Georisk
2008, 1–14, iFirst article

Geostatistics for automatic estimation of environmental variables-some simple solutions

J.O. Skøien[a]*, E.J. Pebesma[b] and G. Blöschl[c]

[a]*Department of Physical Geography, University of Utrecht, Utrecht, The Netherlands;* [b]*Institut für Geoinformatik, Westfälische Wilhelms-Universität Münster, Münster, Germany;* [c]*Institute for Engineering hydrology and Water Resources Management, Vienna University of Technology, Vienna, Austria*

During emergency situations (e.g. pollution peaks, nuclear/radiological accidents, flash-floods) it will often be helpful for decision makers to have maps of the situation. Production of these maps has to be based on automatic procedures, as there will be limited time to analyse the problem. These methods have to be both quick and robust. Although the solutions to such problems tend to be complex, it is easy to forget that simpler methods can be useful. This paper examines some simple geostatistical solutions to two complex mapping problems, showing that these methods can be useful either as part of an automatic mapping procedure, for identification of the most important issues of the method, or as benchmarks for more complex solutions. The first problem is the emergency data set from the SIC2004 exercise, whereas the second problem is related to estimating runoff in flood situations. Two methods are examined-linear variogram and the use of the same variogram in extreme situations as in routine situations. The results from the SIC2004 data indicate that the anisotropy is the most important factor in this case. Both linear variograms and variograms from routine situations give reasonably good results, compared to more sophisticated methods submitted through the exercise. The results from the Austrian runoff data set also indicate that variograms from routine situations can be applied in extreme situations, with reasonably good results. The use of a routing model did not improve the results, and indicated that the flow velocities in flood situations need to be assessed locally.

Keywords: natural hazards; kriging; floods; early warning systems; hazard mapping

Introduction

In case of hazards and emergencies (e.g. pollution peaks, nuclear/radiological accidents, flash-floods), decision managers need updated information about the spatial distribution of the environmental variable of interest. This information will for many variables be given as point observations, which are difficult to interpret. To be better able to asset the situation, the decision makers will benefit from having maps describing the situation as close to real-time as possible. These real-time maps have to be interpolated from the point observations of a monitoring network, using the most recent observations.

Geostatistics provide a set of tools for interpolating maps from a set of point observations (*Cressie,* 1991; *Isaaks and Srivastava,* 1989; *Journel and Huijbregts,* 1978). Although these methods are supposed to give the best linear unbiased estimates (BLUE) (*Journel and Huijbregts,* 1978), they have only to a limited extent been taken into use as tools for real-time mapping. Many geostatisticians will claim that the methods are not reliable without human intervention. The largest problem lies in the fitting of a theoretical variogram model to the sample variogram. Webster (2001) argue for automatic calibration, but suggest that the estimated variogram should be visually examined as well. A number of authors have, however, provided arguments in favour of visual fitting (*Ai-Geostats,* 2004). As visual fitting of variograms is not a feasible approach in an emergency situation due to time constraints, automatic methods have to be used, despite the fact that visual fitting could have improved the maps.

The project Interoperability and Automated Mapping (INTAMAP) started late 2006 and intends to develop an interoperable framework for real time automatic mapping of critical environmental variables by extending spatial statistical methods and employing open, web-based, data exchange and visualization tools. See www.intamap.org for more information. The first case focuses on data from the data base of gamma dose rates in Europe – EURDEP (*De Cort and De Vries,* 1997).

A small subset of this data set was used in the Spatial Interpolation Comparison from 2004 (SIC2004) (*Dubois and Galmarini,* 2005a; b), together with a synthetic data set. Given several sets of routine observations for training, the participants were to

*Corresponding author. Email: j.skoien@geo.uu.nl

ISSN 1749-9518 print/ISSN 1749-9526 online

DOI: 10.1080/17499510802086769
http://www.informaworld.com

develop spatial interpolation for real-time mapping of emergency data. The participants had to submit the results from their real-time mapping short after the emergency data were released, to make sure that the methods could really work in emergency situation.

The added difficulties when modelling an emergency situation is one of the important issues of the INTAMAP project, similar to the SIC2004 exercise. It is typical for an emergency situation that the assumptions of different degrees of stationarity or intrinsic hypothesis necessary for geostatistics are violated to some degree. Methods that perform relatively well in routine situations are not necessarily the best methods for mapping in emergency situations.

As the problem is complex, it is tempting to develop complex solutions. However, an emergency situation is in some cases also often characterized by a rather small number of extreme observations, which easily leads to overfitting when a theoretical correlation model is fitted to the observed data. The use of sample variograms in geostatistics can easily give the impression that the amount of information is larger than it actually is, as will be shown later in this paper.

The objective of this paper is to reconsider the information actually available in extreme situations and to examine some simple solutions to the real-time mapping problem. It is not claimed that these methods are advisable to use for all processes and situations. The methods are applied on the SIC2004 data set and compared to the results of the participants in the exercise. As we had access to the validation results before we interpolated the emergency case from the exercise, we do not claim that we could have achieved a certain result in the exercise. The aim is more to see which results possibly could have been achieved with these simple methods, and to use the methods to identify which parts of the method it is necessary to elaborate further. The aim is also to consider what kind of information that is really available in the data set. In a second example to illustrate the use of simple methods is one of the methods also applied on runoff data from Austria. Estimation of variograms for runoff using the top-kriging method (Skøien and Blöschl, 2007) is too time-consuming for real-time mapping of the runoff situation in in flood situations, and simplifications are necessary.

Data

SIC2004

We have used two sources of data in this paper. The first set of data is from the SIC2004 exercise (*Dubois and Galmarini,* 2005a; b). 1008 stations in a rectangular area from the German gamma monitoring network were selected as the observation locations for the test. The data set was divided into two groups, the test data (200 stations) and the validation data (the remaining 808 stations).

Routine situation data for the 200 stations were released during the training phase for the exercise, when the participants were to develop their methods. An 11th data set was given as the exercise, covering the same 200 stations from the network. In this data set, a release of nuclear material had been simulated, giving considerably higher gamma dose rates at two of the observation locations. The task of the participants was, as good as possible, to interpolate the gamma dose rates levels in this emergency situation at the 808 remaining locations. The different data sets are presented in Figure 1. The validation data were available after the exercise had taken place.

Only two stations in the emergency situation exhibit larger values than in the routine situation, as can be noted from the central pane in Figure 1. However, the observations at these two stations were considerably larger than those of their neighbours, i.e. 1070 nSv/h and 1499 nSv/h respectively, compared to observations in the range 58–196 nSv/h for the remaining 198 stations. The standard deviation for these 198 stations was 19.6 nSv/h.

Austrian runoff data

The second data set in this paper stem from a comprehensive hydrographic data set of Austria. Austria has a varied climate with mean annual precipitation ranging from 500 mm in the eastern lowland region up to about 3000 mm in the western alpine region. Runoff depths range from less than 50 mm per year in the eastern part of the country to about 2000 mm per year in the Alps. Potential evapotranspiration is on the order of 600–900 mm per year. Austria has a dense stream gauge network. Hourly runoff data over the period 1 August 1990 to 31 July 2000 are used in this paper. The raw runoff data were screened to exclude catchments with significant anthropogenic effects, karst and strong lake effects, as the runoff from these catchments are assumed to belong to a different type of processes. The remaining data consisted of 386 stream gauges with catchment areas ranging from 10 to 10,000 km^2.

