



Small-scale (flash) flood early warning in the light of operational requirements: opportunities and limits with regard to user demands, driving data, and hydrologic modeling techniques

Andy Philipp¹, Florian Kerl¹, Uwe Büttner¹, Christine Metzkes^{1,2}, Thomas Singer², Michael Wagner², and Niels Schütze²

¹Saxon State Office for Environment, Agriculture and Geology, Water, Soil, and Waste,
01109 Dresden, Germany

²Institute of Hydrology and Meteorology, Dresden University of Technology, 01069 Dresden, Germany

Correspondence to: Andy Philipp (andy.philipp@smul.sachsen.de)

Published: 12 May 2016

Abstract. In recent years, the Free State of Saxony (Eastern Germany) was repeatedly hit by both extensive riverine flooding, as well as flash flood events, emerging foremost from convective heavy rainfall. Especially after a couple of small-scale, yet disastrous events in 2010, preconditions, drivers, and methods for deriving flash flood related early warning products are investigated. This is to clarify the feasibility and the limits of envisaged early warning procedures for small catchments, hit by flashy heavy rain events. Early warning about potentially flash flood prone situations (i.e., with a suitable lead time with regard to required reaction-time needs of the stakeholders involved in flood risk management) needs to take into account not only hydrological, but also meteorological, as well as communication issues. Therefore, we propose a threefold methodology to identify potential benefits and limitations in a real-world warning/reaction context. First, the user demands (with respect to desired/required warning products, preparation times, etc.) are investigated. Second, focusing on small catchments of some hundred square kilometers, two quantitative precipitation forecasts are verified. Third, considering the user needs, as well as the input parameter uncertainty (i.e., foremost emerging from an uncertain QPF), a feasible, yet robust hydrological modeling approach is proposed on the basis of pilot studies, employing deterministic, data-driven, and simple scoring methods.

1 Introduction

For Saxony, considering the last two decades, the hydrologically most intense and most disastrous events occurred in August 2002, August/September 2010, as well as June 2013 (LfULG, 2015). Total damage for the aforementioned events sums up to 9 billion Euros (ca. 6.1 in 2002, 0.85 in 2010 and 2.0 in 2013). Especially in August/September 2010, flashy events caused large parts of total damages. In this light, the Saxon State Government mandated an independent commission to make suggestions for improving flood risk management actions (Jeschke et al., 2010). One of the commission's recommendations was to line out the potentials and limits of

small-scale flash flood early warning approaches (i.e., based on hydrological forecasts).

As the authority responsible for operational flood forecasting and warning, the Saxon Flood Center drafted a corresponding project with a preferably holistic view on flood risk management procedures, especially, when it comes to small-scale and flashy events. Therefore, a threefold approach is proposed, aiming at (1) the assessment of the demands and requirements of potential users of early warning products; (2) the verification of driving meteorological data for the targeted spatio-temporal scales; (3) checking the usefulness of a preferably broad range of modeling approaches with regard to model skill, robustness, and regional applicability, for

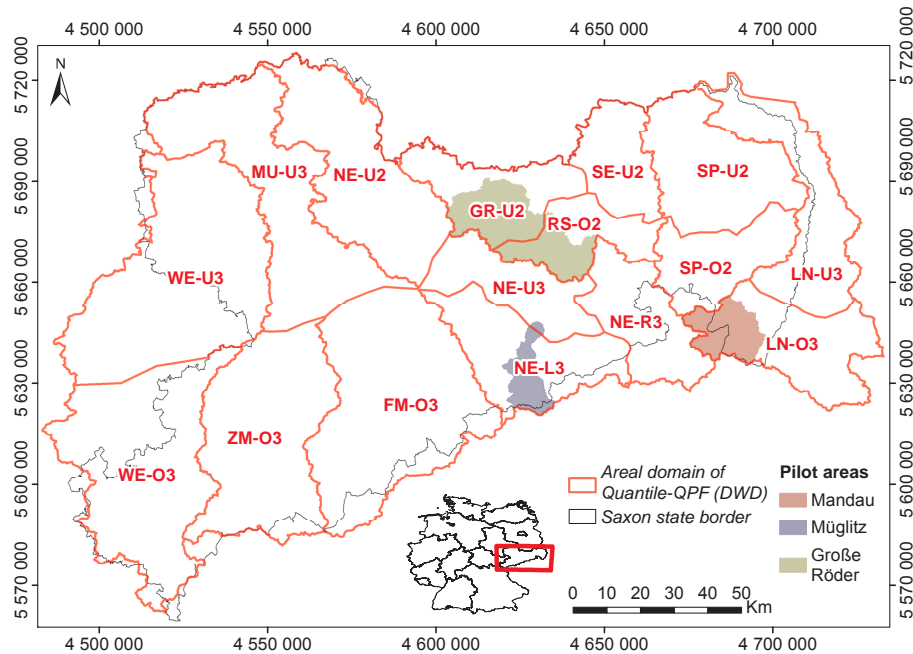


Figure 1. Overview map indicating the areal domain of the Quantile-QPF for Saxony (16 regions; e.g., “FM-O3” indicates the parts of the Freiburger Mulde catchment above 300 m a.s.l.). The area of the regions ranges between approximately 600 and 2700 km². Furthermore, the hydrological pilot areas (cf. Sect. 2.3) are shown. Gauss conformal projection with reference at 12° E (Zone 4). The Thumbnail map is showing the location of Saxony within Germany.

small, potentially ungauged basins. The paper at hand provides a short overview of the current state of work and illustrates a way towards an operational early warning system for small catchments in Saxony.

