**Natural Hazards
and Earth
System Sciences**

Use of past precipitation data for regionalisation of hourly rainfall in the low mountain ranges of Saxony, Germany

T. Pluntke[1], **N. Jatho**[1], **C. Kurbjuhn**[1], **J. Dietrich**[2], and **C. Bernhofer**[1]

[1]Institute of Hydrology and Meteorology, Department of Meteorology, Technische Universität Dresden, Dresden, Germany
[2]Institute of Water Resources Management, Hydrology and Agricultural Hydraulic Engineering, Leibniz Universität Hannover, Hannover, Germany

Abstract. Within the context of flood forecasting we deal with the improvement of regionalisation methods for the generation of highly resolved (1 h, $1 \times 1 \mathrm{km}^2$) precipitation fields, which can be used as input for rainfall-runoff models or for verification of weather forecasts. Although radar observations of precipitation are available in many regions, it might be necessary to apply regionalisation methods near real-time for the cases that radar is not available or observations are of low quality.

The aim of this paper is to investigate whether past precipitation information can be used to improve regionalisation of rainfall. Within a case study we determined typical precipitation Background-Fields (BGF) for the mountainous and hilly regions of Saxony using hourly and daily rain gauge data. Additionally, calibrated radar data served as past information for the BGF generation. For regionalisation of precipitation we used de-trended kriging and compared the results with another kriging based regionalisation method and with Inverse Distance Weighting (IDW). The performance of the methods was assessed by applying cross-validation, by inspection and by evaluation with rainfall-runoff simulations.

The regionalisation of rainfall yielded better results in case of advective events than in case of convective events. The performance of the applied regionalisation methods showed no significant disagreement for different precipitation types. Cross-validation results were rather similar in most cases. Subjectively judged, the BGF-method reproduced best the structures of rain cells. Precipitation input derived from radar or kriging resulted in a better matching between observed and simulated flood hydrographs. Simple techniques like IDW also deliver satisfying results in some occasions. Implementation of past radar data into the BGF-method rendered no improvement, because of data shortages. Thus, no method proved to outperform the others generally. The decision, which method is appropriate for an event, should be made objectively using cross-validation, but also subjectively, using the expert knowledge of the forecaster.

1 Introduction

Several studies about uncertainty in hydrological modelling have shown, that the uncertainty of the meteorological input, especially precipitation, often dominates other sources of uncertainty like model structure or parameters (Sun et al., 2000; Berne et al., 2004; Kuczera et al., 2006). The availability and quality of areal precipitation data is of utmost importance for models used for flood warning. At present, there are essentially three basic systems that provide precipitation observations, which can be used for real time flood forecasting: rain gauges, radar and satellite.

(1) Rain gauges are widely used. These observations can be regarded as those with highest accuracy operationally available (Paulat et al., 2008). There exist time series records that cover several decades, in some cases more than a century. However, gauge networks are normally too sparse to determine the areal rainfall with adequate spatial and temporal resolution, especially for convective events (Sun et al., 2000; Berne et al., 2004).

(2) Detailed insights into spatiotemporal precipitation patterns are possible with radar observations. Rainfall intensities must be calculated from back-scattered radiation. Hereby, sources of error are manifold (Ehret, 2003), e.g. system immanent errors, reflections at ground level (ground clutter), errors due to non-representative sampling space and

calculation of rainfall intensity. The often observed unsystematic offset in precipitation intensity between radar and gauge can be reduced by calibrating radar with hourly rain gauge data (e.g. Bartels, 2004).

(3) A further possibility to obtain precipitation estimates are observations from geostationary or polar orbiting meteorological satellites. The temporal and spatial resolution of satellite observations is generally inferior to radar observations. This remote sensing technique is still in a developing phase, and has not yet reached the quality needed for the implementation in operational flood forecasting systems on small or medium size catchments as shown by Grijsen et al. (1992) for sub-tropical areas.

Since all methods have distinct advantages and limitations, it is reasonable to combine them to take maximum advantage of all available information sources. Great efforts were made especially to combine radar and gauges, for example by Brandes (1975), Alpuim and Barbosa (1999), Todini (2001) and Jatho et al. (2010). The role of satellite data is often limited to extend the region of observation beyond the range of radar, e.g. by providing observations of weather systems approaching from seaward directions or to have any precipitation information in regions of scarce rain gauge measurements.

An operational flood forecasting system needs areal precipitation information at all costs. Electronic systems, schemes and equipment may fail during a flood event. Paulat et al. (2008) reported that on 97 days (6.6%) no radar data were available for the period from 2001 to 2004 for Germany. The task is to ensure that a high quality precipitation field can be provided even in the case of severe data losses or minor quality of one data source.

Our motivation to develop and improve regionalisation methods for hourly precipitation data arose from that point. Areal precipitation information is likewise important for combination with other rainfall data products like radar. Our pretensions were to identify methods that are capable for application in operational flood forecast. The robustness of such methods is important to ensure near real-time operation with minimum need of additional expert knowledge. A priori it is not possible to define whether statistic or non-statistic interpolation methods yield better results for single events. Kriging was often found to be superior to deterministic methods like Inverse Distance Weighting (IDW) or Thiessen-Polygons for precipitation estimation for different temporal resolutions (e.g. Tabios and Salas, 1985; Eischeid et al., 2000; Goovaerts, 2000; Haberlandt, 2007). For spatially dense rain gauge networks Dirks et al. (1998) and for temporal resolution of one to 24 h Dorninger et al. (2008) and Ruelland et al. (2008) demonstrated the equality of deterministic models like IDW. The performance of methods depends among other things on the density and configuration of the gauge net. Additional variables such as altitude, wind direction or radar data can be used to improve interpolation. Examples of methods that include additional variables

are kriging with external drift (Goovaerts, 2000; Haberlandt, 2007) or co-kriging (Seo et al., 1990a, b). The dense net of daily rainfall stations was used as additional variable by Merz et al. (2006) and Paulat et al. (2008), yet their methods can not be operated near real-time.

The aim of this paper is to demonstrate the capability of regionalisation methods for rain gauge data of high temporal resolution in mountainous regions, which are based on kriging and consider knowledge of past precipitation events. We choose the low mountain ranges, because the lead time of flood forecasts is shorter than in the low lands, which makes uncertainties more crucial. Furthermore, it is most demanding to find adequate methods that can deal with the spatial heterogeneity of precipitation in mountainous areas. The temporal variability of precipitation becomes extremely relevant for time steps below the lifecycle of precipitation cells, what makes regionalisation of such events very demanding compared to the regionalisation of daily rainfall amounts. The background is to provide an areal precipitation input for rainfall-runoff modelling of meso- and micro-scale catchments. Therefore we chose a spatial resolution of $1 \times 1\,\text{km}^2$. To achieve a precipitation product that justifies such a resolution, radar data are required. The combination of gauge and radar is part of an operationally working online tool that is developed by Jatho et al. (2010). With intent we did not follow one of the most promising developments, the use of radar data for regionalisation. We aim to provide an areal precipitation field independently of the current radar data, so that a lack of radar data can be compensated near real-time.

For interpolation we used the software package *InterMet*, which includes different regionalisation methods (Hinterding, 2003). Our focus laid on the further improvement of the so called Background-Fields (BGF)-method. We adapted the method establishing BGF for the low mountain ranges of Saxony. For this purpose, past daily rain gauge data of high spatial density respectively past radar data were included. For an hourly precipitation event one of the established BGF is used for regionalisation with de-trended kriging. The performance of the BGF-method was compared to other regionalisation techniques by applying different criteria (cross-validation, inspection, rainfall-runoff-modelling).

2 Investigation area and data

2.1 Investigation area

The investigation areas are situated within the low mountain ranges of the German Federal State of Saxony (Fig. 1). The mountain ranges – from east to west – are: Lusatian Mountains, Elbe Sandstone Mountains, Ore Mountains and *Vogtland*. Saxony is situated, according to genetic climatologic classification, in the West Wind Drift Zone of the Temperate Zone respectively in the zone of transitional climate between maritime West European and continental East
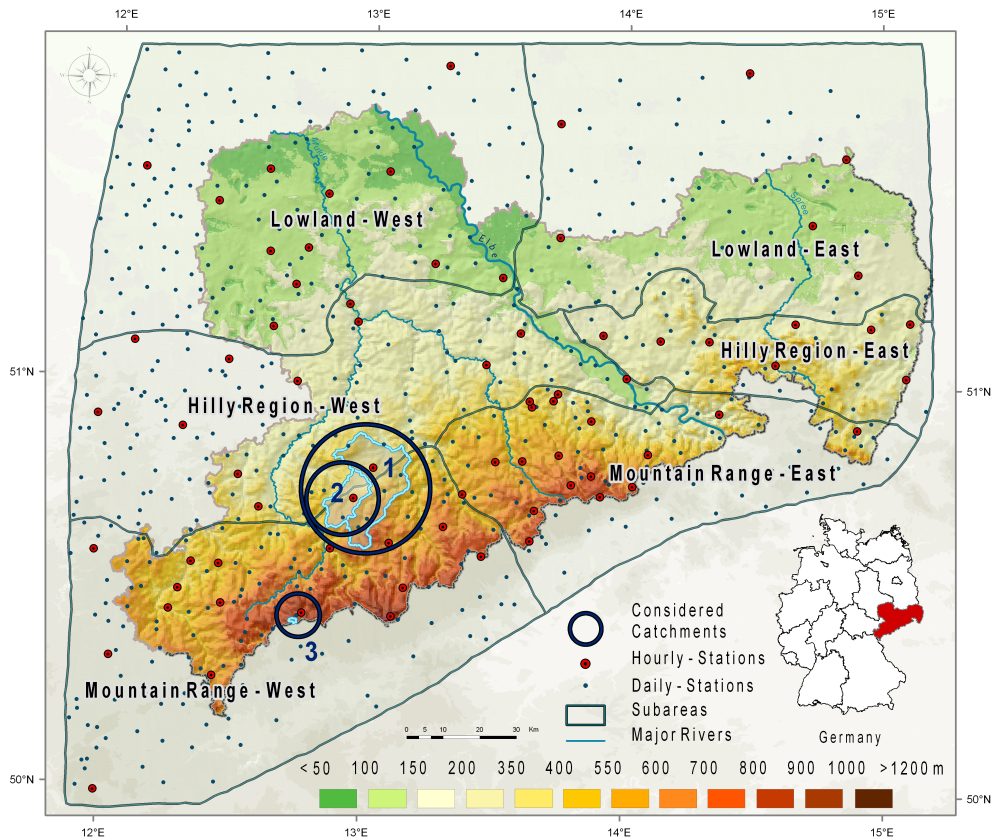
Fig. 1. Investigation area Saxony and its neighbourhood, rain gauges, digital elevation model, partition into six subareas and considered catchments (1. Chemnitz, 2. Würschnitz, and 3. Wiltzsch).