Figure 2 shows the Austrian river network with gauging stations. A further description of the data is given by Skøien and Blöschl (2007). The square represents the the Innviertel region, used for detailed analyses by Skøien and Blöschl (2007) and shown with more details in Figure 3.

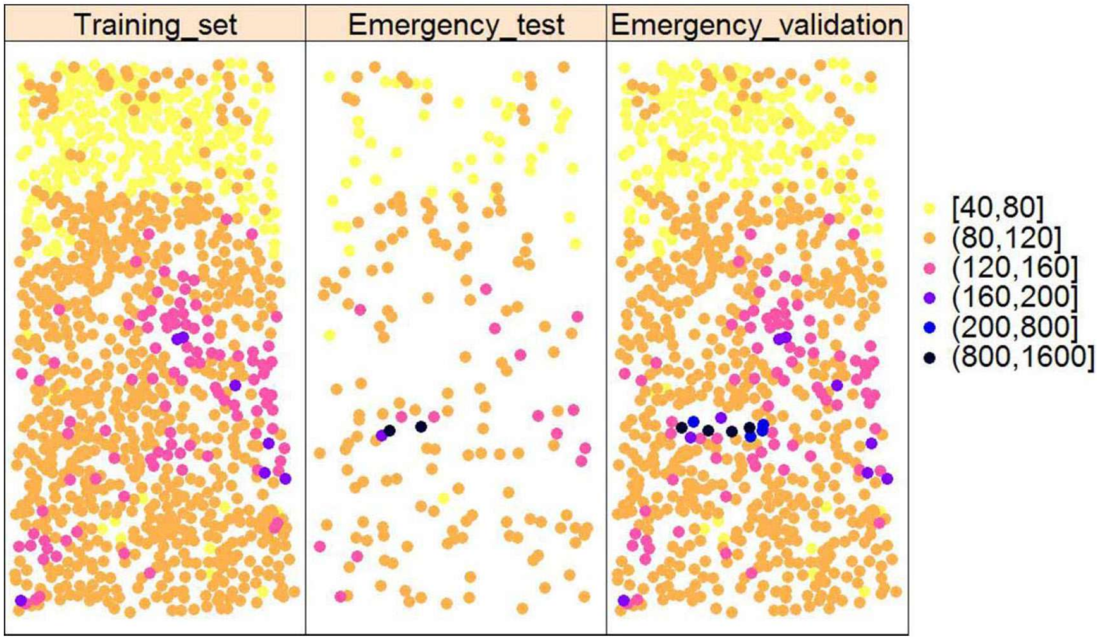

Figure 1. Training set, data set of emergency situation and validation data for the SIC2004 exercise. Units are nSv/h.

We will in this paper focus on estimates of flood events, which we have defined in two different ways in this paper. First, for estimates of Model Efficiencies (see below), we considered the partial time series when the runoff exceeded the 90th percentile of the time series of a station. These events are high flows, but not necessarily large floods. In addition, for other flood characteristics, we only considered events where the maximum runoff exceeded the 99th percentile. For all stations, this last definition gave a total of more than 25000 events in the 386 catchments.

All runoff data given below refer to specific runoff per second, i.e. the runoff divided by the upslope contributing area:

$$q_i(t) = \frac{Q_i(t)}{A_i} \tag{1}$$

where Q (t) i is the runoff measured at a runoff gauge i and i A is the area of this catchment.

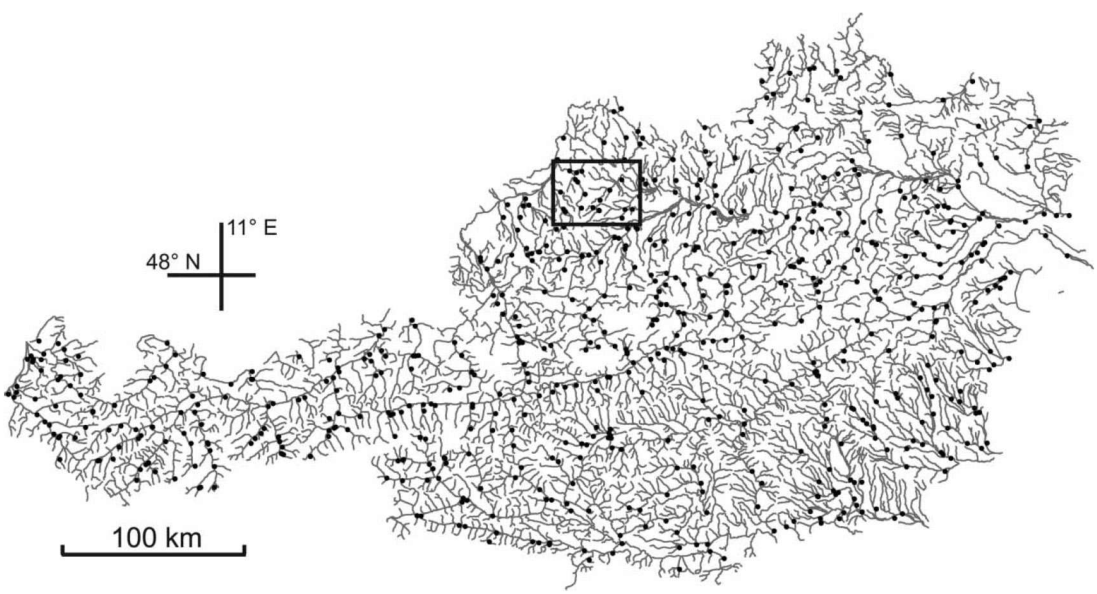

Figure 2. Stream gauges (circles) in Austria used in this paper. The square represents the Innviertel region.

