EnviroInfo 2004 (Geneva)
Sh@ring – EnviroInfo 2004
 ISBN: 28 29 30 275-3

Regionalised cartography of epidemiological data in Central Asia (Issyk Kul region,Kyrgyz Republic)

Anne-Laure Zufferey[1], and Michel Maignan[2]

Abstract

During the last five years, an increasing of diseases has been observed in the Issyk-Kul region in Kyrgyzystan (Central Asia). Environmental factors, radioactivity, water quality and air quality could influence the health condition of people and explain such situation. The present work deals with a GIS study using geostatistical models (spatial predictions) to find relations between population health and environmental factors during the years 1999. 2000 and 2001.
Keywords: GIS. Multivariate analysis, Principal Component Analysis, Cluster Analysis. Kriging, epidemiology, radioactivity

1. Introduction

The purpose of this research is the processing of epidemiological and environmental data using geostatistical model and GIS. The environmental en epidemiological data have been processed differently.

The environmental data are radioactivity measurements. 926 points have been sampled around the lake during the year 2003. Kriging methods have been applied to this dataset.

The epidemiological data have been extracted from dispensaries registers. 70 dispensaries locate around the lake and every point contains information concerning the diseases affecting adults and children.

The multivariate analysis (Principal Component Analysis and the Cluster Analysis) have been applied to the epidemiological data to know the correlation between diseases and the spatial correlation of the points.

To reduce the number of variables, we have applied the Geostatistical tools to the factors scores calculated with the Principal Component Analysis.

[1] University of Lausanne , e-mail: anne-laure.zufferey@igp.unil.ch]
[2] University of Lausanne, e-mail: michel.maignan@img.unil.ch

2. Environmental data

The 926 measurement points of the radioactivity level have been done at 10 centimetres of the ground. The unit used is the "microroentgen/hour". The sampling has been realized inside and outside public building, such as school and dispensaries, near water pump and river and near factories.

The area of study locates around the lake. Variography and Kriging have been applied to the data set. The East part is a flat area with a more homogeneous repartition of the points which allows the application of the Neural Network Residual Kriging.

2.1 Kriging

The basic problem consists in estimating the value of the function at unsampled points. In general this problem has no unique solution.

Univariate linear models are first considered. A linear model means that the unknown value is estimated as a linear combination of the measurement data with unknown weights:

$$Z^*(x_0) = \sum_{i-1}^{N} w_i(x_0) Z(x_i) + w_0(x_0)$$

Where N is the number of data points used for the estimation. N can be very large and usually local neighbourhood is used taking into account behaviour of the variogram. Extreme values have been excluded and the range of values is 1 to 40 microroentgen/hour.

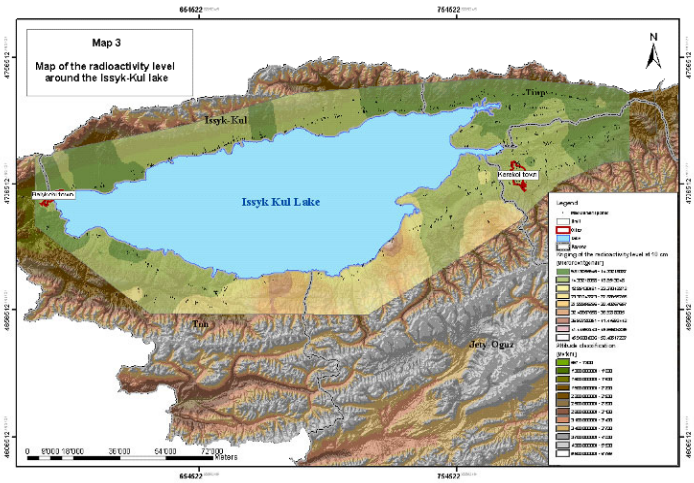

Figure 1: Prediction map of the radioactivity level of the Issyk-Kul oblast

The repartition of the points influences their interpolation. This one is perpendicular to the measurement points and the prediction surfaces of the hot spots are extended. The figure 1 shows the Kriging map obtained.

2.2 Neural Network Residual Kriging (NNRK) [Kanevski & Maignan 2004]

This method has been applied to the measurement point of the East part of the lake. This method gives a more precise modelling of the radioactivity of the area.

The principle of this method is the use of a neural network to find and model the tendency, which is applied to the dataset. The Kriging is applied to the residuals and added to the tendency.

In figure 2 is presented the prediction map realized with the NNRK method.

