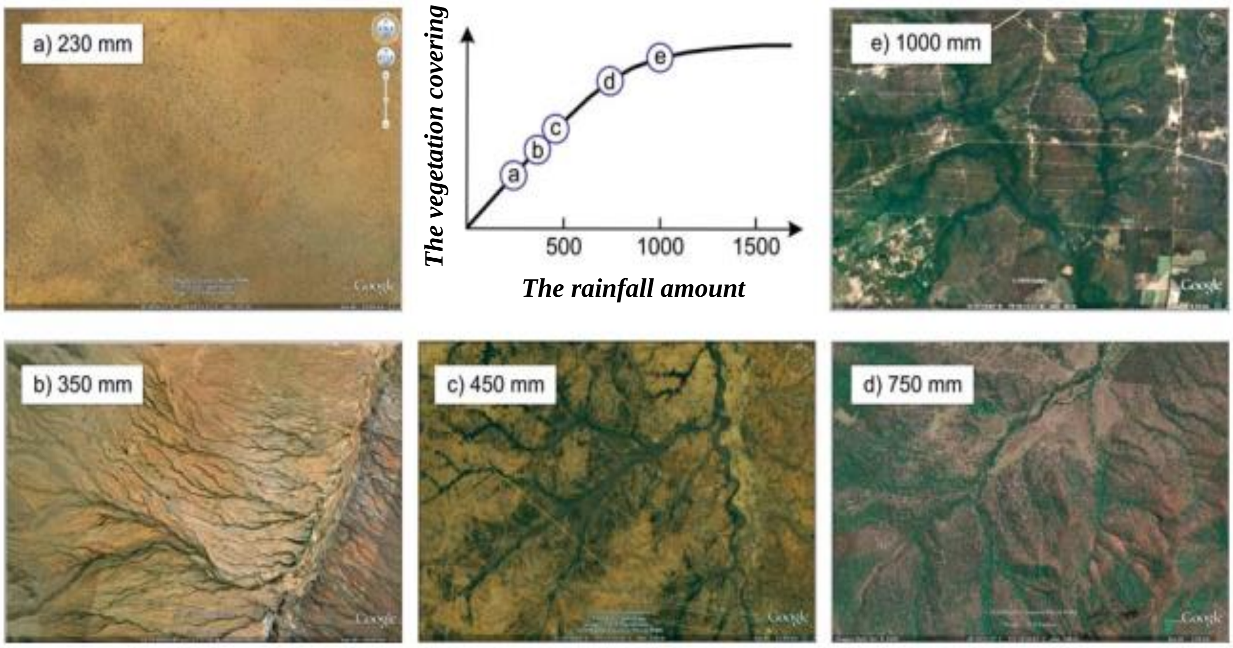


Figure no. 11. The vegetation models for different regions from Australia, with a variation of the annual rainfall from 230 to 1000 mm (l/m^2)

(*Source:* Blöschl, Günter and others, 2013)

In **the figure no. 11** there are presented regions from Australia, where the rainfall amounts differ from one area to another. Under the conditions of an arid climate (a), almost all the rainfall are evaporated. This thing makes as the most hydrological processes vertical being, and the vegetation little being. In proportion as the rainfall increase (b and c), the horizontally flow processes become more and more important, and the vegetation in these areas it is in closely relation with the structure of the drainage network. At the high levels of the rainfall (d and e), the forests occupy the biggest part of the hydrographic basin. However, the vegetation in these areas it can differ between the drainage lines and the mountain areas.

The making of such hydrological models it assumes parameterization of them according to the climatic gradients, for to reflect the modifications of the hydrological processes. By performing of a comparative analysis in these regions, there can make suitable models for the verification of the water flow processes.

5.3. The similarity in the hydrology field

The success of the comparative hydrological approach it depends of the undestanding of the similarity and differentiation notions. If there are compared more basins, some of them there will have similar characteristics and these similarities will influence the emerging models making. The basins are considered similar by hydrological point of view, if there filter equally the rainfall amount. This similarity can be the cause of the similar evolution of the climate, the vegetation, the soils and the landforms from the different hydrographical basins.