2 Methods

2.1 User survey

To investigate the needs and demands of potential users of an envisaged flood early warning system, a quantitative survey was carried out, based on an online questionnaire. The questionnaire comprised 15 questions, with 12 multiple-choice questions, two questions with gradually-scaled answers, and one question for the submission of verbal comments. Strictly speaking, the survey comprised quantitative and qualitative elements. For the sake of brevity, the full questionnaire is not presented herein but can be found in Philipp et al. (2015).

The surveyed sample was selected systematically (i.e., not randomly) and included all legal users (i.e., according to the Saxon Flood Alarm Bylaw; HWMO, 2014) of Flood Center products ($n = 578$) who were reachable via email to be invited for participating in the online survey ($n = 491$). The interviewee affiliation spanned administration/authorities at local/district/state level, fire departments and civil protection agencies, as well as the private sector. It has to be stated that the interviewees do not represent lay people since they partic-

ipate in the official flood management procedures on a legal and regular basis.

The survey results were evaluated using descriptive statistics and subgroup analyses by means of contingency tables. Therefore, given answers were investigated in a user-group specific manner, i.e., more than one variable is considered at a time (multivariate approach). A question to address was whether specific user groups answered differently or not. Such an effect can be induced by strongly differing sizes of sub-samples or can indicate a truly diverse response behavior. The literature suggests χ^2 -based dependency measures to clarify such questions (Sachs, 1999). For the present study, Cramér’s V and χ^2 -based p values were used.

2.2 Verification of QPFs

The verification of meteorological data comprised two Quantitative Precipitation Forecasts (QPFs) which are operationally used by the Saxon Flood Center: the deterministic numerical weather prediction COSMO-DE product (Baldauf et al., 2011) and the probabilistic “Quantile Forecast” (QF) for 16 specific areas in Saxony (cf. Fig. 1), issued by German Met Service’s Regional Center in Leipzig. The two QPFs are compared against a Quantitative Precipitation Estimate (QPE), emerging from rain gauge data, which was spatially interpolated (Ordinary Kriging) to derive areal precipitation estimates. Additionally, weather radar data (Met Service’s RADOLAN-RW product; Sacher et al., 2011) was employed

as another QPE reference. A comprehensive overview of the herein considered QPFs and QPEs is given in Table 1.

The Quantile Forecast represents a probabilistic, qualitative expert estimate of areal precipitation for the next 36 h and consists of three values/quantiles per forecasting time step. Since the forecast is issued for 16 specific areas in Saxony (i.e., river catchments with topographic partitioning according to elevation), verification was based on the comparison of areal rainfall for the mentioned 16 regions, and spanned a period from April 2011 to June 2014.

The comparison of areal rainfall was based on consecutive 6 h sums, starting from 06:00 and 18:00 UTC. 6 h sums were chosen to accommodate the coarsest temporal resolution of the investigated products, given by the Quantile-QPF. The product features areal rainfall totals (for the 16 forecasting regions) with 0.9, 0.5, and 0.1 exceedance probability for two consecutive 6 h and two further 12 h intervals. The forecast is updated twice a day (at 06:00 and 18:00 UTC). However, a main task of the herein presented verification was to evaluate the quality of this product against highly resolved numerical weather prediction output (i.e., COSMO-DE).

A QPF/QPE comparison typically employs a number of tools and methods (Jolliffe and Stephenson, 2012), ranging from simple diagnostic (e.g., time series and totals comparisons, residual and bias analyses, scatter and frequency plots) to integral, quantitative methods. Analyses are often based on threshold-oriented contingency table evaluation and deliver typical verification/skill scores, e.g., False Alarm Rate, Probability Of Detection (FAR, POD) or combined products, e.g., Receiver Operating Characteristic curves (ROC curves; Fawcett, 2006). A prototypical work flow of threshold-oriented skill assessment is shown in Fig. 2. More detailed information on the herein employed QPF/QPE verification methodology (as well as concerning the results) can be obtained from Kerl and Philipp (2015).

2.3 Hydrological modeling approaches

Three different hydrological modeling techniques were implemented and applied for three pilot areas in Saxony (cf. Fig. 1): first, a semi-distributed deterministic model (DeHM), second, a data-driven, neural-network model (DaHM) and, third, a simple classification model, based on the scoring of flood-relevant parameters (ScoHM). Subsequently, the modeling concepts and their application (with regard to calibration, data assimilation, etc.) are briefly described. Only snow-free conditions were regarded for model development and application.

