GI DA JOURNAL OF GEOGRAPHIC INFORMATION AND DECISION ANALYSIS

Journal of Geographic Information and Decision Analysis, vol. 2, no. 2, pp. 65-76, 1998

Mapping Precipitation in Switzerland with Ordinary and Indicator Kriging

Peter M. Atkinson

Department of Geography, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom

pma@soton.ac.uk

Chris D. Lloyd

Department of Geography, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom

cdl195@soton.ac.uk

ABSTRACT **This paper uses the geostatistical methods of ordinary kriging (OK) and indicator kriging (IK) to address the problem of estimating values of precipitation at locations from which measurements have not been taken. Several problems or issues were raised including: (i) lognormality of the data, (ii) non-stationarity of the data and (iii) anisotropy of the spatial continuity. Given that the aim of SIC'97 was to compare a variety of different approaches to estimation. IK (informed using directional indicator variogram models) was selected because it is a means to account for lognormality and it was a method that was unlikely to be used widely within the competition. Accuracy of estimates made using IK were compared with OK estimates. It was observed that the OK algorithm, as implemented here, provided more accurate estimates than IK. This was considered to be due, at least in part, to the method used for tail extrapolation and also the small number of data used in estimation (100 data locations). OK was recommended over IK in this instance as OK provided more accurate estimates and was also easier to implement.**

KEYWORDS: **indicator kriging, ordinary kriging, precipitation**.

ACKNOWLEDGEMENTS The comments of two anonymous referees are gratefully acknowledged.

Contents

1. Introduction

We find ... there is not a single rule, however plausible, and however firmly grounded ... that is not violated at some time or another (Feyerabend, 1975, p. 23)

As a subject such as spatial interpolation reaches maturity it can often be difficult to condense the broad knowledge and experience of many practitioners into a single volume so that novice practitioners can make reasoned judgments about the steps which they should employ. We hope that SIC'97 will provide some insights into the multi-dimensional decisions (often based on judgment) which must be made when embarking on a spatial interpolation problem, specifically by allowing comparisons between different techniques, approaches, decisions and so on. It is with the above knowledge in mind that we are happy to report our method and in doing so reveal our errors.

At first we chose ordinary kriging (OK) as a standard technique for spatial interpolation. The results obtained from OK are described below in section 3. We felt reasonably confident about this technique and, therefore, the results produced. However, with the spirit of the competition in mind we wished to select a technique for spatial interpolation which might be less commonly employed by other contributors. Thus, although we were less confident about the results, we based our spatial interpolation on indicator kriging (IK, section 4). In section 2 below we describe the rationale for choosing the IK approach.

2. Initial data exploration

Our initial exploration of the data revealed five main issues which we wished to address in our choice of technique for spatial interpolation.

2.1 Lognormality of the data

It was quite obvious from the start that the 100 sample data were approximately lognormally distributed (Figure 1). Ordinary kriging is quite robust and so there is some potential for applying OK without modification even when the data do not have a normal distribution. However, there are several alternatives. The most commonly employed is to transform the data to a normal distribution, undertake OK, and then apply a back transform (not the anti-log). This approach has several disadvantages, the main one being that the back-transform can introduce uncertainty to some values. An alternative approach, and the one which we adopted, is IK. Since IK works by decomposing the variable of interest into several binary variables, and in so doing decomposing the univariate distribution function (histogram) into several classes, the dependence on a normal distribution disappears (Journel, 1984). It is mainly for this reason that we decided to adopt IK.

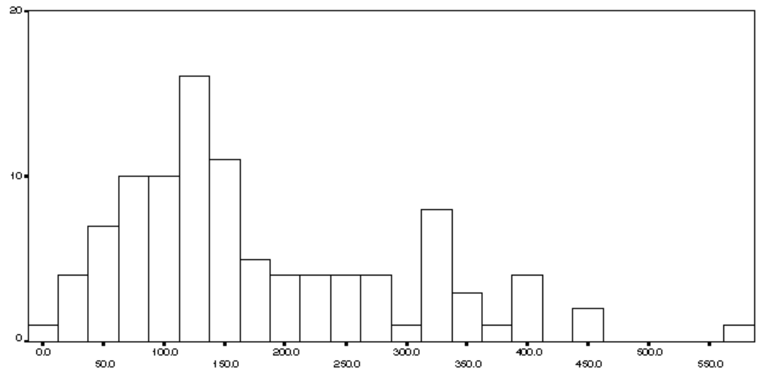


Figure 1. Histogram of the 100 sample data. Units are given in tenths of millimetres.