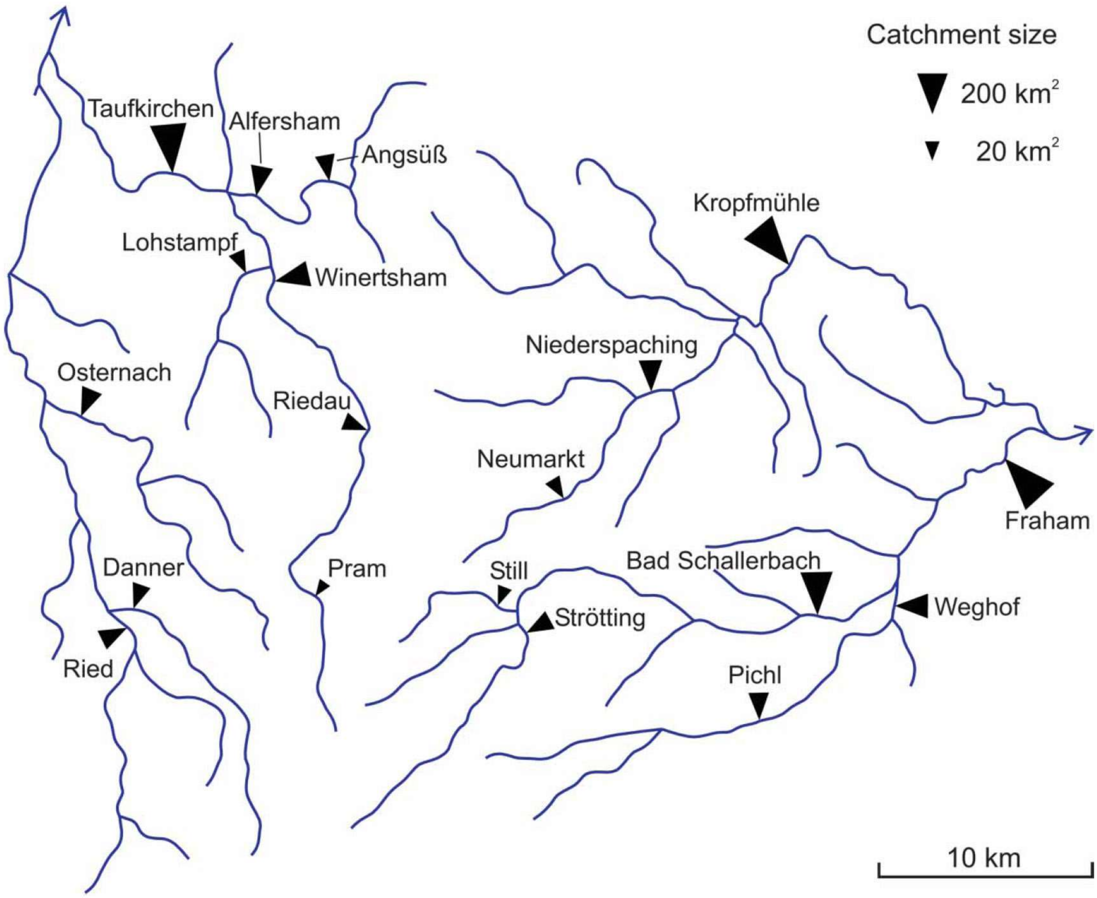

Figure 3.

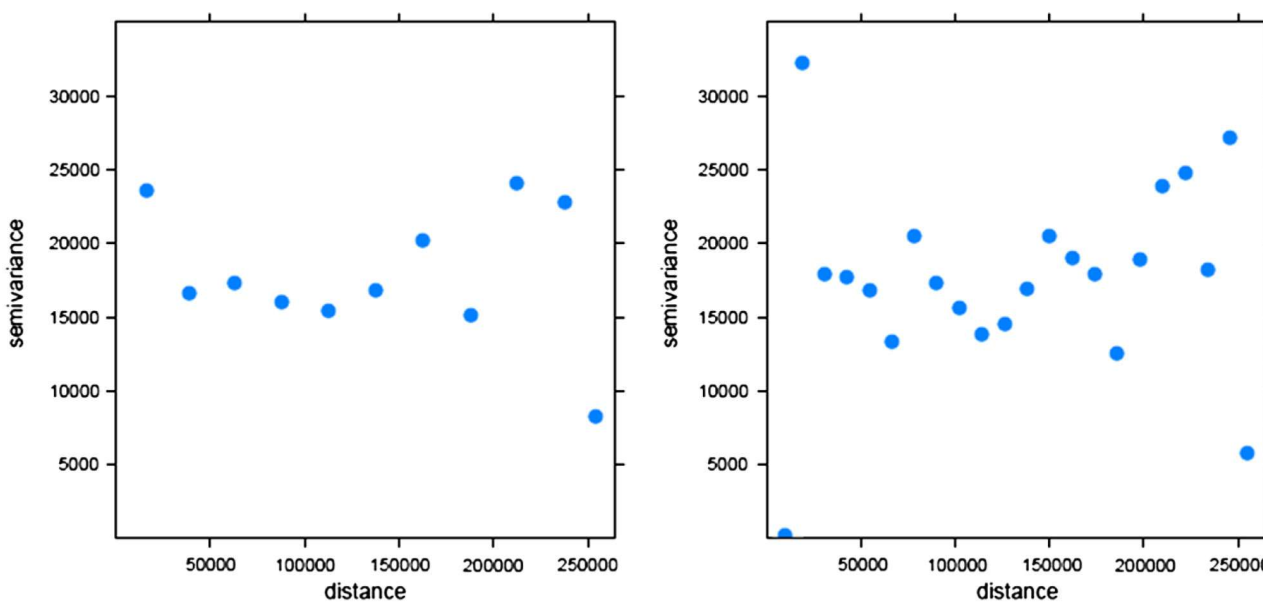

Figure 4. Variograms of emergency situation with different bin sizes. Units of semivariance are $(nSv/h)^2$.

Theoretical considerations

SIC2004 case

The SIC2004 exercise is characterized by the fact that there are two observations in the training set that exhibit considerably larger values than in the routine situation. This will typically be the situation short time after a release of nuclear material. Traditional variograms can be inferred, as presented in Figure 4. Different bin sizes were used, to indicate how the variogram behaviour depends on this rather subjectively chosen value.

The variogram to the left is characterized by an apparent lack of spatial correlation. The semivariance is large also for the shortest distances. The variogram to the right apparently has a spatial correlation for very small distances, but no spatial correlation for larger distances. Still, it would be unwise to try to fit a theoretical variogram model to this sample variogram.

The critical issue in the emergency case is the amount of information that is actually available. If we (as commonly done) try to fit one of the possible variogram models with three parameters (nugget, sill and range) to the sample variogram in the right pane of Figure 4, it is easy to do so in the belief that there are lots of degrees of freedom in the fitting because the theoretical variogram model will be fitted jointly to the different semivariances of the sample variogram. We think this is an assumption which is highly questionable in this situation.

The sample variogram of this emergency case has been inferred from a set of 200 observations. However, as the difference between the two extreme observations and the remaining observations is so large compared to the standard deviation between the remaining observations, this situation can be approximated by a map of a binary process with two 1's and the remaining observations zero. The variogram will be characterized by the following:

- The semivariance for each bin will be determined by how many stations have exactly this distance to one of the two stations with extreme observations.
- Small observed semivariances will only exist for these distance classes that do not include any of the stations with extreme observations.

The last point implies that small semivariances in the short distance part of the variogram is possible only when the separation distances between some of the stations without extreme observations are smaller than the distances between the stations with extreme observations to their closest neighbours. This is the situation in the right pane of Figure 4, where the limit of the first bin is slightly smaller than the distance from the stations with extreme values to their neighbours. Although this variogram behaviour might indicate that a correlation length can be estimated, this correlation length is hence only founded on the network topology and the distribution of distances between stations, hardly at all on observations of the underlying process (Skøien and Blöschl, 2006).

Several goodness-of-fit criteria were considered for both for the routine situations and for the simulated emergency situation. In this paper, we consider two criteria, the correlation coefficient (r):

$$r = \frac{\sum_{i=1}^{n} z_i \hat{z}_i - n\bar{z}\bar{\hat{z}}|}{(n-1)s_z s_{\hat{z}}} \tag{2}$$

and the root mean square error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{z}_i - z_i)^2} \tag{3}$$

where z_i and $\hat{z}_i$ are the observed and predicted value of the process, respectively, n is the number of observations, and s_z and $s_{\hat{z}}$ are the standard deviations of the observations and the predictions, respectively.

