**Natural Hazards
and Earth
System Sciences**

Probabilistic forecast of daily areal precipitation focusing on extreme events

J. Bliefernicht and A. Bárdossy

Institute for Hydraulic Engineering, University of Stuttgart, Germany

Abstract. A dynamical downscaling scheme is usually used to provide a short range flood forecasting system with high-resolved precipitation fields. Unfortunately, a single forecast of this scheme has a high uncertainty concerning intensity and location especially during extreme events. Alternatively, statistical downscaling techniques like the analogue method can be used which can supply a probabilistic forecasts. However, the performance of the analogue method is affected by the similarity criterion, which is used to identify similar weather situations. To investigate this issue in this work, three different similarity measures are tested: the euclidean distance (1), the Pearson correlation (2) and a combination of both measures (3). The predictor variables are geopotential height at 1000 and 700 hPa-level and specific humidity fluxes at 700 hPa-level derived from the NCEP/NCAR-reanalysis project. The study is performed for three mesoscale catchments located in the Rhine basin in Germany. It is validated by a jackknife method for a period of 44 years (1958–2001). The ranked probability skill score, the Brier Skill score, the Heidke skill score and the confidence interval of the Cramer association coefficient are calculated to evaluate the system for extreme events. The results show that the combined similarity measure yields the best results in predicting extreme events. However, the confidence interval of the Cramer coefficient indicates that this improvement is only significant compared to the Pearson correlation but not for the euclidean distance. Furthermore, the performance of the presented forecasting system is very low during the summer and new predictors have to be tested to overcome this problem.

1 Introduction

Short range flood forecasting systems are developed to provide early flood warnings in mesoscale catchments. One requirement of a short range flood forecasting system are precipitation fields with a high resolution in time and space. Precipitation fields provided by a global circulation model cannot be used, because the resolution of these fields is too coarse and the variability of the precipitation process cannot be described sufficiently. To overcome this problem, a downscaling of precipitation is done. Usually, a high-resolved circulation model (regional model) is nested within a global circulation model. This downscaling technique is often referred to dynamical downscaling (Xu, 1999). However, the investigations of Jasper et al. (2002) have shown large deficiencies, if a single forecast of a dynamical downscaling is used as input for a short range flood forecasting system. In their study five different regional models are combined with a hydrological model to forecast several severe flood events in the Lagio Maggiore basin. Finally, they conclude that a single precipitation forecast has to be as accurate as possible, if the flood forecasting should be succesful.

Instead of a single forecast, also an ensemble based forecast can be used to capture the high variability of the precipitation process during extreme events. For example, COSMO-LEPS presents an ensemble forecasting system based on a dynamical downscaling approach (Marsigli et al., 2004, 2005). It uses the ensembles of the global circulation model of ECMWF (European Center For Medium-Range Weather Forecasts) as boundary conditions to run a regional model provided by the german weather service DWD. The ensembles of COSMO-LEPS can be used to provide a probabilistic forecasts. It has been shown, that a forecast based on probabilistic has more potential to reduce economic losses caused by extremes than a single forecast or the mean of ensembles (Richardson, 2003).

Correspondence to: J. Bliefernicht
(jan.bliefernicht@iws.uni-stuttgart.de)

Beside a dynamical downscaling also statistical downscaling approaches can be used to provide high-resolved precipitation fields based on probabilities. They have the advantage that they are easy to implement and not so time-consuming. But the performance of a statistical downscaling depends strongly on the available data.

Several approaches can be selected as statistical downscaling techniques. Bárdossy (2000) distinguishes three different kinds of approaches: regression approaches, resampling methods and conditional probability approaches. All approaches have in common, that a functional relationship is assumed between large-scale variables (predictors) and a local-scale variable (predictand).

A frequently used resampling method is the analogue method, which has been successfully applied in operational flood forecasting (Obled et al., 2002; Bontron and Obled, 2003). This method compares the weather situation of the current day with weather situations of the past. To find similar weather situations, similarity measures like the euclidean distance are used. However, it is not clear, how the performance of a forecasting system of heavy rain events is affected by the choice of similarity criterion. To investigate this question in this work, three different similarity measures are chosen: the euclidean distance, the Pearson correlation and a combination of both criteria. They are applied for three mesoscale catchments located in the Rhine basin using daily areal precipitation from 1958 to 2001.