European climate. The mean temperature is higher than in regions of similar latitude due to the compensating effect of the Atlantic, especially the Gulf Stream. The influence of the Atlantic decreases from west to east, a fact that is noticeable in Saxony as well. The annual temperature amplitude in West-Saxony (18 °C) is 1 °C lower than in East-Saxony (climatological stations Leipzig and Görlitz, respectively). The same holds true for precipitation; in Germany the mean annual precipitation is about 800 mm, whereas in Saxony it is 600 mm (Bernhofer et al., 2008). Besides the distance to the sea, the low mountain ranges have an important influence on temperature and precipitation. Precipitation increases with increasing altitudes. At windward side of the mountain ranges a higher precipitation is observed due to enforced lifting and increased clouding. Precipitation is lower on leeward side as a result of cloud dissipation. Regions with most precipitation are the western mountain sides of the westerly Ore Mountains and of the *Vogtland*.

The investigation area was divided into six subareas (for reasons see Sect. 3.2). Within the scope of this work we focused on four subareas: two southern mountainous and two central hilly subareas (Fig. 1). They have a total area of 18 000 km^2 and an altitude range from 120 to 1214 m.

2.2 Data

Precipitation gauge data from Saxony and neighbouring areas of the Czech Republic (southeast) and four German Federal States were used (Fig. 1). No data were available from Poland, located east of Saxony. Within the four regarded subareas, 380 gauges have daily precipitation records and 67 gauges have sub-daily records. Not all of them operate at present. Most of them are operated by the German Weather Service (resolution: 10, 60 min, and daily), 27 gauges are operated by the Saxon State Ministry of Environment and Agriculture (resolution: 1 and 60 min) and some by the Czech Hydrometeorological Institute (resolution: daily). German data are accessible online. They are continuously assimilated into a database, where they are checked on plausibility (Jatho et al., 2010). We decided to aggregate the highly resolved data to hourly sums (hh:00) in order to be congruent with stations that deliver exclusively hourly values. No correction of systematic errors in precipitation amounts was carried out.

Radar data were used for visual comparison with interpolated rain fields, for the determination of BGF and as direct input into a rainfall-runoff model. The latter was used for evaluating the effect of different regionalisation methods on

runoff. The German Weather Service operationally generates an hourly adjusted radar product, where derived rainfall intensities are calibrated with hourly rain gauge intensities. By this means, the uncertainties associated with the determination of precipitation intensities from measured reflectivity can be overcome to a certain extent. The product is called *RADOLAN* (Bartels, 2004) and is available as a quantitative precipitation field for Germany (temporal resolution: 60 min; spatial resolution: $1 \times 1 \, \text{km}^2$). It is provided at the time hh:50, which causes a time difference between radar und gauge data of 10 min. We did not consider the option to use radar reflectivity (resolution: 5 min) and aggregate them to full hours, because a premise of our project was to utilise publicly available data that are ready to use. Otherwise a calibration of radar data would have been necessary, which is a great effort in an operational mode. The option to use only gauge data with a 10-min resolution and aggregate them to the time hh:50 would entail, that 31 gauges with an hourly resolution could not be used without transformation. The resulting gauge density of one gauge per $500 \, \text{km}^2$ would be far to less to account for variability of rainfall in time and location. Therefore we decided to consider the resulting time difference during the analysis of the results.

All gauge data were projected into a Gauss-Krüger coordinate system (central meridian 15° east). Radar data are originally projected in a polar stereographic grid system of the German Weather Service. For the set up of the BGF they were transformed into a Gauss-Krüger coordinate system. The projection of radar was chosen for visualisation.

3 Methods

In search of regionalisation methods that are robust and operationally applicable, we focus in this study on various geostatistic kriging methods of the software package *InterMet*, especially on the Background-Fields-Methods (BGF). We checked them against a group of kriging models of *InterMet* called *Default* and against the deterministic method IDW. All the methods were successfully applied for mountainous regions (Dirks et al., 1998; Hinterding, 2003; Ahrens, 2006).

3.1 Inverse Distance Weighting

The basic idea of IDW is that an unknown point gets a weighted mean of neighbouring points. The weight depends on the inverse distance between unknown and known points. The power of the distance in this function describes the magnitude of dependency. A higher power results in a lower weighting of remote points and assigns higher weights to local dependencies. Lower values result in a more regional view of dependencies due to the increased weighting of remote points. Dirks et al. (1998) showed that it is not crucial to find the optimum power for the regionalisation of highly resolved precipitation. Thus, we followed their recommendation and used the power two.

3.2 Kriging

Kriging belongs to a group of geostatistical techniques that interpolate the value of a random field at an unobserved location from observations at nearby locations. Kriging computes the best linear unbiased estimator based on a stochastic model of the spatial dependence quantified either by the variogram or by the expectation and the covariance function of the random field. It is assumed that the random variable consists of a large-scale deterministic and a small-scale stochastic fraction. The large-scale fraction can be constant, but also dependencies on external factors can be described. Often altitude is used for regionalisation of precipitation via linear regression. Different kriging options can be applied depending on the stochastic properties of the random field. Simple kriging assumes a known trend and ordinary kriging an unknown constant trend. Universal kriging assumes a general linear trend model. External drift kriging is a special form of universal kriging, where for example the altitude can be used as an external drift. For details on theory and general application of kriging, see Cressie (1991) and Chilès and Delfiner (1999). Kriging was applied for interpolation of precipitation measurements, e.g. by Atkinson and Lloyd (1998), Goovaerts (2000) and Hinterding (2003).

Our aim was to find methods for operational application. Therefore, we used a software package called *InterMet* that was developed for automatic interpolation of meteorological variables (Hinterding, 2003). Static and dynamic non-stationarities in the investigation area are considered by so called homogeneity areas and by an adaptation of the interpolation model to the random field of each homogeneity area. We partitioned the investigation area into six subareas (Fig. 1) to account for the static part of non-stationarity. It is envisioned to consider existing precipitation gradients (recall Sect. 2.1). The predominant west wind causes that the eastern part of the Ore Mountains lies in the rain shadow of the western part (Flemming, 2001), yet a gradient exists also in the lowland. The dynamic part of non-stationarity depends on the current event. Generally, the application of kriging methods is affected due to the fact that precipitation patterns cover only a part of a domain. Therefore, areas with precipitation are distinguished from areas without precipitation applying indicator kriging. Stations with precipitation intensity higher than $0.1 \, \text{mm} \, \text{h}^{-1}$ get assigned one and the remaining stations zero. Combination of static and dynamic non-stationarity leads to the homogeneity areas. Because all homogeneity areas are modelled separately, fuzzy methods are deployed in *InterMet* in order to avoid strong gradients along the borders of homogeneity areas.

The experimental variogram is determined automatically, using the robust variogram estimator after Genton (1998). It was chosen, because the classical estimator proposed by Matheron (1962) is not robust against non-normal distributed data and outliers. Variance $2\hat{\gamma}(h)$ is calculated from the k-th quantile of the sorted set of all absolute differences of the
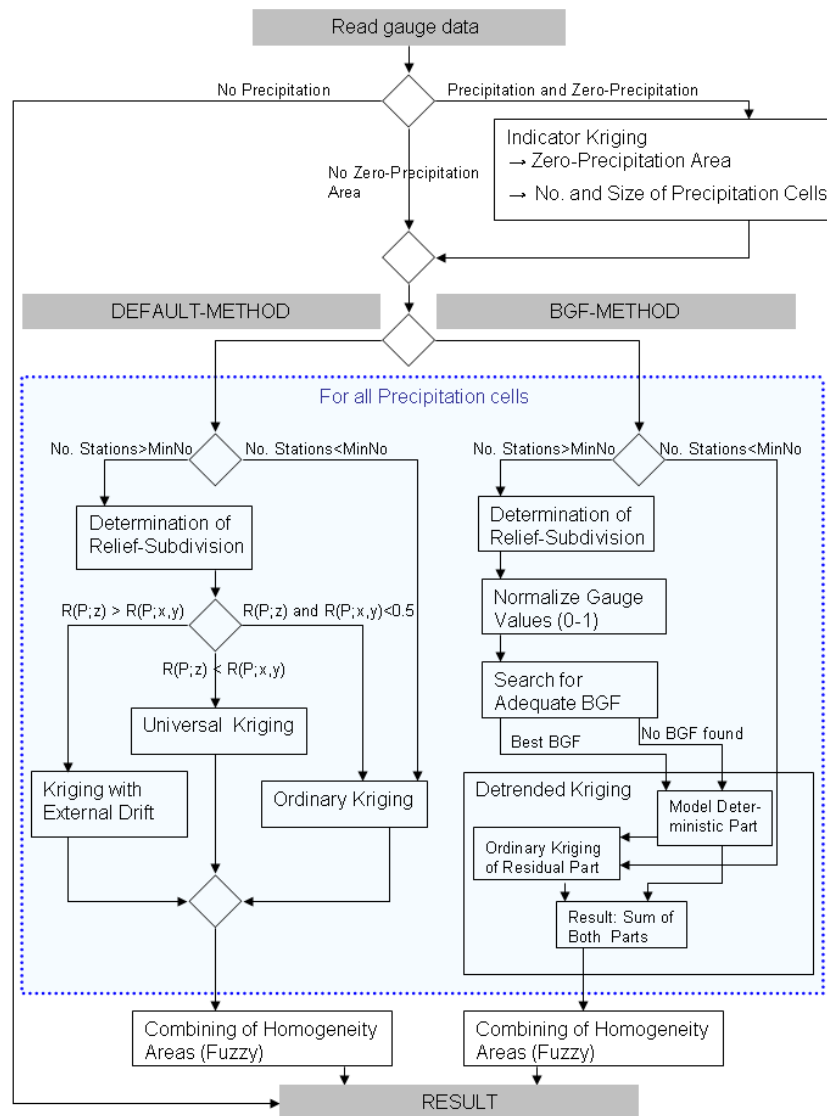
Fig. 2. Flow chart of the methods *Default* and BGF of the software package *InterMet*. (Used abbreviations: MinNo – minimum number of stations; R – correlation coefficient; P – Precipitation; x, y, z – space coordinates).

samples $V_i(h) - V_j(h)$:

$$2\hat{\gamma}(h) = (2.2191\{|V_i(h) - V_j(h)|; i < j\}_{(k)})^2 \quad (1)$$

$$k = \left(\frac{[N_h/2] + 1}{2}\right) \quad (2)$$

The factor 2.2191 is chosen to ensure that the estimator is unbiased for the case of normal distribution. The term $[N_h/2]$ denotes the integer part of half of the sample size N_h.