According with [1], in the arid basins the rainfall are low and the most of them there are evaporating and there is a low grade of infiltration of the water into the soil. Instead, in the hydrographical basins where the rainfall are significant, there are a high amount of water, too, which will seep into the soil.

The similarity of the processes it is difficult to identify in a real context. This is due to the fact that there are only partially knowed the hydrological processes. Due to the fact that water flow it is the result between the climatic processes and the hydrographic basin characteristics, there can make similarities of the flow, the climate and the hydrographical basin. The comparative hydrological approach consists in the similarities and differences research between the hydrographical basins regarding their climate, the hydrographical basins characteristics and the water flow into these basins. It is assumes that the similar hydrographical basins regarding the climate characteristics and of the hydrographical basin, there will behave similarly, too, by hydrological point of view. This hypothesis can be verified in the case of the ungauged basins, where there can be researched the ratio between the water flow and the characteristics of the climate and of the capture. There can be used the the concept of similarity for to transpose the made research results in the gauged basins, for the water flow estimation in the ungauged basins.

5.3.1. The climate similarity

The climate similarity assumes the similarity of the relevant climate characteristics for hydrology. Thus, those hydrographical basins which have the aridity index higher than 1 there are considered limited in water, and those with the aridity index smaller than 1 there are considered to be limited for the electrical energy production. If the aridity indexes are similar, the basins are considered similar regarding the relative reserve of the water and of the energy.
According with [7], the aridity index is determined using the following calculation relation:

$$I = \frac{12 \cdot p}{t + 10} \quad (5.1), \text{ where:}$$

I represents the aridity index;

p represents the average monthly rainfall, expressed in mm or ${}^{l}/_{m^2}$;

t represents the average monthy air temperature, expreseed in $°C$.

The climate similarity can be defined, also, like the similarity in the interannual variability of the rainfall, if we are interested of the long-term flow fluctuations.

5.3.2. The similarity of the hydrographical basins

The similarity of the hydrographical basins it assumes the similarity of those characteristics of the basins which supervise the flow processes. The basins characteristics which refer to their segmentation, there are including the soil properties which allow the water infiltration, such as the hydraulic conductivity. The hydraulic conductivity represents a global parameter of the flow capacity of the groundwater into the permeable lands, according with [12].
The hydrographical basins characteristics which refer to transmission there are those which present the water flow ways, as it is specified within [1].
The hydrographical basins characteristics which refer to their water storage capacity there are the geology and the soil properties, such as the soil thickness. Also, the area it is used as an indicator of the hydrographical basin storage. Thus, in the bigger basins there are more groundwaters. These basins have deeper flow ways and greater possibility of water storage.

5.3.3. The similarity of the water flow

In the case of the flow similarity it is followed how similar or different there are the water flow characteristics within the hydrographical basins. If we are interested of the long-term hydrologic behavior, the basins are considered similar if "their annual flow signature" do not differ very much. From the perspective of the floods, the hydrographical basins can be considered similar if their floods signatures, such as the floods frequency curve, it presents similarities. The similarity do not means that all signatures are indentical.

For to calculate the flow similarity indexes within hydrographical basin, it is needed of data. In the case of the ungauged basins, there are not very much availabled data regarding to the water flow. The flow similarity approach uses the similarity of the climate and of the hydrographical basin characteristics, for to deduce the hydrologic similarity and for to help at the flow estimation in the ungauged basins.

5.4. Predictive statistical methods for the water flow in the ungauged basins

The statistical methods used for the water flow estimation in the ungauged basins there are basing on the supposition that the flow signatures are the random variables. The statistical methods do not base on the balance equations of the mass, of the impulse and of the energy. There are consisting of linear or nonlinear relations between flow and the characteristics of the climate and of the hydrographical basin. Usually, the model structure is assumed, without a preliminary research. The model parameters are estimated based on the data in the region of interest, as it is specified within [1]. The statistical methods for to predict the water flow there are presenting in the following.