2.2 Non-stationarity of the data

The maps produced initially using straightforward techniques such as inverse distance weighting squared interpolation indicated that there may be some justification to adopting a non-stationary model of the spatial variation in precipitation. Of course, whether a variable is modelled as stationary or otherwise depends among other things on the scale of the analysis and the choice of the investigator. The justification for our choice was that the variation appeared relatively smooth (indicating that a deterministic model may be appropriate).

We tried to fit several three-dimensional polynomials of order up to and including three, but none provided a satisfactory fit. A higher order polynomial may well have provided a better fit. However, we felt that the need for such high order polynomials indicated that a stationary model may well be appropriate despite the fact that the variation was apparently smooth. We also considered segmenting the region into two or more areas of like variation (mean values within a neighbourhood). However, with only 100 sample observations in total, segmentation would reduce these numbers further and make the characterization of spatial variation necessary for kriging unreliable. For the above reasons we adopted a stationary model.

2.3 Anisotropy of the spatial continuity

As with all standard mapping applications of the kriging family of algorithms it was necessary to check for anisotropy. An initial simple structural analysis revealed the directions of maximum and minimum variation (geometric anisotropy) to be 45° and 135° approximately. Therefore, it was necessary to model this anisotropy and include the model in any technique chosen for spatial interpolation.

2.4 Lack of correlation with the elevation data

Since the original 100 sample data were supplied with a co-registered digital elevation model (DEM) of Switzerland we considered using regression type statistical techniques such as simple regression, co-kriging and artificial neural networks. However, despite our efforts we did not find any satisfactory relation between precipitation and elevation.

Given that the spatial distribution of precipitation is likely to have been driven by wind orientation and relief (with rain occurring most on the windward side of slopes and least on the leeward side) we decided to estimate slope aspect as a potential covariate from the elevation data. However, there was little observable relation which persisted across the whole region. The assessment of correlation between variables in this manner is problematical but although the relationship may have been non-linear no obvious association of any kind was observed. The principal problem may be related to scale. For example, Daly *et al*. (1994) observed large positive correlations between elevation and precipitation for the Willamette River basin in Oregon, but not for the western United States as a whole. Even when we smoothed the elevation and aspect maps to attempt to match the scale of the processes which result in precipitation, little relation was observed. There did appear to be two different relations, one for elevated areas and one for low-lying areas, but modelling them separately would have meant segmenting the region. Given the small number of sample data we decided against this.

2.5 Preferential clustering of the data

Depending on the technique used for spatial interpolation it may sometimes be necessary to decluster the data where they are preferentially located in areas of large or small values. Since this is necessary for IK, we chose to decluster the data (decreasing the mean from 180.15 to 173.78), using the GSLIB routine declus (Deutsch and Journel, 1992), to estimate the form of the histogram prior to applying the IK algorithm (Figure 2).

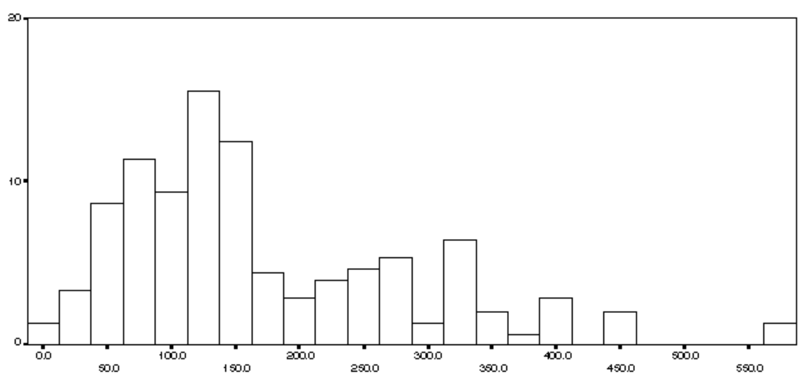
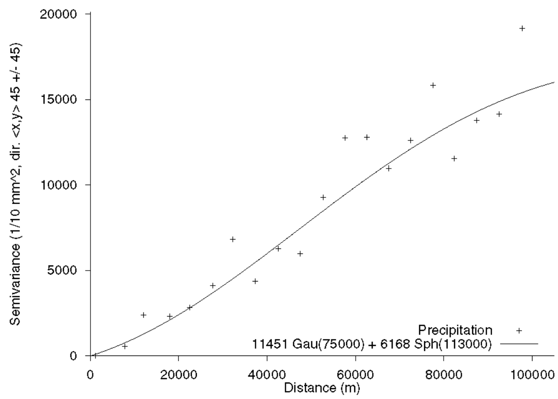