2.3.1 Deterministic hydrological model (DeHM)

DeHM model's topology is based on a nodal representation of sub-catchments. Runoff generation is portrayed by the SCS Curve Number method. Runoff concentration is either modeled via an arbitrarily long cascade of linear reser-

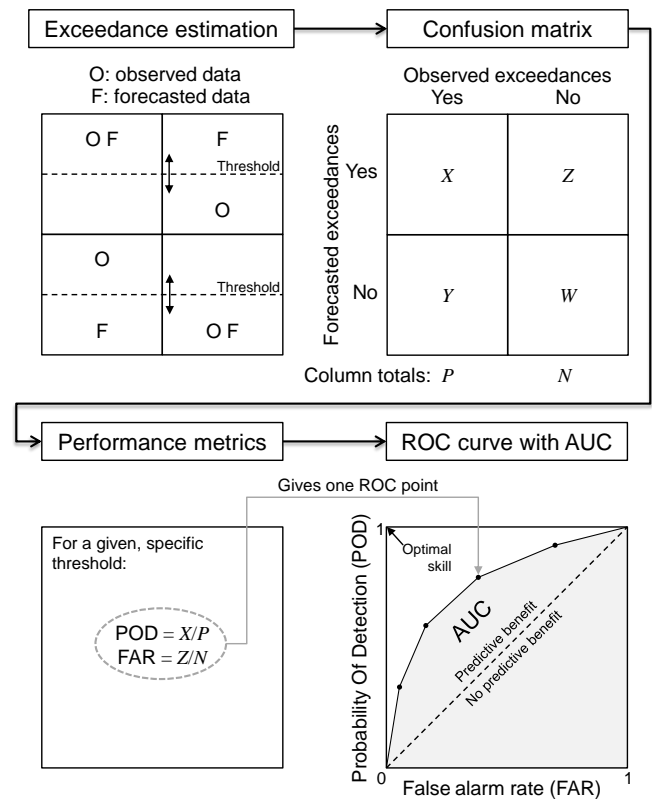


Figure 2. Typical work flow for deriving threshold-exceedance based skill scores, e.g., False Alarm Rate, Probability Of Detection (FAR, POD) or combined products, e.g., Receiver Operating Characteristic (ROC) curves and Area Under Curve (AUC) values.

voirs or via response-function convolution. Channel routing is described with either a time-lag function, a cascade of linear reservoirs, Muskingum method, or a translation-diffusion model. Since there are a number of multi-purpose and flood-retention reservoirs in the pilot areas, flood control was specifically included in the model.

Model calibration was based on event-specifically masked hydrograph data and employed a mixed performance criterion after Li et al. (2015). Data assimilation/state updating was realized with a simplified Kalman filter with error variances, following Blöschl et al. (2014). More details on the DeHM model and its application can be found in Schwarze et al. (2015).

2.3.2 Data-driven hydrological model (DaHM)

DaHM is an artificial neural network model, employing a feed forward two-layer perceptron (Hagan et al., 2002). The input vector features flow, rainfall, and cumulative rainfall data with the general 15-element form I : $[Q_{t-[0...3]}; P_{t-[0...3]}; P^c_{t-[0...6]}]$ (with hourly values of flow Q , rainfall P , and cumulative rainfall P^c). Adding to that, and depending on the considered lead time in the forecast-

Table 1. Overview of the considered QPF and QPE products.

Product	Provider	QPF/QPE	Type	Temporal resolution	Spatial resolution	Lead time	Update cycle
COSMO-DE	German Met Service (DWD)	QPF	Deterministic numerical weather prediction (gridded)	1 h	2.8×2.8 km	21/27 h ^a	3 h
Quantile Forecast (QF)	DWD-RWB-LZ ^b	QPF	Probabilistic forecast of mean areal precipitation	6/12 h ^c	Forecast regions from ca. 600 to 2700 km ²	36 h	12 h
Interpolated rain gauge data	DWD	QPE	89 stations for the area of Saxony plus 25 km buffer	1 h	1×1 km ^d	–	1 h
RADOLAN-RW	DWD	QPE	Rain gauge adjusted weather radar estimate (gridded)	1 h	1×1 km	–	1 h

^a 27 h since February 2014. ^b DWD's Regional Service Center (Regionale Wetterberatung) in Leipzig. ^c Product comprises two consecutive 6 h and two further 12 h intervals. ^d Data gridded via Ordinary Kriging.

ing case, inputs for the rainfall forecast were included, e.g., for forecasting Q_{t+6} , the input P_{t+6} is added, for Q_{t+12} , P_{t+6} ; $t+12$, respectively, whereas the P_{t+x} values represent specific QPF lead times.

