Hydrol. Earth Syst. Sci. Discuss., 9, 6857–6887, 2012
www.hydrol-earth-syst-sci-discuss.net/9/6857/2012/
doi:10.5194/hessd-9-6857-2012

Hydrology and Earth System Sciences Discussions

This discussion paper is/has been under review for the journal Hydrology and Earth System Sciences (HESS). Please refer to the corresponding final paper in HESS if available.

Long-range hydrometeorological ensemble predictions of drought parameters

F. Fundel, S. Jörg-Hess, and M. Zappa

Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Zürcherstrasse 111, 8903 Birmensdorf, Switzerland

Received: 3 May 2012 – Accepted: 7 May 2012 – Published: 1 June 2012

Correspondence to: F. Fundel (felix.fundel@wsl.ch)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Low streamflow as consequence of a drought event affects numerous aspects of life. Economic sectors that may be impacted by drought are, e.g. power production, agriculture, tourism and water quality management. Numerical models have increasingly been used to forecast low-flow and have become the focus of recent research. Here, we consider daily ensemble runoff forecasts for the river Thur, which has its source in the Swiss Alps. We focus on the low-flow indices duration, severity and magnitude, with a forecast lead-time of one month, to assess their potential usefulness for predictions. The ECMWF VarEPS 5 member reforecast, which covers 18 yr, is used as forcing for the hydrological model PREVAH. A thorough verification shows that, compared to peak flow, probabilistic low-flow forecasts are skillful for longer lead-times, low-flow index forecasts could also be beneficially included in a decision-making process. The results suggest monthly runoff forecasts are useful for accessing the risk of hydrological droughts.

1 Introduction

Droughts, which can occur on a wide temporal range, can be defined through various parameters. Typical indicators are a temporally and spatially extended lack of precipitation (meteorological droughts), reduced soil moisture (agricultural droughts) and low levels of runoff or groundwater (hydrological droughts) (Heim Jr., 2002). This variety in the way droughts are defined is a direct consequence of the range of socioeconomic impacts they have on different interest groups. In this study, we assess the quality of monthly forecasts of hydrological droughts, characterized by low streamflow (low-flow). Streamflow is an appealing measure for droughts as it combines different catchment aspects, ranging from the input of precipitation to storage and transfer processes. Low-flow forecasts on a monthly time scale are therefore potentially useful for hydropower generation, agriculture (irrigation), conventional power production (supply of heat exchange water), water quality, navigation and tourism. Skillful forecasts of low-flows could help to prevent or mitigate the consequences of water shortage in those sectors (Steinemann, 2006).

Commonly, two approaches to predict properties of low-flow events in the long range can be distinguished. First, stochastic approaches, which relate the current state of a catchment and potential predictors to what has been observed in the past, to infer the likelihood of low-flow within the prediction period. These include regression techniques (Cebrián and Abaurrea, 2011; Moreira et al., 2008; van Ogtrop et al., 2011), time series models (Bordi and Sutera, 2007; Chung and Salas, 2000; Lohani and Loganathan, 1997; Mishra and Desai, 2005), and neural network techniques (Mishra and Desai, 2006; Kim and Valdes, 2003; Morid et al., 2007). Also procedures identifying correlations of drought events with tele-connection patterns (Tadesse et al., 2005; Özger et al., 2012) or certain weather types (Fleig et al., 2011) can be used to indicate potential drought events. Cancelliere et al. (2006) and Hwang and Carbone (2009) used autoregressive models not only to predict drought parameters but additionally to quantify the uncertainty of their prediction. Drought parameters inferred from statistically downscaled atmospheric models Cacciamani et al. (2007) also have proven predictive quality.

The second, less common approach for the long-range prediction of droughts involves a coupled atmospheric-hydrological model. Wood et al. (2002) employ monthly forecasts from a global atmospheric model to drive a grid-based hydrological model that produces reasonable predictions of low-flow up to several months in advance. The refined system of Li et al. (2008) and Luo and Wood (2007) was able to predict average monthly drought conditions up to three months ahead.