Figure 2. Histogram of the 100 declustered data. Units are given in tenths of millimetres.

3. Ordinary kriging

Ordinary kriging has been described as the 'anchor algorithm of geostatistics' (Deutsch and Journel, 1992, p. 64) because of its remarkable robustness under a range of conditions. On account of this robustness we decided to apply OK in the first instance to map precipitation in Switzerland at the 367 locations for which values were held back.

Directional variograms were estimated from the sample data and the directions of maximum and minimum variation (geometric anisotropy) were estimated as 45° and 135° approximately. Sample variograms were estimated for these directions and these were fitted with a Gaussian plus spherical model using the weighted least squares functionality of the GSTAT software (Pebesma and Wesseling, 1998). The coefficients were subsequently modified by eye (Figure 3 and 4).



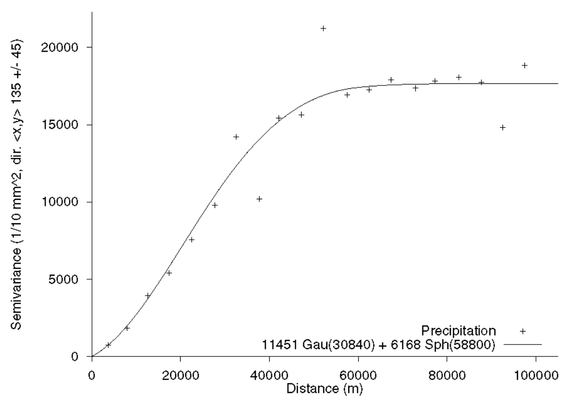


Figure 3. Sample variogram for 45° (+ symbols) with fitted model (solid curve).

Figure 4. Sample variogram for 135° (+ symbols) with fitted model (solid curve).

The above variogram models were used in OK (using the GSTAT software) to map precipitation at the unobserved 367 locations and the remainder of the study area. We chose a search radius of 45 km, and minimum and maximum numbers of data to use in kriging of 1 and 16. The isolines for the kriged estimates are shown in Figure 5. Figure 6 maps the errors from the OK estimates. There is no clear pattern of under or over estimation suggesting that even if a trend model had been used it would probably not have increased significantly the accuracy of OK.