5.4.1. The regression method

In the case of the regression method, the flow signature, $\hat{y}$ (for example, the flood of a certain area, with a certain probability) is estimated based on the characteristics of the hydrographical basin and / or of climatic ones, x_i, with the sampling error, ε. For the flow signature determination, $\hat{y}$, it is used the following calculation relation:

$$\hat{y} = \beta_0 + \sum_{i=1}^{p} \beta_i \cdot x_i + \varepsilon + \eta \quad (5.2), \text{ where:}$$

i represents different characteristics;

β_i represents the model parameters (the regression coefficients);

η represents the mathematical model error.

There are more techniques for to estimate the model parameters in the case of a linear model. It can be used a regression model for the entire field of interest (named global regression) or it can be divided the field into regions (such as in **the figure no. 12. b**) and to be applied the distinct regressive models for each region (named regional regression).

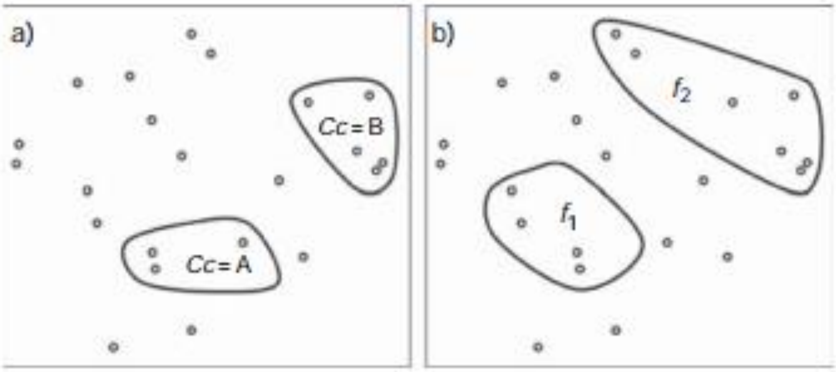


Figure no. 12. The map of an imaginary country, in which the river basins (shown by points) are grouped into regions: a) the grouping into regions, depending on the similar characteristics of the river basins; b) the grouping into regions in which the regionalization methods (such as: the regression equations) are similar

(*Source:* Blöschl, Günter and others, 2013)

The regression coefficients must be able to be interpreted by hydrologic point of view. This thing is due to the fact that such interpretations increase the probability as the regression equation to be applied, too, in the case of the ungauged hydrographical basins, which have not been researched, for to estimate the coefficients.

The regression equations are very simple representations of complex processes. This processes can involve the coevolutiv aspects of the hydrographical basin. Therefore, as it is specified within [1], the interpretration of the regression equations coefficients it is based on the researching of the mode in which there are evolving the landforms, the climate, the soils and the vegetation.

5.4.2. The indexes method

The indexes method is based on the scaled properties of the hydrographical basin. For example, the water flow duration curve (**the figure no. 13**) can be reduced with the help of the annual average flow. The indexing method assumes the existing of an uniformed flow curve. Another example is the “Budyko” curve method, where the ratio between the annual average evaporation and the average rainfall it is expressed as a function of the aridity index, namely the ratio between the annual average potential evaporation and the annual average rainfall.