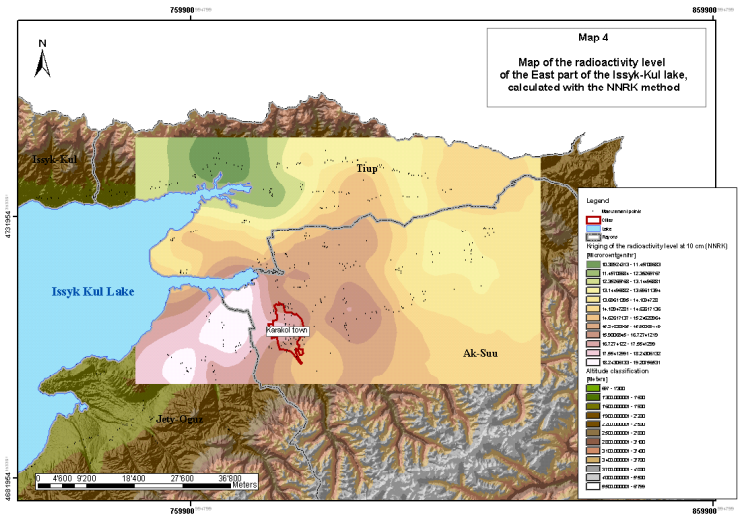

Figure 2: Prediction map of the East part of the Issyk Kul lake, realized with the NNRK method

2.3 Comparison between the maps

We can see well the hot spot on the figure 2, but the extension of the prediction surfaces creates some artefacts without signification. The figure 3 shows more precise limit between the areas of interpolation.

A lot of properties of data, such quantity and a homogeneous repartition of points, should be gathered to apply the NNRK. In our case, only the East part of lake has these properties.

3. Epidemiological data

The data extracted from the dispensaries registers are not stationary. Each point contains the number of diagnosis of 26 diseases, calculated for 100'000 persons. The dataset concerns the years 1999, 2000 and 2001. Each point represents the level of 26 diseases of an area depending of the distance and the topography between the dispensaries.

The difficulty to model the level of diseases results from the no stationary and from the repartition of the points around an extended lake.

3.1 Multivariate analysis

3.1.1 Cluster analysis

This method is a classification of multivariate data. The results show the clustering of similar observations (the dispensaries).

In our work, we have used the cluster analysis to find groups of dispensaries with similar diseases rate. The graphic result is a dendrogram which shows the different clusters with their observations. We visualize well the excluded points of the PCA.

To know if a spatial correlation exists between the values of points (diseases levels of the dispensaries), we have created a map of the clusters. The dispensaries have the same colour as the cluster calculated (figure 3 and 4).

We can see in the repartition of the colours that a spatial correlation could exist. The East shows well this correlation.

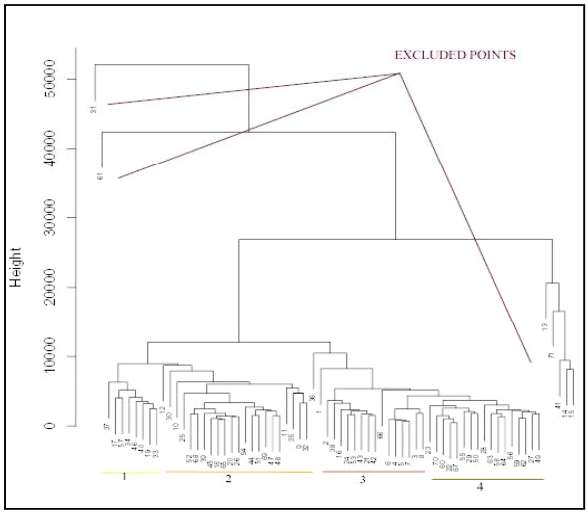

Figure 3: Dendrogram of the cluster analysis

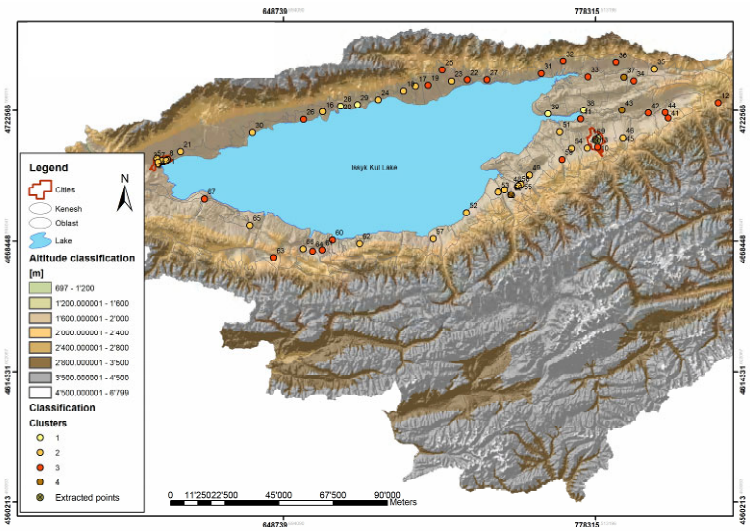

Figure 4: Dispensaries presented with their clusters colours

3.1.1.1 Factorial Kriging

The Cluster Analysis has shown that a spatial correlation exists. It is also possible to apply geostatistical tools to the epidemiological dataset. The problem we have met is the number of variables which represent the health situation of the region. To reduce the number of variables, we have used the PCA results: the factors scores. They are the converted values of points (the dispensaries) to be projected in the new space created by the two first principal components. The projection is done by multiplying the observed values by the corresponding elements of the eigenvectors. We obtain only one value for every point corresponding to the first principal component which represents the best the basis cloud of points. The corresponding geographical coordinates of the dispensaries have been added to the factors scores file.

