Analysis of Environmental Data Using Spatial Statistics Methods

Krejčíř, Pavel
2000

INSTITUTE OF COMPUTER SCIENCE

ACADEMY OF SCIENCES OF THE CZECH REPUBLIC

Analysis of environmental data using spatial statistics methods

Pavel Krejčíř

Technical report No. 827

December 2000

Institute of Computer Science, Academy of Sciences of the Czech Republic
Pod vodárenskou věží 2, 182 07 Prague 8, Czech Republic
phone: (+4202) 6884244   fax: (+4202) 8585789
e-mail: krejcir@cs.cas.cz

INSTITUTE OF COMPUTER SCIENCE

ACADEMY OF SCIENCES OF THE CZECH REPUBLIC

Analysis of environmental data using spatial statistics methods

Pavel Krejčíř

Technical report No. 827
December 2000

Abstract

We describe the spatial statistics analysis done on the environmental data in this document. The analysis were carried out on meteorological and air pollution data. Temperature and ozone has been examined especially. The aim of the research was to explore the spatial covariance structure of the data in order to predict missing data or data at non-measured spots. The covariance structure is important for kriging, which is the best linear spatial predictor. We have also compared the results achieved by kriging to two ather methods.

Keywords
Statistics, spatial statistics, covariance function, correlation, variogram, quadratic beta-splines, ozone

1 Estimation of the Covariance Structure of the Spatial Data

1.1 Introduction

Despite of the results of the last report concerning with the spatial covariance structure of the meteorology or air pollution data, we have made a further progress on this field. We succeeded in calculation of variograms for ozone data and have been able to apply kriging for the prediction of the missing or unmeasured values.

We also have made some exploration of the spatial distribution data and checked the possible Gaussian distribution of the data. Finally we have been testing three methods for spatial prediction of the ozone data and the results are compared in this document.

1.2 Theoretical background

We have been dealing with the theory of spatial statistics in previous report. So let me just remind that kriging mean the best linear spatial prediction, for calculating of kriging we need to know the covariance structure of the underlying spatial stochastic process and the covariance structure is usually examined by means of variogram.

Variogram is function of the distance of two spatial point and in general it should look like the curve on the Fig. 1

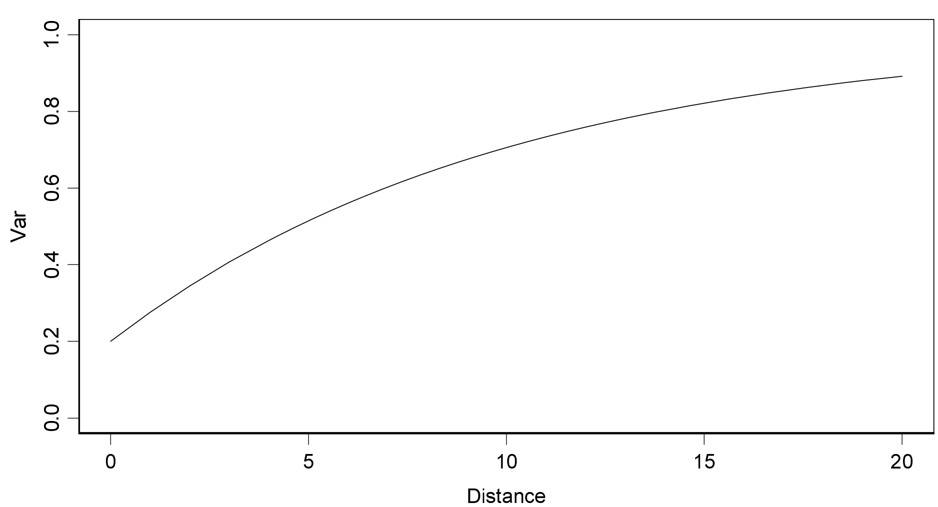

Fig. 1 *The general shape of the variogram*

The variogram is usually estimated by so called sample variogram. We have been calculating sample variograms and the results are discussed in the next section.