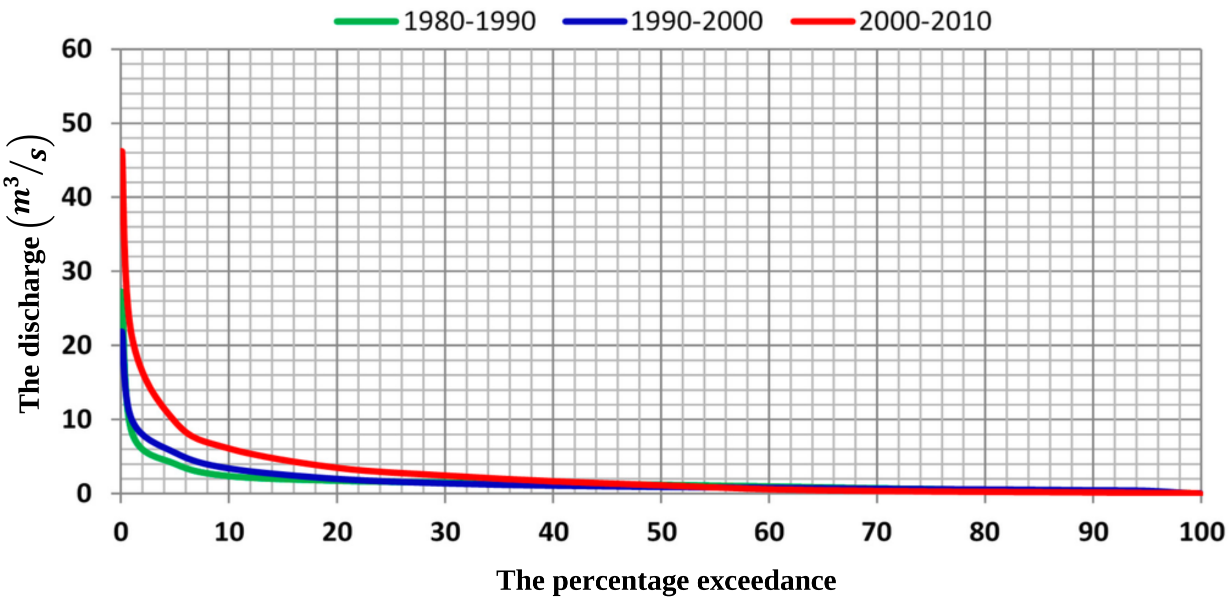


Figure no. 13. Exemple of a water flow duration curve

(*Source:* [15])

5.4.3. The geostatistical methods

The geostatistical methods assumes the correlation of the flow signatures in space. In the geostatistical approach, the flow signature in the hydrographic basin which interests us it is considered to be a weighted average of the flow signatures in the neighboring basins. In the case of the geostatistical method it is given a lower weight of the observations within the hydrographic basins which are closer to each other. According with [1], this thing is due to the fact that the observations are correlated, so there contain fewer information about the random variable.

6. The floods risk maps making in Romania

In [6] the floods are defined as being "the temporary covering with water of a land which there are not usually covered with water". The floods are the effect of complex factors, difficult of evaluating. The floods are natural phenomena, but their effects are amplified by human by placing of the socio-economic objectives (human settlements, economic objectives, roads, railways, bridges, aso.) in the rivers meadows.

In many countries, the floods represent the most important natural hazards. At European level, Romania it is considered one of the countries with the most severe and frequent floods, together with Russia, Turkey and Great Britain, according with [3].

In this context, in the 2007 year there was approved "Directive 2007/60/EC of the European Parliament and of the Council, on the assessment and management of flood risks". By this directive it is imposed of the member states the making of a preliminary assessment of the floods risk, the hazard and the risk floods maps making, as well as the floods risk management plans making. The assessment process is a cyclical one. The floods risk maps updating must be done at every 6 years.

6.1. Methods for evaluation of the floods risk. The floods risk matrix

According with [4], the risk represents a possible danger, the possibility to reach in a danger, to have to endure a distress or to confront with a damage. In scientifical terms, the risk represents the effect of a danger, characterized in terms of probability. So, for to exist a risk it must exist a danger (a hazard). In the case of the floods, the hazard is represented by a dangerous natural phenomenon, by the possible damages and, in some cases, by the lack of the capacity to face the floods.

Within [6], the floods risk is defined as being "the combination between the probability of the occurrence of some floods and the potential adverse effects for the human health, environment, cultural heritage and the economic activity associated with the beginning of a flood".