The kriging processing has been applied to this new dataset. We obtain a factorial kriging map which shows prediction surfaces. They represent the similarities between the observations which signify the similarities between the diseases rates recorded by the dispensaries.

We have observed that the surfaces with low values of factor scores are hot spots. The same diseases rates are always higher in dispensaries of low factors scores values than high factors scores values for the three years and for children and adults separately.

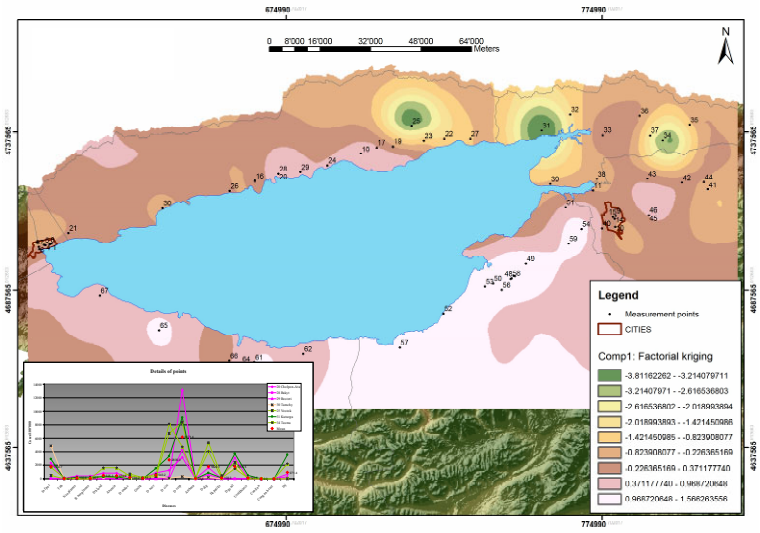

Figure 5: Factorial kriging map

This method is useful for visualizing area "with problems". To know the exact kind of problem, the details of each point should be focus on. We have used Microsoft Excel graphs to visualize the difference between dispensaries situated on green prediction surfaces (low factors scores level) and pink surfaces (high factors scores level). The figure 5 shows the step.

4. Regionalization

Multivariate Analysis and Factorial Kriging allowed us to model the health situation of people around the Issyk Kul Lake.

Hypothesis concerning environmental factors could explain some diseases levels. To model the interactions between environment and health we have separate some regions characterized by geographical and geological parameters, which are:

- Altitude
- Distance of less than 500 meters of a river
- Distance of less 1 kilometre of the lake
- Region with radioactivity

The dispensaries locating in these areas could be influenced by these parameters and have recorded associated diseases. Their diseases levels have been focus on and they have been searched in the Factorial Kriging and the Multivariate Analysis to know which position they have.

Some diseases are characteristic of region. For example the infectious and parasitic diseases are associated to the river closeness and the iron deficient anemia to the area with radioactivity.

Dispensaries stating in the same prediction surface are often in the same cluster. The multivariate analysis can be used in the modelling steps of the environmental and health situation of a region.

The Factorial Kriging could be useful for the interpretation. The resulting maps show the similarities between dispensaries. If the similarities are bounded to the environment, predictive surfaces could visualize health problems bounded to the environment.

5. Discussion

People of the Issyk Kul oblast don't move long far from their villages. The dispensaries network is regularly distributed around the lake and in mountains villages. The few moving of people induces a stationary of the measurements points. This property is very important for the application of geostatistical tools. If it would be the case, no spatial correlation had been found.

The Kriging and NNRK methods give good results to the radioactivity levels measurements. Only the geographical extension of the lake creates artefacts.

The epidemiological data are not stationary. This property is important because the socio-economical situation of the region allows us to consider the dispensaries as stationary points. The cluster analysis has shown that a spatial correlation exists. The Factorial Kriging method has been used to create a global model of the health situation concerning several diseases. In our study the diseases were very different. It would be interesting to apply this method to diseases of same groups or with the same etiology and to diseases which could be associated to the same environmental factors.

Such processing can not be applied to our countries. The diseases studied are no more a problem in our countries. The interesting diseases are contagious and the moving of people (pendular workers) induces diffusion of diseases instantaneous.

Bibliography

Collet C, (1992), Système d'information géographique en mode image

Dalgaard, P. (2002), Statistics and Computing

Davis J. C. (1986), Statistics and Data Analysis in Geology

Haining R. (1990), Spatial data analysis in the social and environmental sciences

Isaaks E. H, & Srivastava R. M.(1989),An introduction to applied Geostatistics

Kanevski M. & Maignan M. (2004), Analysis and modelling of spatial environmental data.

McCoy J. & Johnston K. (2001), Using ArcGis Spatial Analyst

Morgenthaler S. (1997), Introduction à la statistique

Indicators as suggested by participants of the Consultation on Environment and Health Indicators for Use with a Health and Environment Geographic Information System (HEGIS) for Europe, Bilthoven, Netherlands, 11-13 March 1993. WHO,1994

State agency of geology and mineral resources under the government of the Kyrgyz Republic, (2000), Explanation note and Map of mineral resources of the Kyrgyz Republic.