1.3 Ozone data covariance structure

We have chosen a part of the ozone data from August 1998 and tried to calculate the sample variograms. The result did not resemble the shape from the Fig. 1 much. We suspected that this was caused by the non-constant trend of the data, so we tried to estimate the trend using the quadratic beta splines and subtract it from the data. Then we calculated the variograms again. The result is summarized in the Fig. 2. The upper left plot shows the variogram with no trend estimated, the others represent the vartiograms, where trends were estimated and subtracted. The trends were estimated using quadratic Beta-splines and the number of splines in longitude and latitude direction respectively is shown in the caption of each plot.

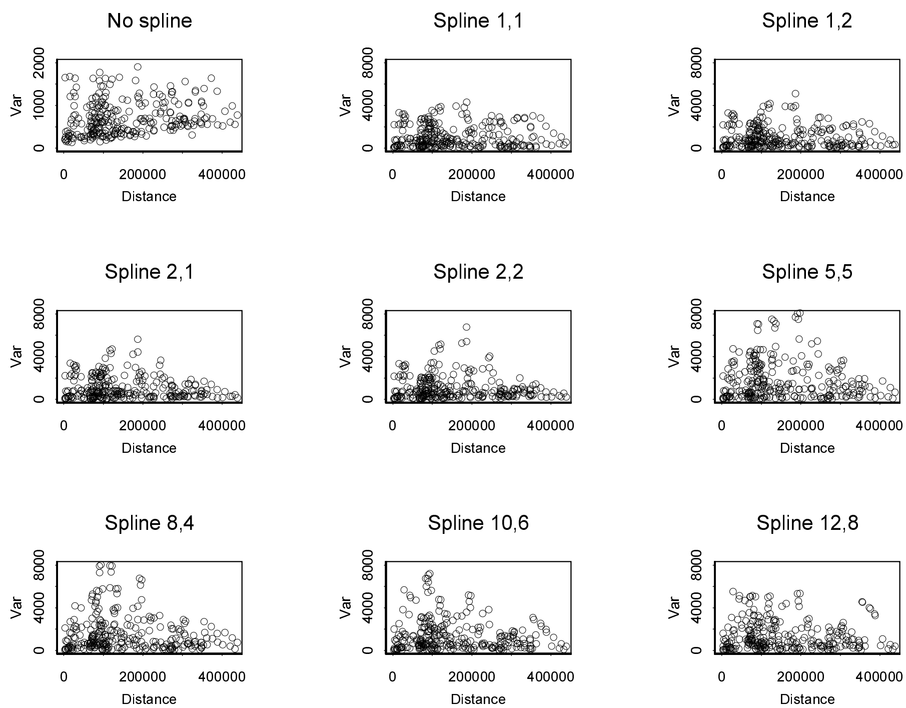

Fig. 2 *Sample variograms for ozone data from August 1998*

The sample variogram didn't allow us to estimate the model variogram, therefore we have calculated the sample correlations for all the stations and the results, ordered by the distance of the stations, are shown in the Fig. 3.

Form the correlation coefficients we have been able to select a set of stations with mutual high correlation coefficient and try to recalculate the variograms for the selected stations. The result is shown in the Fig. 4. The variograms with subtracted trend were not useful yet, but it was possible to use the variogram without trend. We have estimated the spherical model variogram of this data and the result is shown in the Fig. 5.

We tried to fit several variogram models, but only the spherical fit could have been solved.