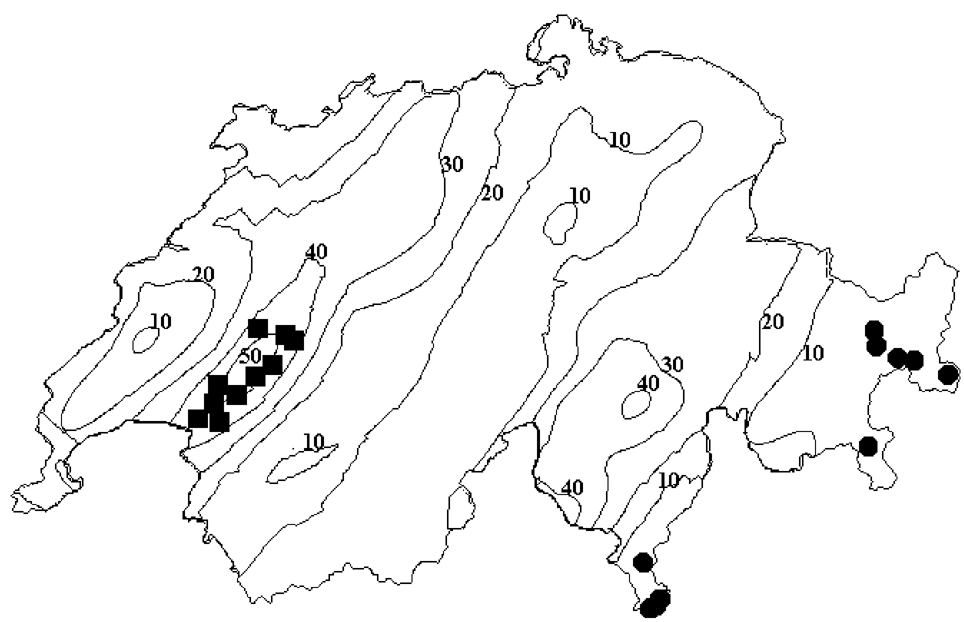


Figure 5. Isolines of OK estimates, with a 10 mm interval. The squares and circles show the locations of the ten maximum and ten minimum estimates respectively.

4. Indicator kriging

Despite our relative confidence in the OK algorithm we decided to apply IK for the reasons given in section 2. First, we declustered the data to obtain a slightly modified histogram as described in section 2.5 (Figure 2). This distribution was divided subsequently with nine cut-offs, (that is, we chose cut-offs based on the nine deciles of the distribution) and these cut-offs were applied to the sample data to estimate indicator variograms using the GSLIB software (Deutsch and Journel, 1992). In retrospect, we feel that we were expecting too much of the data by dividing the distribution into so many classes because there were only 100 observed values in the sample. Nevertheless, the variograms obtained appeared to be fairly well behaved, exhibiting the kind of variation for each cut-off that we might expect.

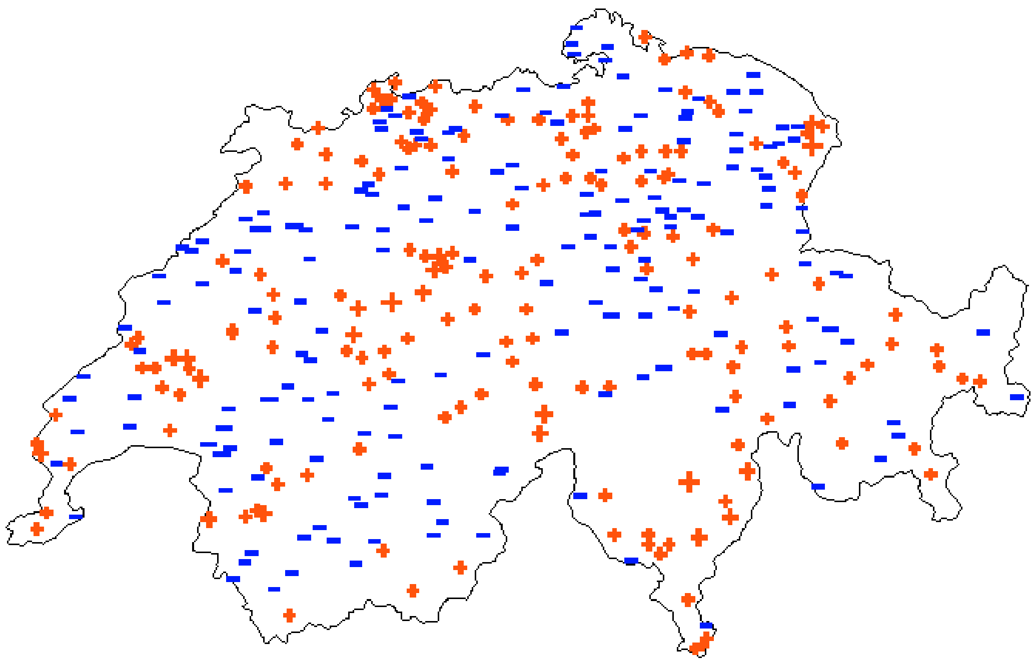
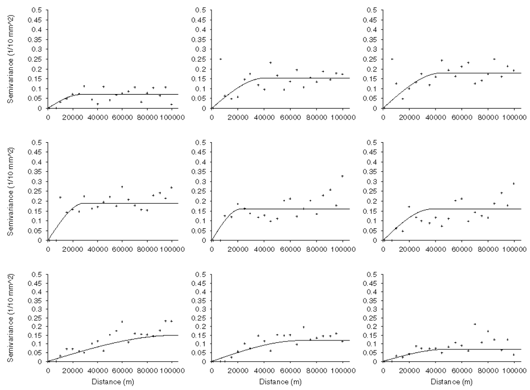


Figure 6. Map of OK errors. The + and – symbols show the locations of over and under estimation respectively.