By coupling meteorological and hydrological models, useful peak flow predictions for the short- to medium-range are possible. As the predictability of an event mainly depends on its life-time (Hirschberg et al., 2011), peak-flow forecasts rarely show skill beyond 10 days. This although depends on the catchment characteristics, as well as the quality of the models involved and the observations needed for an appropriate initialization (Fundel and Zappa, 2011; Webster et al., 2010). As low-flow events are generally rather persistent phenomena, predicting them could still be valuable, and it might be worthwhile investigating their properties at lead-times when peak flow predictions lost their value long ago. The value of forecasts can be increased if additionally the prediction uncertainty is quantified, e.g. by using an ensemble prediction system (EPS) or a multi-model ensemble to drive the hydrological model (Cloke and Pappenberger, 2009)

The objective of this paper is to assess the value of daily predictions of low-flow up to a lead-time of one month, employing the VarEPS ensemble reforecast (Vitart et al., 2008b) from the European Centre for Medium Range Weather Forecast (ECMWF) as forcing for the hydrological model PREVAH (Viviroli et al., 2009). Daily mean catchment runoff forecasts of 32 days lead-time were produced in the period from 1991–2008. The forecasts are evaluated here with regard to the lower end of the flow regime. Ensembles of commonly applied low-flow indices, namely duration, severity and magnitude, are produced and verified against indices taken from the observed runoff. The resulting forecasts are then evaluated in terms of their economic value for different interest groups.

2 Data and methodology

2.1 Domain

The domain chosen for this study is the pre-alpine catchment of the Thur river, located in the north-eastern part of Switzerland, which discharges into the Rhine river. The catchment is 1696 km^2 in area and its orography extends from 356 to 2503 m a.s.l. The climatic conditions are relatively cool and the runoff generating processes from autumn to spring are affected by snowfall and melt processes. The annual average precipitation amount is about 1500 mm, and mainly falls during the summer months (Gurtz et al., 1999). This catchment, which is relatively large for Alpine conditions, was chosen as in smaller catchments low-flow over a longer period very seldom occurs. In small catchments, low-flow events are easily interrupted by small-scale precipitation events, which complicates the evaluation of longer-lasting events. A large catchment is therefore more appropriate to demonstrate the value of forecast of longer events happening within a monthly forecast.

2.2 Meteorological forcing

The unified variable resolution ensemble prediction system VarEPS (Vitart et al., 2008a,b) produces each week a global, 51-member forecast with a lead-time of 32 days. Its horizontal resolution is 50 km for the first 10 days an 80 km in the remaining 11–32 days. From day 10 onwards, the atmospheric model is coupled to an oceanic model. The motivation for varying the resolution is to benefit from the higher resolution in the early forecast range at longer lead-times. At the time of this study, the forecasts were issued for Thursday at 00:00 UTC and were archived in time-steps of 6 h. With each run of the operational VarEPS, an ensemble reforecast is started for the same day of the year over the past 18 yr, and also ranges over 32 days. The reforecasts share the same model version as the forecasts and are meant to capture the same model errors, e.g. to allow for an efficient post-processing. Compared to the operational forecast, the VarEPS reforecast dataset consists of 954 ensemble forecasts, allowing for a more robust evaluation of forecast skill. Unlike the operational EPS with 51 ensemble members, the reforecasts consist of only 5 members and initial states for each reforecast are taken from the ECMWF global reanalysis ERA-Interim or from ERA40. More detailed information about the history of model developments is available on http://www.ecmwf.int/products/data/technical/model_id/index.html.

As forcing for the hydrological model, VarEPS 5-member reforecast fields of wind speed, 2 m temperature, 2 m dew point, sunshine duration, surface albedo and solar radiation were used. To meet the grid size of the hydrological model of 500 m × 500 m, a downscaling was performed, based on a bilinear interpolation. Temperature

HESSD
9, 6857–6887, 2012

was adjusted according to elevation, assuming a lapse rate of 0.65°C/100 m (Jaun and Ahrens, 2009).