In [11], with examples from 19 European countries, S.U.A. and Japan, the floods maps are divided into two categories, namely:

- ✓ the floods hazard maps, which cover those areas which could be flooded depending on different scenarios;
- ✓ the floods risk maps, which present possible negative consequences associated with floods within these scenarios.

In the "FAME" report ("The Flood risk and damage Assessment using Modelling and Earth observation techniques"), it is proposed the risk assesssment through a matrix (**the figure no. 14**), as a function of the hazard level (P1, P2, P3) and the exposure class (E1, E2, E3, E4).

The exposure class	The hazard level		
	P1 100 < T < 1000	P2 10 < T < 100	P3 T < 10
E1	R0	R1	R1
E2	R1	R2	R3
E3	R2	R3	R4
E4	R2	R4	R4

Figure no. 14. The floods risk matrix

(*Source:* Chendeș, Viorel and others, 2014)

In the case of the floods risk matrix, R0 represents a very small floods risk; R1 represents a small floods risk; R2 represents a medium floods risk; R3 represents a high floods risk and R4 a represents very high floods risk. Similarly, E1 represents a small floods exposure; E2 represents a medium floods exposure; E3 represents a high floods exposure and E4 represents a very high floods exposure.

The consequence of the floods can be expressed qualitatively, as in the case of the presented matrix, or quantitatively. The most approaches use the qualitative classes, because the quantitative classes need detailated information.

6.2. The identification of the areas with significant potential floods risk

Between the 2010 ÷ 2012 years it has been implemented the first phase within [6], respectively the floods risk preliminary assessment. This stage have two sub-stages, respectively:

- the selection of the significant historical floods, inclusively their spatial localization;
- the indentification of the areas with significant potential floods risk.

The selection of the significant historical floods it was made based on of the criteria elaborated by the "I.N.H.G.A." („The National Institute of Hydrology and Waters Management"). These criteria were taken into account both the phenomenon magnitude (the exceedance probability of the maximum flows and the expansion of the flooded areas), and the consequences and the damages of the floods.

It is made a database with the significant floods, which there were been divided at the level of the water courses (main rivers and tributaries). The floods were been analyzed in detail, using a GIS application, both by the point of view of the maximum flows, and by the point of view of the damages caused at the level of the localities.

The information relating to the significant historical floods, as well as other available data, such as those related to the dammed river sectors (especially those along which the probabilities of the maximum registered flows were exceed the insurance degree of the designed works), the information about the urbanization degree and the socio-economic objectives existing and the data regarding the infrastructure density, there were used for to identify the areas with significant potential floods risk (**the figure no. 15**).

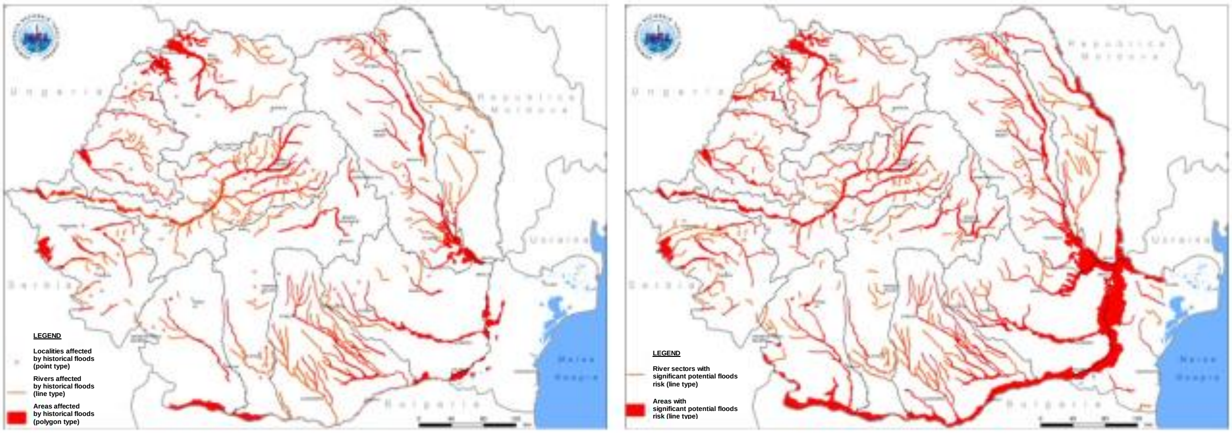


Figure no. 15. The localization of the identified areas in the preliminary floods risk evaluation phase: **a)** the affected areas by significant historical floods; **b)** the areas with significant potential floods risk