As for OK we wished to model the obvious anisotropy in the variable of interest. We adhered to the orientations of maximum and minimum variation found for OK to keep the analysis simple. The directional variograms for each of the nine deciles of the histogram are shown in Figures 7 and 8. These variograms were fitted with a variety of models (shown in the figures), the coefficients of which are given in Table 1. The means of the two sills for each cut-off were obtained and the anisotropy was modelled as geometric.



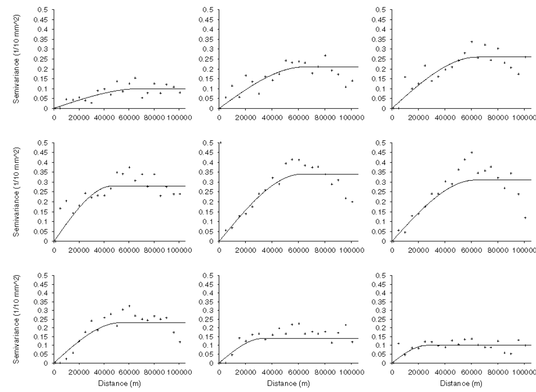


Figure 7. Sample indicator variograms for 45° (+ symbols) with fitted models (solid curves).

Figure 8. Sample indicator variograms for 135° (+ symbols) with fitted models (solid curves).

Table 1. Indicator variogram model coefficients for 45° and 135°.

45°			135°		
a_1	Model	c_1	a_1	Model	c_1
0.070	Sph	26750.8	0.100	Sph	69551.3
0.154	Sph	40729.9	0.210	Sph	64587
0.187	Sph	44882.1	0.260	Sph	64502.1
0.192	Sph	29430.5	0.285	Sph	46560.9
0.167	Sph	22196.2	0.341	Sph	62393.2
0.160	Sph	37392.1	0.314	Sph	62756.2
0.155	Sph	102875	0.239	Sph	53176
0.120	Sph	68383.5	0.140	Sph	32996
0.075	Sph	48854	0.100	Sph	26250.6

Table 2. Summary statistics for 367 data. All values are given in tenths of millimetres.

	Minimum	Maximum	Mean	Median	Std. dev.
Observed values	0	517	185.359	162.000	111.015
OK estimates	–27.92	510.67	181.87	154.387	107.15
IK estimates	29.549	489.625	186.911	162.485	100.707

The IK algorithm provided in GSLIB was used with some minor modification to estimate the values at the 367 unobserved locations from the 100 sample data. We chose a search radius of 45 km and minimum and maximum numbers of data to use in kriging of 1 and 16 for each decile. In the absence of prior knowledge we chose the linear method of estimating the tails of the distribution, although these choices may have been sub-optimal (see for example, Goovaerts, 1997). The isolines for the IK estimates are shown in Figure 9.

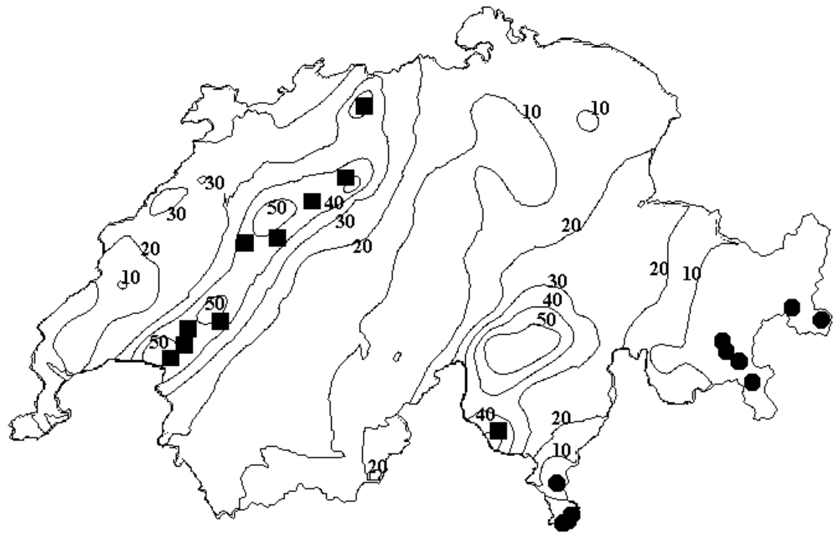


Figure 9. Isolines of IK estimates, with a 10 mm interval. The squares and circles show the locations of the ten maximum and ten minimum estimates respectively.