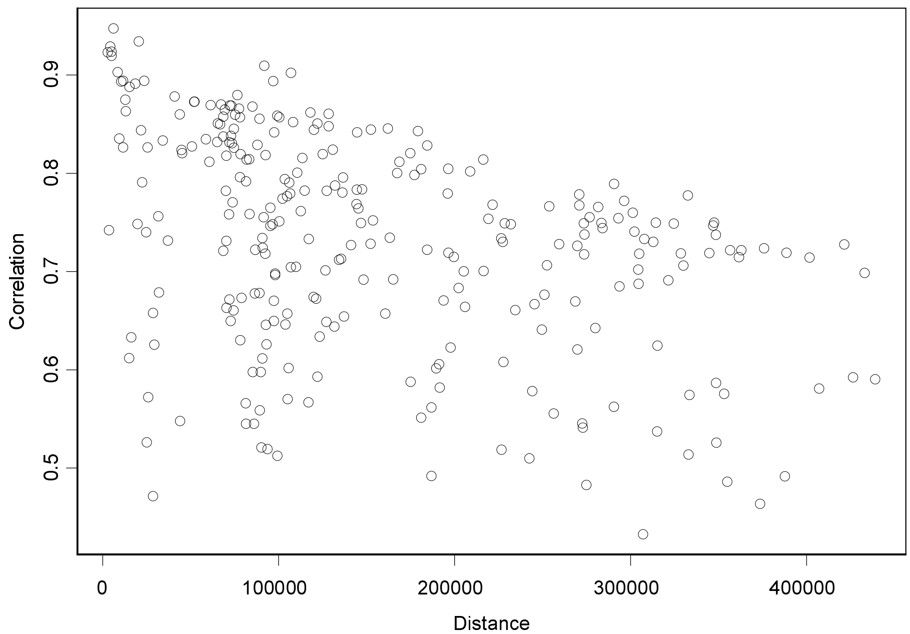

Fig. 3 *Correlation coefficients for the ozone data*

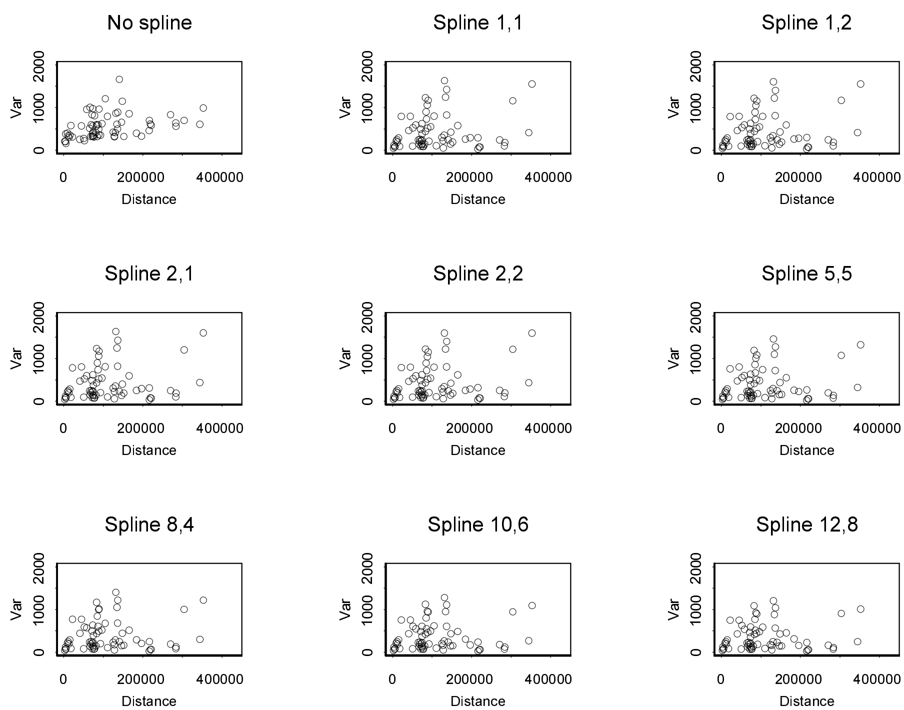

Fig. 4 *Variograms calculated for the selected stations*

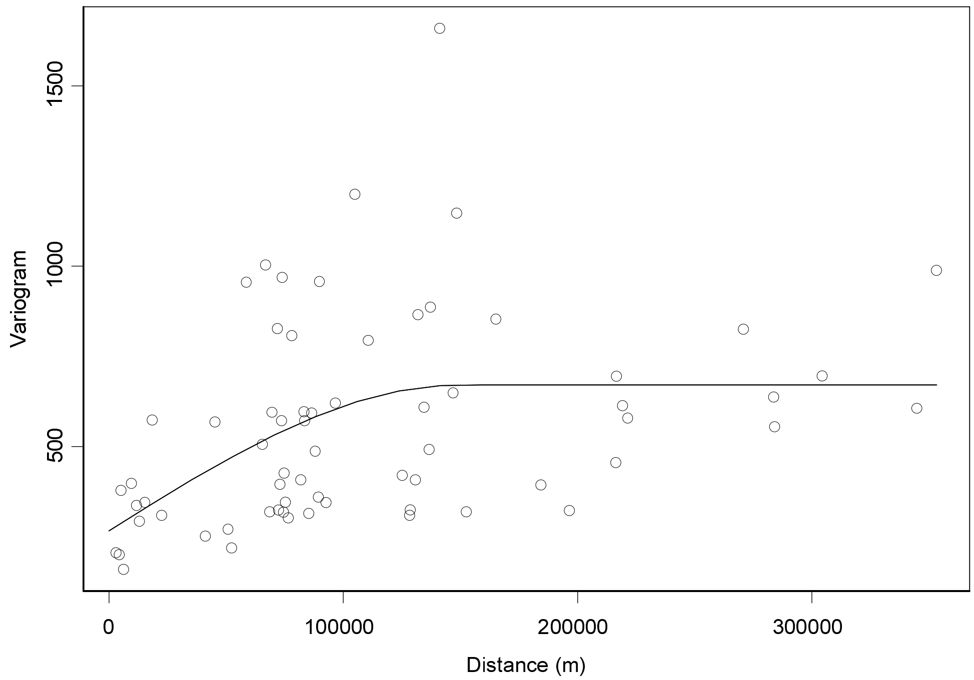

Fig. 5 *The spherical variogram model*

1.4 Normality check

We have also been calculating the Chi square goodness of fit test of the normality of the ozone data. The normality is important for the model and procedures concerning with it, e.g. variogram, kriging. The normality has not been confirmed by the test, but the sample of the station was not large enough. Some histograms are shown in the Fig. 6.