The number of lags and lag-size is determined in dependence on the distribution of available gauges. The experimental variogram is converted into a monotonically increasing function using the pool adjacent violators algorithm proposed by Barlo el al. (1972). Hereof, nugget c_0, sill c and

range a can be determined. The first value of this function is fixed as the nugget. The sill is reached, if the experimental variogram is constant in the last part. Otherwise the sill and the range will be infinite. The range is fixed at the lag value, where the sill is reached, minus half the lag size. Subsequently, an exponential theoretical function is adjusted.

$$\gamma(h) = c_0 + c \cdot \left(1 - \exp\left(\frac{-3h}{a}\right)\right) \quad (3)$$

Although anisotropy occurs in hourly precipitation data, it is not regarded in *InterMet*. Precipitation cells are frequently divided and modelled separately due to subdivision of the investigation area. Thus anisotropic effects are reduced (Hinterding, 2003).

Table 1. Number of determined BGF$_{gauge}$ and BGF$_{radar}$ for five wind directions and two seasons in the investigated subareas.

Method	Subarea	No. hourly stations	No. daily stations	BGF for five wind directions and two seasons*										
				U		NE		SE		SW		NW		
				Su	Wi	Su	Wi	Su	Wi	Su	Wi	Su	Wi	Total
BGF$_{gauge}$	Hilly-Region-East	7	35	9	8	8	7	6	4	27	22	28	29	148
	Hilly-Region-West	11	117	10	10	3	8	10	8	23	10	10	26	118
	Mountain-Range-East	10	79	10	6	10	7	5	3	27	22	29	30	149
	Mountain-Range-West	11	130	10	10	7	10	5	4	27	24	28	29	154
BGF$_{radar}$	Mountain-Range-West	17	565**	6	4	3	5	1	2	4	6	5	4	40

* Abbreviations: U – undefined, NE – northeast, SE – southeast, SW – southwest, NW – northwest, Su – summer, Wi – winter.

** These are no real daily stations, but virtual stations based on radar information. For details see Sect. 3.2.

Four different interpolation models for precipitation are implemented in *InterMet*. We used the two following (compare Fig. 2):

- BGF-method: a de-trended kriging is applied. Typical precipitation fields are considered as trend. Past hourly and daily gauge data are used to determine BGF. Furthermore, we tested an option of the BGF-method that uses radar data instead of daily gauge data to establish the BGF. In the subsequent sections we give more details.

- *Default*-method: various linear kriging models are implemented and applied for homogeneity areas. The most adequate kriging model is automatically chosen based on an analysis of the current event. If the correlation between precipitation and altitude of the gauge exeeds 0.5, kriging with external drift is used. Universal kriging is applied, if the spatial (x, y) correlation exeeds 0.5. The maximum correlation coefficient determines the method of choice. Otherwise ordinary kriging is used. For more details see Hinterding (2003).

3.2.1 Background-fields method using rain gauge data (BGF$_{gauge}$)

This interpolation method is based on the assumption that typical precipitation fields exist for different weather conditions and seasons. These fields are the result of gradients caused by altitude, continentality or due to large or small scale wind- and leeward effects. We determined typical precipitation fields (BGF$_{gauge}$) for each of the four considered subareas for regionalisation of hourly precipitation events. For preparation of BGF$_{gauge}$ past hourly and daily gauge data are required. The precipitation dataset of the period from 1992 to 2005 was split into subsamples considering winter and summer and general weather conditions. For the latter we used five wind directions (northeast, southeast,

southwest, northwest, undefined) of the *objective weather type classification of the German Weather Service* (Bissolli and Dittmann, 2003). To obtain a more detailed view of the precipitation events during the most frequently observed *objective weather types* southwest and northwest, the dataset was further subdivided into three (south/southwest/west and west/northwest/north, respectively) weather types that were determined after the subjective classification method from Hess and Brezowsky (1977). Applying cluster analysis (software: SPSS 14.0) with the hourly rainfall data, we determined several typical precipitation events for each of the weather types in winter and summer, in total 569 typical events (Table 1). To increase information density regarding spatial variability of these chosen precipitation events, the dense net of daily rain gauges was considered as past knowledge. Daily values were disaggregated into hourly intervals preserving temporal variability of nearby gauges. Integration of these values lead to a spatial refinement of the chosen typical precipitation events. Up to this step the typical precipitation events were punctual information. In a final step they were interpolated onto a previously defined grid of 1×1 km^2 that is used in *InterMet*. The final BGF$_{gauge}$ are normalised with the maximum precipitation to yield an areal precipitation field with values between 0 and 1 (see Fig. 3 for an example). Additionally, for each combination of weather type and season a so called basic model is generated. In this special BGF$_{gauge}$ all grid cells are allocated with 1. This implies that no meaningful secondary knowledge (deterministic component) enters into the regionalisation method. The gauge values are regionalised with ordinary kriging exclusively. For further details on the determination of BGF the reader is referred to Hinterding (2003).

A de-trended kriging is applied in *InterMet* to use the BGF$_{gauge}$ for regionalisation of hourly events. For a current data set, which can even have different stations than the ones used for the establishment of the BGF$_{gauge}$, the best fitting BGF$_{gauge}$ is automatically chosen (Fig. 2). Information regarding the current weather type is downloaded from

the World Wide Web pages of the German Weather Service. Gauge data are normalised with maximum precipitation. The normalised data are compared with the corresponding BGF$_{gauge}$ grids of the current season and weather type. The BGF$_{gauge}$ with the best least squares fit to gauge data is chosen. It represents the deterministic component of the current precipitation field. Existing differences between BGF$_{gauge}$ and current gauge values represent the stochastic rainfall component and are interpolated with ordinary kriging. The final rainfall field is the sum of both components. An advantage of this method is the possibility to consider detailed spatial information of past precipitation events. Therefore, current data shortages are better compensable (Hinterding, 2003).

3.2.2 Background-fields-method using radar data (BGF$_{radar}$)

Radar data provide a higher density of spatial information than other observations of precipitation. Within this study we investigated, if this additional information can be advantageously used by the BGF-method. Instead of considering daily rainfall observations to improve the establishment of typical precipitation fields, hourly calibrated radar data with a high spatial resolution were used. A further advantage is the possibility to use hourly radar fields directly without the necessity to disaggregate them from daily values. This potentially reduces inaccuracies caused by the disaggregation procedure.

Calibrated radar data were available only from 2005 to 2007. We delimited our work on the south-western subarea (Mountain Range-West; see Fig. 1), because of the high effort in preparing radar data. The procedure of determining typical precipitation events based on hourly rain data is the same as explained in Sect. 3.2.1, except that hourly data of only three years (2005–2007) were considered here. Furthermore, only the objective weather type classification was used, because a further splitting of data is not meaningful for small datasets. As a result of a cluster analysis of hourly gauge data we found 40 typical events for the selected subarea (Table 1). That is about one third of the BGF$_{gauge}$-method. Furthermore, the quality of the typical events was low in comparison to the period that could be used for the setup of the BGF$_{gauge}$. Radar data were integrated into the process of determining the BGF$_{radar}$ as virtual stations. The radar pixel of $1 \times 1 \, \text{km}^2$ is represented by a point. Not all points could be included in the analysis due to restrictions of *InterMet*. Therefore, the net was equally thinned out; only every 20th pixel of the net was used. In total, data of 565 virtual stations were deployed. This is still four times more than the number of daily gauge stations in the subarea and implies a significant gain of spatial information.

To be able to compare all three methods we had to regionalise the chosen events exclusively in the south-westerly sub-
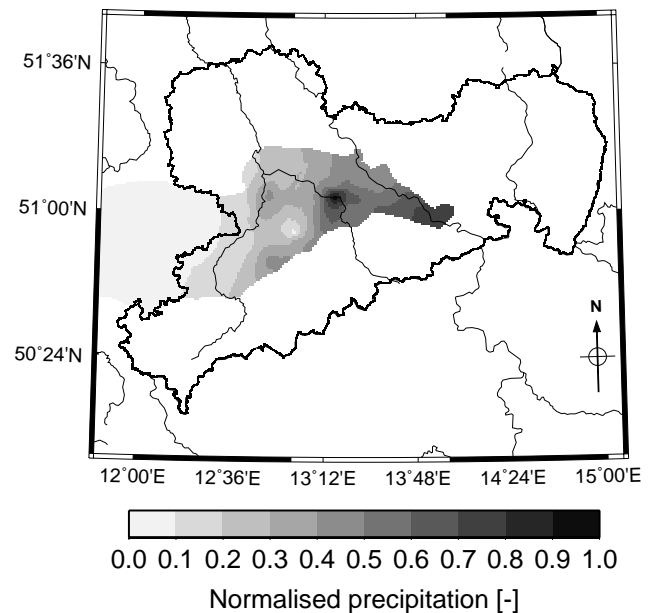

Fig. 3. BGF$_{gauge}$ for the north westerly subarea for summer and an undefined wind direction (objective weather type). This BGF$_{gauge}$ was used for the interpolation of the event 27 May 2006, 17:00 UTC.

area with the *Default*- and the BGF$_{gauge}$-method. Differences in the number of gauges would otherwise lead to differences at the margins of the regionalised subarea.

3.3 Performance criteria

To evaluate the performance of the applied regionalisation methods we used three criteria:

(1) The standard method to assess the quality of regionalisation methods is leave-one-out cross-validation (Wackernagel, 1995). For data pairs of observed (P_o) and regionalised precipitation (P_m), the mean absolute error (MAE), the mean squared error (MSE) and the Pearson product-moment correlation coefficient (R) were calculated (Eqs. 4–6).

$$\text{MAE} = \frac{1}{n} \sum_{k=1}^{n} |P_o - P_m| \tag{4}$$

$$\text{MSE} = \frac{1}{n} \sum_{k=1}^{n} (P_o - P_m)^2 \tag{5}$$

$$R(P_o, P_m) = \frac{\text{cov}(P_o, P_m)}{\sqrt{\text{Var}(P_o)} \cdot \sqrt{\text{Var}(P_m)}} \tag{6}$$

MAE and MSE values close to zero and R close to one indicate a high accuracy of the method. The MSE is very sensitive to large differences between observed and predicted values due to squaring of the differences. A good correlation coefficient can misleadingly indicate a high quality of

a method, although the predicted values are biased. A detailed look onto the data pairs can avoid this. That is why all criteria have to be considered. In the decision process, the highest weight was given to the correlation coefficient, whereas both other criteria must not be considerably worse. A critical point of cross-validation is that the performance of regionalisation methods changes, if one value is excluded. The importance of these changes depends on the size of the dataset, the eccentricity of the point and the regionalisation method. In particular the regionalisation of convective rainfalls may be affected by high uncertainty, if convective cells are detected by just one gauge. The BGF-method is more sensitive, because exclusion of a crucial gauge can lead to a selection of a different BGF.