(*Source:* Chendeș, Viorel and others, 2014)

6.3. The floods hazard maps making

According to [3], between the 2013 ÷ 2014 year it was made the second phase within [6], respectively the hazard and the risk floods maps making for the identified areas as having significant potential floods risk.

According to the requirements within [6], there were modeled 3 different scenarios, respectively:

- ✓ floods with a small probability of maximum flow exceedance (once at every 1000 years, and in the case of the Olt Basinal Water Administration, once at every 500 years);
- ✓ floods with a medium probability of maximum flow exceedance (once at every 100 years);
- ✓ floods with a high probability of maximum flow exceedance (once at every 10 years).

The most of the hazard maps were been done by the using of the results obtained within the National Program “The Plan of Prevention, Protection and Diminution of The Floods Effects” (“P.P.P.D.E.I.”) (**the figure no. 16**). Within this program there were made high accuracy Digital Terrain Models (M.D.T.), with a vertical precision between ± 20 ÷ ± 50 cm and a resolution with values between 1 ÷ 5 m. For the making of these Digital Terrain Models there were used both LIDAR scan technologies, and precision topo-bathymetric measurements.

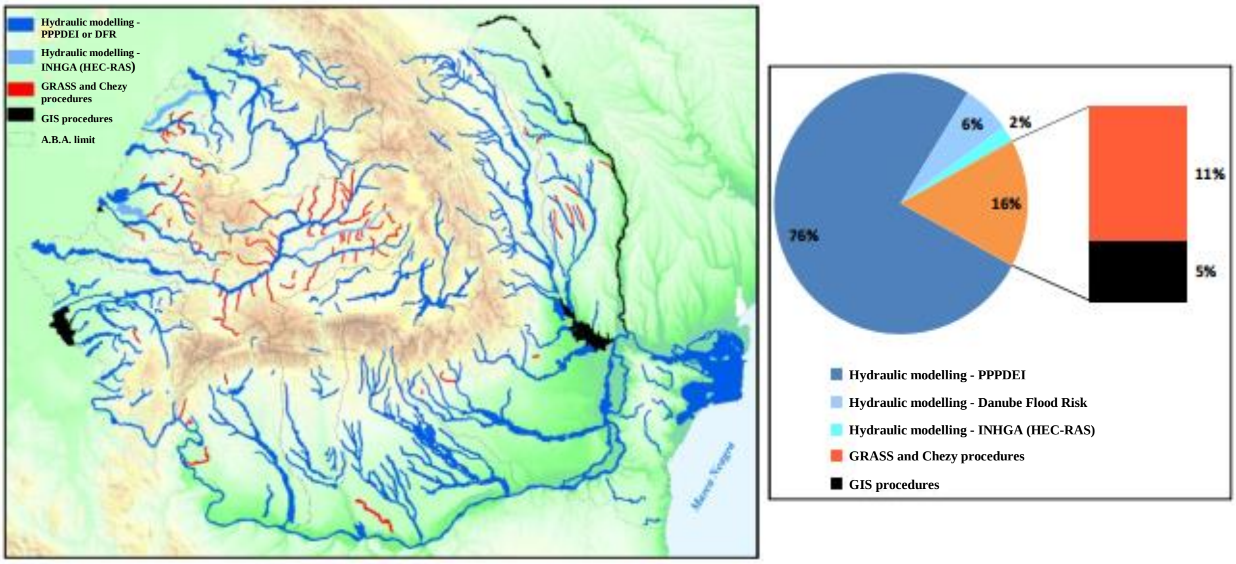


Figure no. 16. Types of methods and data sources for the risk maps making

(*Source:* Chendeș, Viorel and others, 2014)

For the not included areas in "P.P.P.D.E.I." there was used simplified methods for the floods hazard maps making. Due to the lower accuracy of the results, this methods was been applied only for the middle scenario (once every 100 years). This scenario is mandatory to report at European level, but also the most important from a practical point of view.