The Levenberg-Marquardt algorithm was applied for network training, whilst allowing the number of hidden neurons range from 3 to 13. Event-wise masked hydrograph data and hourly areal rainfall were used for training. 15 training runs were evaluated for each specific hidden-neuron configuration and the best network was selected. Schwarze et al. (2015) give more details on the training and validation of the DaHM model.

2.3.3 Scoring model (ScoHM)

The basic concept of scoring models is – in contrast to deterministic and data-driven concepts – not to simulate or reproduce the development of process variables (e.g., flow) but to empirically determine the current and/or expected further state of a variable by means of a simple classification-based, additive assessment of influencing parameters (i.e., scoring). The employed scoring model resembles the Flooding Susceptibility Assessment approach proposed by Collier and Fox (2003). The method is twofold; first, a baseline susceptibility is derived, based on morphological features, e.g., slope, land cover, etc. Second, a time-variant, dynamic susceptibility is calculated, incorporating the Standardized Precipitation Index (SPI; Edwards and McKee, 1997), cumulative precipitation measures, and the response of a linear reservoir being charged with hourly precipitation.

The scoring is carried out according to Table 2; baseline sub-scores and the SPI sub-score are mapped linearly, according to the range of each respective feature. For the remaining dynamic susceptibility sub-scores, frequency analyses were applied to deliver specific percentiles that are in turn

connected to specific sub-score values, e.g., P -sums within the 75th–90th percentile-range of the data result in a sub-score of 1, etc. The method requires only one effective parameter, namely the recession constant of the incorporated linear reservoir, which was manually adjusted to a global value of 8 h.

In contrast to the DeHM and DaHM models, the ScoHM approach does not rely on observed flow data at all; neither in the sense of directly including auto-correlative signals, as applies for the data-driven DaHM model (in form of the Q_{t-x} inputs), nor indirectly via data assimilation/state updating, as applies for the deterministic DeHM model. Therefore, the ScoHM approach might offer a robustly transferable methodology for hydrological prediction in small, ungauged basins.

3 Results

3.1 User survey

Subsequently, the most important results of the user survey (cf. Sect. 2.1) are presented in a concise manner; a more detailed report can be found in Philipp et al. (2015). The response rate was 76 % ($n = 373$), which is extraordinarily high and is mainly a result of the systematic sampling. For 11 out of 15 questions, user-group specific replies were not distinguishable in a statistical sense. The outcomes of the statistical analysis of the survey data can be summarized as follows:

Information and pathways: (1) The interviewees request selective, event-related information or inform themselves on an event-related basis (rather than on a regular basis). (2) 37 % of all users trust that a more regular and more frequent distribution of warning products will provide increased security for their management decisions, even if the meteorological and hydrological trend remains unchanged. (3) All

Table 2. ScoHM scoring system.

	Parameter description	Upper parameter limits	Sub-score range
Baseline susceptibility	Mean catchment slope	0.02/0.08/0.14/0.20/∞	0 to 4
	Catchment shape factor ^a	0.20/0.40/0.60/0.80/1.00	0 to 4
	Degree of surface sealing	0.05/0.20/0.35/0.50/1.00	0 to 4
	Proportion of fast runoff components ^b	0.10/0.23/0.37/0.50/1.00	0 to 4
Dynamic susceptibility ^c	SPI over the last 30 days ^d	−3/−2/−1/0/1/2/∞	−3 to 3
	Precipitation sum over the last 7 days	Sub-score percentiles ^e	0 to 4
	Precipitation sum over 12/24/48 h ^f	based on actual data from	0 to 4
	Linear reservoir outflow ^g	01/2010 to 09/2015	0 to 4
Total susceptibility score		−3 to 31	

^a Catchment being more circular for values near unity. ^b According to Peschke et al. (1999). ^c In contrast to Collier and Fox (2003), snow-specific dynamic sub-scores were not considered. ^d SPI values rounded to integers. ^e Percentiles: 75th/90th/95th/99th/100th. ^f Only highest sub-score is considered. ^g Linear reservoir being charged with hourly precipitation.

groups, except the group “private persons”, attach greatest importance to the internet in contrast to other communication channels (e.g., fax, video text, voice mail). The official flood warnings issued by fax or email are also used for information by a majority of users. (4) A high availability of warning services and products is deemed important by a vast majority of users, especially in case of flooding.

Flood warning products: (1) A short-termed, but more precise warning is preferred over a long-term estimation, carrying presumably more uncertainty. (2) The majority of users (> 65 %) are interested in receiving a possibly reliable forecast of the peak water level. 45 % of users would appreciate being informed about the peak timing. (3) Most popular products for fulfilling early warning purposes are forecasted hydrographs with uncertainty bands (about 50 % of all persons interviewed), as well as catchment-oriented classification products (“traffic light”, approximately 40 % of all persons interviewed).