5. Assessing the estimates

Five summary statistics are given in Table 2 for both the observed values, the OK estimates and the IK estimates. Clearly, IK has underestimated the maximum values and overestimated the minimum values as one would expect of a weighted averaging technique (see also the standard deviation). That IK has larger errors than OK is to some extent disappointing given the extra effort required for IK, but is likely to be due to having to extrapolate to estimate the tails of the histogram and to the small number of data available. Also, the use of linear extrapolation for estimating the tails may have been a sub-optimal approach. The maximum OK estimate is closer to the observed maximum but the fact that OK has produced negative estimates is an obvious problem.

The histogram of the IK errors was approximately normally distributed and the mean, 0.155 mm, was closer to zero than that for OK (–0.349 mm) (Figure 10 and Table 3). The proportion of large errors was greater for IK (hence the larger standard deviation of the error distribution for IK (6 mm) compared to OK (5.96 mm)).

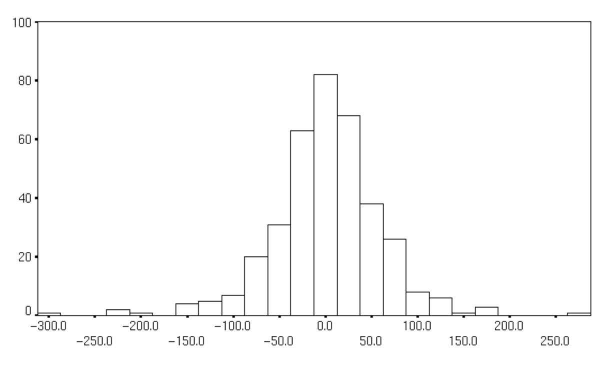


Figure 10. Histogram of the IK errors. Units are given in tenths of millimetres.

The root mean square error (RMSE) for estimates made using IK was 6 mm whereas for OK it was 5.97 mm. Though the difference was not great the OK errors were clearly smaller as a whole than those for IK. That OK achieved estimates closer to the observed values than IK is again illustrated by the mean absolute error (MAE). For IK the MAE was 4.26 mm whereas for OK it was 4.1 mm. Dividing the absolute errors by the observed values gives the mean relative error (MRE). For IK the MRE was 0.051 whereas for OK it was 0.034. The MRE was calculated after the removal of the five observed values of zero to avoid contributions of infinity to the sum.

Table 3. Summary statistics for the observed values minus the OK and IK estimates. All values are given in tenths of millimetres.

	Minimum	Maximum	Mean	Median	St. dev.
Observed–OK	–263.54	349.08	–3.49	0.688	59.69
Observed–IK	–288.82	287.02	1.552	4.591	60.04

Table 4. Number of locations of the 10 minimum and maximum actual values that are also among the locations of the estimated 10 minimum and maximum values. Kriging honours values in the sample of 100 observations and these are tabulated separately.

Technique	Maximum 10 estimated	Maximum 10 honored	Maximum 10 estimated	Maximum 10 honoured
OK	6	1	2	3
IK	3	1	3	3

Table 5. The ten largest and smallest values and their IK estimates. All values are given in tenths of millimetres.

Observed	OK	IK	Observed	OK	IK
0.00	54.84	83.24	426	301.13	285.36
0.00	37.79	45.42	429	239.78	226.09
0.00	33.37	38.86	432	489.52	442.97
0.00	–4.52	122.89	434	399.46	453.90
0.00	–6.23	74.44	434	328.99	331.55
1.00	–5.78	42.27	444	408.13	462.27
5.00	–19.65	67.01	445	375.84	433.77
6.00	–25.94	70.25	493	460.87	397.74
8.00	–27.92	70.87	503	268.28	267.33
13.00	32.67	72.35	517	253.46	228.18

The errors of the IK estimates are plotted against the observed values in Figure 11. The r for IK was 0.433 whereas for OK it was 0.333. While the correlation coefficient indicates little correlation Figure 11 illustrates a tendency for the IK errors to decrease with an increase in the observed values.

