

***Epidemiological and environmental GIS of the Issyk-Kul region, Kyrgyz Republic**

ZUFFEREY Anne-Laure; Octobre 2004

Supervisor: Prof. M. Maignan, Institut de Minéralogie et Géo chimie

This work concerns a statistical, geostatistical and spatial analysis of epidemiological and environmental data of a region in Central Asia.

The studied area locates in Kyrgyzstan, a republic of the ex-Soviet Union, in the Issyk-Kul oblast, around a lake with the same name. This work is a SCOPE project and had been followed with kyrgyz scientific partners.

Environmental processed data is measurements of radioactivity, sampled around the lake. Epidemiological data have been extracted from the dispensaries network stating in the Issyk-Kul oblast and are levels of 26 diseases for the years 1999, 2000 and 2001. These dispensaries are called "Family Physician Groups" or "FPG" and are numerous (around 70).

Different methods of analysis have been used: the multivariate analyses (Principal Component Analysis and clustering), the Geostatistics (Variography, Kriging, Neural Network Residual Kriging) and the Factorial Kriging. This uses the results of the Principal Component Analysis and the Geostatistics tools.

The creation of maps has been realized with the GIS.

The geostatistical processing of the radioactivity data allows us creating prediction maps of the level of radioactivity around the lake.

The epidemiological data are not stationary in sensus stricto. Every point corresponds to a dispensary which lists all patients come in consultation and the diagnosis established. The patients residence is near the dispensary because people don't move a lot; they use in most case horses or donkeys. Because the few moving of people, we can consider the points as stationary.

To know the validity of the application of the geostatistical methods and the correlation and similarities between diseases and dispensaries, we have used the multivariate analysis. The Principal Component Analysis allows knowing the correlations between diseases and the cluster analysis the similarities between the dispensaries concerning the level of diseases. If, for example, two dispensaries have the same rate of diseases, they will be in the same cluster resulting of the cluster analysis. Causes bounded to the environment could be invoked if the dispensaries locate in an environment with the same characteristics (basis of a valley, littoral, near a river or in altitude). To know if such repartition exists, we have made a representation of the dispensaries with different colours corresponding to the cluster analysis. If groups with the same colour appear on a geographical representation of the points, it signifies that a spatial structure exists. In this case, the geostatistical tools can be used. The results have shown that such spatial structure exists and the prediction maps have been done with ordinary kriging for different diseases.

A prediction map could be done only for one disease. We have process 26 diseases for three years and two age categories (adults and children). The maps are too numerous to realize. To obtain global information concerning the level of diseases of each dispensary, we have applied the Factorial Kriging. This method uses the values calculated with the Principal Component Analysis for every point (the

dispensaries. These values, the factors "scores", correspond to the values of the points projected in the new space created by the two first Principal Components. The geographical coordinates of the dispensaries have been added. The last step is the geostatistical processing with the ordinary kriging. The maps show area "with problems" and their diseases rates should be focus on to know which diseases create the "hot spots".

A regionalization had been done with GIS to extract the dispensaries located in a characteristic environment. The Kyrgyzstan has strong environmental and socio-economical factors which influence the people health. The Geology, the Tectonic, the closeness to the lake and river, the altitude and the urban situation are factors focus on with the GIS. We have used the geographical and environmental data and a digital elevation model.

The complete geostatistic study and results were obtained with Geostat Office, a scientific software package for geostatistical and machine learning analysis of data, the clustering and PCA study have been realized with R languages (R Project for statistical computing) and the GIS project has been created with ArcGis from ESRI (Environmental Systems Research Institute).