Spatial Interpolation Comparison 97.

Introduction and description of the data set.

G. Dubois

REM, IES, DG JRC, EC

1. INTRODUCTION

Spatial interpolation is an essential feature of many Geographic Information Systems, or GIS. It is a procedure for estimating values of a variable at unsampled locations. A map with isolines is usually the visual output of such a process and plays a crucial role in decision making. Based on Tobler's Law of Geography, which stipulates that observations close together in space are more likely to be similar than those farther apart, the development of models attempting to represent the way close observations are related can sometimes be very problematic. The approaches can be divergent and may therefore lead to very different results. As a consequence, an understanding of the initial assumptions and methods used is the key to the spatial interpolation process.

Surprisingly, when spatial interpolation tools are integrated within GIS, they are often implemented in such a way that users have no real choice in selecting the best possible methods; and if they do have a choice, required input parameters are sometimes fixed, without any possible way to modify them. To my knowledge, there are currently practically no GIS in the market that can be used to perform a rigorous spatial correlation analysis and interpolation of the data. Frequently, additional statistical packages will be required. Without such additional tools, the spatial interpolation becomes a black box instead of a clear process where all parameters can be properly defined. This means that it becomes almost impossible to provide the experts and decision-makers with basic information on the method used. Finally, it is also difficult to provide the end user with information on the uncertainty associated with the estimates generated during the interpolation.

One reason for the frequent blind use of spatial interpolation methods, and spatial statistics in general, has probably its origins in teaching. Despite the large variety of its applications, the discipline has been confined to those fields where it has seen its major developments. The progress made in spatial statistics is therefore usually presented only in journals dedicated to statistics, mining, and petroleum engineering. As a consequence, GIS users who have a different technical background often do not have an in-depth knowledge of such spatial interpolation techniques. Furthermore, since the conventional tests used in basic statistics usually generate some kind of categorical answer, the prerequisite experience and statistical knowledge necessary for the proper use of spatial interpolation techniques are often discouraging to the neophyte. Nevertheless, during the last couple of years, the diversity of the applications of these methods has stimulated the publication of new books and new case studies, as well as a number of conferences on the subject.

To better understand spatial interpolation techniques so that practitioners can grasp the "science behind the technology", a kind of contest has been organised on the AI-GEOSTATS mailing list (www.ai-geostats.org) in 1997. Called "Spatial Interpolation Comparison 97 (SIC 97)", this scientific exercise where participants were invited to estimate daily rainfall measurements at unsampled sites as well as their extreme values, was organised in the frame of the Radioactive Environmental Monitoring institutional support programme of the Environment Institute (Joint Research Centre, EC, Ispra). The expected objectives of SIC97 were manifold. The first one was to give a more general overview of existing spatial interpolation methods and to highlight the latest developments in spatial statistics. If daily rainfall measurements have been distributed to the participants, a variable everyone is familiar with, their conclusions can nevertheless be easily extended to the problem of mapping radioactivity in the environment. Daily rainfall and radioactivity deposited on the ground after an accidental release in the atmosphere do not only behave in a similar way, i.e. the variables present usually a global trend and strong local fluctuations, but they are also intimately correlated since rainfall is the main factor affecting the deposition of radioactivity released to the atmosphere. The participants were therefore also invited to evaluate their methods for the monitoring of the radioactivity in the environment, this in situation of routine or emergency.

2. COMPARING THE CONTRIBUTIONS TO SIC97

The participants to SIC97 were invited to describe the methods they used and the related decision-process. To facilitate the comparison of the results they obtained, the participants were asked to provide the reader with the following information:

1. the minimum, maximum, mean, median and standard deviation of the 367 estimated values;
2. the root mean squared error;
3. the bias of the errors,;
4. the mean relative and absolute errors;
5. the correlation between the errors;
6. a correlation plot of the estimated values against those observed;
7. a map with isolines showing the levels of the estimated values.