As it is presented within [3], there was used three (3) methods for the floods hazard maps making, namely:

- modelling with fuzzy systems;
- simplified hydraulical calculus;
- the restoration, with the help of the GIS technology, of the registered levels at the recent floods, with the probability of exceedance flows about 1%.

Conclusions

From the studied materials, there can draw the following conclusions:

- The fields in which the predictive modelling can use there are varied, including the medicine, the archeology, the geology, the town planning, the hydrology or the physics.
- Due to the complexity of the involved notions in the case of the modelling of the evolution of a phenomenon, it is needed of the cooperation some specialists from different domains of activity. Thus, for the prevention of the landslides it is needed of geologists, hydrologists and specialists in the soils field.
- For to be able to model a phenomenon and to can estimate its evolution over time, it must know the factors which influence that phenomenon. Knowing this factors, there can estimate their evolutions and so there can, also, predict the evolution of the phenomenon as a whole.
- The predictive modelling has at the base the mathematical statistics, especially the notions related of regression and parameterization of equations.
- For the modelling of phenomenon it must be analyzed a sufficiently large number of cases, for to be able to drag a truthful conclusion concerning of the evolution of that phenomenon. As a result of the made analysis, there are writting equations for to model that phenomenon. This equations can be parameterized in function of the characteristics of the phenomenon which are the subject of study.
- The basis of predictive modelling in the risk management field there are the risk maps. The risk maps can be done knowing the mode in which a phenomenon was evolved, the affected and possibly affected areas. In the case of predictive modelling there must know both the evolution in the past of the phenomenon, and especially, based on the analysis of the factors which influence that phenomenon, its evolution in the future. Thus, in the case of the floods risk maps making, the predictive modelling has the decisive role, through the prediction of the phenomenons values, such that to result the estimations as close to reality as possible.
- For to be able to predictively model a phenomenon, there are needed of high performance geographical information systems, which to allow the performing of statistical calculations. Also, it must to be a well structured database, from which can to extract the information regarding the evolution of that phenomenon, at any time.

Bibliography:

1. Blöschl, Günter, Murugesu Sivapalan, Thorsten Wagener, Alberto Viglione, Hubert Savenije (2013), *Runoff prediction in ungauged basins. Synthesis across Processes, Places and Scales*, Cambridge University Press, ISBN 978-1-107-02818-0

2. Bolboacă, Sorana Daniela (2014), Corelația și regresia liniară, accessible at the address: http://sorana.academicdirect.ro/pages/doc/TD2014/Curs09.pdf (consulted at 06.08.2017);

3. Chendeș, Viorel, Daniela Rădulescu, Sorin Rândașu, Bogdan Ion, Diana Achim, Alexandru Preda (2014), Aspecte metodologice privind realizarea de hărți de risc la inundații raportate în cadrul directivei 2007/60/CE, publicat în *Hidrotehnică*, no. 10-11, accessible at the address: http://www.researchgate.net/profile/Viorel_Chendes/publication/298305602_Aspecte_metodologice_privind_realizarea_hartilor_de_risc_la_inundatii_raportate_in_cadrul_Directivei_200760EC/links/56e7de7b08ae438aab8a9ba1/Aspecte-metodologice-privind-realizarea-hartilor-de-risc-la-inundatii-raportate-in-cadrul-Directivei-2007-60-EC.pdf (consulted at 27.08.2017);

4. *DEX - Dicționarul Explicativ al Limbii Române*, accessible at the address: https://www.dexonline.ro/