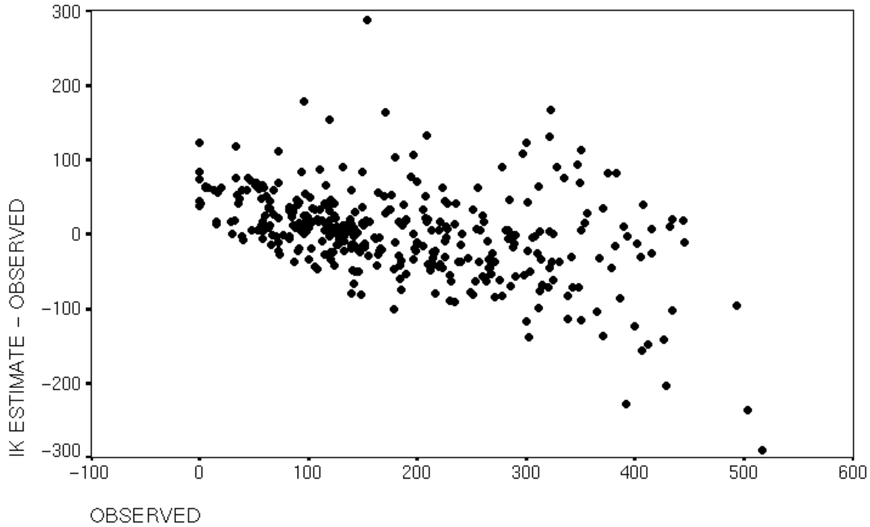


Figure 11. The IK errors plotted against the observed values. Units are given in tenths of millimetres.

The scatterplot of the observed values against the IK estimates is given in Figure 12. The *r* for IK is 0.842 whereas the value for OK is 0.85.

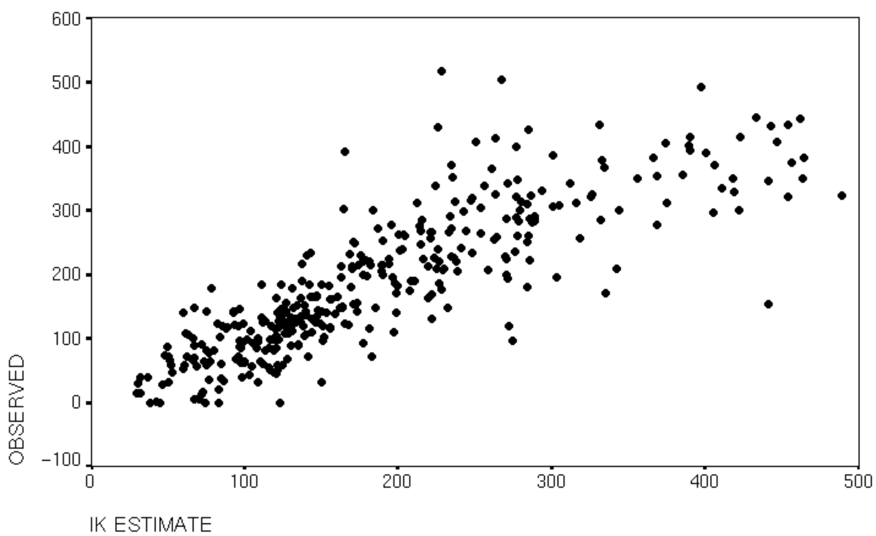


Figure 12. The observed values plotted against the IK estimates. Units are given in tenths of millimetres.

Isolines of the IK estimates are given in Figure 9. The distribution of the highest and lowest estimates are notably different for OK (Figure 5) and IK (Figure 9). The maximum positive and negative errors are given in Table 6. The errors for both OK and IK are large and indicate that neither technique can be considered to have estimated accurately locally.

Table 6. The ten maximum negative and positive errors using IK. Errors are given in tenths of millimetres.