2 Methods

2.1 Analogue method

The analogue method was first proposed by Lorenz (1969). If the analogue method is applied as a downscaling approach for precipitation, it can be divided into two steps (Obled et al., 2002):

1. The weather situation of the current day is compared with weather situations of the past.

2. If an analogue situation is identified, the precipitation which occurred at that day is chosen as forecast.

To find analogue situations several similarity measures can be chosen. A good overview about several similarity measures is given by Drosdowsky and Zhang (2003). For instance, the S1-Score (also known as Teweles-Wobus-Score) can be selected as similarity measure (Obled et al., 2002; Bontron and Obled, 2003).

If the euclidean distance e_1 is applied as similarity measure to compare two fields of a predictor $\mathbf{x}$ at time step t_1 and t_2, it can be defined as follows:

$$e_1(\mathbf{x}(t_1), \mathbf{x}(t_2)) = \sqrt{\sum_{i=1}^{k} (x_i(t_1) - x_i(t_2))^2} \qquad (1)$$

where x_i presents the absolute value of one predictor element at location i. k is the number of elements. After Eq. (1) the euclidean distance is the root of the quadratic sum of the residuals between both fields. However, the euclidean distance is limited, because it measures only the closeness between two spatial fields but sometimes two fields are similar according to their pattern. This property can be measured by the Pearson correlation r:

$$r(\mathbf{x}(t_1), \mathbf{x}(t_2)) = \frac{\frac{1}{n} \sum_{i=1}^{k} (x_i(t_1) - \bar{x}(t_1)) (x_i(t_2) - \bar{x}(t_2))}{s(t_1) s(t_2)} \qquad (2)$$

where $\bar{x}(t_1)$ and $\bar{x}(t_2)$ represents the arithmetic mean and $s(t_1)$ and $s(t_2)$ the standard deviation of the predictor at corresponding time steps.

The combination of the euclidean distance and the correlation is not straightforward. Therefore a distance related to the correlation is first introduced. At first, the absolute values of a predictor are standardized for a given day t:

$$\mathbf{x}'(t) = \left(\frac{x(t) - \bar{x}(t)}{s(t)} \right) \qquad (3)$$

The euclidean distance e_2 of two predictors in the normed space is

$$e_2(\mathbf{x}'(t_1), \mathbf{x}'(t_2)) = \sqrt{\sum_{i=1}^{k} \left(x_i'(t_1) - x_i'(t_2) \right)^2} = \qquad (4)$$

which can be simplified to

$$= \sqrt{\sum_{i=1}^{k} \left(x_i'(t_1)^2 - 2x_i'(t_1)x_i'(t_2) + x_i'(t_2)^2 \right)} =$$

$$= \sqrt{2K - 2\sum_{i=1}^{k} \left(x_i'(t_1)x_i'(t_2) \right)} \qquad (5)$$

Thus the following relationship exists between the euclidean distance and the Pearson correlation in the transformed space:

$$e_2(\mathbf{x}'(t_1), \mathbf{x}'(t_2)) = \sqrt{2K \left(1 - r(\mathbf{x}'(t_1)\mathbf{x}'(t_2)) \right)} = \qquad (6)$$

The two distances e_1 and e_2 can be be combined to a new distance e_n:

$$e_n(\mathbf{x}(t_1), \mathbf{x}(t_2)) = e_1(\mathbf{x}(t_1), \mathbf{x}(t_2)) + \alpha e_2(\mathbf{x}'(t_1), \mathbf{x}'(t_2)) \qquad (7)$$

where $\alpha > 0$ balance between closeness and pattern similarity.

2.2 Forecast verification

The performance of a probabilistic forecast system cannot be expressed by common deterministic verification measures like the Pearson correlation or the root mean square error. To

Table 1. Possible outcomes of a binary forecasting system. yes = event is forecast (observed), no = no event is forecast (observed).

Forecast	Observation Yes	No	Σ
Yes	Hit a	False alarm b	$a+b$
No	Miss c	Correct rejection d	$c+d$
Σ	$a+c$	$b+d$	n

Table 2. Catchment properties of Pruen, Sieg and Nahe. Arithmetic mean P_m, 95%-quantile P_{95} and the 99.5%-quantile $P_{99.5}$ of daily areal precipitation are shown.

