Analysis of Rainfall Data by Simple Good Sense: is Spatial Statistics Worth the Trouble ?

M. G. Genton[1] and R. Furrer[2]

[1]*Department of Mathematics*
Massachusetts Institute of Technology
Cambridge
MA 02139-4307, USA

E-mail: genton@math.mit.edu

[2] *Chair of Applied Statistics*
Swiss Federal Institute of Technology
CH-1015 Lausanne
Switzerland

E-mail: reinhard.furrer@epfl.ch

Abstract. This paper discusses the use of simple good sense in the spatial prediction of rainfall measurements in Switzerland. The method consists of a forecast based on the values of the available observations combined with topographic knowledge of the Swiss territory. Comparison of our subjective estimates with the true measurements is completed and yields surprisingly good results.

Keywords: Interpolation; Swiss Bayesian prior; Rationality; SPLUS.

1. INTRODUCTION

The analysis and interpretation of spatial data sets forms an important part of geostatistics and is unfortunately highly human dependent. For instance, it is well known that different individuals will take different approaches (Englund, 1990), yielding a large assortment of distinct solutions. This is partly due to the variety of available spatial interpolation methods, ranging from simple intuitive predictions to more sophisticated and complex procedures. Some of the more commonly used interpolation methods (Cressie, 1993) include:

- Inverse distance weighting and nearest neighbor
- Polynomial trend surfaces and splines
- Kriging
- Likelihood and Bayesian analysis
- Neural networks

Most of these methods are in fact only classes of procedures, which contain an incredible number of possible variations accessible to the researcher. Moreover, several sources of variability arise from the data themselves and involve subjective decisions on the part of the individual. These include data transformations, detection and handling of outliers or even more nasty inliers, choice of estimators for dependent observations, variability of estimators, model selection, choice of hardware and software. An attempt of robustness in variogram estimation and fitting for kriging is proposed by Genton (1996, 1998a, 1998b, 1998c), and an example of application to sediments data of Lake Geneva, Switzerland, can be found in Furrer and Genton (1998). Having such a large diversity of methods, yielding so many different results (Englund, 1990), one can legitimately ask whether spatial statistics is worth the trouble ? In this paper, we try to give some possible answers to this question.

2. SIMPLE GOOD SENSE PREDICTION

In this section, we describe briefly the simple good sense method that the first author used for the spatial prediction of rainfall in Switzerland. First, the n=100 available locations of rainfall are plotted, and each corresponding value of amount of rainfall is labeled with the function `identify` in SPLUS. Second, the 367 locations where rainfall has to be predicted are plotted, and the corresponding amounts of rainfall are estimated by eye. We used the information given by the 100 known rainfall amounts, as well as some subjective knowledge of the Swiss territory, which can be viewed as a Swiss Bayesian prior. For example, we are conscious of the sunny micro-climate of the county of Wallis, as well as the locations of mountains and plains. Moreover, one of the fundamental principle of geostatistics has been applied: observations which are closely located in space are more likely to be similar than observations which are far away. It is also well known that the precipitations on the lee side of mountains is much higher than on the luv side. Thus the 100 known observations helped to determine on which side of the mountains it rained. Note that in our prediction procedure, no additional information from any other spatial statistical method has been used: this is pure good sense !

3. DISCUSSION OF THE RESULTS

The overall performance of our method in predicting the remaining 367 rainfall data is summarized in Table 1. We consider the true values $Z(\mathbf{x})$, the estimated values (by simple good sense) $\hat{Z}(\mathbf{x})$, the errors $e(\mathbf{x}) = \hat{Z}(\mathbf{x}) - Z(\mathbf{x})$, the absolute errors $|e(\mathbf{x})|$ and the relative errors $|e(\mathbf{x})| / Z(\mathbf{x})$. For each of these quantities, the minimum, the maximum, the mean, the median and the standard deviation is computed.

	Min	Max	Mean	Median	Std. Dev.
True values	0	517	185	162	111
Estimated values	13	450	176	150	94
Errors	-318	223	-9	2	72
Absolute errors	0	318	50	36	52
Relative errors	0	10.00	0.40	0.23	0.78

Table 1. *This table presents the minimum, the maximum, the mean, the median, and the standard deviation for the 367 true rainfall values, the estimated values, the errors, the absolute errors, and the relative errors.*

The distribution of the estimated values by simple good sense is in agreement with the distribution of the true values. This is confirmed by a plot of estimated values (horizontal) against true values (vertical) in Figure 1.