The performance of the methods used had to be discussed as well in terms of computational time and ease of automatisation in view to assess their potential use in the frame of an emergency-mapping system.

3. DESCRIPTION OF THE DATA

3.1. *Observed rainfall data*

The data distributed to the SIC97 participants were 100 daily rainfall measurements made in Switzerland on the 8th of May 1986 which were randomly extracted from a dataset of 467 measurements. The participants had to estimate the rainfall at the 367 remaining locations. The measurements were in units of 1/10th of a mm, but values ranging from 1 to 10 were often observed due to air condensation. Figure 1 shows the locations of the 367 measurements for which the values had to be estimated (diamonds) while those of the 100 measurements used for the estimation are presented with proportional symbols (filled circles).

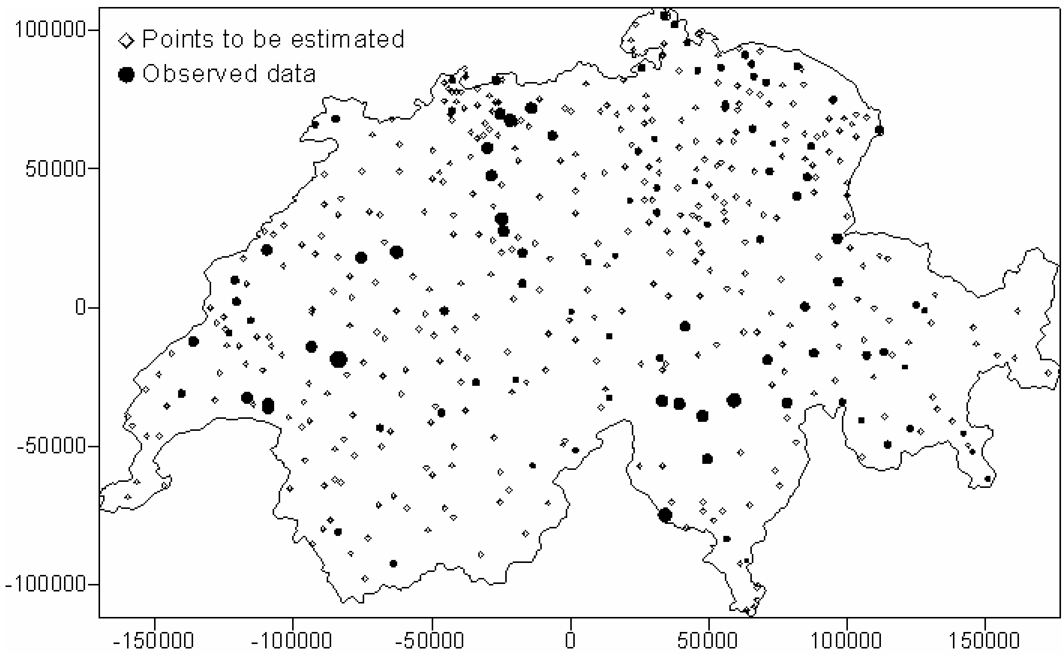
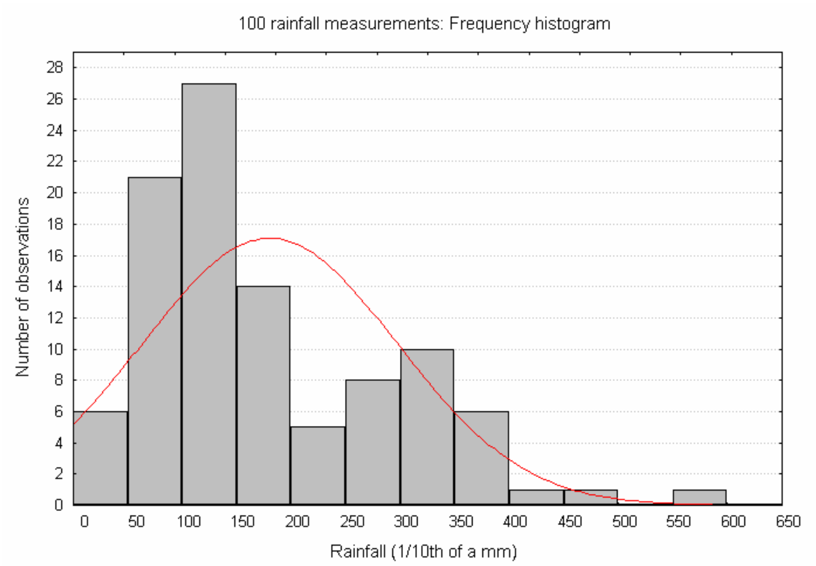