Property	Pruen	Sieg	Nahe
Size [km^2]	600	750	2500
P_m [mm/d]	2.74	3.01	2.12
P_{95} [mm/d]	12.7	14.2	11.0
$P_{99.5}$ [mm/d]	29.4	30.7	24.5

verify a probabilistic forecasts, scores are needed which can describe the cumulative distribution function of the predictand.

Murphy (1973) proposed that the performance of a probabilistic forecast depends on two attributes: reliability and resolution. He defines that a forecast system is reliable if it provides unbiased estimates of the predictand with different forecast probabilities. If it is able to separate between situations under which an event occurs in the future or not, then it has resolution. To determine both attributes, the Brier score is a frequently used measure (Brier, 1950):

$$\text{BS} = \frac{1}{n} \sum_{t=1}^{n} (p_t - o_t)^2 \tag{8}$$

where p_t presents the probability of the forecast at time t and o_t the observed frequency at the same time. n is the number of forecasts. If an event occurs than the observed frequency $o_t=1$, otherwise $o_t=0$. The Brier score depends on the selection of the threshold which is used to separate whether an event occurs or not. For instance, daily precipitation thresholds like 0 mm or 20 mm can be chosen. Then, the probabilistic forecast is evaluated if it is either able to separate among dry and wet days or to forecast extreme events.

The Brier score can be decomposed into three terms (Murphy, 1973):

$$\text{BS} = \frac{1}{n} \sum_{i=1}^{k} N_i (p_i - \bar{o}_i)^2 -$$
$$\frac{1}{n} \sum_{i=1}^{k} N_i (\bar{o}_i - \bar{o})^2 + \bar{o}(1 - \bar{o}) \tag{9}$$

After Eq. (9) forecasts are divided according to the forecast probabilities p_i into k classes and for each class the arithmetic mean of the observed frequencies $\bar{o}_i$ is calculated. N_i presents the number of forecasts in class i and $\bar{o}$ is the arithmetic mean of the observations expressed by binary elements. The first term of Eq. (9) describes the reliability and the second term the resolution of a probabilistic forecast. The third term determines the uncertainty, which is not influenced by the forecast. It depends only on the variability of the observations. For a perfect reliable system, the reliability term is zero. The larger the second term, the larger the resolution of a forecast system is.

Beside the Brier score also the ranked probability score RPS can be chosen to verify a probabilistic forecast (Epstein, 1969):

$$\text{RPS} = \frac{1}{k}\frac{1}{n} \sum_{i=1}^{k} \sum_{t=1}^{n} (p_t - o_t)^2 = \frac{1}{k} \sum_{i=k}^{k} \text{BS}_k \tag{10}$$

Therefore, the ranked probability score represents the mean of Brier scores for k selected precipitation thresholds. However, both RPS and BS, cannot be used to decide whether the forecast system has more skill compared to a reference forecast or not. For instance, the arithmetic mean of the observation (climatological mean) represents a reference forecast, which is a simple way to forecast the precipitation amount of future days. To solve this problem the corresponding skill scores, the Brier skill score BSS (Wilks, 1995):

$$\text{BSS} = 1 - \frac{\text{BS}}{\text{BS}_0} \tag{11}$$

and the ranked probability skill score RPSS are selected:

$$\text{RPSS} = 1 - \frac{\text{RPS}}{\text{RPS}_0} \tag{12}$$

BS$_0$ and RPS$_0$ represent the scores of a reference forecast. Both skill scores range between minus infinity and 1. A skill score lower than zero indicates, that the forecast system is not better than the chosen reference forecast. The forecast is perfect, if the skill score is 1. The RPSS has the disadvantage that the skill is sensitive to the number and choice of thresholds. But it has also the nice property that it focuses more on extreme events, if corresponding thresholds are selected.

It is also possible to create relative measures of the reliability and the resolution term of Eq. (9), if each term is divided by the uncertainty. To get negatively oriented measures the relative measure of the resolution term is subtracted from one (Toth et al., 2003):

$$B_{\text{rel}} = \frac{\frac{1}{n} \sum_{i=1}^{k} N_i (p_i - \bar{o}_i)^2}{\bar{o}(1 - \bar{o})} \tag{13}$$

$$B_{\text{res}} = 1 - \frac{\frac{1}{n} \sum_{i=1}^{k} N_i (\bar{o}_i - \bar{o})^2}{\bar{o}(1 - \bar{o})} \tag{14}$$

Table 3. Downscaling performance of the analogue method using three different similarity measures (euclidean distance E, Pearson correlation P, combined criterion EP). In addition, the Pearson correlation r, the Heidke Skill Score for the summer and winter half year (HSS_s and HSS_w) and the contingency table of the binary forecast for the 80 largest rain events (fa = false alarm, ih = inverse hit) is given. Note that bold indicates the best results for each catchment.