2.3 Hydrological model

Runoff predictions were generated using the semi-distributed hydrological model Precipitation Runoff EVApotranspiration Hydrotope (PREVAH, Viviroli et al., 2009), with VarEPS as the meteorological forcing. PREVAH consists of hydrologic response units (HRUs) and a runoff generation module based on the HBV model (Bergström and Forsman, 1973), taking account of the spatial distribution. Information on PREVAH physics, parameterization and the downscaling method is given in Gurtz et al. (1999) and Viviroli and Gurtz (2007). PREVAH's parameter setting was conditioned by matching the produced runoff to observations for an extended reference period. This optimization was performed with a focus a on the average flow volume. The setup of PREVAH adopted for the Thur basin is the same as the one used in Fundel and Zappa (2011) and Zappa and Kan (2007). In the latter publications the calibration and verification of the modelled runoff against the observed runoff, the water balance components, and hydrographs are presented.

Initial conditions for each forecast run were obtained from a continuous reference simulation forced with meteorological surface observations, containing information about the water storage in the different modules of each HRU. To highlight the quality of the initial conditions, Fig. 1 shows a comparison of the reference simulation and the runoff observed for the low runoff regime. The seasonal cycle and the runoff volume concur fairly well. The evaluation of daily mean runoff from the reference simulation during the study period 1991–2008 results in a mean error of −0.003 (−3 %).

2.4 Drought characterization

Hydrological droughts are generally characterized by their duration (time between onset and offset), severity (cumulative water deficit) and magnitude (severity/duration)

HESSD
9, 6857–6887, 2012

(Fleig et al., 2006; Hisdal et al., 2001; Hisdal and Tallaksen, 2000; Mishra and Singh, 2010; Nalbantis and Tsakiris, 2008; Smakhtin, 2001; Tallaksen et al., 1997; Yoo et al., 2011; Zaidman et al., 2002). Defining hydrological droughts solely by considering the runoff requires the assignment of a runoff threshold. Whenever the predicted or observed runoff drops below that threshold, this counts as a low-flow event. Figure 2 illustrates the low-flow indices drawn from a observed or forecast hydrograph. It is, however, not obvious where the threshold should be set. Some low-flow sensitive stakeholders may be interested in a constant threshold, e.g. a power plant that requires a certain amount of water for cooling. Others might be affected by droughts only in certain seasons, e.g. tourism. To meet both concerns, a seasonally variable, quantile-based, low-flow threshold was selected.

In addition, a separate threshold for forecast and observed runoff was used, based on the non-exceedance probability of runoff. This is a very simple way to correct for systematic, additive error (bias) in runoff predictions. Choosing the same threshold for observations and forecasts would possibly result in a systematic over- or underforecasting of low-flow events. By choosing a different threshold for observed and for forecast runoff based on the same frequency of occurrence, we can assure this bias will not affect the verification results of low-flow duration. Possible nonlinearities in the prediction bias could, however, influence the verification results of low-flow severity and magnitude. A lead-time dependency of the low-flow threshold was implemented for the forecasts. After all, a bias could be removed by, e.g. statistical post-processing, but as well-calibrated forecasts of rare events require a large training dataset, which is why we preferred to use this simple method.

Within the 32 day forecast period, the events of observed or forecast runoff falling below the low-flow threshold were detected and the low-flow indices were calculated. If more than one event was detected within a forecast interval of 32 days, the longest consecutive period of low-flow was considered as the forecast/observed event. Forecast and observed events do not necessarily have to overlap, neither between forecast and observation nor between the ensemble forecast members. The forecast system is reduced in order to answer the questions: what is the longest expected low-flow event within the next 32 days and was such an event observed? In the case of probabilistic forecasts, the question is different: what is the probability of exceeding a low-flow event of x days duration (y mm severity; y/x mm/day magnitude) within the next 32 days? By doing so, the lead-time is no longer a possible source of forecast error.

Low-flow duration, severity and magnitude are clearly not independent as they are all calculated from the same set of events. For example an event of long duration is likely to be very severe as well. Their forecast performances are therefore expected to be similar. Still, the different indices should reflect the demands of different interest groups to a low-flow forecast. For hydropower production, for example, the severity might be of paramount interest, whereas for power-plant cooling the duration is more crucial.