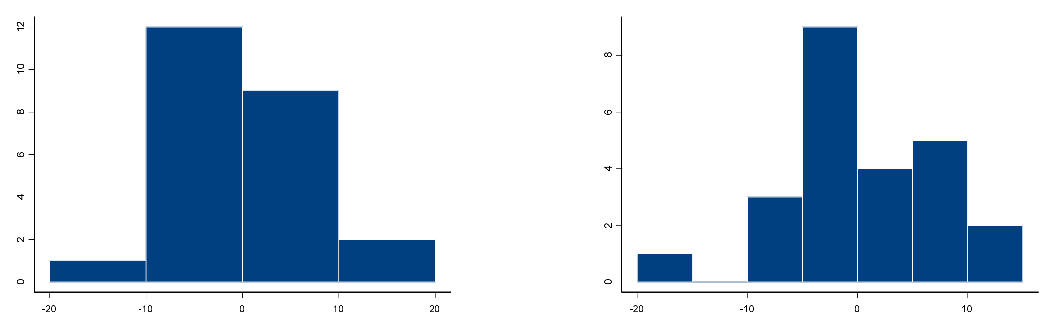

Fig. 6 *Histogram of the ozone data from 1.8.1998, 12:00*

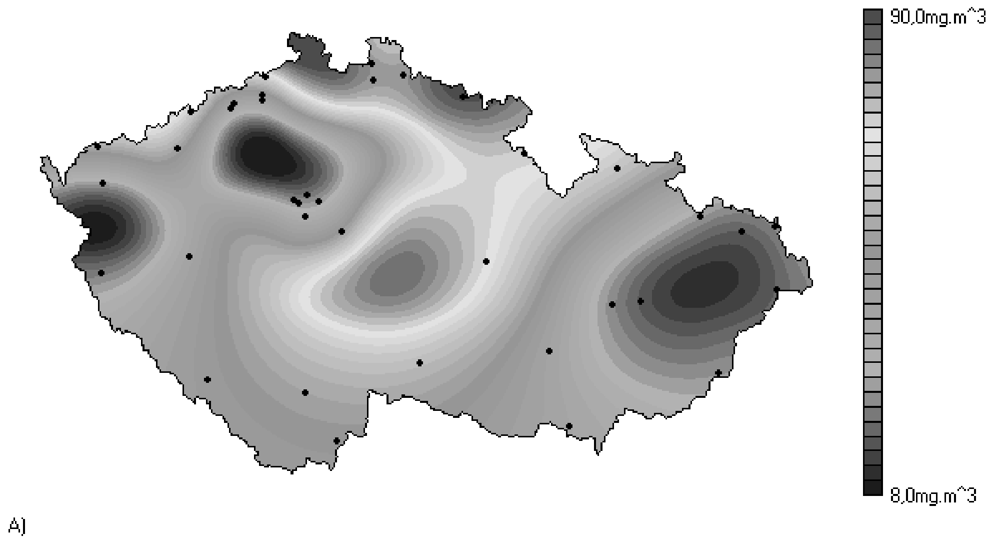

Fig. 7 *Map of the ozone concentrations created by the spline method*

1.5 Methods of spatial prediciton

In order to choose an appropriate method for spatial prediction of the data, we decided to test three possible methods. The methods were as follows:

1. Nearest neighbour method
2. Spline method
3. Kriging

We have calculated prediction for each station from the rest of stations with measurements and for several time snaps. Then we evaluated the results by cross validation method, where all the predictions were compared to the known true values. The results are shown in the Tab. 1 in form of mean sum of squares.

Method	Nearest neighbour	Kriging	Spline	Spline + Kriging
Mean sum of sq.	462.22	367.33	405.66	1589.52

Tab. I *Comparison of tested methods*

We can see that the results are comparable however, the kriging provides the best fit. In this case the mean error represents 12.5% of the data range, which varies over observed time and space approximately from 10 to 160 (μg/m^3).

In the end the spline method and kriging are visually compared at the Figs. 7 and 8.

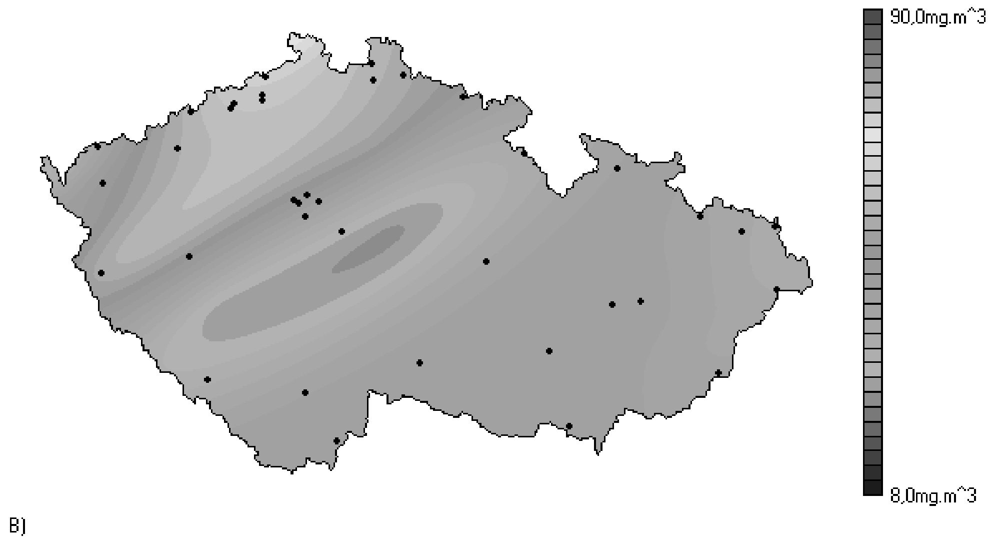

Fig. 8 *Map of the ozone concentrations created by krigingu*

1.6 Conclusion

We have been examining the covariance structure of the spatial ozone data during the last period. We have been able to calculate kriging spatial prediction and compare it to other useful methods. The kriging provides best spatial prediction from three tested method, however we believe the results are not optimal yet.

In the future we would like to improve the spatial prediction in two ways:

1. By better splitting the stations into several categories with high correlation and hence the better variogram estimators.

2. To introduce the temporal characteristic into the prediction methods.