Catchment	criterion	r	BSS	B_{rel}	B_{res}	RPSS	HSS	HSS_s	HSS_w	hit	fa	miss	ih
	E	0.67	0.22	**0.010**	0.77	0.23	**0.34**	0.24	**0.43**	**27**	52	53	15937
Pruen	P	0.65	0.22	**0.010**	0.77	0.17	0.18	0.12	0.24	24	157	56	15832
	EP	**0.68**	**0.23**	0.011	**0.76**	**0.25**	**0.34**	**0.25**	**0.43**	**27**	**50**	**47**	**15949**
	E	0.67	0.21	0.012	**0.76**	0.23	0.24	**0.18**	0.30	26	83	54	15906
Sieg	P	0.65	0.17	0.015	0.82	0.18	0.17	0.09	0.25	20	108	60	15881
	EP	**0.68**	**0.22**	**0.006**	**0.76**	**0.24**	**0.25**	0.16	**0.33**	**28**	**82**	**50**	**15906**
	E	**0.66**	0.20	0.014	0.79	**0.23**	0.24	**0.15**	0.34	**17**	42	**63**	15947
Nahe	P	0.64	0.14	0.015	0.85	0.16	0.13	0.08	0.19	12	92	68	15897
	EP	**0.66**	**0.21**	**0.013**	**0.78**	**0.23**	**0.26**	0.14	**0.38**	**17**	**36**	**63**	**15953**

B_{rel} is the relative measure of reliability and B_{res} of resolution. Both measures ranges between 0 and 1. If both scores are zero, a probabilistic forecast is perfect. They are related to the Brier skill score in the following way (Toth et al., 2003):

$$\text{BSS} = 1 - B_{rel} - B_{res} \tag{15}$$

However, a forecast system for extreme events cannot only be verified by using RPSS and BSS. If an extreme event is forecast, someone has to decide whether an alarm is given or not. Such a decision causes always two right and false decisions (Table 1). Then, a probabilistic forecast is reduced to a simple binary (yes = 1, no = 0) forecast and verification measures are needed which are able to describe a binary forecast. For example, the Heidke skill score HSS represents a widely used measure (Murphy and Daan, 1985):

$$\text{HSS} = \frac{PC - PC_0}{1 - PC_0} \tag{16}$$

PC is the proportion correct, which is the number of correct forecasts (hit and correct rejection) divided by the number of forecasts. PC_0 is the proportion correct of a reference forecast, which here is the random forecast. The Heidke skill score ranges between -1 and 1, where a skill score of 1 indicates a perfect forecast. A forecast system is not better than a reference forecast, if the skill score is 0 or less.

As second binary verification measure the Cramer coefficient V is chosen. It is a χ^2-based association measure, which is defined for a 2-by-2 contingency table as (Hartung, 1999):

$$V = \sqrt{\frac{\chi^2}{n}} \tag{17}$$

with χ^2 as:

$$\chi^2 = \frac{n\,(a \cdot b - c \cdot d)^2}{(a+b)\,(c+d)\,(a+c)\,(b+d)} \tag{18}$$

The Cramer coefficient is not a skill score. It ranges from 0 to 1, where a value of one indicates a perfect forecast. The coefficient decreases, if the number of correct forecasts also decreases. If the number of correct forecast is equal to the number of false forecast, the score is zero. Note that a number of false forecast higher than a number of correct forecast would also yield a score larger than zero.

Compared to the other scores the Cramer coefficient has the advantage that a confidence interval can be derived in a simple way. To determine the confidence intervals the standard deviation of the coefficient σ_V is calculated at first. A detailed description of this estimation is given by Hartung (1999). Due to the large sample size, it is assumed that the distribution of the Cramer association coefficient can be approximated by a normal distribution. Thus, the lower boundary V_u and the upper boundary V_o of a $1 - \alpha$ confidence interval is estimated by

$$V_u = V - \sigma_V \cdot u_{1-\alpha/2} \tag{19}$$

$$V_o = V + \sigma_V \cdot u_{1-\alpha/2} \tag{20}$$

The quantity $u_{1-\alpha/2}$ presents the corresponding $1 - \frac{\alpha}{2}$ quantile of the standard normal distribution.