Real time mapping of runoff

River runoff is a spatio-temporal process which exhibits large variations in space and time. As the measured runoff at a location along a river is the aggregated value of the rather complicated rainfall-runoff generation upstream this location, straightforward geostatistical methods are not possible to use. Instead, newer methods that take this aggregation into account have been developed (Gottschalk, 1993a,b; Gottschalk, et al., 2006; Sauquet, 2006; Skøien et al., 2006, Skøien and Blöschl, 2007).

These methods include different kinds of integrations of a point variogram. As such integrations can be rather time consuming, it may be difficult to apply these methods in a real-time mapping situation. However, there are simplifications that can be made.

Kyriakidis (1999) pointed out that there are three options when representing a spatio-temporal random variable: full space-time models, simplified representations as vectors of temporally correlated spatial random fields, and simplified representations as vectors of spatially correlated time series. The latter reduces to a spatial estimation problem for each time step and is of interest for variables with observations that are rich in time but poor in space as is the case of runoff (*Rouhani and Hall*, 1989). Full spatio-temporal kriging is more complicated than the two simplifications as the kriging system needs to be solved simultaneously for both spatial and temporal kriging weights (*Kyriakidis and Journel,* 1999).

When estimating runoff time series in Austria, *Skøien and Blöschl* (2007) assumed that the spatio-temporal pattern of runoff could be modelled as spatially correlated time series. A simple routing model was added to model the flow from upstream to downstream gauges. Although we also examined different routing velocities, we only present the results without routing in this paper. The proposed routing model was also dependent on measurements at downstream gauges at later time steps, what is not available in a real-time mapping situation. As a further simplification, they assumed that the correlation between the time series was constant in time, i.e. one theoretical point variogram model was fitted to the data and used for different runoff situations. However, the further away from routine situation, the more likely it is that the process has a different correlation structure than the one of a routine situation.

Unlike Skøien and Blöschl (2007), we now focus on estimates of floods events. The Nash-Sutcliffe coefficient (*Nash and Sutcliffe,* 1970), equivalent to R^2 in regression (*Pebesma, et al.,* 2007) and used by Skøien and Blöschl (2007), is a measure of the model efficiency (ME) that is usually applied on continuous time series. It is of less interest when it comes to floods, but we will still apply it for the partial time series of floods above the 90th percentile. The ME for a certain catchment i is then:

$$ME_i = 1 - \frac{\sum_{\omega=1}^{\Omega}(q_i(\omega) - \hat{q}_i(\omega))^2}{\sum_{\omega=1}^{\Omega}(q_i(\omega) - \overline{q(\omega)}_i)^2} \quad (4)$$

where Ω is the number of time steps estimated, $q_i(\omega)$ is the observation at time step ω, $\hat{q}_i(\omega)$ is the estimate at time step ω and $\overline{q_i(\omega)}$ is the average of the observations at these time steps. This model efficiency indicates a perfect estimate if $ME = 1$. The estimate is no better than the mean of the observations if $ME = 0$, and even poorer if $ME < 0$.

As a comparison, a review study by Merz et al. (2007) suggested that typical rainfall-runoff models would give ME for ungauged catchments in the order of 0.6–0.7. We are not aware of similar reviews of the efficiency of rainfall-runoff models for flood situations. However, it can be assumed that the efficiency will be lower for these situations.

Two other variables are better for the analyses of large floods, i.e. the difference between observed and estimated time of peak (t_{PK}) and the difference between observed and estimated peak flow (q_{PK}). Due to the large number of events, it was necessary to automatically identify events and to pair a peak from the estimated time series with the correct observed peak. We used a rather empirical approach for this identification, which in some cases resulted in wrong pairing of peaks. We discarded all pairs with $t_{\mathrm{PK}} > 40$ hours, as larger time differences most likely could be addressed to misclassification of events belonging together.

We normalized the deviation between estimated and observed peak flow, for better comparison of such a large number of events. For each peak flow, we estimated the relative difference as

$$\Delta q = \frac{q_{PK} - \hat{q}_{PK}}{q_{PK}} \quad (5)$$

where $\hat{q}_{PK}$ is the estimated peak flow. Equation 5 gives $\Delta q \in [-\infty,1]$ where $\Delta q = 0$ means correct estimate, $\Delta q < 0$ represents overestimation and $0 < \Delta q < 1$ represents underestimation.

Are geostatistical methods still applicable?

The two cases above describe situations where the assumptions behind geostatistical interpolation are violated. In the first of these cases is the initrinsic hypothesis violated (expected variance is a function of separation distance) and the observations are far from having a Gaussian distribution. In the second case is the correlation structure non-stationary over time, whereas the use of one single variogram necessitates a stationary correlation structure.

Despite the fact that these violations indicate that geostatistical analyses will be deceiving, we will still suggest that geostatistical interpolation methods can be useful. The fact that the two observations with extreme values in the SIC2004 case are neighbours suggests that there is a spatial correlation also in the area affected by the release, although it is difficult to find through variogram analyses. There are also reasons to assume that there will be a spatial correlation between observations also in a flood situation, although the correlation structure may be different than in a routine situation.

It is general knowledge that the quality of the predictions is relatively insensitive to the choice of variogram (see e.g. *Lark,* 2000), at least as long as there are several observations within the range of the variogram. This insensitivity is not without limits though, but we will still test a couple of simple geostatistical solutions for interpolation of the two data sets introduced earlier in this paper. The simplicity is assumed to increase the robustness of the result. The simple methods can also be used to identify which part of the modeling needs to be closer examined. In this paper, we examine the following simple solutions:

- Use a variogram inferred for a routine situation for the emergency case. The correlation length will most likely be too long in this case, but at least this is a variogram describing a process that has similarities with the extreme situation to be mapped (*Pebesma,* 2005).
- Use a linear variogram. This takes away the necessity of fitting sill or correlation length of the process.

The disadvantage for both of these solutions is that we are not able to give reliable quantitative estimates of the prediction errors. These are normally useful results from geostatistical interpolation. However, although the level of the prediction errors is not correct, we will examine the correlation between these prediction errors and the model efficiency for the case with Austrian runoff data.

The use of a linear variogram model can be seen as a slightly more sophisticated method than an inverse distance weighting (IDW) approach. Kriging automatically takes into account the correlation between the neighbours, and compensates for a possible clustering of the observations that IDW is not able to take into account.

In addition, it is possible to assume anisotropy for kriging, i.e., that the correlation is larger in one direction than in the other. The emergency case of SIC2004 suggest anisotropy, as the two stations with extreme values are further apart in the x-direction than the unaffected neighbours in y-direction. The anisotropy consist of two parameters; the direction of the larger correlation and the ratio between the correlations in the direction of the larger correlation and the correlation in the perpendicular direction.

We will examine both methods for the SIC2004 case. For the case with Austrian runoff data, we will only examine the use of fixed variogram. Whereas Skøien and Blöschl (2007) estimated the complete time series, we will in this paper examine the extreme values of the estimated hydrographs more closely, as this is the situation where real-time mapping of runoff will be of largest interest. We will for each station only examine the part of the hydrographs which exceeds the 90th percenentile of the station, as mentioned above.