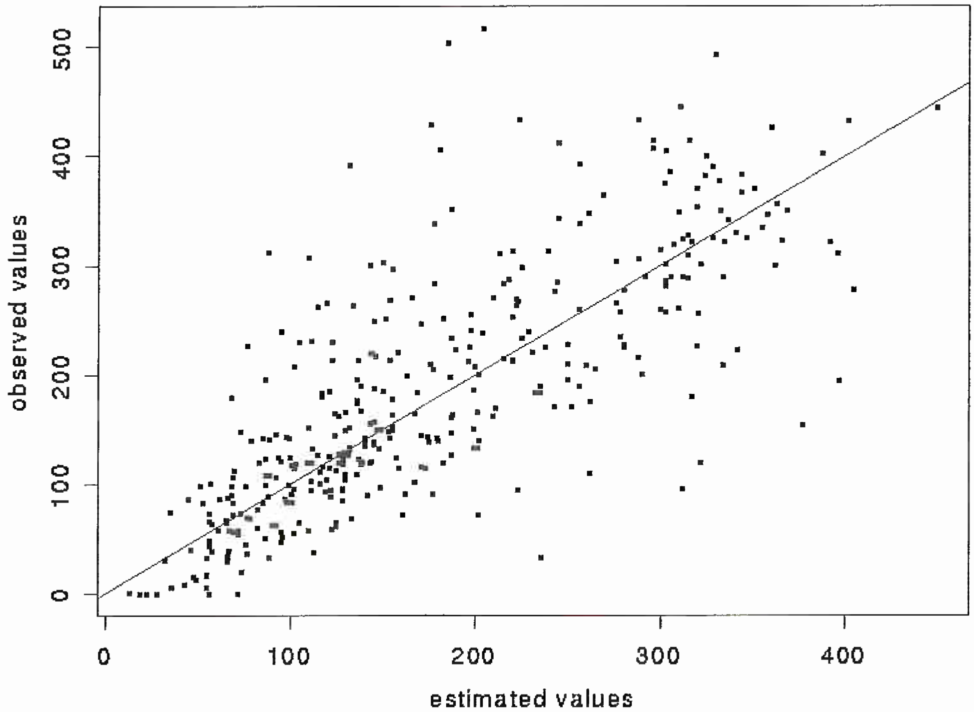

Figure 1. *Plot of estimated values (horizontal) against true values (vertical).*

A small positive bias is however revealed. A plot of observed (true) values against residuals in Figure 2 indicates that small values are generally overestimated whereas large values are underestimated.

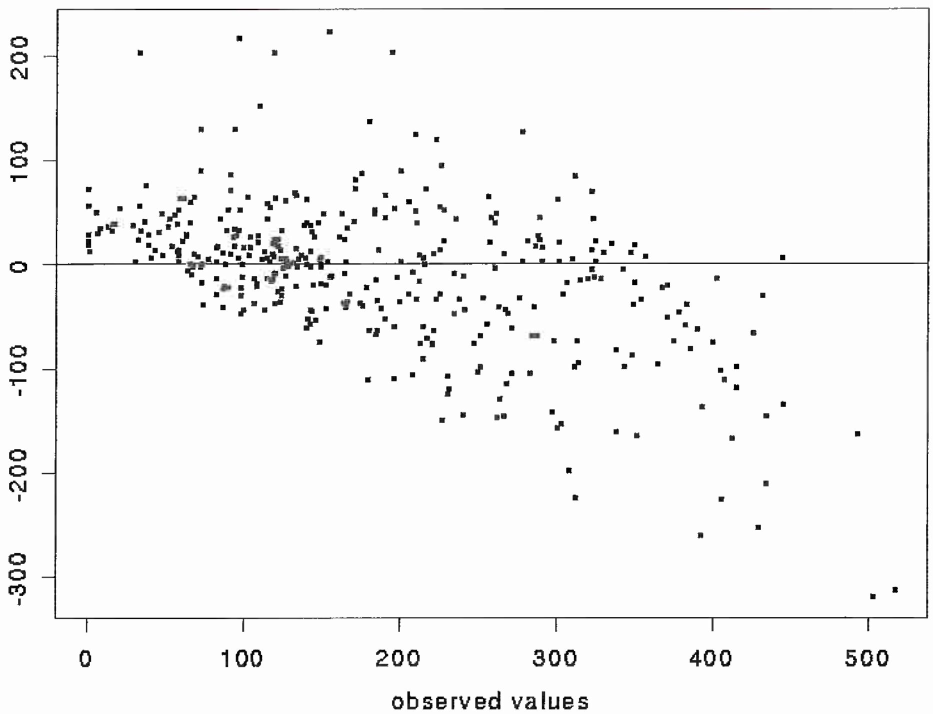

Figure 2. *Plot of observed (true) values (horizontal) against residuals (vertical).*

Figure 3 shows the maps for the rainfall prediction by simple good sense and the corresponding absolute errors.