2.5 Verification scores

The focus of the verification of ensemble low-flow index forecasts is set on the value for potential forecast users. A score, designed to give the economic value of a forecast depending on the vulnerability of a forecast user to a certain event, is the relative value or value score, (Murphy, 1977; Richardson, 2000; Roulin, 2007). The user is supposed to take preventive action whenever a forecast is issued with a probability exceeding the user's personal cost-loss ratio. Loss is the customers expense when being struck by an event without any preparation. Cost is what the user would spend on taking preventive action. The value score then gives the relative economic gain for the user when following the advice of the forecast, compared to having only the climatological event frequency as a basis for decision-making. The value score varies between ≤ 0 (no additional forecast value) and 1 (perfect forecast). Note that even if the value score equals 1, the user still has to bear the costs for taking preventive action. The value score is calculated for probabilistic forecasts of exceeding a threshold. It gives a value for each possible probability the prediction system can issue, depending on the cost-loss ratio. The upper envelope of all value curves gives the value of the prediction system.

HESSD
9, 6857–6887, 2012

Another score, designed to give an intuitive measure of the prediction system performance, is the Generalized Discrimination Score or Two Alternatives Forced Choice Score (2AFC) (Mason and Weigel, 2009; Weigel and Mason, 2011). The score reflects the forecasts' performance in discriminating between different observations. An appealing property of the score is that it can be used for ensemble, probabilistic, dichotomuos, polychotomous or continuous forecasts and corresponding observations. The 2AFC score is not affected by systematic biases in the forecast system. Therefore it gives a measure of the potential quality of the forecasts if they were well calibrated. A 2AFC score above 50 % is reached if the forecast is better than a guess based on climatology and 100 % for a perfect forecast. Here, the 2AFC score is used to evaluate ensemble predictions against continuous observations, or probabilistic predictions against dichotomous observed outcomes, depending on the context.

3 Results

Here we first evaluate the forecast quality for different flow regimes. A low-flow detection threshold is then selected to subsequently evaluate forecasts of low-flow indices and estimate their economic value.

3.1 Extended lead-time forecast quality

One basic hypothesis in this study is that low-flow events are more predictable than peak flow events. Thus, long-range forecasts of low-flow are potentially very useful. In the first part it is tested how well the forecast system can predict the probability to exceed different thresholds, from very low to very high runoff. As described, a number of seasonally varying quantiles from the complete gauged Thur runoff time-series available are used as thresholds. Figure 3 shows the 2AFC score for these thresholds. The verification results support the hypothesis. Peak flow forecasts, i.e. exceedance probabilities for the 80th quantile and above, show skill for up to about day 15. Low-flow

HESSD
9, 6857–6887, 2012

forecasts, e.g. the 20th quantile, are skillful for up to the end of the forecast range. At low thresholds, the forecast quality decreases less rapidly with growing lead-time.

The long-term predictability of low-flow events might be attributed to the persistence of the initial state of the model. Conceptually, the recession rate of a hydrograph is a function of the runoff (e.g. Kirchner, 2009), and any further reduction of already low-flow is a very slow process. Consequently, a good initialization of the prediction model is especially crucial for forecasts of low-flow events. If the initialization leads to a surplus of runoff, for example, the low-flow threshold might not be crossed within the forecast range or, at least, the forecast timing would be poor. The dependence of forecast quality on the quality of the initial state for high- and low-flow regimes is shown in Fig. 4 using the 2AFC score. It is assumed that a good initial state, taken from the reference run, results in a predicted runoff close to the observed runoff. In order to get a clear signal, we consider the initial state as good if the relative absolute error of the predicted runoff at initialization is smaller than 25 %, and as poor if the relative error is higher than 75 %. For both groups the probability of exceeding the 85th and the 15th quantile is verified. In the low runoff regime, a good initial state is beneficial over the complete forecast range of one month. The effect of a good initialization on the forecast of higher runoff is lost after about one week.