ID	OK negative errors	ID	OK positive errors	ID	IK negative errors	ID	IK positive errors

150	–263.54	63	349.08	150	–288.82	63	287.02
437	–234.99	70	221.47	350	–235.67	70	179.01
350	–234.72	58	175.99	437	–226.73	44	166.63
171	–189.22	327	148.07	171	–202.91	327	164.22
438	–151.85	88	147.05	114	–155.05	58	153.56
385	–146.07	47	136.62	184	–148.21	306	133.27
38	–144.72	44	134.10	189	–140.64	175	132.04
361	–141.33	445	114.40	438	–138.25	367	122.89
184	–134.39	288	106.14	41	–136.23	98	122.62
189	–124.87	76	103.67	161	–123.11	288	117.55

Figure 13 shows the map of the IK errors. As for the map of OK errors (Figure 6) there is no clear pattern of over or under estimation across the region. This suggests that, whilst far from ideal, a stationary model was acceptable in this situation.

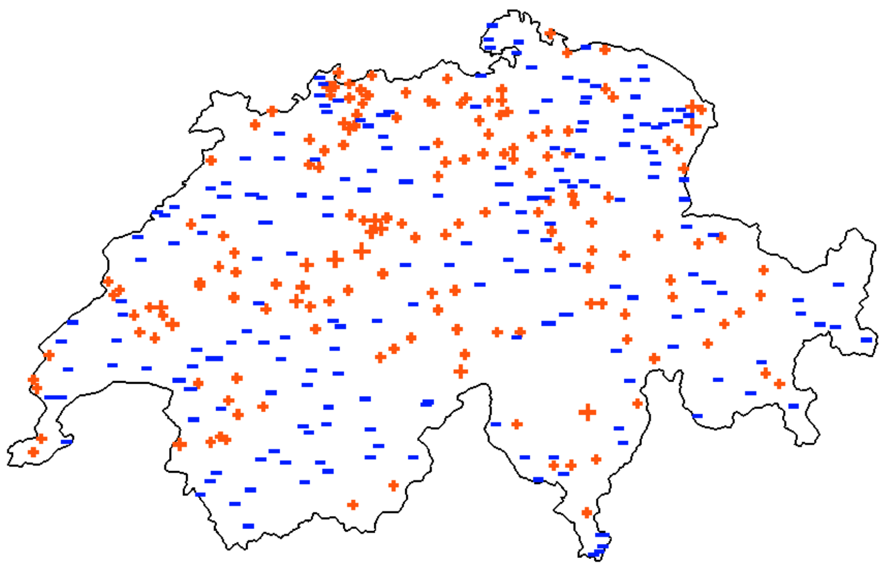


Figure 13. Map of IK errors. The + and – symbols show the locations of over and under estimation respectively.

6. Summary

The summary statistics presented consistently indicated that the simpler OK algorithm performed more accurately than IK. For mapping in an emergency we would recommend OK because it is relatively quick to implement and in any case was more accurate. However, it is clear that neither OK nor the primary technique used, IK, could be considered accurate estimators of precipitation. It has been noted with respect to IK that this is probably partly due to over-optimistic use of IK with a small data set. Additionally, the use of other methods for tail extrapolation may have improved the estimates. However, we anticipate that any conventional geostatistical approach would be likely to produce similar results.

References

Daly, C., Neilson, R. P. and Phillips, D. L. (1994) A statistical-topographic model for mapping climatological precipitation over mountainous terrain, *Journal of Applied Meteorology*, 33; 140–158.
Deutsch, C. V. and A. G. Journel (1992) *GSLIB: Geostatistical Software Library and User's Guide.* New York: Oxford University Press.
Feyerabend, P. (1975) *Against method*. London: Verso.
Goovaerts, P. (1997) *Geostatistics for Natural Resources Evaluation.* New York: Oxford University Press.
Journel, A. G. (1984) *Indicator Approach to Toxic Chemical Sites*. Project No. CR-811235-02-0. Stanford, CA: Leland Stanford Jr. University and U. S. Environmental Protection Agency.
Pebesma, E. J. and C. G. Wesseling (1998) GSTAT, a program for geostatistical modelling, prediction and simulation, *Computers and Geosciences*, 24; 17–31.

JGIDA vol. 2, no. 2
JGIDA Home