Lead time and miscellaneous: (1) The minimum required lead times amount to ≤ 3 (9 % of users), ≤ 6 (27 %), ≤ 12 (50 %), ≤ 24 (83 %), ≤ 72 h (98 %). (2) A lead time of ≤ 12 h is deemed to be adequate by a slim majority of users in small catchments (<200 km²). (2) The interviewed user groups vary significantly in terms of the replies given when being asked for the requested updating frequency of flood warnings and their communication via email or fax. (3) Furthermore, the interviewees of various user groups specifically replied to the questions concerning the quality of current products and the quality of the work of the Saxon Flood Center. (4) Moreover, no significant differences in the response behavior of the various user groups could be identified by statistical means.

3.2 Verification of QPFs

The investigated QPFs (COSMO-DE and Quantile Forecast) were compared against areal precipitation estimates, based

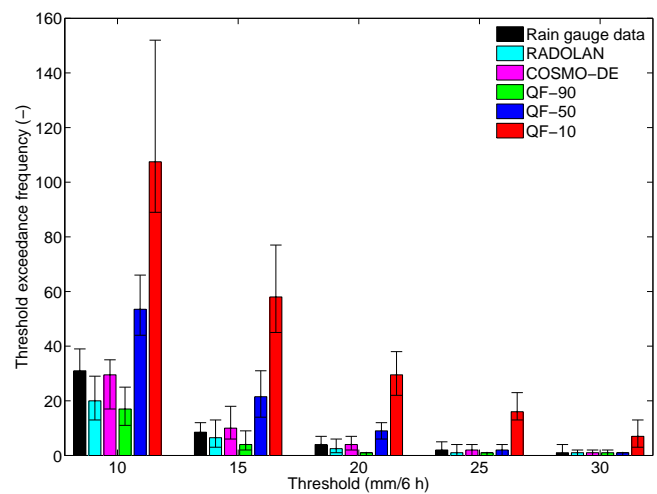


Figure 3. Threshold exceedance frequencies of 6-hourly areal precipitation sums for QPEs (gridded rain gauge data, RADOLAN-RW) and QPFs (COSMO-DE, Quantile Forecast) from April 2011 to June 2014. The bars show the median of exceedance frequencies for the respective precipitation products for the 16 forecast areas (cf. Fig. 1). The whiskers illustrate the minimum and maximum values.

on gridded rain gauge data, and, additionally, a radar-based QPE (RADOLAN-RW product). First, threshold exceedance frequencies were derived from the QPEs and QPFs for threshold values from 10 to 30 mm/6 h (cf. Fig. 3). COSMO-DE delivers exceedance frequencies which are close to the ones obtained from rain gauge data. RADOLAN slightly underestimates the threshold exceedance frequencies from rain gauge data, whereas the chance of underestimation is higher at lower thresholds, and vice versa. Threshold exceedances drawn from the Quantile Forecast’s 50th and 10th percentiles are generally more frequent than the observed ones (i.e., from rain gauge data), whereas the 90th percentile underestimates observed frequencies.

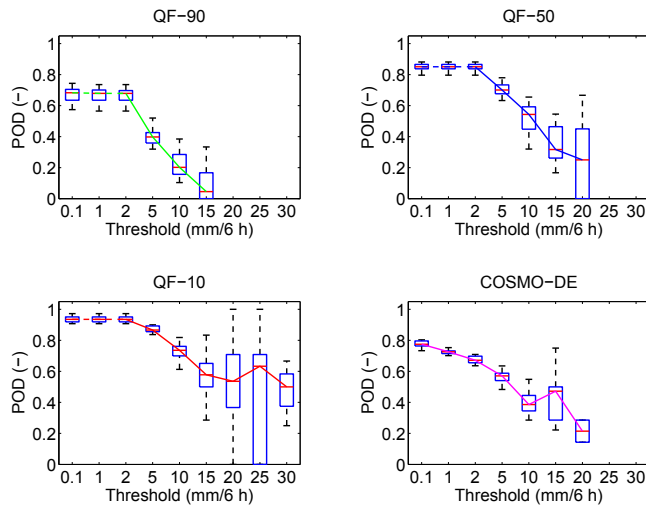


Figure 4. Probability Of Detection (POD) according to thresholds of areal precipitation sums ranging from 0.1 to 30 mm/6 h for the Quantile Forecast and COSMO-DE from April 2011 to June 2014. The box plots indicate the spread of POD over the 16 forecast areas.