3.2 Choice of the low-flow detection threshold

The next aspect includes evaluating the performance of low-flow index ensemble predictions according to the choice of thresholds used to define a low-flow event. The result of this analysis should give an indication about which threshold would be best for the subsequent verification of the low-flow index forecasts. Figure 5 shows the verification results of ensemble duration, severity and magnitude forecasts for low-flow detection thresholds from the seasonally varying 5th to 50th quantile of Thur runoff. For all indices, the overall forecast performance appears to be better if a lower threshold is chosen. Duration forecasts degrade most strongly with higher thresholds, but stay skillful with a 2AFC score of 67 % when using the 50th quantile. Severity forecasts

HESSD
9, 6857–6887, 2012

reach a 2AFC score of about 70 % for the 50th quantile, and the magnitude forecast skill stays fairly constant within the uncertainty bounds when varying the threshold. The decrease in forecast skill with increasing threshold can be attributed to there being greater uncertainty in the forecast itself. The heigher uncertainty in the score at lower thresholds is due to the lower number of events. For the subsequent evaluation of low-flow index forecasts, a threshold based on the seasonal 15th quantile is applied. This is a compromise between the number of low-flow events and the significance of the findings regarding low-flow. The runoff associated with the 15th quantile for our research catchment can be deduced from Fig. 1. A distinct seasonal runoff maximum occurs in April due to the contribution of snowmelt. The yearly minimum is reached in October or November, when snow accumulation starts.

3.3 Low flow prediction/observation

The quality of a low-flow forecast relies greatly on the quality of the representation of the initial states in the hydrological model. These initial states are taken from a reference run, forced with meteorological surface observation. Usually no measurements of the state variables of hydrological models are available, which means a good reproduction of runoff by the reference run is interpreted as an indication the system's state has been well reproduced. Figure 6 shows the occurrence (therefore indirectly the duration), severity and magnitude of low-flow events at the runoff gauge in Andelfingen during the study period 1991 to 2009, as observed or predicted by the reference run. By definition, the events are distributed evenly over the year (because a low-flow detection threshold of varying quantiles is used). The most severe events and the events of greater magnitude however mostly occur in late spring/ early summer, when the runoff reaches the yearly maximum due to the contribution of melting snow. The high quality of the reference run already mentioned is also reflected in the predictions of the low-flow indices. A high degree of agreement in the occurrence and timing of the events can be seen. High severities and magnitudes in the observations are also strong events in

HESSD
9, 6857–6887, 2012

the reference run, although the absolute number can vary. Altogether, the initial states seem to provide a reasonable basis for the initialization of a forecast.

Ensemble forecasts of low-flow duration, severity and magnitude, utilizing a low-flow threshold based on the seasonal 15th quantile, are shown in Fig. 7, together with the corresponding observations. It is striking how well the duration and severity of the observed low-flow is contained within the range of the ensemble. The signals of duration, severity and magnitude are very similar to each other as they are derived from the same low-flow events. However, severity and magnitude are sensitive to the amplitude of runoff and thus capture different aspects of the forecasts. The forecasts of low-flow indices are biased, as can be seen in the rank histograms (Talagrand et al., 1997) in Fig. 8. A disproportionately large part of observations fall in the lower bins spanned by the ensemble members. This lack of reliability is most distinct for magnitude forecasts, and less for forecasts of duration and severity. However, the discrepancy between the forecast probability and the observed frequency is an error that could possibly be corrected for with a statistical post-processing. Such a model bias was partly corrected for by using separate detection thresholds for observations and for forecasts. No seasonal dependency of low-flow events is apparent. This was expected as the detection threshold varies with season. One noticeable feature that can be seen in the ensemble forecasts and the observations is the distinct signal left by the 2003 drought caused by a heat wave that affected large parts of Europe (Beniston, 2004; Schär et al., 2004; Zappa and Kan, 2007).

3.4 Relative economic value

The economic value of probabilistic forecasts of low-flow exceeding different levels of duration, severity and magnitude was calculated for forecast probabilities using the 15th quantile low-flow threshold (Fig. 9). Value scores >0 are shown for a variety of forecast users, characterized by their individual cost-loss ratio. It can be seen that, for all thresholds and low-flow indices, valuable forecasts for certain user groups can be produced. Especially risk avers forecast users with low cost-loss ratios would benefit from ensemble forecasts for events of long duration or high magnitude and severity. In contrast, forecasts of longer, more severe events have no additional value for users with higher cost-loss ratios. These users would benefit from probabilistic forecasts of less intense events. For all users, value scores > 50 % are possible when using the 15th quantile as the low-flow detection threshold. Higher value scores are possible with the 5th quantile, and lower scores with the 50th quantile as the low-flow detection threshold. The highest value score for each threshold is reached where the cost-loss ratio concurs with the climatological frequency of the event. This explains the shift to the left with increasing thresholds. The maximum value score reached is generally independent of the duration, severity or magnitude of the event.