(2) Since not all criteria are objectively ascertainable, we visually compared regionalised precipitation fields and radar. An evaluation of radar images was possible with gauges that were not used for radar calibration (27 stations of the Saxon State Ministry of Environment and Agriculture). There is a temporal offset of ten minutes between hourly gauge values and radar. This can cause a shift of rain between two time steps with the amount of a 10-min rainfall intensity. The highest observed intensities in Saxony during our investigation period was 20 mm, but in most cases the shift is far below these values.

(3) Cross-validation is useful as an objective measure for the assessment of regionalisation methods. It extracts the method that processes the limited information best (in our case: precipitation of few hourly rainfall stations). However, in comparison to radar observations, interpolated gauge data often poorly represent the structure of rain cells. On the other hand we have to admit that we never know exactly the real spatial rainfall distribution, even not with calibrated radar observations.

The stream flow observed at a gauging station integrates runoff from a river catchment. The transformation of precipitation into runoff is the result of highly non-linear processes, which can be simulated with rainfall-runoff models. On the one hand, these models can clearly indicate errors in the spatio-temporal rainfall pattern (Casper et al., 2009; Kneis and Heistermann, 2009). On the other hand, the difference in the precipitation information used during the calibration and the validation periods can cause errors in hydrological modelling. A better interpolated field can lead to a worse model performance (Bàrdossy and Das, 2008). Thus, rainfall-runoff modelling can be used as an additional means to evaluate the performance of different interpolation methods, but the quantitative result has to be carefully interpreted.

Interpolated rain fields were used as input for a calibrated rainfall-runoff model, where the simulated river runoff was compared with the observed. The idea is the following. The more accurate the spatio-temporal variability of the real rain field is reproduced by the regionalised rain field, the lower are the differences between observed and simulated runoff. Consequently, we interpret the error of runoff mo-

delling as performance criteria of the input rain field. Similar approaches were successfully applied by Sun et al. (2000) and Pessoa et al. (1993). We used the conceptual hydrological model *ArcEGMO* (Becker et al., 2002), which is in wide use for its short computational time, its flexibility in spatial, temporal and structural resolution and the mainly GIS-based assignment of parameters. The model was calibrated for the Mulde catchment for flood events from 1954 to 2006 using precipitation data from a) recording stations with high temporal resolution and b) disaggregated time series from stations with daily records (Dietrich et al., 2008). The upper Mulde catchment is situated in the Ore Mountains, whereas several sub-catchments drain from South to North. The following sub-catchments in the Mountain Range-West (Fig. 1) were examined for this study (river/runoff station/catchment size):

- Chemnitz/Chemnitz 1/532 km^2,

- Würschnitz/Jahnsdorf 1/136 km^2, and

- Wiltzsch/Carlsfeld ZP1/1.5 km^2.

The period from 26 July 2006 to 11 August 2006 was computed with a default set of model parameters, which resulted from prior calibration and proved to be efficient for historic flood events for the Chemnitz and Würschnitz catchment from 1994 on. For the very small Wiltzsch sub-catchment, a contributing area of a reservoir, the model was only calibrated for the flood event evaluated within this case study. All hourly simulations started seven days before the chosen event for an optimal adjusting of initial model conditions, which had been computed by a continuous long time model run with daily time steps before (Dietrich et al., 2008). We calculated the following objective indicators of performance for the events four and five:

1. Ratio of simulated to observed runoff volume.

$$ VR = \frac{R_\mathrm{m}}{R_\mathrm{o}} \tag{7} $$

2. The Nash-Sutcliffe model efficiency coefficient (Nash and Sutcliffe, 1970) is applied to quantitatively describe the accuracy of model outputs.

$$ E = 1 - \frac{\sum\limits_{t=1}^{T} \left(R_\mathrm{o}^t - R_\mathrm{m}^t \right)^2}{\sum\limits_{t=1}^{T} \left(R_\mathrm{o}^t - \overline{R_\mathrm{o}} \right)^2} \tag{8} $$

In Eqs. (7) and (8) R_o is the observed and R_m is the modelled runoff at time t. Nash-Sutcliffe efficiencies can range from $-\infty$ to 1. If E equals 1, it corresponds to a perfect match of modelled to observed runoff. If the model predictions are as accurate as the mean of the

Table 2. Analysed precipitation events and details on observed rainfall amounts and weather conditions.

Event No.	Date	Considered duration [h]	Precipitation gauge [mm h^{-1}]		Precipitation radar [mm h^{-1}]		Precipitation type/Weather condition
			mean	max	mean	max	
1	27 May 2006	2	1.5	8.5	1.2	25.3	Advective/NW – anticyclonal
2	16 June 2006	2	1.2	26.0	1.1	103.0	Convective/SW – cyclonal
3	19 June 2006	3	0.9	12.4	0.9	46.7	Convective/SW – cyclonal
4	31 July 2006	3	0.1	9.7	0.5	29.4	Convective/SW – anticyclonal
5	5–8 August 2006	84	0.6	16.4	0.41	34.4	Advective/NO, NW – anticyclonal
6	18 January 2007	3	2.1	11.9	1.9	23.8	Convective/SW – cyclonal

observed data the efficiency equals 0. Nash-Sutcliffe efficiency less than 0 indicate that the observed mean is a better predictor than the model. The coefficient evaluates especially the simulation of discharge peaks, what is of major interest in flood forecast.

3. Relative error of peak runoff [%].

$$R_{\text{peak}} = 100 - \frac{R_{\text{o-peak}}}{R_{\text{m-peak}}} \cdot 100 \qquad (9)$$

4. Temporal offset of peak runoff [h].

The Nash-Sutcliffe coefficient was given most weight in the process of determining the best input field, because it considers both, the peak adjustment and the average runoff. However, the other criteria are subjectively considered for the required interpretation of the model results as well.

4 Results

We choose six rainfall events (four convective and two advective) in the recent past to analyse the performance of the applied regionalisation methods (Table 2). Our focus laid on convective events. Since convective rainfalls are often spatially highly variable, it is more demanding to find appropriate regionalisation methods. To determine the type of a precipitation event we used a method that bases on radar data (Ehret, 2003). The performance check using a rainfall-runoff model was carried out for events No. 4 and 5. The BGF$_{\text{radar}}$-method was less intensively analysed, because differences to BGF$_{\text{gauge}}$-method were minimal.

Cross-validation results of the applied regionalisation methods are listed in Table 3. The absolute values of the criteria MAE and MSE are not comparable between the events, because they depend on type and volume of the respective precipitation event. MAE ranges from 0.51 to 1.17 for advective and from 0.64 to 1.24 for convective events. Although no clear differences in the MAE values of the six events between advective and convective events are recognizable, the level is lower for advective events. MSE is remarkably higher for

Table 3. Cross-validation results of three regionalisation methods for six precipitation events. Given are means of the mean absolute error (MAE), mean squared error (MSE) and correlation coefficient (R). Best results for each event are typed in bold letters.

Event No.	Method	MAE [mm]	MSE [mm^2]	R
1 (adv.)	IDW	1.17	3.08	0.48
	Default	1.09	2.92	0.54
	BGF$_{\text{gauge}}$	**1.06**	**2.78**	**0.55**
2 (conv.)	IDW	1. 21	**9.00**	**0.55**
	Default	1.21	10.15	0.50
	BGF$_{\text{gauge}}$	1.21	10.56	0.45
3 (conv.)	IDW	1.21	**6.16**	**0.28**
	Default	1.24	7.32	0.17
	BGF$_{\text{gauge}}$	**1.18**	7.20	0.14
4 (conv.)	IDW	0.83	2.73	0.23
	Default	0.70	2.30	0.55
	BGF$_{\text{gauge}}$	**0.64**	**2.12**	**0.58**
5 (adv.)	IDW	**0.51**	**1.28**	0.40
	Default	0.61	1.42	0.45
	BGF$_{\text{gauge}}$	0.59	1.42	**0.45**
6 (conv.)	IDW	**1.00**	**3.49**	0.33
	Default	1.11	3.53	**0.46**
	BGF$_{\text{gauge}}$	1.18	4.20	0.41

convective events (advective: 1.28–3.08; convective: 2.12–10.56). Large differences between observed and interpolated precipitation result in a high MSE due to squaring of differences. The often observable small extent of convective cells frequently leads to high MSE values. The correlation coefficient R ranged from 0.4 to 0.55 for the advective events and from 0.14 to 0.58 for the convective events. All in all, cross validation results confirmed the expectation that regionalisation of advective events shows a better performance than regionalisation of convective events (Ehret, 2003). Variability of all error criteria depended on the constellation of
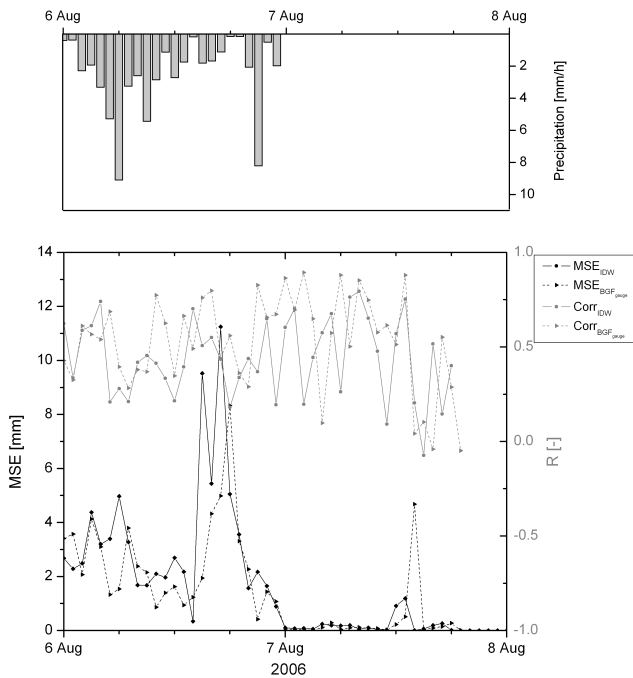
Fig. 4. Performance criteria of cross validation (Mean Squared Error, MSE, and correlation coefficient R) of 48 h of event 5 for the methods IDW and BGF$_{gauge}$. Displayed catchment rainfall is radar based.

gauges to precipitation cells and their spatial extension. If a precipitation intensity range was measured by more than one gauge, the leave out of one gauge did not lead to a remarkable decrease of the cross-validation results (event 4). Otherwise, if a precipitation intensity range was measured by just one gauge, the leave out of this gauge leads to poor cross-validation results (event 3). This explains the high variability of the performance criteria within a single event. Exemplarily, the fluctuations of IDW and BGF$_{gauge}$ of 48 h of event 5 are visualised in Fig. 4. There is no overall best performance of one method recognisable. In discrete hours the performance of IDW and BGF$_{gauge}$ is often contrary. Especially MSE can vary by one order of magnitude within two hours. Highest errors occurred when catchment rainfall and spatial extension of rain cells was low. There is no connection between MSE and R.