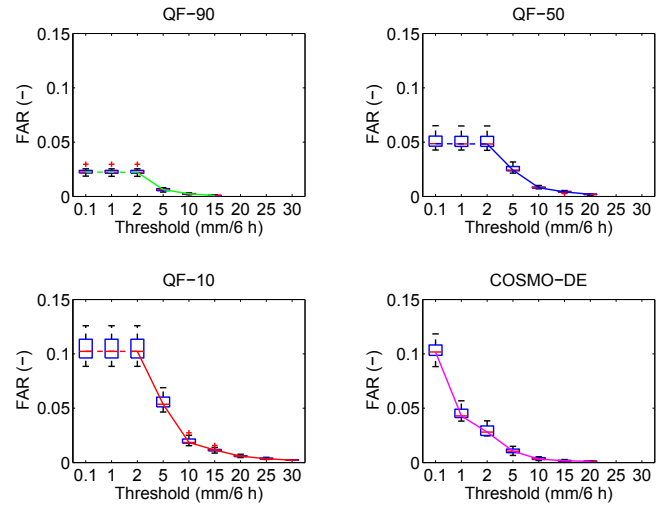


Figure 5. False Alarm Rate (FAR) according to thresholds of areal precipitation sums ranging from 0.1 to 30 mm/6 h for the Quantile Forecast and COSMO-DE from April 2011 to June 2014. The box plots indicate the spread of FAR over the 16 forecast areas.

Second, for a more in-depth view at the regarded QPFs, the contingency-based measures POD and FAR were evaluated down to thresholds of 0.1 mm/6 h (Figs. 4 and 5). Due to product-specific conventions of the Quantile Forecast (areal precipitation sum < 4.5 mm/6 h is set to zero), the results are constant for thresholds < 4.5 mm. Following Winterrath et al. (2012), a minimum of 10 observed or predicted threshold exceedances should be required for the calculation of skill scores. Therefore, POD and FAR were not always evaluated for higher thresholds. Generally, higher precipitation thresholds are connected with lower POD and lower FAR values, and vice versa. Furthermore, for POD, the skill variance amongst the forecast areas increases with increasing precipitation thresholds. $POD = FAR$ indicates a boundary for which the considered QPF has no predictive benefit anymore. This boundary is not reached for both QPFs, concerning the investigated thresholds. Finally, for the regarded QPFs, COSMO-DE exhibits the highest performance with regard to POD/FAR relations and skill variance amongst the forecast areas.

3.3 Hydrological model validation

The three presented models (DeHM, DaHM, and ScoHM) were applied for the three aforementioned pilot areas (cf. Fig. 1). The herein investigated QPEs (gridded rain gauge data and RADOLAN data) and QPFs (COSMO-DE and Quantile Forecast; cf. Sects. 2.2 and 3.2) were used as meteorological drivers (for the current state of work, on the QPF side, ScoHM was charged with the Quantile Forecast only). Validation for the DeHM and DaHM models is straightforward since modeled hydrographs are simply compared

against observed ones. Model evaluation is a bit more delicate for the ScoHM results, since the ScoHM output (i.e., dimensionless scores) does only qualitatively correlate with observed flow values. Therefore, a quantile-mapping procedure (Piani et al., 2009) was applied to relate thresholds of Q with corresponding total-score values.

Model performance was evaluated on the basis of threshold-oriented contingency table analyses, i.e., it is checked if modeled output matches/exceeds a certain observed flow level or not. More specifically, the variation of threshold values delivers a set of corresponding skill scores, e.g., POD values with corresponding FARs. These POD/FAR tuples were used to establish catchment-specific Receiver Operating Characteristic curves (Fawcett, 2006). The curves were finally integrated to deliver the Area Under Curve (AUC), with values near unity for a near-perfect model prediction and near 0.5 for no predictive skill (cf. Fig. 2 and Sect. 2.2). For brevity, results are presented and discussed for the Mandau catchment only, featuring four river gauges.

Generally, different combinations of lead times and update cycles (i.e., the time after which a new forecast is processed) were investigated; herein, results for an update cycle length of 12 h are presented. Event-specifically masked, hourly hydrograph data and hourly rainfall observations were used during model validation. Data which were employed in model calibration/training were not used for validation purposes. Calibration/training data originated from the period of 2006 to 2011 (11 events), validation data from 2010 to 2015 (10 events). Figure 6 comprehensively shows the validation results for the Mandau pilot area. For the Quantile Forecast, results for the 50th percentile are exemplarily shown.