2.3 Application

Three mesoscale catchments located in the Rhine basin (Pruen, Sieg and Nahe) are selected. External drift kriging is used to calculate daily areal precipitation from 1958 to 2001 in a spatial resolution of 5 km×5 km. The daily areal precipitation is selected for the validation of the forecast system for the same period. It is defined as the sum of the precipitation amount of each grid point within a catchment divided by the number of these grid points. A short overview about the catchment characteristics is given by Table 2.

Three predictors are chosen to describe a weather situation at a certain time step: geopotential height at 700
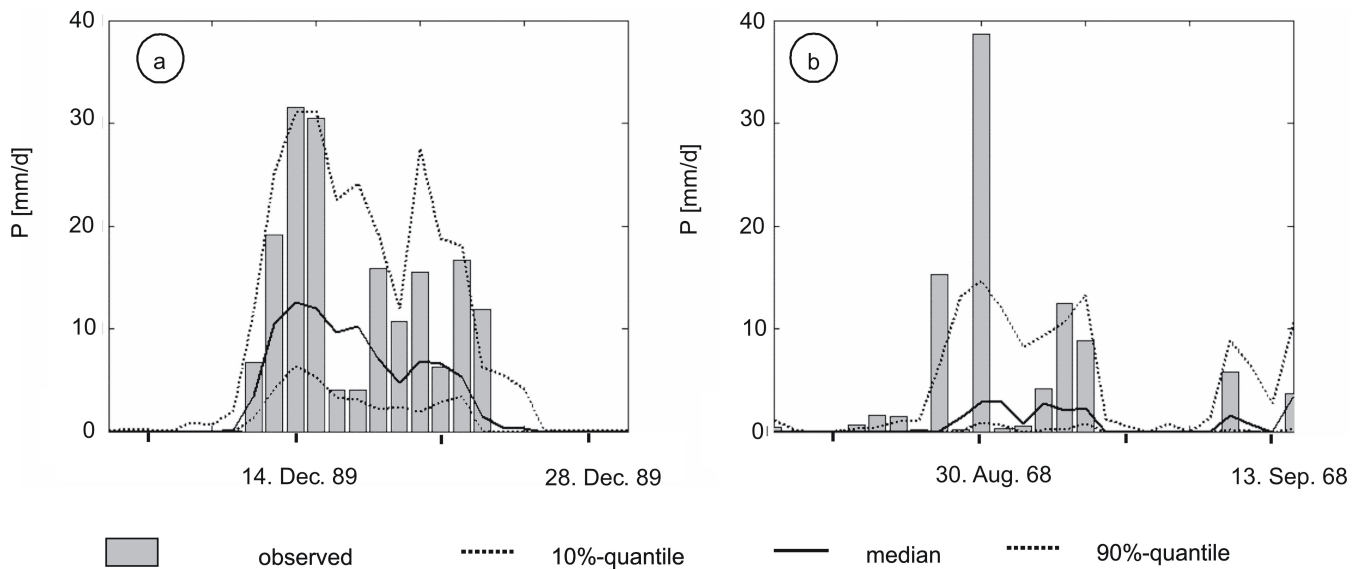
Fig. 1. Probabilistic forecast of an advective **(a)** and a convective rain event **(b)** occurred in the Pruen catchment.

and 1000 hPa-level (GPH 700 and GPH 1000) and westerly moisture flux at 700 hPa-level (MFLUX). If is defined that MFLUX is the product between the specific humidity (SHUM) at 700 hPa-level and westerly wind (UWND) at the same level. The selected predictors (GPH 700, GPH 1000, SHUM and UWND) are delivered by the NCEP/NCAR-Reanalysis project (Kalnay et al., 1996). The use of reanalysis data has the advantage that possible candidates for a downscaling can be tested over a long time period (1948 to present), if the predictand covers the same period. The spatial resolution of the reanalysis data is $2.5° \times 2.5°$. Furthermore, the predictors can be selected for each day at four different time steps (00:00 UTC, 06:00 UTC, 12:00 UTC and 18:00 UTC). Therefore, the optimal settings of the space and time domain of each predictor is determined. The optimal domain size of GPH 700 and GPH 1000 is $20° \times 15°$. The center of both domains is $5°$ E and $47.5°$ N. The space domain of MFLUX has also the same center but a different domain size ($20° \times 10°$). The optimal time step is 12:00 UTC for all predictors.