Results

SIC2004

We tested the ability of four different variogram models to make predictions at the locations of the validation example. We also tested the effect of using a local kriging neighbourhood. The following models were examined:

1. Isotropic linear variogram
2. Linear variogram with eye-ball fitted anisotropy parameters (largest correlation in x-direction, ratio 1:3)
3. Linear variogram with fitted anisotropy parameters
4. Spherical variogram fitted to one of the routine data cases

It is only the first and last of the models above that can really be compared to the automatic interpolation results of the SIC2004 participants. The second model is based on a subjective consideration of the case, whereas the third model necessitates knowledge of the validation data which is of course not available in advance in an automatic modelling exercise.

For all models, the gamma dose rate was interpolated for the locations of the validation data and compared to the validation data in terms of the correlation coefficient (r) and the root mean squared error (RMSE). These were among the goodness-of-fit criteria applied to the estimates of the participants of the SIC2004 exercise. Figure 5 presents the results from the models tested in this paper and the results from the SIC2004 exercise, taken from Table 4 in Dubois and Galmarini (2005b).

In the SIC2004 exercise, 13 participants submitted 31 solutions to the automatic mapping problem. Some of the extra solutions were intended more as comparisons, and not as the best possible solution to the problem. Hence, the best submission from each of the participants has been given a separate colour in Figure 5.

It can be noted from the figure that estimation using the isotropic linear variogram performs poorest of the models examined in this paper. Still, this model performs better than several of the submitted solutions in the SIC2004, even better than one of the best efforts from one of the participants in terms of root mean square error and four of the participants in terms of correlation coefficient. The estimates from an exponential variogram model fitted to one of the data sets from a routine situation performs slightly better than the linear model, in terms of both of the goodness of fit measures we consider in this paper.

A considerably larger improvement comes when anisotropy is introduced in the linear model. Model 2 is then among the best of the participants of the SIC2004 exercise, for both goodness of fit measures. The result is further improved when the anisotropy parameters are fitted, and the result is better than all

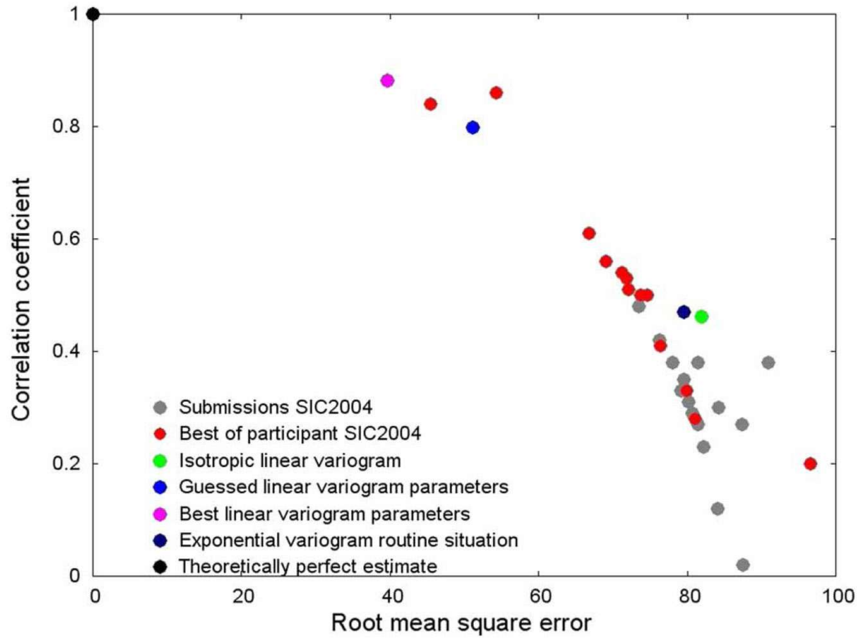

Figure 5. Results from SIC2004 exercise and the models presented in this paper. Units of Root mean square error are nSv/h.

the participants of the SIC2004 exercise. The results hence indicate that properly estimated anisotropy parameters are more important than the variogram parameters in the SIC2004 case.

It was also tested whether it would improve the results using only local observations. Although also the number of neighbours could be optimised, the result using the complete data set was better than most choices of a local neighbourhood.

Austrian runoff data

Skøien and Blöschl (2007) presented several hydrographs of flood situations from the Innviertel region. We will therefore not reproduce a series of hydrographs in this paper, but include Figure 6 as an example (Figure 8 in Skøien and Blöschl, 2007). In this figure, the runoff of the catchment Ried (69 km^2) has been estimated from the observations of its neighbours, including a simple routing model. This catchment has neither nested upstream nor downstream neighbours. In the three day period in fall 1998 two events occurred. The first event (shown at time 18 hours) was properly estimated. The second event (shown at time 48 hours) was not. This is because it was apparently the result of a local precipitation event. None of the surrounding catchments experienced more than a small change in runoff for this period. It is interesting to note that the local event was also missed by the rain gauges in the region, so rainfall-runoff models could not have simulated this event either.

As a first statistic for evaluation of the model, we will focus on the model efficiencies, measured by the Nash-Sutcliffe coefficient (Nash and Sutcliffe, 1970). Figure 7 presents the cumulative distribution functions (cdf) for the model efficiencies of top-kriging for large runoff. The results for three data sets are shown, "All" represents the cdf for all 386 catchments, "Parajka et al." represents the 208 stream gauges in common with Parajka et al. (2005), whereas "Innviertel" represents the 19 stream gauges in the Innviertel region.

The results are indeed poorer than the results from Skøien and Blöschl (2007). The number of interpolated catchments with ME <0 has increased from a few to about 20 percent for the catchments in common with Parajka et al. (2005) and to around 30 percent of all catchments. The median for all catchments is 0.37, whereas it was 0.75 for the same catchments in the paper of Skøien and Blöschl (2007) and 0.52 for the catchments in common with Parajka et al. (2005), wheres the median was 0.82 for these catchments in the paper of Skøien and Blöschl (2007) The median for the catchments in the Innviertel region was still high, 0.78 compared to 0.87.

The results of Skøien and Blöschl (2007) were improved by the inclusion of a routing model. For real-time mapping of a flow situation, this is only possible to take into account for the upstream stream gauges. The routing model was still tested assuming different flow velocities, mostly with poorer results.

The kriging error is in geostatistical methods a measure of the estimated prediction uncertainty,