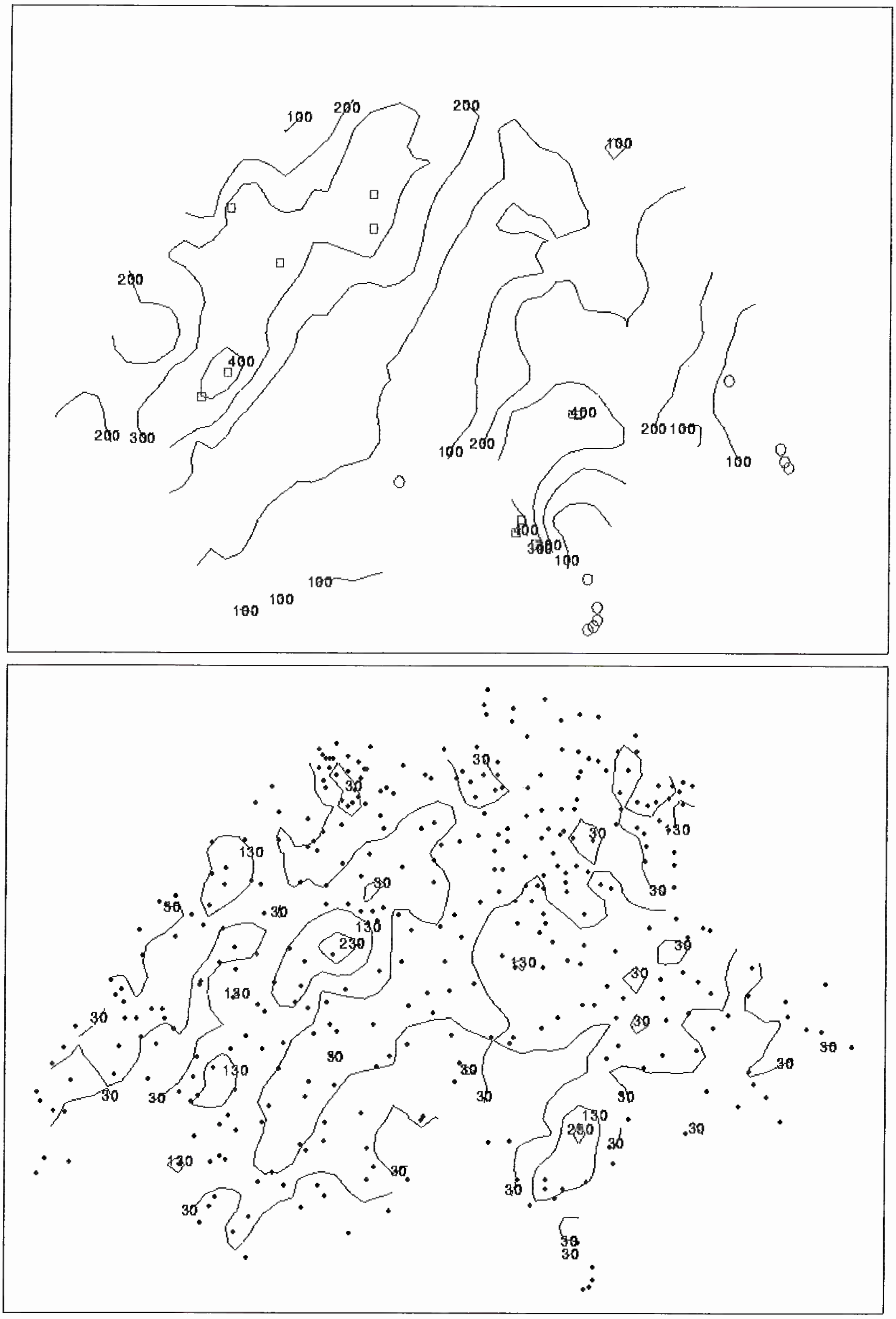

Figure 3. *Maps for the rainfall prediction by simple good sense (the 10 lowest values are represented by circles and the 10 highest by squares) and the corresponding absolute errors.*

Proportional plots of the absolute errors and relative errors in Figure 4 indicate the locations of the smaller or higher errors. It seems to be correlated with the smaller or higher rainfall measurements. The root mean squared error is *RMSE*=72 and should be compared with other predicting methods.