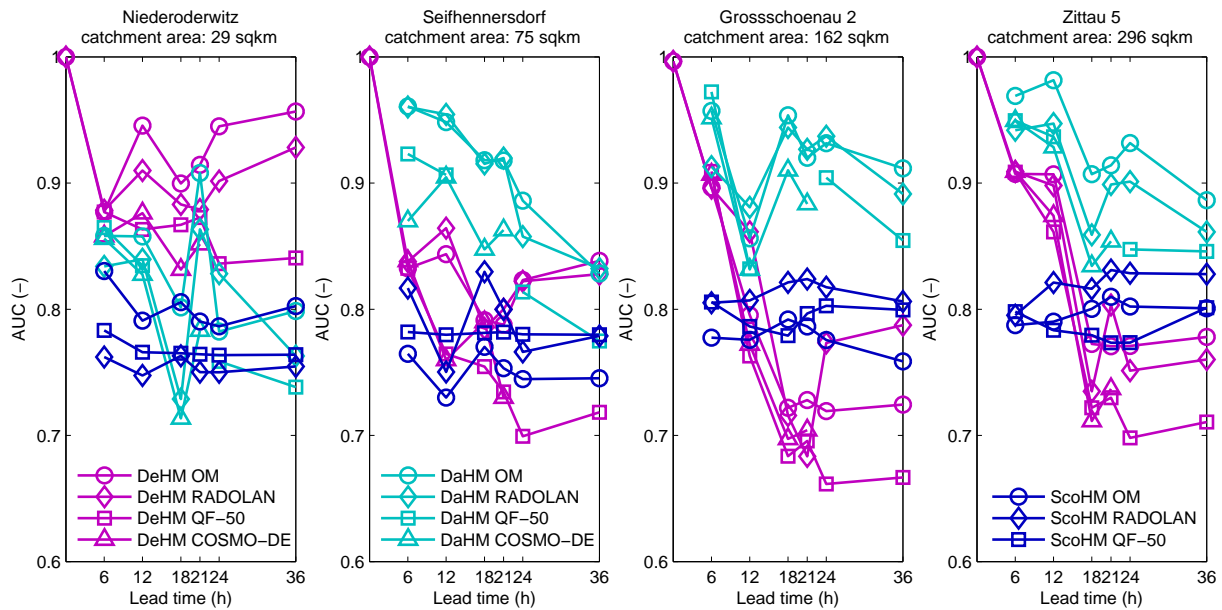


Figure 6. Results of hourly, threshold-oriented evaluation for DeHM, DaHM and ScoHM output in the Mandau pilot area, based on Area Under Curve values. Lead times range from 6 to 36 h, update cycle is 12 h. OM: ombrometer data (i.e., gridded rain gauge data); RADOLAN: QPE from weather radar scans; QF-50: 50th percentile of Quantile Forecast; COSMO-DE: numerical weather prediction output. Skill for DeHM at a lead time of zero is based on true model output after assimilation/updating and can be slightly smaller than unity (e.g., apparent for Großschönau 2).

For the smallest sub-catchment, Niederoderwitz (29 km²), DeHM performs best; for the three larger sub-catchments, DaHM features the highest Area Under Curve values. However, DeHM and DaHM performance trends to decrease with increasing lead time; ScoHM features a quite constant/robust skill development. The reason for this might be that for shorter lead times (6 h) the auto-correlative Q_{t-x} signal, included directly or indirectly in the DeHM and DaHM model (cf. Sect. 2.3), leads to improved performance. This does not apply for the ScoHM results, since the model is not dependent on observed flow data. Generally, ScoHM exhibits Area Under Curve values around 0.8 which indicates a good overall predictive skill, foremost, when keeping in mind the generality and straightforwardness of the model approach.

It can be further seen from Fig. 6 that QPE data delivers highest predictive skill with a tendency of RADOLAN outperforming the rain gauge data. Predictive skill under QPF data (Quantile Forecast and COSMO-DE) is mostly lower. For different QPFs as drivers, resulting skills do not differ greatly. Apparently, the observed differences in QPF quality (cf. Sect. 3.2) do not systematically impact hydrological model skill. Furthermore, it is important to say that validation was carried out on the basis of hourly values; a more general evaluation, e.g., comparing only the highest values within a specific temporal window (e.g., 6 h), would yield considerably higher skill scores.

Finally, it should be stated that the results for the other investigated pilot areas are consistent with the herein presented

findings for the Mandau pilot region when focusing on catchments with areas of up to 200 km². For larger scales, when wave translation and diffusion impact flood expression, the deterministic and data-driven models outperform the scoring approach since it does not account for such processes.

4 Conclusions and outlook

In this study, user demands, driving data, and hydrologic modeling techniques were evaluated within a real-world application context in order to illustrate a way towards a flash flood early warning strategy for (sub-)mesoscale catchments in Saxony. First, results suggest that the majority of potential users of flood warnings would be satisfied with forecasting lead times of up to 24 h and that users are foremost interested in predicted peak water/alarm levels (rather than peak timing). Second, on the basis of meteorological verification results, highly resolved numerical weather prediction data seem to provide the best predictive skill, compared to more general, areally integrated products. Third, differences in the quality of meteorological driving data do not greatly influence hydrological model skill. Fourth, a clear statement on the superiority of one hydrological model over another cannot be made.

In fact, if simple classification models would be sufficient to satisfy warning needs (e.g., providing the information whether or not a specific threshold is likely to be exceeded in the next forecasting interval), results show that such a

modeling approach (i.e., ScoHM) performs with favorable skill, compared to more sophisticated modeling techniques, and without introducing cumbersome parameter estimation problems and limited (DeHM) or even non-existent (DaHM) regional transferability. However, overall forecasting skill always decreases with increasing randomness of driving events and conditions, i.e., the more rare/focused/intense the flood-causing processes and/or the longer the lead time, the smaller the chance of correct detection/warning.