5. Dickey, Alan David (2012), Introduction to Predictive Modeling with Examples, SAS („Scholars Academic and Scientific"), Mumbai, India, accessible at the address: http://support.sas.com/resources/papers/proceedings12/337-2012.pdf (consulted at 21.06.2017);

6. *Directiva 2007/60/EC a Parlamentului European și a Consiliului* din 23 octombrie 2007, privind evaluarea și gestionarea riscurilor de inundații, publicată în *Jurnalul Oficial al Uniunii* Europene, accessible at the address: http://www.inhga.ro/documents/10184/57279/Directiva+60+CE+din+2007.pdf/334e57ef-aaea-4966-9889-aae1e3089c80 (consulted at 27.08.2017);

7. Domuța, Cristian, Cornel Domuța, Gheorghe Ciobanu, Maria Șandor, Alina Dora Samuel, Ioana Borza, Radu Brejea, Adrian Vușcan (2009), Influența irigațiilor asupra cantității și calității producției de porumb în Câmpia Crișurilor, Analele "INCDA Fundulea", volumul 77, ISSN-0253-1682 http://www.incda-fundulea.ro/anale/77/77.13.pdf (consulted at 26.08.2017);

8. Felicísimo, M. Ángel (2003), Use of Spatial Predictive Models in forested areas territorial planning", prezentat în cadrul celei de a IV - a Conferințe Internaționale de Planificare Spațială http://www6.uniovi.es/~feli/pdf/CIOT2003.pdf (consulted at 21.06.2017);

9. Lee, Bryan S. și David J. Buckley (iunie 1992), Forestry Canada Applies GIS Technology to Forest Fire Management, publicat în *Earth Observation*, accessible at the address: https://cfs.nrcan.gc.ca/pubwarehouse/pdfs/19144.pdf (consulted at 28.06.2017);

10. Rădulescu, Mihaela Șt. (2011), *Metodologia cercetării științifice. Elaborarea lucrărilor de licență, masterat, doctorat. Ediția a II-a*, editura Didactică și Pedagogică, R.A., București, ISBN 978-973-30-2894-9, (consulted at 05.09.2017);

11. RWS - RIKZ (septembrie, 2007), *Atlas of Flood Maps*, accessible at the address: https://circabc.europa.eu/webdav/CircaBC/env/wfd/Library/floods_programme_1/b_wg_f_on_floods/meeting_19102007/Rapport%20-%20Draft%2011.pdf (consulted at 27.08.2017);

12. Scrădeanu, Daniel, Alexandru Gheorghe (2007), *Hidrogeologie generală*, Editura Universității din București, ISBN 978-973-737-367-0, accessible at the address: https://www.scribd.com/document/268674516/Curs-Hidrogeologie-Generala (consulted at 26.08.2017);

13. Verhagen, Jacobus Wilhelmus Hermanus Philippus (2007), *Case studies in archaeological predictive modelling*, Leiden University Press, ISBN 978-90-8728-007-9, accessible at the address: https://openaccess.leidenuniv.nl/handle/1887/21069 (consulted at 15.08.2017);

14. https://datascience.stackexchange.com/questions/5023/what-is-a-discrimination-threshold-of-binary-classifier (consulted at 26.08.2017);

15. http://www.mdpi.com/remotesensing/remotesensing-06-12544/article_deploy/html/images/remotesensing-06-12544-g011-1024.png (consulted at 28.08.2017);

16. https://profs.info.uaic.ro/~val/statistica/AsocVarDisc.pdf (consulted at 19.08.2017);

17. https://profs.info.uaic.ro/~val/statistica/StatWork_10.pdf (consulted at 13.08.2017);

18. http://www.rasfoiesc.com/inginerie/comunicatii/CARACTERISTICILE-IMAGINILOR-SA14.php (consulted at 22.06.2017)

19. https://softstatistica.wordpress.com/2015/04/07/curba-roc-in-spss/ (consulted at 25.06.2017).