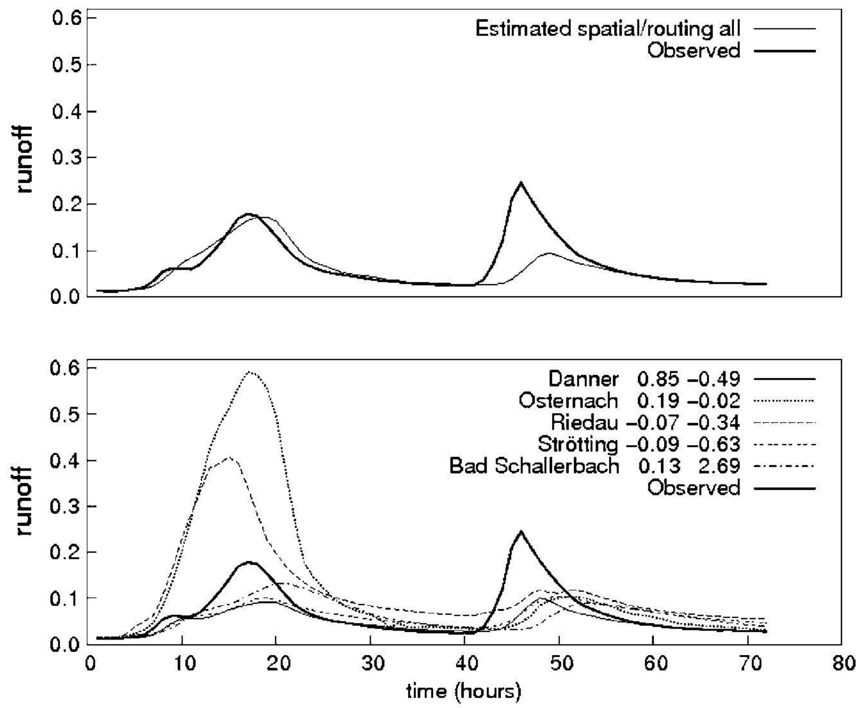

Figure 6. Observed and estimated hydrographs (including routing) for Ried (69 km^2) (top panel) and hydrographs of the neighbors (bottom panel) for the period Oct. 29–Nov. 1, 1998. The first number after neighbor name represents the kriging weight of the neighbor, the second number the lag of the routing model in hours. Units of runoff are $m^3s^{-1}km^{-2}$. Figure from Skøien and Blöschl (2007).

which can be in terms of the standard deviation of the estimates. When, as in our case, using a variogram inferred from a mixture of situations for a flood situation, we must acknowledge that the kriging error cannot be referred to as the standard deviation of the flood estimates. Still, we can examine how the differences in kriging error compares to the estimated ME. Figure 8 gives a comparison of estimated prediction uncertainty in terms of standard deviation of the estimates and the model efficiencies for the 208 catchments in common with Parajka et al. (2005). The model efficiencies have been grouped in bins with

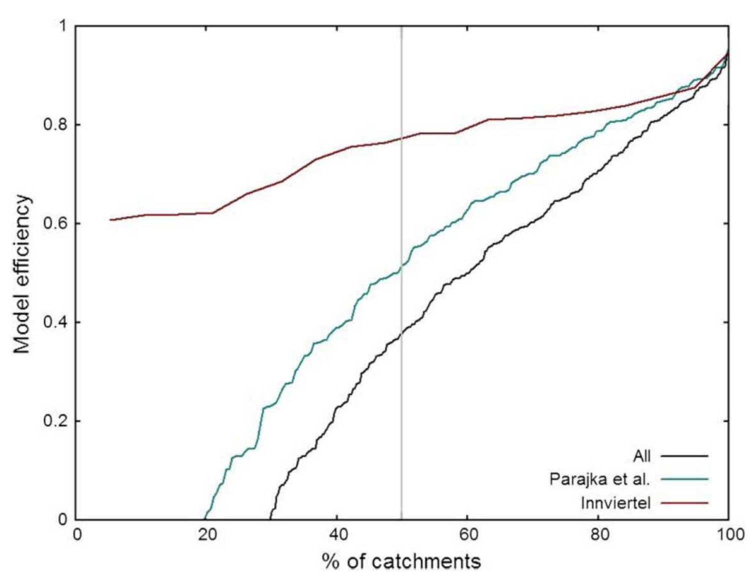

Figure 7. Cumulative distribution function of model efficiencies ME of hourly runoff estimated by top-kriging for 386 (all), 208 (Parajka et al.) and 19 (Inn-viertel) catchments in Austria.

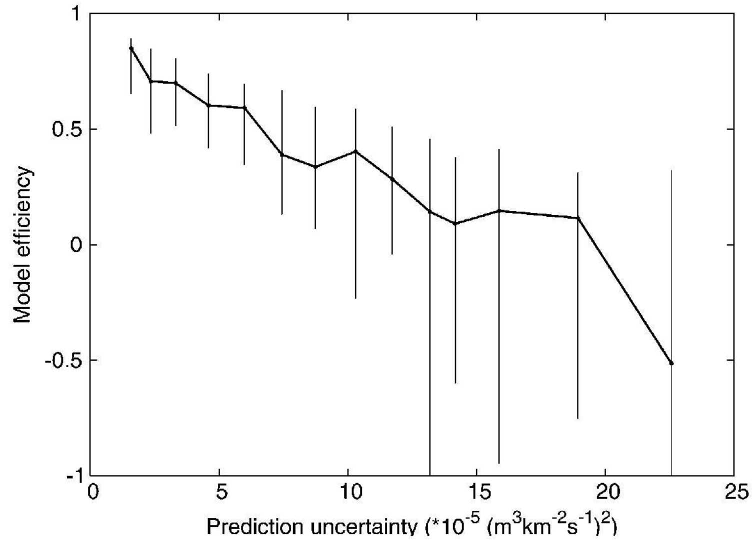

Figure 8. Model efficiencies of hourly flood estimates from top-kriging plotted as a function of the estimated kriging error for 386 catchments in Austria. The thick line represents the median ME and the error bars represent the 25 and 75 percentiles. Both model efficiencies and the kriging variance relate to the ungauged catchment case obtained in a cross-validation mode.

similar prediction uncertainty, similar to figure 16 of Skøien and Blöschl (2007). The line represents the median within such a bin, the error bars represent the 25 and 75 percentiles within the bin. The model efficiency is high for prediction uncertainties up to 8×10^{-5} $(\mathrm{m}^3\mathrm{km}^{-2}\mathrm{s}^{-1})^2$.

Although we cannot expect the kriging error to give reliable quantitative information about the prediction errors, Figure 8 does indicate a rather strong relationship between the kriging error and the model efficiency. The internal ranking of the prediction uncertainties of the estimates does indeed give a qualitative indication of the model efficiency for a certain catchment.

The second test statistic that we examine is the time between the peaks of the observed runoff and of the estimated runoff. Figure 9 shows the box plot of the time differences between the peaks of the observation and the peak of the estimated time series. The centre of the box indicates that estimates are on the average unbiased, as the median of the differences are zero and the mean indicates that the method estimates the peaks 0.5 hours too late. The borders of the box refer to the 25 and 75 percentiles, respectively. Also these percentiles suggest that the method estimates the time of the peak relatively well, but with the peaks slightly too late. This is probably a result of the lack of a routing routine, and a large number of the catchments being headwaters without upstream neighbours. Approximately 10 percent of the events are outside the whiskers of the plot at -6 and 5, respectively. A considerably part of the

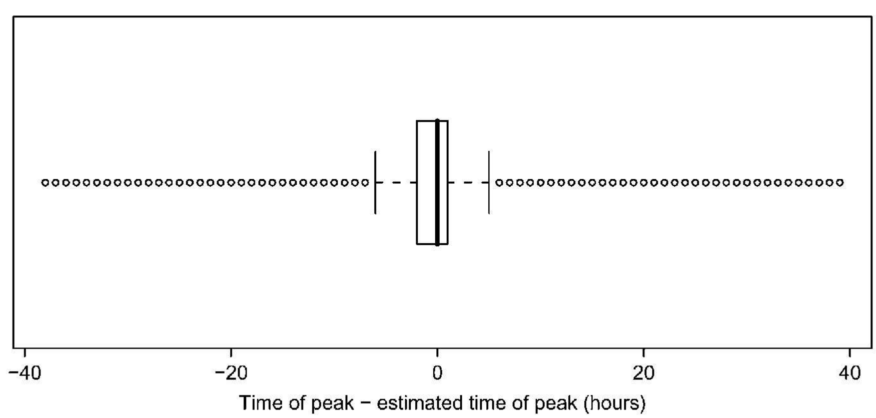

Figure 9. Boxplot of time differences between peak times of observed and estimated time series.

wrong peak estimates can also be assumed to come from wrong pairing of peaks from observations and simulations.