We counted how often each method performed best for the single events. The kriging method BGF$_{gauge}$ as well as IDW were in two cases the best methods (Table 3). For the events five and six it is not clear from the cross validation criteria, which method performed best. Giving the correlation coefficient most weight, the BGF$_{gauge}$-method and the *Default*-method are the most appropriate. In general, the overall differences between the applied regionalisation methods were small. There was no special aptitude of regionalisation methods either for convective or advective events.

For a visual comparison one time step of the regionalised precipitation events 1 and 2 and the radar products is exemplarily illustrated in Figs. 5 and 6. Generally, we first looked at congruence between precipitation intensities of gauges and radar. If there were only minor differences, we assumed that radar represents well the spatial rainfall distribution, but not necessarily the exact areal precipitation intensity.

The radar image reveals a nearly area-wide rainfall with embedded high-intensity cells during the advective event on 27 May 2006, 17:00 UTC (Fig. 5). The delineation of precipitation areas was better succeeded by the methods BGF$_{gauge}$ and *Default*. In the northern part radar indicates an area without precipitation that was not measured by gauges and could therefore not be reproduced. Observed higher precipitation intensities were regionalised with IDW as single precipitation cells that have a slightly higher elongation in north-south direction (Fig. 5c). The *Default*, and even more the BGF$_{gauge}$-method, merged the cells that were adjacent and reproduced better the east-west orientated structures of the rain field (Fig. 5a, b). The highest radar based precipitation intensities were not observed by gauges. A satisfying congruence between radar and regionalised gauge data was therefore not possible. However, the structures that are recognisable in the radar image were better reproduced by the kriging methods. Within the *Default*-method, universal kriging was automatically chosen as appropriate for the north-westerly subarea. In the other subareas ordinary kriging was applied, because there was neither a strong correlation to sea-height nor a spatial trend in the data. Proper BGF$_{gauge}$ were found and applied for the north-westerly (Fig. 3) and the south-easterly subarea. Apart from that, basic BGF$_{gauge}$ were used for the other subareas, i.e. no secondary information entered the regionalisation method.

The precipitation event 16 June 2006, 14:00 UTC was identified as a mesoscale convective system. The sharp delineation of the rain areas was reproduced well by the *Default*- and BGF$_{gauge}$-method (Fig. 6a, b). IDW produced more extensive rainfall areas (Fig. 6c). Reproduction of cell structure was limited due to the low number of observations. The radar image (Fig. 6d) shows a large north-south stretched cell with high rain intensities and two smaller cells in the western part with lower intensities. The smaller cells were detected only by gauges that are located at the margin of the cells. That is why they could not be reproduced properly by the applied regionalisation methods. The large cell was detected by 10 gauges. The size of the northern part of the cell was regionalised satisfactorily, but maximal intensity differed about 50%. The differences between radar and regionalised fields were large in the southern part. The maximum captured intensity by a gauge was $26 \, mm \, h^{-1}$, and radar values in the vicinity reached values of more than $100 \, mm \, h^{-1}$. Two gauges in the centre of the high intensity cell (Fig. 6d) measured 0.7 and $9.6 \, mm \, h^{-1}$. Large discrepancies exist also around the cell (gauge observations: 4.4, 1.8, 3.3, and $10.3 \, mm \, h^{-1}$). All these stations belong to the
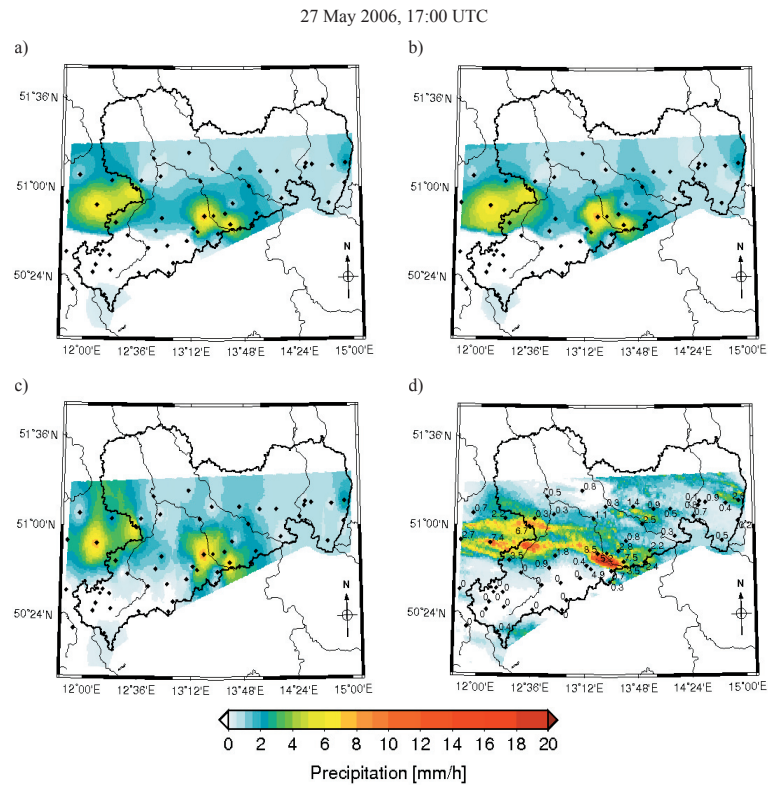
27 May 2006, 17:00 UTC

Fig. 5. For the advective event 27 May 2006, 17:00 UTC the regionalised precipitation fields **(a)** BGF$_{gauge}$, **(b)** *Default*, **(c)** IDW, and **(d)** the radar field are presented.
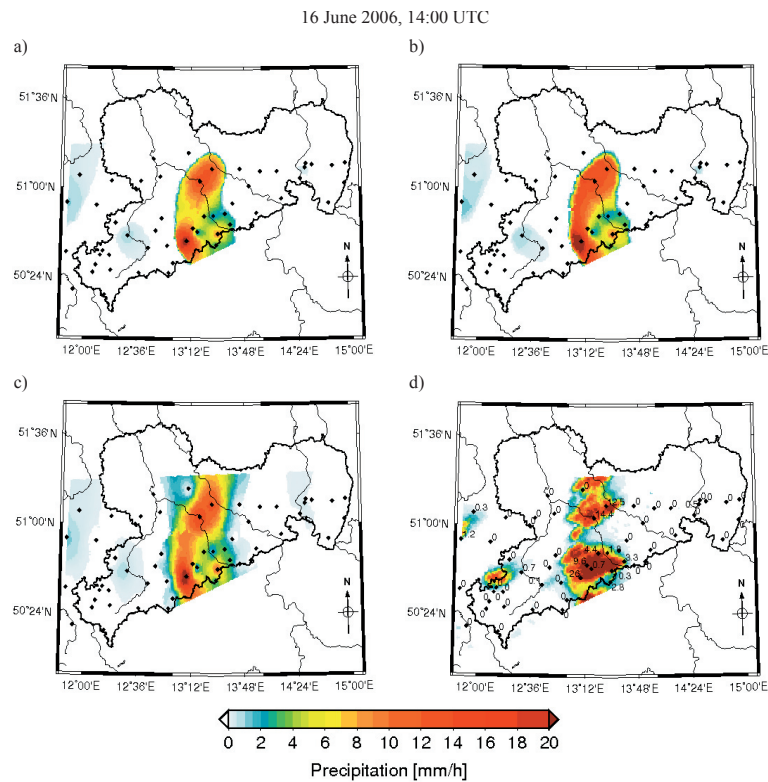
16 June 2006, 14:00 UTC

Fig. 6. For the convective event 16 June 2006, 14:00 UTC the regionalised precipitation fields **(a)** BGF$_{gauge}$, **(b)** *Default*, **(c)** IDW and **(d)** the radar field are presented.

Table 4. Performance criteria for the runoff simulation of two precipitation events. Four precipitation fields served as input. Performance criteria were calculated for the sub-catchments Chemnitz 1 (Ch.), Jahnsdorf 1 (Ja.), and Carlsfeld (Ca.). Best results for each event are typed in bold letters.

Precipitation input	Volume ratio			Nash-Sutcliffe coefficient			Relative error of peak runoff (%)			Temporal offset of peak runoff (h)		
	Ch.	Ja.	Ca.	Ch.	Ja.	Ca.	Ch.	Ja.	Ca.	Ch.	Ja.	Ca.
Event no. 4 (convective)												
Radar	**1.04**				−0.16			**3.4**			**0**	
Default	**1.04**				−0.24			5.5			**0**	
BGF$_{gauge}$	**1.04**				−0.23			5.3			**0**	
IDW	0.89				**0.13**			−13.6			−1	
Event no. 5 (advective)												
Radar	1.06	1.16	0.61	0.91	**0.95**	0.79	**−0.5**	**−3.6**	−32.7	1	2	0
Default	1.06	1.13	0.57	**0.92**	0.92	0.76	−1.2	−9.9	−33.8	1	**1**	0
BGF$_{gauge}$	**1.04**	0.84	0.57	**0.92**	0.85	0.76	−3.7	−33.6	−33.8	1	2	0
IDW	1.18	**1.03**	**0.71**	0.78	0.9	**0.89**	20.9	−11.5	**−16.4**	**0**	**1**	0

station net of the Saxon State Ministry of Environment and Agriculture, and were not used for radar calibration. We conclude that in this area an erroneous radar precipitation is very probable. That is why radar is not suited for a visual comparison at this location. Within the area of lower intensities in the centre of the large cell (Fig. 6d), no gauge observations were available. This area was best regionalised by the BGF$_{gauge}$-method (Fig. 6a). No appropriate BGF$_{gauge}$ was found for regionalisation, and therefore ordinary kriging was applied for all 10 delimited homogeneity areas. Within the *Default*-method no spatial correlation and no correlation to sea height were determined in the four delimited homogeneity areas. Ordinary kriging was applied in all homogeneity areas. Note, only the different subdivision of the investigation area (BGF$_{gauge}$: 10; *Default*: 4) led to differences in the regionalised precipitation field.