Low-flow index predictions no longer contain information about the lead-time. A direct comparison with probabilistic predictions is therefore not possible. The loss of predictive skill of probabilistic forecasts with growing lead-time has already been addressed (Fig. 3). In comparison, the maximum score of probabilistic low-flow index exceedance remains fairly constant if the index increases. As the timing of the low-flow events is not exactly specified, but is somewhere between day 1 and day 32, the value score for the events ranges around the lead-time averaged score of the probabilistic forecasts, independent of the duration, severity or magnitude of the event.

3.5 Timing

One drawback of forecasting low-flow indices instead of the full hydrograph is the lack of information obtained about the timing of the event. However, the information is not lost, as the start and the end of each event can be forecast and evaluated as well. Here, we evaluate the mean lead-time of forecast low-flow events, as described in Sect. 2.3. This lead-time is restricted to 32 days, whereas this limit can only by reached by events starting on the last day of the forecast. A concentration of events will have a lead-time around 16 days because the longer the event, the closer the mean lead-time has to be to the center of the forecast range. Of all forecast events, 46 % had a counterpart in the observations (hit rate), and 32 % were missed or false alarms. Figure 10 shows how the timing of the observed event relates to the median of the predicted timing for the 42 % of hits. The data exhibit a lot of scatter, but, as can be seen in the 2-D density map, the forecast timing can explain 22 % of the variance in the observed timing. In order to attribute a forecast event to an observed event, the events should at least for one day overlap. This is the case for 49 % of all hits.

4 Discussion and conclusions

We evaluated 18 yr of hydrological ensemble forecasts of daily mean runoff for the Thur catchment in Switzerland that were weekly initialized, with 32 days lead-time. We focused on their potential to provide skillful, and thus valuable information about low-flow events. The basic assumption that the predictability of low-flow is greater than that of peak flow was confirmed by the verification results. We attribute this to the fact that long-lived processes dominate the recession behavior of the runoff. For higher flow, the quality of the runoff forecasts is strongly dependent on the correct timing and the amount of precipitation given by the meteorological forcing model. The positive effect of a good initial state on the forecast is quickly lost. This was also confirmed for catchments prone to flash-floods, where the influence of initial conditions was found to be already lost after just a few hours (Zappa et al., 2011). Additionally, a good representation of the system states in the hydrological model when the forecast is initialized is essential. In this study, a comparison of the observed runoff with the runoff from a reference run indicates the initial states had a good quality, which is also supported by the good agreement found between the observed and the modeled low-flow events. In a low-flow regime, a good initialization can be beneficial for the forecast for a much longer range. Shukla and Lettenmaier (2011) similarly stress the importance of the initial conditions for lead-times of up to one month, depending on the climatic conditions.

An ensemble forecast of the low-flow indices duration, severity and magnitude could be beneficial for various interest groups, even though it is biased, and have an economic advantage mostly independent of the duration, severity or magnitude of the event. Such forecasts are of reduced complexity, as they no longer contain information about the timing of the low-flow event. Hence, the forecast system is not penalized when the on- and offset of low-flow events are lagged compared to observations, which enhances the value of the forecasts. A comparison of the forecast timing of events with the observed timing showed that this information only gives a rough estimate. The forecasts of low-flow indices can provide useful information about the characteristics of an upcoming event. The timing prediction should, however, be handled with caution.

The ensemble forecasts evaluated in this study were performed for a relatively large catchment. This was chosen mainly in order to ensure the occurrence of longer low-flow periods, as in larger catchments runoff is not as strongly affected by small or local precipitation events. However, we also tested low-flow predictions for smaller, Alpine catchments. There, only short low-flow events could be evaluated and the value of these forecasts was found to be lower. Nevertheless, low-flow index forecasts can still be useful for smaller catchments.