Further research is currently carried out regarding the statewide implementation and comparative evaluation of the herein considered approaches to gain more insight into the dependencies of meteorological drivers, hydrological models, spatio-temporal scaling effects, and regional transferability. Meteorological verification will be carried out for smaller spatio-temporal scales and with a temporally extended data set. Additionally, the set of QPFs will be extended to German Met Service's 21-member ensemble product, COSMO-DE-EPS. Thus, allowing a statewide, comprehensive probabilistic verification and validation of the presented hydrological models.

Acknowledgements. The meteorological driving data were provided by courtesy of German Meteorological Service (Deutscher Wetterdienst). The authors would like to thank one anonymous referee for thoroughly reviewing the manuscript. The manuscript was funded by Saxon State Ministry of the Environment and Agriculture Grant/project number: 45-8904.20/1.

References

- Baldauf, M., Förstner, J., Klink, S., Reinhardt, T., Schraff, C., Seifert, A., and Stephan, K.: Short description COSMO-DE (LMK) and its data bases on DWD's data server, Technical Report, DWD, 2011 (in German).
- Blöschl, G., Nester, T., Parajka, J., and Komma, J.: Flood forecasting on the Austrian Danube and data assimilation, *Hydrol. Wasserbewirts.*, 58, 64–72, 2014 (in German).
- Collier, C. G. and Fox, N. I.: Assessing the flooding susceptibility of river catchments to extreme rainfall in the United Kingdom, *International Journal of River Basin Management*, 1, 225–235, 2003.
- Edwards, D. and McKee, T.: Characteristics of 20th century drought in the United States at multiple time scales, *Atmospheric Science Paper*, 634, 1–155, 1997.
- Fawcett, T.: An introduction to ROC analysis, *Pattern Recogn. Lett.*, 27, 861–874, 2006.
- Hagan, M. T., Demuth, H. B., Beale, N., and De Jesus, O.: *Neural Network Design*, Published by Martin Hagan, 2nd Edn., 2002.
- HWMO: Verwaltungsvorschrift des Sächsischen Staatsministeriums für Umwelt und Landwirtschaft zum Hochwassernachrichten- und Alarmdienst im Freistaat Sachsen (Saxon Flood Alarm Bylaw), Saxon State Government, 2014 (in German).
- Jeschke, K., Greiff, B., Kolf, R., Burk, H.-P., Merker, H., Bogatsch, C., Fritzsche, C., and Vogel, M.: Commission's report on the assessment of the flood information and warning procedures in Saxony during the August 2010 flood event, Technical Report, Saxon State Government, 2010 (in German).
- Jolliffe, I. T. and Stephenson, D. B. (Eds.): *Forecast Verification: A Practitioners Guide for Atmospheric Science*, Wiley, 2012.
- Kerl, F. and Philipp, A.: Verification of operationally available QPF and QPE products for the area of Saxony (Germany) from 04/2011 to 06/2014 (in German), Technical Report, Saxon State Office for Environment, Agriculture and Geology, 2015.
- LfULG: Event analysis of the 2013 Flood (in German), Technical Report, Saxon State Office for Environment, Agriculture and Geology, 2015.
- Li, Y., Ryu, D., Western, A. W., and Wang, Q. J.: Assimilation of stream discharge for flood forecasting: updating a semi-distributed model with an integrated data assimilation scheme, *Water Resour. Res.*, 51, 3238–3258, 2015.
- Peschke, G., Etzenberg, C., Müller, G., Töpfer, J., and Zimmermann, S.: The expert system FLAB: a tool for the delineation of landscape units with similar runoff generation mechanisms, Technical Report, IHI Zittau, 1999 (in German).
- Philipp, A., Kerl, F., and Müller, U.: Demands by potential users for a flood early warning system for Saxony, *Hydrol. Wasserbewirts.*, 1, 4–22, 2015 (in German).
- Piani, C., Haerter, J. O., and Coppola, E.: Statistical bias correction for daily precipitation in regional climate models over Europe, *Theor. Appl. Climatol.*, 99, 187–192, 2009.
- Sacher, D., Weigl, E., Podlasly, C., and Winterrath, T.: RADOLAN/RADVOR-OP: description of the composite format, Technical Report, DWD, 2011 (in German).
- Sachs, L.: *Applied Statistics*, Springer Berlin Heidelberg, 9th Edn., 1999.
- Schwarze, R., Singer, T., Stange, P., Wagner, M., and Schütze, N.: Development and implementation of deterministic and data-driven methods for hydrological forecasting in small, partly ungauged basins for the purpose of deriving flood early warnings, Technical Report, Dresden University of Technology, 2015 (in German).
- Winterrath, T., Weigl, E., Reich, T., Rosenow, W., and Stephan, K.: RADVOR-OP: radar-based, real-time precipitation nowcasting for operational purposes, Technical Report, DWD, 2012 (in German).