The period from 26 July 2006 until 11 August 2006 was simulated with a rainfall-runoff model. It covers the small scale convective event 4 for the Jahnsdorf stream flow gauge and the large scale advective event 5 for three gauges of different catchment size (see Sect. 3.3 and Table 2). Table 4 compares the results from rainfall-runoff modelling with four different driving precipitation fields, namely the three regionalised fields and the calibrated radar field.

For the largest catchment (Chemnitz 1) the regionalisation methods BGF$_{gauge}$, *Default* as well as radar produced similar results for the advective rainfall event, which are within the uncertainty range of the hydrological model as known from calibration and validation. However, differences to IDW showed up clearly, e.g. the flood peak was overestimated by 20%.

The radar input was best suited to reproduce the runoff in the smaller catchment Jahnsdorf 1 for both events. It caused a volume that was too high, but simulated very well both dis-

charge peaks. A similar performance resulted from the *Default* and IDW input. BGF$_{gauge}$ had the worst performance, e.g. more than 30% error in estimating the peak runoff.

In the smallest catchment Carlsfeld the simple method IDW performed best. A possible reason is the proximity of the rain gauge, which is situated at the outlet of the small catchment. The regionalisation methods do not produce relevant differences so close to a station. Most probable, IDW turned out to be best by coincidence. Since for the Carlsfeld catchment there was only one event observed, we transferred the model parameters from an adjacent catchment with similar characteristics to the Carlsfeld catchment without recalibration. This may explain the underestimation of runoff volume at the Carlsfeld gauge and weakens the conclusions about the regionalisation for this catchment.

Figure 7 shows exemplarily observed and simulated runoff for the Würschnitz/Jahnsdorf 1 catchment. The runoff that was simulated based on radar input had the highest congruence to measured runoff for the convective event No. 4 (Fig. 7a). It met the peak discharge best, had a low discrepancy to the observed volume and no temporal offset to the peak. Differences to *Default* and BGF$_{gauge}$ are not relevant, but the IDW input led to a significantly worse discharge simulation. For this relatively small flood event (far below the first alert level and out of the calibration range of the model), the influence of hydrologic model uncertainty is large.

Runoff peaks that resulted from the advective rainfall period (event 5) were best reproduced by the radar input field (Fig. 7b). The peak discharge was simulated very well, but runoff volume was overestimated (Table 4), because radar overestimated precipitation at the beginning of the event. Input fields from *Default* and IDW caused lower peak discharges. Only the BGF$_{gauge}$ rain fields resulted in a simulated peak flow that was too low. Both observed peaks are
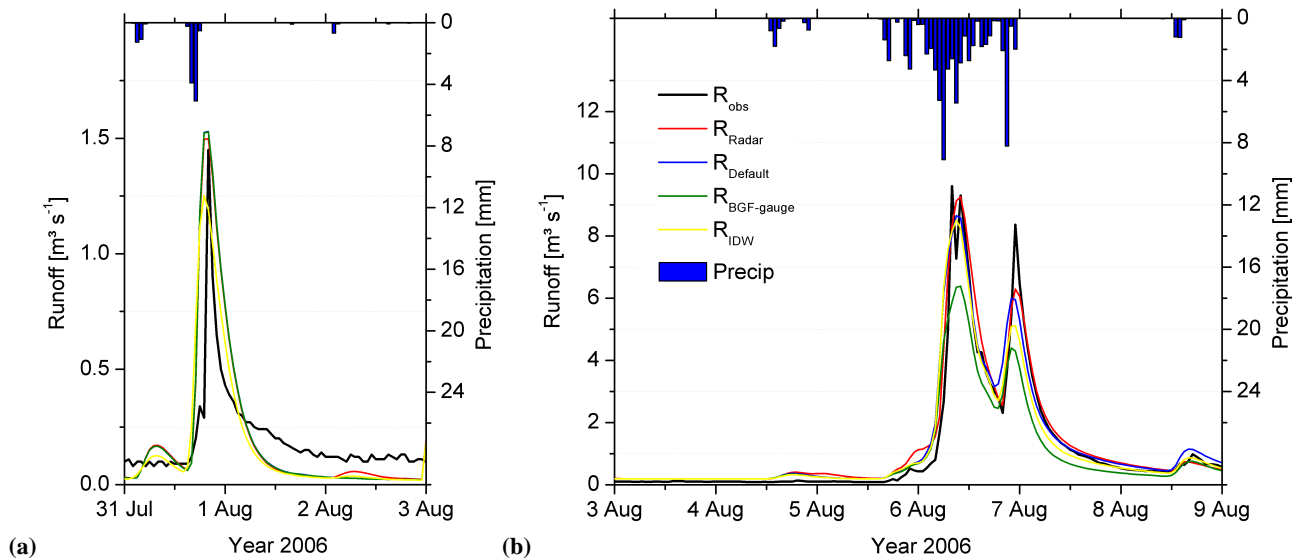
Fig. 7. Observed and simulated runoff for the Würschnitz catchment/gauge Jahnsdorf 1 for **(a)** a convective period and **(b)** an advective period (corresponds to events 4 and 5). Displayed catchment-rainfall is radar based. In (a) the *Default*-method is masked by the BGF_{gauge}-method.

about the same level, whereas the second was caused by less precipitation. The runoff of the first peak consists mainly of near to surface flows. At the second peak of hourly rainfall the near to surface flow was overlain by flows of the fast groundwater storage, which were caused by the first part of the rainfall event. Simulation of this second peak was therefore more demanding. Event 5 was one of only two observed events in the past, where the near to surface flow was dominating other runoff components. The recalibration of the concentration time of the near to surface flow resulted in a more behavioural "default" parameter set for the entire set of flood events by integrating event 5 into the calibration period of the hydrological model. Note that we did not calibrate a single event parameter set for the Jahnsdorf and Chemnitz catchments, but one parameter set with optimal performance for all observed events.

We tested a variation of the BGF-method for the southwesterly subarea (Fig. 1), where past radar data were used for preparation of the BGF_{radar}. Cross-validation results of this method are compared with the applied kriging methods *Default* and BGF_{gauge} in Table 5. Results are presented with three digits to see any differences between both BGF-methods. Differences were very small and from our opinion negligible. Some minor improvements of BGF_{radar} over BGF_{gauge} are identifiable for the events no. 4 and 6. A deterioration of performance did not occur.

The main reason that the BGF_{radar}-method did not perform better is the insufficient number of available past BGF_{radar}. We found nearly four times less BGF_{radar} when using radar data instead of exclusively using rain gauge data (BGF_{gauge}). Here, the limited amount of radar data was crucial. The quality of the determined BGF_{radar} was lower. Because of

the limited number of available events, typical precipitation events with a lower quality had to be chosen within the cluster analysis. The subarea is, depending on number of precipitation cells, further divided into homogeneity areas for a current event. For each homogeneity area an adequate BGF_{radar} is needed. If there is no adequate BGF_{radar}, the basic model is used. In most hours of the chosen events regionalisation was carried out with the basic model. The advantage of having entered much more details of spatial variability of rainfall fields via radar data had not taken effect. However, even in hours, where adequate BGF_{radar} for one or more homogeneity areas were available, the method BGF_{gauge} was mostly better than BGF_{radar}. This clearly indicates the poor quality of the applied BGF_{radar}.

Runoff of a catchment in this subarea was not modelled with the rainfall-runoff model, because cross-validation results did not let us expect satisfying results.

5 Discussion and conclusions

In the context of flood forecasting we aimed at the improvement of regionalisation methods for precipitation in mountainous catchments. Although radar measurements are now available in some regions, we have to face the fact that radar data could be unavailable or could be of low quality during operational use. Therefore, we focussed on the development of regionalisation methods that consider spatially detailed information of past precipitation fields for the interpolation of hourly events.

As a result of cross-validation we found that the quality of regionalisation depends on the type of the precipitation event. Results were slightly better for advective than for convective

Table 5. Cross-validation results of the regionalisation methods *Default*, BGF$_{gauge}$, and BGF$_{radar}$ for six rainfall events for the south-westerly subarea. Given are the means of the mean absolute error (MAE), the mean squared error (MSE) and the correlation coefficient (R). Best results for each event are typed in bold letters.

Event No.	Method	MAE [mm]	MSE [mm^2]	R
1 (adv.)	Default	**0.392**	**0.962**	**0.340**
	BGF$_{gauge}$	0.433	1.295	0.221
	BGF$_{radar}$	0.433	1.291	0.221
2 (conv.)	Default	**0.867**	22.535	0.000
	BGF$_{gauge}$	**0.867**	22.533	–
	BGF$_{radar}$	**0.867**	22.533	–
3 (conv.)	Default	1.348	**9.494**	0.158
	BGF$_{gauge}$	**1.322**	9.552	**0.180**
	BGF$_{radar}$	**1.322**	9.552	**0.180**
4 (conv.)	Default	0.173	0.539	0.034
	BGF$_{gauge}$	0.157	0.515	0.155
	BGF$_{radar}$	**0.156**	**0.514**	**0.163**
5 (adv.)	Default	0.135	0.273	0.015
	BGF$_{gauge}$	**0.131**	**0.263**	**0.059**
	BGF$_{radar}$	0.132	**0.263**	0.016
6 (conv.)	Default	1.074	2.533	0.336
	BGF$_{gauge}$	0.791	**1.493**	0.302
	BGF$_{radar}$	**0.789**	**1.493**	**0.303**

events, coinciding with findings from Ehret (2003). The applied methods IDW, *Default* and BGF$_{gauge}$ showed no general differences in the aptitude for specific rainfall types. For single events, differences between regionalisation methods were not large. In two of six cases BGF$_{gauge}$ showed the best performance and in two cases IDW. No clear results appeared in two cases. There is a high temporal variability of the performance criteria that indicates that besides the event type the present constellation of rain cells and gauges is important.