The parameters of the hydrological model PREVAH were found by optimizing the predicted runoff subject to average flow volume. This is certainly not the best approach for low-flow forecasts and might introduce biases in predictions of the lower flow regime. Our findings, however, suggest that an operationally used hydrological prediction system, which is meant to give warnings primarily of peak flow, can also be useful in forecasting low-flow. Biases due to inadequate model parameterization can be partly addressed by defining the low-flow threshold for the observed runoff from an observation climatology and for forecast runoff from the model predictions. This can be seen as a simple approach to statistical post-processing. More complex methods could be considered, but this was beyond the scope of this study.

The downscaling of the meteorological model could be further improved. In this study, the relatively coarse horizontal grid of the meteorological forcing model was downscaled to the grid of the hydrological model in a simple bilinear interpolation. A dynamic downscaling involving one ore more nested regional models would be preferable. No

HESSD
9, 6857–6887, 2012

such approach is, however, available for region and the forecast range of one month considered here.

The operational version of the meteorological forcing model VarEPS would offer 51 ensemble members compared to only 5 members from the reforecast. As a low number of ensemble members introduces low reliability in the forecast (Weigel et al., 2007), the here shown verification scores reflect the lower limit of what can be achieved with the full ensemble. Despite the limitations, ensemble forecasts of low-flow indices could provide a valuable basis for decision-making and be of economic value for forecast users.

Acknowledgements. Support from the Swiss National Research Program Sustainable Water Management (NRP 61 project DROUGHT-CH) is gratefully acknowledged. We also express out gratitude to ECMWF and MeteoSwiss for providing the necessary data.

References

Beniston, M.: The 2003 heat wave in Europe: A shape of things to come? An analysis based on Swiss climatological data and model simulations, Geophys. Res. Lett., 31, L02202, doi:10.1029/2003GL018857, 2004. 6868

Bergström, S. and Forsman, A.: Development of a conceptual deterministic rainfall-runoff model, Nord. Hydrol., 4, 147–170, 1973. 6862

Bordi, I. and Sutera, A.: Drought monitoring and forecasting at large scale, in: Methods and Tools for Drought Analysis and Management, edited by: Rossi, G., Vega, T., and Bonaccorso, B., Chap. 1, 3–27, Springer Netherlands, doi:10.1007/978-1-4020-5924-7, 2007. 6859

Cacciamani, C., Morgillo, A., Marchesi, S., and Pavan, V.: Monitoring and forecasting drought on a regional scale: Emilia-Romagna region, in: Methods and Tools for Drought Analysis and Management, edited by: Rossi, G., Vega, T., and Bonaccorso, B., Chap. 2, 29–48, Springer Netherlands, doi:10.1007/978-1-4020-5924-7, 2007. 6859

Cancelliere, A., Mauro, G. D., Bonaccorso, B., and Rossi, G.: Drought forecasting using the Standardized Precipitation Index, Water Resour. Manag., 21, 801–819, doi:10.1007/s11269-006-9062-y, 2006. 6859

Cebrián, A. C. and Abaurrea, J.: Risk measures for events with a stochastic duration: an application to drought analysis, Stoch. Env. Res. Risk A., in press, doi:10.1007/s00477-011-0521-5, 2011. 6859

Chung, C.-H. and Salas, J. D.: Drought Occurrence Probabilities and Risks of Dependent Hydrologic Processes, J. Hydrol. Eng., 5, 259, doi:10.1061/(ASCE)1084-0699(2000)5:3(259), 2000. 6859

Cloke, H. L. and Pappenberger, F.: Ensemble flood forecasting: A review, J. Hydrol., 375, 613–626, doi:10.1016/j.jhydrol.2009.06.005, 2009. 6860

Fleig, A. K., Tallaksen, L. M., Hisdal, H., and Demuth, S.: A global evaluation of streamflow drought characteristics, Hydrol. Earth Syst. Sci., 10, 535–552, doi:10.5194/hess-10-535-2006, 2006. 6863

Fleig, A. K., Tallaksen, L. M., Hisdal, H., and Hannah, D. M.: Regional hydrological drought in north-western Europe: linking a new Regional Drought Area Index with weather types, Hydrol. Process., 25, 1163–1179, doi:10.1002/hyp.7644, 2011. 6859