The jackknife method is chosen as technique to validate the forecast system. It presents the most extreme form of a cross-validation. It allows to incorporate nearly all data into the calibration and validation procedure. The predictor settings are optimised for the Pruen catchment and transferred with the same settings to the other catchments. 30 analogue situations are selected for each forecast day.

To verify the forecast, the 95%-quantile of the observed frequency is chosen as precipitation threshold to determine the Brier skill score, the resolution and the reliability. The ranked probability skill score is calculated by using six thresholds. Similar to Wetterhall (2005) the α-quantiles 0.90,

Table 4. Cramer association coefficient V and the 95% confidence intervals $[V_u; V_o]$ derived by using the values of the contingency table of Table 3. Note that bold indicates the best results for each catchment.

Catchment	criterion	V	V_u	V_o
	E	0.34	0.24	0.43
Pruen	P	0.19	0.13	0.26
	EP	**0.35**	0.26	0.45
	E	0.27	0.19	0.36
Sieg	P	0.19	0.12	0.27
	EP	**0.30**	0.21	0.38
	E	0.24	0.15	0.34
Nahe	P	0.13	0.06	0.19
	EP	**0.26**	0.16	0.36

0.925, 0.95, 0.975, 0.99 and 0.995 of the observed frequencies are chosen to focus more on extreme events. To calculate the Heidke skill score, the 99.5%-quantile of the observed frequencies and the 85%-quantile of the forecast probabilities is chosen. Therefore the forecast system is evaluated, whether it is able to forecast the 80 largest rain events of the past (1958 and 2001) or not.

3 Results and discussion

The comparison of the skill scores and the Cramer coefficient shows, that the highest performance is achieved, if the combined similarity measure is chosen (Tables 3 and 4). Then,

the presented forecast system is able to predict more extreme events. Also the resolution and and the reliability of the probabilistic forecast increases indicated by the decreasing relative measures. However, the improvement in predicting extreme events using the combined criteria is not very large compared to the euclidean distance. An example showing the robustness of the estimation of a verification measure is given in Table 4. The confidence interval of the Cramer coefficient indicates, that the improvement compared to the euclidean distance could be also caused by random.

The skill scores also show that the forecast system is more successful in predicting rare extreme events than a simple reference forecast. Both probabilistic and binary forecasts provide valuable information for a forecaster. How large the benefit is cannot be decided yet and economic decision model have to be introduced to determine the economic benefit. Furthermore, the presented forecast system has to be tested using forecast data instead of reanalysis data. However, Obled et al. (2002) have shown, that the use of operational forecast data produces only a minor decrease in the performance.

The results show also a significantly weakness of the forecast system. It fails in more than 60 percent of the cases, which cannot be significantly reduced by using the combined similarity measures. The cause of the high failure rate is that convective rainfall events like local thunderstorms cannot be simulated correctly (Fig. 1). Only rainfall events with a more advective character are described fairly well. Therefore, the selected predictors are not sufficient to describe the variability of the rainfall process. To overcome this problem, other meteorological variables have to be included in the downscaling method. For instance, Harpham and Wilby (2005) have shown that also vorticity is an important downscaling variable for heavy rain events. Furthermore, predictors of other reanalysis-projects, like the reanalysis project of ECMWF (ERA 40), can be used instead of the corresponding NCEP/NCAR-variable.

4   Conclusions

The analogue method can be used to provide high-resolved daily precipitation fields for a short-range flood forecasting system. The performance of the downscaling method is affected by the similarity criterion, which is used to identify similar weather situations. It has been shown, that a combination of euclidean distance and Pearson correlation yields the best performance in predicting heavy rain events. However, this improvement is only significant compared to the Pearson correlation but not for the euclidean distance. This combined measure can also not increase the downscaling performance of convective rainfall processes. To solve this problem, other predictor variables will be tested. Future research will also focus on economic value showing which end

user gets an economic benefit from the presented forecast system.