Figure 1. *Proportional symbols of the 100 observed data and locations of the values to be estimated*

The statistics of the 100 measurements are given in Figure 2.



N	100
Min.	10
Max.	585
Mean	180.15
Var.	13614.47
Std. dev.	116.68
Skew.	0.956
Kurt.	0.467

Figure 2. *Statistics of the 100 observed rainfall value*

3.2. *Digital elevation model and country borders*

A digital elevation model (DEM) with a resolution of around 1 km x 1 km was provided as secondary information so as country borders used to define the area under study (Figure 3).

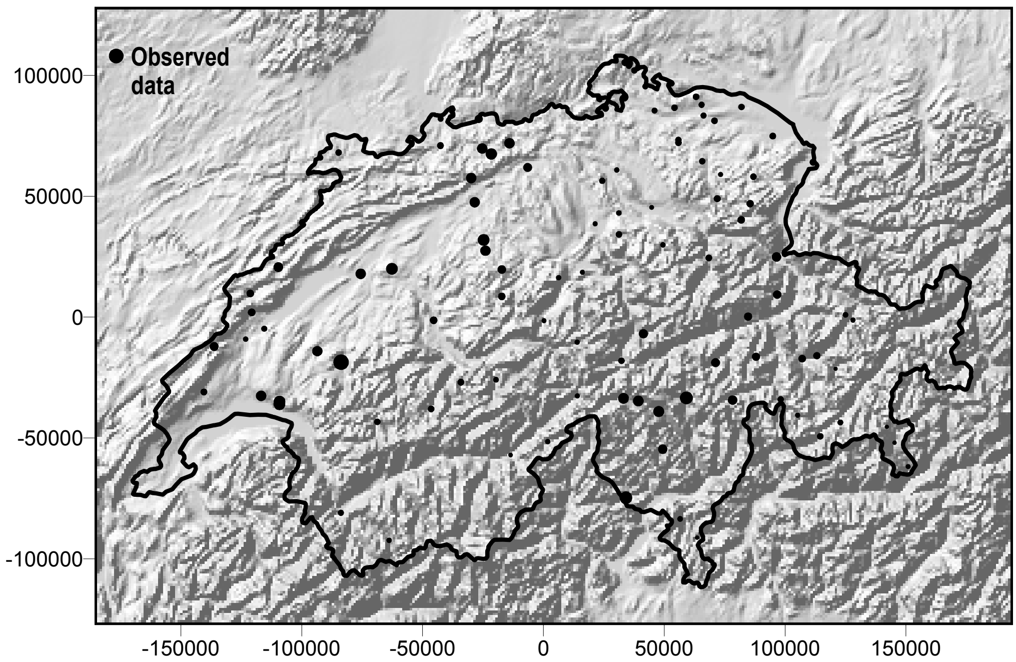
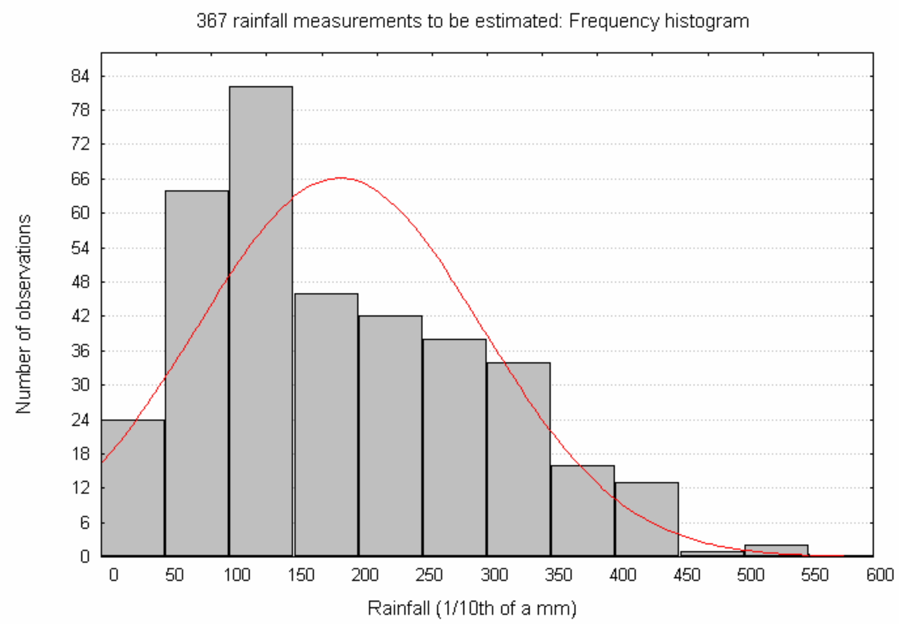