Figure 10 presents a similar box plot for the relative difference of peak flow, calculated as in Eq. 5. It can be observed that the median of the relative differences is positive, $\Delta q = 0.18$. This indicates that the method tends to underestimate the peak, compared to the observation. A better result would maybe be possible applying a routing model, or by assuming that the specific peak flow is negatively correlated with catchment area, as assumed in Skøien et al. (2006).

Figure 11 presents scatter plots of observed and simulated peaks for different catchment areas. Although the number of events decreases with catchment area does the figure clearly indicate that there is a strong relationship between the catchment area and the correspondence between observed and estimated peak flow. The scatter is large for small catchments, with a tendency of underestimation indicating that floods in small catchments are not properly estimated using this method. Floods in the two largest groups of catchment areas are better estimated, although there is a larger scatter for the most extreme floods.

The findings above were confirmed when examining the correlations between observed and simulated peak flow. The correlation for the smallest catchments was 0.65, increasing to 0.81 and 0.94 for the two groups with the largest catchments.

Conclusions

The results in this paper indicate that it is possible to use geostatistical methods also when the assumptions behind these methods are unlikely to hold, but that the quality of the results decrease with increasing deviations from these assumptions. The results also support that it may be possible to achieve good results also with simple methods. Only two examples were handled in this paper, but they were of very different character; one example was characterised by two extreme values whereas the true correlation structure in the other example could be assumed to be non-stationary over time.

For the SIC2004 case, the results indicated that the most important parameters to estimate were the anisotropy parameters. If these were reasonably well estimated, even the simple linear model was able to give better estimates than most of the participants in the SIC2004 exercise. It is not straightforward to estimate these parameters automatically, and particularly not in a case with only two extreme values. Ecker and Gelfand (1999) developed a method using a Bayesian updating metod. More recent developments includes the covariance tensor identity method of Hristopulos (2005). This method is further developed by Chorti and Hristopulos (2008) within the INTAMAP project. Although we have not tried to incorporate their results in this paper, these methods are promising for the type of problems encountered with the SIC2004 data.

The results from estimating floods in Austria by top-kriging (*Skøien and Blöschl,* 2007) using the same variogram as in routine situations did indeed give poorer results than what Skøien and Blöschl (2007). A reduction was expected, as spatial variation is likely to increase with increasing floods. The estimation of the catchments used for comparison with Parajka et al. (2005) performed a bit poorer than the rainfall-runoff models examined by Merz et al. (2007), but it should be noted that they examined the model efficiency for all situations. Hence it would be expected that also the rainfall-runoff models would perform poorer for floods, and the results presented

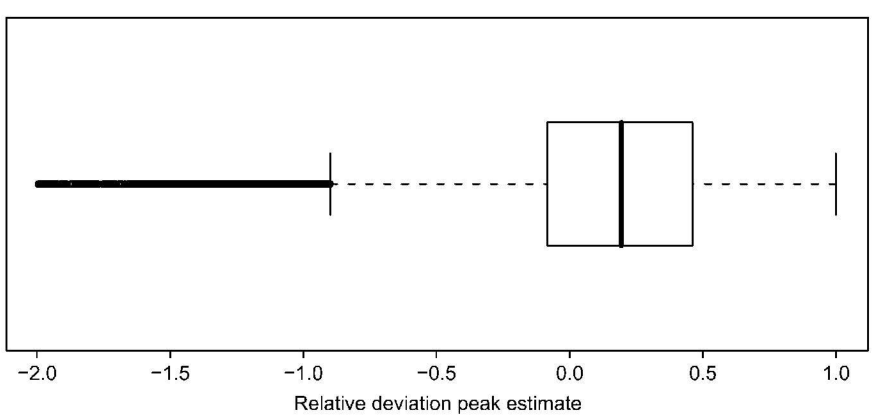

Figure 10. Boxplot of relative deviations from peak estimate.

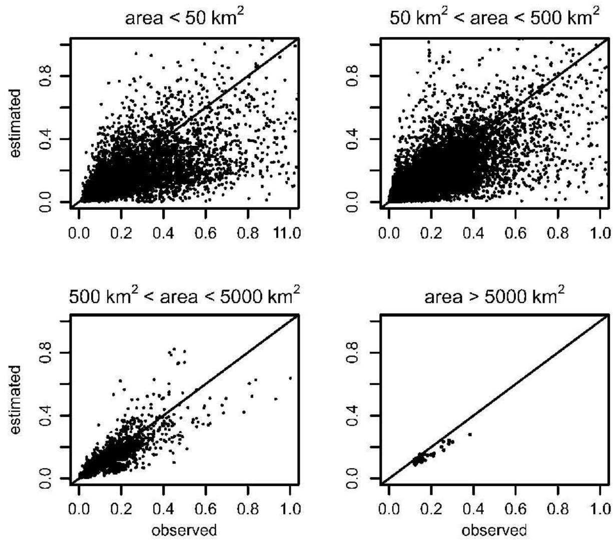

Figure 11. Scatterplots of observed and simulated peaks for different catchment areas. Units are $m^3 s^{-1} km^{-2}$.

here are acceptable. Comparisons of estimated runoff hydrographs and estimated hydrographs for flood situations by Skøien and Blöschl (2007) indicated that top-kriging performed well also in flood situations, though it should be noted that these hydrographs were all taken from the Innviertel area.

Comparisons between observed and estimated peak flows in this paper indicate that the top-kriging method is able to estimate peak flow for medium to large catchments, but have larger difficulties with the smaller catchments. One explanation for this lies in the source of the floods; whereas floods in larger catchments are typically a result of more regional precipitation, floods in smaller catchments are typically of a more local behaviour. For a method based on regionalisation of data, these floods are then difficult to estimate.

Geostatistical methods are known to smooth the output compared to the observations. This is probably a part the reason for the underestimation of large floods, seen in Figure 10. There is also a negative correlation between catchment area and peak flow, as noted by Merz and Blöschl (2005) and used by Skøien et al. (2006) when estimating flood statistics. A similar approach for correcting flood peaks could also prove useful in a real-time case.

It was also interesting to notice the difference when analyzing different parts of the data set. The order (estimates of Innviertel catchments best, estimates of all catchments worse) is the same as in Skøien and Blöschl (2007), but the differences seems to increase. Skøien and Blöschl (2007) suggested that Parajka et al. (2005) had been more restrictive in selecting high quality runoff records. The results in this paper support that suggestion, and indicates that a proper examination of the catchments and the runoff records is even more important when estimating floods. Figure 8 also indicates that the prediction uncertainty obtained from the top-kriging estimates is important for the model efficiencies, even though the number itself is not a reliable quantitative descriptor of the uncertainty of the estimates. Hence, the top-kriging method with one variogram for all situations can be used in regions with a high stream gauge density.