References

Bárdossy, A.: Stochastic Downscaling Methods to Assess the Hydrological Impacts of Climate Change on River Basin Hydrology. Climate Scenarios for Water-Related and Coastal Impacts. Proceedings of the EU Concerted Action Initiative ECLAT-2 Workshop 3. KNMI, Netherlands, May 10-12th 2000, Climatic Research Unit, Norwich, 18–34, 2000.

Bontron, G. and Obled, C.: New Developments in Quantitative Precipitation Forecasts by Analog Sorting Techniques. In: Mediterranean Storms. Proc. Of the 5th EGS Plinius Conference held at Ajaccio, Corsica, France, October 2003.

Brier, G. W.: Verification of Forecasts expressed in Terms of Probability, Mon. Wea. Rev., 78, 1–3, 1950.

Drosdowsky, W. and Zhang, H.: Verification of Spatial Fields, in: Forecast Verification – A Practioner's Guide in Atmospheric Science, edited by: Jolliffe, I. T. and Stephenson, D. B., John Wiley and Sons Ltd, 121–136, 2003.

Epstein, E. S.: A Scoring System for Probability Forecast of Ranked Categories, J. Appl. Meteorol., 8, 985–987, 1969.

Harpham C. and Wilby, R. L.: Multi-site Downscaling of Heavy Daily Precipitation Occurrence and Amounts, J. Hydrol., 312, 235–255, 2005.

Hartung, J.: Statistik. Lehr- und Handbuch der angewandten Statistik, R. Oldenbourg Verlag München, 12th edition, 1999.

Jasper, K., Gurtz, J., and Lang, H.: Advanced Flood Forecasting in Alpine Watersheds by Coupling Meteorological Observations and Forecasts with a Distributed Hydrological Model, J. Hydrol., 269(3–4), 253–254, 2002.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-Year Reanalysis Project, Bull. Amer. Meteor. Soc., 77, 437–471, 1996.

Lorenz, E. N.: Atmospheric Predictability as Revealed by Natural Occurring Analogues, J. Atmos. Sci., 26, 636–646, 1969.

Marsigli, C., Montani, A., Nerozzi, F., and Paccagnella, T.: Probabilistic High-Resolution Forecast of Heavy Precipitation over Central Europe, Nat. Hazards Earth Syst. Sci., 4, 315–322, 2004, http://www.nat-hazards-earth-syst-sci.net/4/315/2004/.

Marsigli, C., Boccanara, F., Montani, A., and Paccagnella, T.: The COSMO-LEPS Mesoscale Ensemble System: Validation of the Methodology and Verification, Nonlin. Processes Geophys., 12,

527–536, 2005,
http://www.nonlin-processes-geophys.net/12/527/2005/.

Murhpy, A. H.: A New Vector Partition of the Probability Score, J. Appl. Met., 12, 534–537, 1973.

Murphy, A. H. and Daan, H.: Forecast Evaluation, in: Probability, Statistics and Decision Making in the Atmospheric Sciences, edited by: Murphy, A. H. and Katz, R. W., Boulder, Colorado: Westview Press, 379–437, 1985.

Obled, C., Bontron, G., and Garçon, R.: Quantitative Precipitation Forecasts: A Statistical Adaption of Model Outputs through an Analogues Sorting Approach, Atmos. Res., 63, 303–324, 2002.

Richardson, D. S.: Economic Value and Skill, in: Forecast Verification – A Practioner's Guide in Atmospheric Science, edited by: Jolliffe, I. T. and Stephenson, D. B., John Wiley and Sons Ltd, 165–188, 2003.

Toth, Z., Talagrand, O., Candille, G., and Zhu, Y.: Probability and Ensemble Forecast, in: Forecast Verification – A Practioner's Guide in Atmospheric Science, edited by: Jolliffe, I. T. and Stephenson, D. B., John Wiley and Sons Ltd, 165–188, 2003.

Wetterhall, F.: Statistical Downscaling of Precipitation From Large-Scale Atmospheric Circulation – Comparison of Methods and Climate Regions. Dissertation, Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology, ISSN 1651-6214, 93, Uppsala University, 2005.

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences: An Introduction. San Diego, Academic Press., 1995.

Xu, C.-Y.: From GCMs to River Flow: A Review of Downscaling Methods and Hydrologic Modelling Approaches, Water Resour. Manage., 13, 369–382, 1999.