Figure 3. *Digital Elevation Model of Switzerland* (*1 km* * *1 km*) *and country borders*

3.3. *Description of the data to be estimated*

The co-ordinates of the sampling locations of the remaining 367 daily rainfall measurements were distributed to the participants in order that estimates could be made. The observed rainfall values for these locations were made available only after reception of the estimations of all the participants. The statistics of the 367 measurements are given below (Figure 4).



N	367
Min.	0
Max.	517
Mean	185.09
Var.	12254.70
Std. dev.	110.70
Skew.	0.569
Kurt.	- 0.373

Figure 4. *Statistics of the 367 rainfall data points to be estimated.*

3.4. *Description of the full data set*

Figure 5 presents the complete 467 rainfall data in terms of proportional symbols and summarised with their statistics and associated histogram (Figure 6).

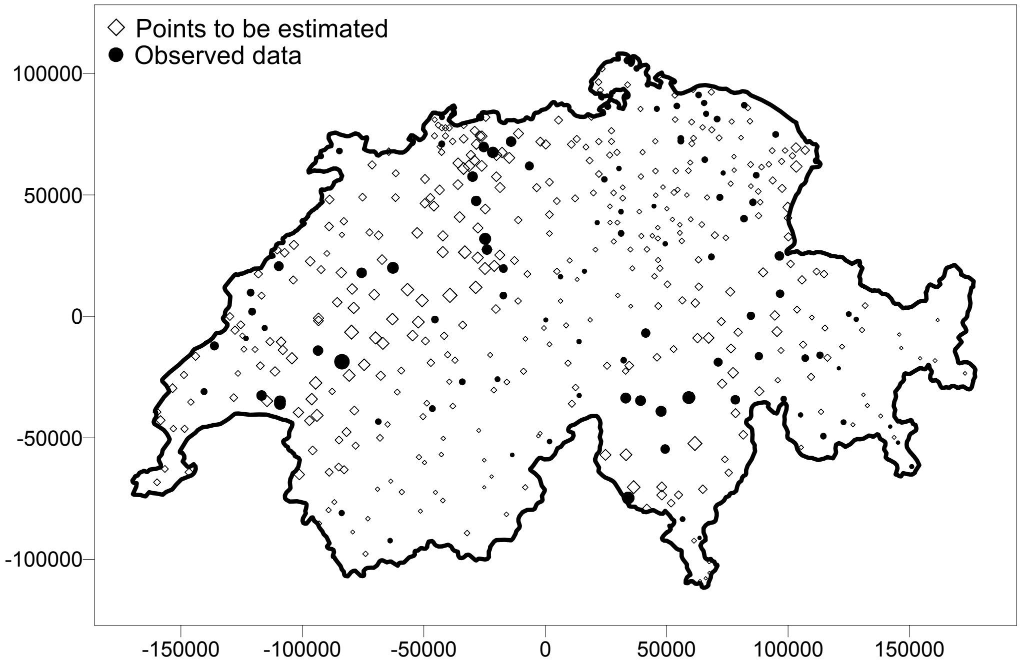
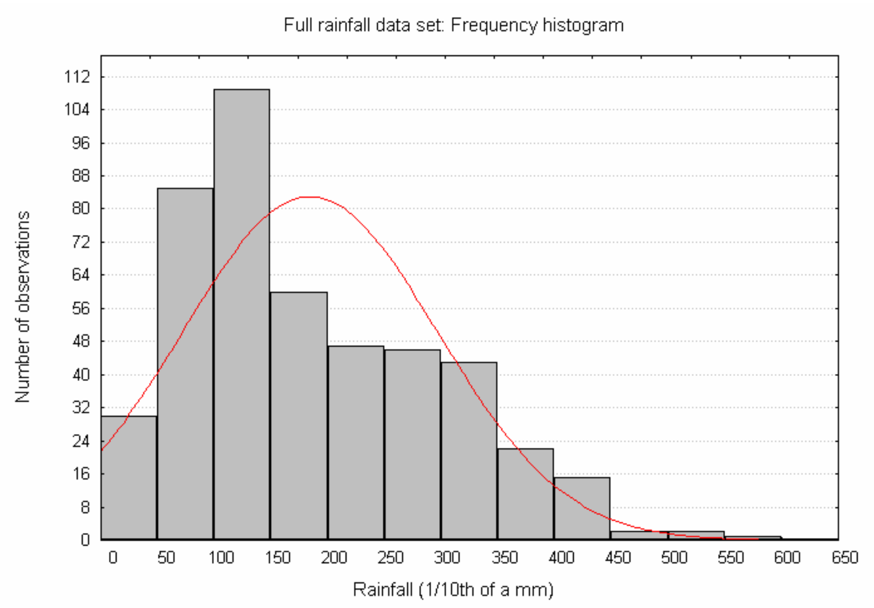


Figure 5. *Proportional symbols for the whole rainfall data set*



N	467
Min.	0
Max.	585
Mean	184.24
Var.	12603.13
Std. dev.	112.26
Skew.	0.661
Kurt.	- 0.193

Figure 6. *Statistics of the complete rainfall data set.*

4. SOURCES OF THE SIC97 DATA SETS

The gathering of the rainfall data, provided by Giovanni Graziani from the Environment Institute of the Joint Research Centre (Ispra, Italy), has been undertaken, under JRC-Ispra funding, by the Air pollution Group at Imperial College, London.

The Digital Elevation Model has been provided by EROS Data Centre from the U.S. Geological Survey (USGS). http://edcwww.cr.usgs.gov/

The country borders are extracted from ESRI's Digital Chart of the World (DCW) provided by ESRI.

If you wish to test your methods and experience on the same case study, the data have been made available on this web site (see the "Data" section of ai-geostats). The point data set correspond to a simple ASCII file with, for each measurement, an identifier given as an integer, the co-ordinates X and Y in meters and the rainfall measurement given in 1/10th of a mm. The data have been projected with a Lambert azimuthal projection system. Country borders of Switzerland are available as an Autocad Interchange Drawing file (.dxf), the DEM is provided as an ASCII grid projected in order to match with the point data sets and the country borders.