The visual comparison between regionalisation methods and the radar image revealed that both kriging methods of *InterMet* (*Default* and BGF$_{gauge}$) delineated the rain areas better than IDW. A proper delineation of rain areas and hence regionalisation areas is important, e.g. to avoid negative rainfall estimations. The BGF$_{gauge}$-method reproduced structures of rain areas best. Differences to the *Default*-method were not grave, but differences were noticeable to IDW. The advantages of the kriging methods – consideration of stochastic properties of the rain field as well as of catchment characteristics (topography and typical precipitation patterns) – showed up clearly. In general, regionalisation methods provide a more homogeneous rainfall distribution than radar due to low data density. For instance, if precipitation is observed only at the margin of a small rain cell, it leads to a regionalised intensity that is too low in the proximate vicinity, and the resulting cell is probably larger due to scarce gauge net. However, the reverse happens too. If only the highest in-

tensity of a small cell is captured, the regionalised cell is too large and intensities are too high. These effects can lead to unrealistic areal precipitation estimates, particularly in smaller catchments. Often the effects neutralise each other and result in a satisfying areal precipitation field, especially in larger catchments.

An alternative attempt to judge over regionalisation methods was to use different precipitation input fields for rainfall-runoff modelling. Past studies showed that radar data improved flood estimation, but only in combination with gauge data (e.g. co-kriging, Sun et al., 2000). In our case study, the input fields from calibrated radar data and from kriging methods resulted in the best fitting of the simulated runoff hydrograph against observed runoff in three cases. For one event the BGF input gave worst performance. Runoff of the smallest catchment was best reproduced with the IDW input, most probably caused by the short distance to the next rain gauge. We expected that calibrated radar data are the most appropriate basis for flood simulations, because of their ability to resolve the spatial variability of rainfall. Why was radar not clearly superior? In our opinion, the main reason is the limited number of rain gauges available for radar calibration. Calibration is difficult in areas with strong precipitation gradients, e.g. during convective events (Bartels, 2004). A fundamental problem of radar calibration is that radar and gauges measure in different spatial and temporal scales (Pessoa et al., 1993). Radar measures reflectivity within a volume in the atmosphere in a split second that may have a projection on the ground of several square kilometres. Rain gauges quantify continuously cumulative precipitation in a sampling area of e.g. 200 cm^2. The spatial variability of rain gauges can be extremely high. Jensen and Pedersen (2005) detected variability of up to 100% between neighbouring gauges within an area of 500×500 m^2 in a flat region over a four day period. Furthermore, at any point rainfall may change within minutes or less. Therefore, gauge observations may not be representative for the area beneath the radar-sampled volume. A radar calibration with the observed gauge values does not guarantee a complete approximation to real areal rainfall. We disposed of 27 additional gauges in the investigation area, whose observations were not used for calibration of radar. Obviously, using the denser net for regionalisation produced – in comparison to radar observations – equivalent areal rainfall estimates in most occasions. Especially in small areas, such as the Carlsfeld catchment, the existence of a nearby gauge can improve significantly the estimate of areal precipitation.

The performance check using the rainfall-runoff model indicated a predominance of the applied kriging methods over IDW. This is consistent with the results of the cross-validation for the events four and five (recall Table 3). The low performance of the BGF$_{gauge}$-method for event 5 at Jahnsdorf 1 gauge needs to be discussed. For the calibration and validation events of the rainfall-runoff model between 1994 and 2008, the Nash-Sutcliffe efficiency had a

median of 0.93 (worst: 0.67), the median of the error in volume ratio was 2% (worst: 16%) and the median of the relative peak error was 4% (worst: 23%). The evaluation criteria for the BGF$_{\text{gauge}}$-method for event 5 are at the lower end of the range from calibration/validation, the error of the peak discharge is significantly worse than one could expect from known model uncertainty. Even if we assume that total model uncertainty is larger than the uncertainty known from the observed period, there is an indication that the BGF method did not perform well for the combination of sub-catchment and event. However, differences between methods were mostly not pronounced in their impact on runoff. In addition it should be mentioned, that small changes in model parameters could easily favour a different method of the ones with rather similar performance. A further aspect is that the applied regionalisation methods differ not only in their ability to reproduce the temporal and spatial variability of precipitation, but differ also in their long-term estimation of the areal precipitation. This can be crucial for rainfall-runoff simulations. Only precipitation inputs, that reproduce the flood generating event and the long-term mean areal rainfall, lead to runoff simulations that approximate the observed runoff (Kneis and Heistermann, 2009). With respect to the rareness of the rainfall pattern of event 5, as noticed during model calibration, this could explain the low performance of the BGFgauge-method.

Due to the highly non-linear relationship between rainfall and runoff, the analysis of the hydrographs should not be used as the only criterion for the assessment of the quality of different regionalisation methods. However, it proved to be an important additional tool to demonstrate the relevance of the differences between the regionalisation methods for flood modelling. The rainfall-runoff model is currently integrated into the operational flood forecast system for the Mulde River.

The attempt to use radar data for the establishment of BGF$_{\text{radar}}$ rendered no improvement of the BGF regionalisation method. The main reason was shortage of past radar data. We had three years of calibrated radar data, which turned out to be too short to yield enough high-quality BGF$_{\text{radar}}$. In most cases no adequate BGF$_{\text{radar}}$ was found for regionalisation of the events and ordinary kriging was applied. We estimate that a pool of data of at least eight years is required to overcome this problem. It is also imaginable to use radar data that are not calibrated with gauge data.

Why did IDW sometimes outperform the statistical methods? Superiority of geostatistical methods was found for monthly precipitation totals in low-density gauge networks, e.g. by Goovaerts (2000) and Creutin and Obled (1982). The major advantage over simpler methods is that the sparsely sampled observations can be complemented by secondary variables that exist in a higher density. The benefit of using kriging methods, that use e.g. elevation or BGF, decreases as correlation to these variables and the spatial dependence between observations weakens (Goovaerts, 2000). Both facts

applied in our cases. Correlations between precipitation and elevation as well as spatial dependence for convective rainfalls were weak. Universal kriging was applied as the most appropriate model of the *Default*-method in 16% of all considered hours in this work, because spatial correlation was higher than 0.5. In only 4% a correlation to altitude greater than 0.5 existed, and a kriging with external drift was applied. Exemplarily, we calculated the correlation coefficients between precipitation and elevation for 71 h of our investigated events. The mean was 0.13 and the standard deviation 0.23. Correlation is low due to the high temporal resolution considered in this investigation. Even negative correlations were found for hourly rainfalls in the Italian Alps by Allamano et al. (2009). In contrast, Bernhofer et al. (2008) showed clearly the dependency of precipitation from altitude for climatological periods in Saxony. In our study it turned out that IDW is often equivalent to geostatistical methods for short time rainfall events of high spatial and temporal variability. Also Dorninger et al. (2008) and Ruelland et al. (2008) found that more sophisticated analysis tools just led to slightly better results than simpler methods in spite of the higher effort.

Although the uncertainty of the regionalised precipitation fields is not a focus of this work and quantitative statements are not possible for all steps during regionalisation, we want to give some information regarding the following points:

(1) Measurement errors of gauges occur due to the following reasons: deformation of the wind field, loss from wetting, evaporation losses, splash-out and splash in of precipitation, read out errors and errors in the digital transmission path of the data. Error corrections require the knowledge of climatological conditions. Especially local wind measurements are important, but rarely available. The mean error of monthly precipitation data in Saxony is around 12% (Richter, 1995). However, highly resolved data can have errors that are much higher. We did not correct the observation for systematic measurement errors.

(2) The representativeness of a gauge for an event can be determined, if data of a dense gauge network are available. The normalised spatial sampling error NSSE can be determined with Eq. (10), being P_{aver} the areal precipitation of the dense network (simplest case: arithmetic mean), P_{o} the observed precipitation at a single gauge and N the number of gauges in the domain.

$$\text{NSSE} = 100 \cdot \frac{\sqrt{\frac{\sum_{i=1}^{N} \left(P_{\text{aver}} - P_{\text{o}_i} \right)^2}{N}}}{P_{\text{aver}}} \, (\%) \tag{10}$$

The only network that is suitable to calculate NSSE in the investigation area is located in the city of Dresden (16 gauges, 330 km^2). For the convective event on 16 June 2006, 17:00 UTC (we had to calculate it 3 h later than event 2, because the cell arrived later in Dresden) and for the advective event on 27 May 2006, 17:00 UTC the NSSE is around 80%.

Such high values stand for a high micro-scale variation and were found for other intense rainfalls in the investigation area too. Points that constrain the quality of these calculations are that gauges were not set up according to standards and rainfall variability in cities is often higher than outside.

(3) The 10-min time difference between radar and hourly gauge values can affect the quality of the BGF$_{radar}$-method. Statistically, the error is around 17%, but much higher errors can occur.

(4) Another error source is the regionalisation of point data. For kriging methods the krige variance is an often used criterion to assess estimation uncertainty, but there is no equivalent for IDW. Because of the overall low gauge density, there were no independent control datasets available to check for regionalisation errors. Therefore, we used cross-validation results and calculated a normalised root mean square error NRMSE (equation similar to Eq. 10). Errors were computed from the observed and simulated data and were normalized to the mean of the observed values. For the above mentioned convective event NRMSE is 246, 261, 266% and for the advective event 118, 115, 112% for the methods IDW, Default, BGFgauge, respectivly. NRMSE accounts for all uncertainties during regionalisation process, as e.g. the limited amount of data and the usage of daily data for set up of BGF, uncertainties during automatic variogram determination, and the chosen exponent 2 for IDW. A possibility to quantify the uncertainty to determine areal precipitation fields presented Bliefernicht et al. (2008) applying stochastic simulation. He showed that uncertainty can be remarkable and depends on station density and rainfall amount.

The sum of all errors can lead to enormous differences between regionalised and true areal rainfall, especially when focusing on small catchments. To overcome the often most striking error sources – the micro-scale variations and regionalisation – dense gauge networks have to be built up or past detailed rainfall information or radar data have to be considered during regionalisation.

Our findings:

– The low density of temporally highly resolved rain gauges is a serious problem in the context of regionalisation. In combination with the high spatial variability of hourly rain events, high uncertainties can result in the areal rain field. This can be compensated to some extent applying secondary information. The use of information of past precipitation fields resulted in the most realistic areal rainfall estimates in comparison to IDW or other kriging variants in most cases.

– There is no method that clearly outperformed the other methods under all conditions and no special aptitude for specific precipitation types. Consequently, the applied methods BGF$_{gauge}$, *Default* and IDW belong to the methods that are capable to regionalise events of high temporal and spatial variability.

– Rainfall-runoff modelling turned out to be a valuable tool that can help to assess the quality of regionalisation methods from a more integrative perspective and for longer terms.

– The implementation of past radar data into the BGF-method seems to be promising, but only with a dataset of more than eight years.

– We recommend providing a pool of regionalisation methods for real time applications, e.g. flood forecast. The decision, which method is most appropriate, can be made automatically by objective criteria of cross-validation. Though, the decision should be checked by inspection. Rain gauge data, which were not used for radar calibration, are of particular importance in this context.