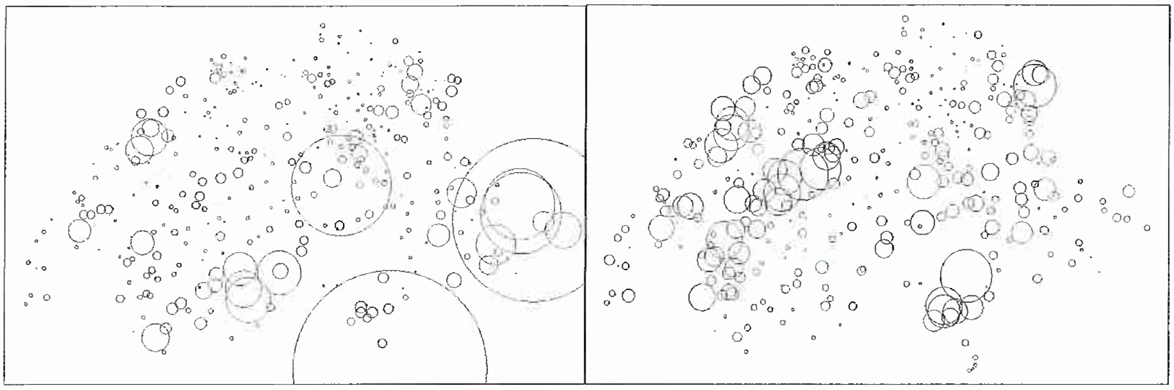

Figure 4. *Proportional plots of the absolute errors and relative errors for the simple good sense prediction of rainfall amounts. Positive errors are represented by red circles, and negative errors by blue circles.*

Table 2 compares the prediction of the ten lowest values and the ten highest values of the initial data set with the corresponding estimated values. This method identified four respectively five locations of the ten highest respectively ten lowest values of the initial data set, which are written in bold in Table 2

Ten lowest values		ten highest values	
true values	estimated values	true values	estimated values
0	56	434	288
0	**22**	434	224
0	**19**	**441**	**441**
0	**28**	**444**	**450**
0	72	445	311
1	**13**	**452**	**452**
5	55	493	330
6	36	503	185
8	43	517	204
10	**10**	**585**	**585**

Table 2. *This table compares the prediction of the ten lowest values and the ten highest values of the initial data set with the corresponding estimated values. The simple good sense method identified four respectively five locations of the ten highest respectively ten lowest values of the initial data set, which are written in bold.*

The performance in predicting the lowest and the highest 10 rainfall measurements can also be summarized by the root mean squared $RMSE_{min}$=16 and $RMSE_{max}$=10 respectively. It seems that higher values are more accurately predicted than lower ones. This method of rainfall prediction can be useful for the monitoring of accidental releases of radioactivity in the environment, because it doesn't require computations at all, but only some knowledge of the Swiss territory. The procedure is straightforward and almost automated. As rainfall is strongly correlated with radioactive fallout of accidental releases, this method can easily be used in emergency situations of nuclear accidents for making fast decisions. In the extreme case, it is sufficient to consult a meteorological map to predict the rainfall and winds. Using this information, citizens can be informed within a matter of hours. Of course, this *ad hoc* method is imprecise but much faster than any other.

4. CONCLUSIONS

In this paper, the use of simple good sense in the spatial prediction of rainfall measurements in Switzerland has been discussed. The method consists of a forecast based on the values of the available observations combined with topographic knowledge of the Swiss territory. Comparison of our subjective estimates with the true measurements has been completed and has yielded results which are not too far from the true values. Fortunately, the results of this method are worse than those from a robust spatial statistics methodology used by the authors in another analysis. This would lead to support that spatial statistics is worth the trouble. However, it would be very interesting to compare the simple good sense method with methods used by other participants, for example by mean of the RMSE. We believe that simple good sense would not be the worst.

References

1. Cressie, N. (1991). *Statistics for Spatial Data*, Wiley.
2. Englund, E. J. (1990). A Variance of Geostatisticians, *Mathematical Geology,* **22**, 417-455.
3. Furrer, R., Genton, M. G., (1998). Robust Spatial Data Analysis of Lake Geneva Sediment with S+ SpatialStats, *International Journal of Systems Research and Information Science*, special issue on Spatial Data Analysis and Modeling, in press.
4. Genton, M. G. (1996). "*Robustness in Variogram Estimation and Fitting in Geostatistics*", Ph.D. Thesis #1595, Department of Mathematics, Swiss Federal Institute of Technology.
5. Genton, M. G., (1998a): Highly Robust Variogram Estimation, *Mathematical Geology*, **30**, 213-221.
6. Genton, M. G., (1998b): Variogram Fitting by Generalized Least Squares Using an Explicit Formula for the Covariance Structure, *Mathematical Geology*, **30**, 323-345.
7. Genton, M. G., (1998c): Spatial Breakdown Point of Variogram Estimators, *Mathematical Geology*, **30**, 853-871.