It could be seen as surprising that the best results were achieved without applying a simple routing model. We assume that there are two reasons for this, first that a higher velocity should be used than by Skøien and Blöschl (2007), second that that the different velocities throughout Austria is of larger importance than for the continuous case. Hence, these velocities should be assessed through more local analyses, e.g. as done by Merz and Blöschl (2003). A routing model can only be applied for the upstream stream gauges for real-time mapping.

One of the methods suggested in this paper was used in both cases; the use of a variogram estimated from a routine situation also for the extreme situation. The results indicated that this worked better in the runoff case than in the radioactivity case. The reason is most likely that the deviations from the assumptions behind kriging were larger for the SIC2004 data. Also the analyses of the runoff case indicated that the results from using only one variogram for different situations depend on how extreme the situation is.

Although geostatistical methods are suited for direct predictions of a process, it is also important to remember that the methods presented here can be applied on the residuals of a deterministic model. For the case of radioactivity, there are some deterministic models that also can provide predictions of the locations of the plume; the geostatistically based methods can then be applied on the residuals, attempting to model the difference between the deterministic model and the observations.

We cannot take for granted that simple methods as the ones presented in this paper will be satisfying for all estimation purposes, despite acceptable results for two very different case studies. The advantage of more complex methods may be that they are more able to deal with problems of different complexities. The participants in the SIC2004 exercise were not aware of the possible difficulties of the data sets before they had to submit their solutions. We both knew the distribution of the data and were able to make some subjective judgements of the interpolation procedure. We can hence not compare our results with the results of the participants of the SIC2004 procedure.

However, we find the results an important reminder that simple methods in many cases can perform as good as the more complex methods. Even if they will not be used as such in the final implementation of an automatic mapping procedure, can they be useful for identification of critical parts of the procedure, and for benchmarking of the more complex models. According to the law of parsimony and Occam's razor should a simpler model be preferred to more complex models if their results are otherwise the same.

Acknowledgements

This work has been funded by the European Commission, under the Sixth Framework Programme, by the Contract N. 033811 with the DG INFSO, action Line IST-2005-2.5.12 ICT for Environmental Risk Management, and from The Austrian Academy of Sciences, project HÖ 18. The views expressed herein are those of the authors and are not necessarily those of the funders.

References

Ai-Geostats. 2004. Discussion on "Automated variogram modelling" [Mailing-list], msg 1560–1574 [online]. Available from: Available at: http://www.mail-archive.com/ai-geostats@jrc.it [Verified 1. March 2008]

Chorti, E., and D. T. Hristopulos. 2008. Systematic identification of anisotropic correlations in spatially distributed data sets. IEEE Transactions on Signal Processing, Submitted.

Cressie, N., 1991. *Statistics for spatial data*. New York, NY: Wiley.

De Cort, M. and De Vries., G., 1997. The European Union radiological data exchange platform (EURDEP): Two years of international data exchange experience. *Radiation Protection Dosimetry*, 73 (1–4), 17–20.

Dubois, G., and S. Galmarini. 2005a. Introduction to the spatial interpolation comparison (SIC) 2004 exercise and presentation of the datasets. *Applied GIS*, 1, (2), 09–01 to 09–11.

Dubois, G. and Galmarini., S., 2005b. Spatial interpolation comparison (SIC) 2004: introduction to the exercise and overview of results. *In*: G. Dubois, ed. *Automatic mapping algorithms for routine and emergency monitoring data*. Luxembourg: Office for Official Publications of the European Communities, 150.

Ecker, M. D. and Gelfand., A. E., 1999. Bayesian modeling and inference for geometrically anisotropic spatial data. *Mathematical Geology*, 32 (1), 67–82.

Gottschalk, L., 1993a. Correlation and covariance of runoff. *Stochastic Hydrology and Hydraulics*, 7, 85–101.

Gottschalk, L., 1993b. Interpolation of runoff applying objective methods. *Stochastic Hydrology and Hydraulics*, 7, 269–281.

Gottschalk, L., Krasovskaia, I., Leblois, E., and Sauquet., E., 2006. Mapping mean and variance of runoff in a river basin. *Hydrology and Earth System Sciences*, 10, 469–484.

Hristopulos, D. T. 2005. Identification of spatial anisotropy by means of the covariance tensor identity. In: G. Dubois, editor.. Automatic mapping algorithms for routine and emergency monitoring data: Office for Official Publications of the European Communities, p. 150.

Isaaks, E. H. and Srivastava., R. M., 1989. *An introduction to applied geostatistics*. New York: Oxford University Press.

Journel, A. G. and Huijbregts., C. J., 1978. *Mining geostatistics*. London, UK: Academic Press.

Kyriakidis, P. C. and Journel., A. G., 1999. Geostatistical space-time models: A review. *Mathematical Geology*, 31 (6), 651–684.

Lark, R. M., 2000. Estimating variograms of soil properties by the method-of-moments and maximum likelihood. *European Journal of Soil Science*, 51, 717–728.

Merz, R. and Blöschl., G., 2003. A process typology of regional floods. *Water Resour. Res.*, 39 (12), SWC51–SWC520.

Merz, R. and Blöschl., G., 2005. Flood frequency regionalisation-Spatial proximity vs. catchment attributes. *Journal of Hydrology*, 302 (1–4), 283–306.

Merz, R., G. Blöschl, and J. Parajka. 2007. Regionalisation methods in rainfall-runoff modelling using large catchment samples. In: V. Andréassian, editor., et al. *Large sample basin experiments for hydrological model parameterization: Results of the model parameter exmperiment-MOPEX* (Proceedings of the Paris (2004) and Foz de Iguacu (2005) workshops: IAHS Publ. 307, pp. 117–125.

Nash, J. E. and Sutcliffe., J. V., 1970. River flow forecasting through conceptual models. Part I—a discussion of principles. *Journal of Hydrology*, 10 (3), 282–290.

Parajka, J., Merz, R., and Blöschl., G., 2005. A comparison of regionalisation methods for catchment model parameters. *Hydrology and Earth System Sciences*, 9, 157–171.

Pebesma, E. J. 2005. Mapping Radioactivity from Monitoring Data-Automating the Classical Geostatistical Approach. *Applied GIS*, 1, (2).

Pebesma, E. J., P. Switzer, and K. Loague. 2007. Error analysis for the evaluation of model performance: rainfall-runoff event summary variables. *Hydrological Processes*, In Press.

Rouhani, S., and T. J. Hall, 1989. Space-time kriging of groundwater data. paper presented at 3rd International Geostatistics Congress. Kluwer Academic Publishers, Dordrecht. Avignon, France.

Sauquet, E., 2006. Mapping mean annual river discharges: Geostatistical developments for incorporating river network dependencies. *Journal of Hydrology*, 331 (1–2), 300–314.

Skøien, J. O. and Blöschl., G., 2006. Scale effects in estimating the variogram and implications for soil hydrology. *Vadose Zone Journal*, 5 (1), 153–167.

Skøien, J. O. and Blöschl., G., 2007. Spatio-temporal Top-kriging of runoff time series. *Water Resour. Res.*, 43, W09419.

Skøien, J. O., Merz, R., and Blöschl., G., 2006. Top-kriging-geostatistics on stream networks. *Hydrology and Earth System Sciences*, 10, 277–287.

Webster, R. and Oliver., M. A., 2001. *Geostatistics for environmental scientists*. Chichester: Wiley.