The present study concerns only a particular case, since it applies only to our investigation area and the methods used. Especially the software package *InterMet* is due to its complexity not comparable to, e.g. an application of a kriging outside of *InterMet*. Any generalisation would require the use of much more rainfall events as well as more catchments with different orographic and climatic characteristics.

Further investigation is planned to verify, whether these conclusions hold for further rainfall events other than the six analysed in this paper. A comprehensive comparison of the results of cross-validation and rainfall-runoff modelling was difficult in this paper, because techniques were applied for the whole domain and catchments, respectively. We think both methods could complement each other more efficiently, if they would be applied for the same particular area.

Edited by: M. Disse
Reviewed by: S. Bàrdossy and two other anonymous referees

References

Ahrens, B.: Distance in spatial interpolation of daily rain gauge data, Hydrol. Earth Syst. Sci., 10, 197–208, 2006

Allamano, P., Claps, P., Laio, F., and Thea, C.: A data-based assessment of the dependence of short-duration precipitation on elevation, Phys. Chem. Earth, 34(10–12), 635–641, 2009.

Alpuim, T. and Barbosa, S.: The Kalman filter in the estimation of area precipitation, Environmetrics, 10(4), 377–394, 1999.

Atkinson, P. M. and Lloyd, C. D.: Mapping Precipitation in Switzerland with Ordinary and Indicator Kriging, Journal of Geographic Information and Decision Analysis (GIDA), 2(2), 65–76, 1998.

Bàrdossy, A., and Das, T.: Influence of rainfall observation network on model calibration and application, Hydrol. Earth Syst. Sci., 12, 77–89, 2008, http://www.hydrol-earth-syst-sci.net/12/77/2008/.

Barlo, R., Bartholomew, D., Brenner, J., and Brunk, H.: Statistical Inference under Order Restrictions, John Wiley and Sons, New York, USA, 1972.

Bartels, H.: Projekt RADOLAN. Routineverfahren zur Online-Aneichung der Radarniederschlagsdaten mit Hilfe von automatischen Bodenniederschlagsstationen (Ombrometer), Projekt-Abschlussbericht, available at: http://www.dwd.de, (search for: Radolan) (last access: 10 February 2010), 2004.

Becker, A., Klöcking, B., Lahmer, W., and Pfützner, B.: The hydrological modelling system ARC/EGMO, in: Mathematical models of large watershed hydrology, edited by: Singh, V. P. and Frevert, D. K., Water Resour. Publ., Littleton, Colorado, USA, 2002.

Berne, A., Delrieu, G., Creutin, J. D., and Obled, C.: Temporal and spatial resolution of rainfall measurements required for urban hydrology, J. Hydrol., 299, 166–179, 2004.

Bernhofer, C., Goldberg, V., Franke, J., Häntzschel, J., Harmansa, S., Pluntke, T., Geidel, K., Surke, M., Prasse, H., Freydank, E., Hänsel, S., Mellentin, U., and Küchler, W.: Sachsen im Klimawandel – Eine Analyse. Sächsisches Staatsministerium für Umwelt und Landwirtschaft, Eigenverlag, Dresden, Germany, 2008 (in German).

Bissolli, P. and Dittmann, E.: Objektive Wetterlagenklassen, Klimastatusbericht des DWD 2002, Germany, 2003.

Bliefernicht, J., Bàrdossy, A., and Ebert, C.: Stochastic simulation of hourly precipitation fields for extreme events on the rivers Freiberger Mulde, Oberer Main, and Fränkische Saale, Hydrol. Wasserbewirts., 52(4), 168–172, 2008.

Brandes, E. A.: Optimizing rainfall estimates with the aid of RADAR, J. Appl. Meteorol., 14, 1339–1345, 1975.

Casper, M. C., Herbst, M., Grundmann, J., Buchholz, O., and Bliefernicht, J.: Influence of rainfall variability on the simulation of extreme runoff in small catchments, Hydrol. Wasserbewirts., 53(3), 134–139, 2009.

Chilès, J. P. and Delfiner, P.: Geostatistics: modelling spatial uncertainty, John Wiley and Sons, New York, USA, 1999.

Cressie, N.: Statistics for spatial data, John Wiley and Sons, New York, USA, 1991.

Creutin, J. D. and Obled, C.: Objective analyses and mapping techniques for rainfall fields: an objective comparison, Water Resour. Res., 18(2), 413–431, 1982.

Dietrich, J., Trepte, S., Wang, Y., Schumann, A. H., Voß, F., Hesser, F. B., and Denhard, M.: Combination of different types of ensembles for the adaptive simulation of probabilistic flood forecasts: hindcasts for the Mulde 2002 extreme event, Nonlin. Processes Geophys., 15, 275–286, 2008, http://www.nonlin-processes-geophys.net/15/275/2008/.

Dirks, K. N., Hay, J. E., Stow, C. D., and Harris, D.: High-resolution studies of rainfall on Norfolk Island, Part II: Interpolation of rainfall data, J. Hydrol., 208(3–4), 187–193, 1998.

Dorninger, M., Schneider, S., and Steinacker, R.: On the interpolation of precipitation data over complex terrain, Meteorol. Atmos. Phys., 101, 175–189, 2008.

Ehret, U.: Rainfall and Flood Nowcasting in Small Catchments using Weather Radar, Mitteilungen Instit. Wasserbau, Universität Stuttgart, Germany, 2003.

Eischeid, J. K., Pasteris, P. A., Diaz, H. F., Plantico, M. S., and Lott, N. J.: Creating a serially complete, national daily time series of temperature and precipitation for the Western United States, J. Appl. Meteorol., 39, 1580–1591, 2000.

Flemming, G.: Angewandte Klimatologie von Sachsen – Basis und Zustandsklima von Sachsen. Tharandter Klimaprotokolle, 4, Technische Universität Dresden, Germany, 2001 (in German).

Genton, M. G.: Highly Robust Variogram Estimation, Math. Geol., 30(2), 213–221, 1998.

Goovaerts, P.: Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall, J. Hydrol., 228, 113–129, 2000.

Grijsen, J. G., Snoeker, X. C., Vermeulen, C. J. M., Mohamed El Amin Moh. Nur, and Yasir, A. M.: An Information System for Flood Early Warning, in: Floods and Flood Management, edited by: Saul, A. J., Kluwer Academic Publishers, 263–289, 1992.

Haberlandt, U.: Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event, J. Hydrol., 332(1–2), 144–157, 2007.

Hess, P. and Brezowsky, H.: Katalog der Großwetterlagen Europas 1881–1976. Berichte des Deutschen Wetterdienstes, 113, Eigenverlag Deutscher Wetterdienst, Offenbach a. M., Germany, 1977 (in German).

Hinterding, A.: Entwicklung hybrider Interpolationsverfahren für den automatisierten Betrieb am Beispiel meteorologischer Größen, Inst. f. Geoinformatik, Westfälische Wilhelms-Universität Münster, IfGIprints, 19, Münster, Germany, 2003 (in German).

Jensen, N. E. and Pedersen, L.: Spatial variability of rainfall: Variations within a single radar pixel, Atmos. Res., 77, 269–277, 2005.

Jatho, N., Pluntke, T., Kurbjuhn, C., and Bernhofer, C.: An approach to combine radar and gauge based rainfall data under consideration of their qualities in low mountain ranges of Saxony, Nat. Hazards Earth Syst. Sci., in review, 2010.

Kneis, D. and Heistermann, M.: Bewertung der Güte einer radarbasierten Niederschlagsschätzung am Beispiel eines kleinen Einzugsgebiets, Hydrol. Wasserbewirts., 53(3), 160–171, 2009 (in German).

Kuczera, G., Kavetski, D., Franks, S., and Thyer, M.: Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters, J. Hydrol., 331, 161–177, 2006.

Matheron, G.: Traite de geostatistique appliquee, Tome I: Memoires du Bureau de Recherches Geologiques et Minieres, 14, Editions Technip, Paris, 1962 (in French).

Merz, R., Blöschl, G., and Parajka, J.: Raum-zeitliche Variabilität von Ereignisabflussbeiwerten in Österreich, Hydrol. Wasserbewirts., 50(1), 2–11, 2006 (in German).

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models, part I – A discussion of principles, J. Hydrol., 10(3), 282–290, 1970.

Paulat, M., Frei, C., Hagen, M., and Wernli, H.: A gridded dataset of hourly precipitation in Germany: Its construction, climatology

www.nat-hazards-earth-syst-sci.net/10/353/2010/

Nat. Hazards Earth Syst. Sci., 10, 353–370, 2010

and application, Meteorol. Z., 17(6), 719–732, 2008.

Pessoa, M. L., Rafael, L. B., and Earle, R. W.: Use of weather radar for flood forecasting in the Sieve river basin: a sensitivity analysis, J. Appl. Meteorol., 32, 462–475, 1993.

Richter, D.: Ergebnisse methodischer Untersuchungen zur Korrektur des systematischen Messfehlers des Hellmann-Niederschlagsmessers, Berichte des Deutschen Wetterdienstes 194, Eigenverlag Deutscher Wetterdienst, Offenbach a. M., Germany, 1995 (in German).

Ruelland, D., Ardoin-Bardin, S., Billen, G., and Servat, E.: Sensitivity of a lumped and semi-distributed hydrological model to several methods of rainfall interpolation on a large basin in West Africa, J. Hydrol., 361, 96–117, 2008.

Seo, D. J., Krajewski, W. F., and Bowles, D. S: Stochastic interpolation of rainfall data from raingauges and radar using co-kriging: 1. Design of experiments, Water Resour. Res., 26(3), 469–477, 1990a.

Seo, D. J., Krajewski, W. F., and Bowles, D. S: Stochastic interpolation of rainfall data from raingauges and radar using co-kriging: 2. Results, Water Resour. Res., 26(5), 915–924, 1990b.

Sun, X., Mein, R. G., Keenan, T. D., and Elliott, J. F.: Flood estimation using radar and raingauge data, J. Hydrol., 239(1–4), 4–48, 2000.

Tabios, G. Q. and Salas, J. D.: A comparative-analysis of techniques for spatial interpolation of precipitation, Water Resour. Bull., 21(3), 365–380, 1985.

Todini, E.: A Bayesian technique for conditioning radar precipitation estimates to rain-gauge measurements, Hydrol. Earth Syst. Sci., 5, 187–199, 2001,
http://www.hydrol-earth-syst-sci.net/5/187/2001/.

Wackernagel, H.: Multivariate geostatistics: an introduction with applications, Springer, Heidelberg, Germany, 1995.