Fundel, F. and Zappa, M.: Hydrological Ensemble Forecasting in Mesoscale Catchments: Sensitivity to Initial Conditions and Value of Reforecasts, Water Resour. Res., 47, 1–15, doi:10.1029/2010WR009996, 2011. 6860, 6862

Gurtz, J., Baltensweiler, A., and Lang, H.: Spatially distributed hydrotopebased modelling of evapotranspiration and runoff in mountainous basins, Hydrol. Process., 13, 2751–2768, doi:10.1002/(SICI)1099-1085(19991215)13:17<2751::AID-HYP897>3.3.CO;2-F, 1999. 6860, 6862

Heim Jr., R. R.: A Review of Twentieth-Century Drought Indices Used in the United States, Bu. Am. Meteorol. Soc., 83, 1149–1165, doi:10.1175/1520-0477(2002)083<1149:AROTDI>2.3.CO;2, 2002. 6858

Hirschberg, P. A., Abrams, E., Bleistein, A., Bua, W., Monache, L. D., Dulong, T. W., Gaynor, J. E., Glahn, B., Hamill, T. M., Hansen, J. a., Hilderbrand, D. C., Hoffman, R. N., Morrow, B. H., Philips, B., Sokich, J., and Stuart, N.: A Weather and Climate Enterprise Strategic Implementation Plan for Generating and Communicating Forecast Uncertainty Information, B. Am. Meteorol. Soc., 92, 1651–1666, doi:10.1175/BAMS-D-11-00073.1, 2011. 6859

Hisdal, H. and Tallaksen, L. M.: Technical Report to the ARIDE project No. 6 Drought Event Definition, Tech. Rep. 6, University of Oslo, 2000. 6863

Hisdal, H., Stahl, K., Tallaksen, L. M., and Demuth, S.: Have streamflow droughts in Europe become more severe or frequent?, Int. J. Climatol., 21, 317–333, doi:10.1002/joc.619, 2001. 6863

Hwang, Y. and Carbone, G. J.: Ensemble Forecasts of Drought Indices Using a Conditional Residual Resampling Technique, J. Appl. Meteorol. Clim., 48, 1289–1301, doi:10.1175/2009JAMC2071.1, 2009. 6859

Jaun, S. and Ahrens, B.: Evaluation of a probabilistic hydrometeorological forecast system, Hydrol. Earth Syst. Sci., 13, 1031–1043, doi:10.5194/hess-13-1031-2009, 2009. 6862

Kim, T.-W. and Valdes, J. B.: Nonlinear Model for Drought Forecasting Based on a Conjunction of Wavelet Transforms and Neural Networks, J. Hydrol. Eng., 8, 319, doi:10.1061/(ASCE)1084-0699(2003)8:6(319), 2003. 6859

Kirchner, J. W.: Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward, Water Resour. Res., 45, 1–34, doi:10.1029/2008WR006912, 2009. 6866

Li, H., Luo, L., and Wood, E. F.: Seasonal hydrologic predictions of low-flow conditions over eastern USA during the 2007 drought, Atmos. Sci. Lett., 9, 61–66, doi:10.1002/asl.182, 2008. 6859

Lohani, V. K. and Loganathan, G. V.: AN early warning system for drought management using the palmer drought index, J. Ame. Water Resour. As., 33, 1375–1386, doi:10.1111/j.1752-1688.1997.tb03560.x, 1997. 6859

Luo, L. and Wood, E. F.: Monitoring and predicting the 2007 U.S. drought, Geophys. Res. Lett., 34, 1–6, doi:10.1029/2007GL031673, 2007. 6859

Mason, S. J. and Weigel, A. P.: A Generic Forecast Verification Framework for Administrative Purposes, Mon. Weather Rev., 137, 331–349, doi:10.1175/2008MWR2553.1, 2009. 6865

Mishra, A. K. and Desai, V. R.: Drought forecasting using stochastic models, Stoch. Env. Res. Risk A., 19, 326–339, doi:10.1007/s00477-005-0238-4, 2005. 6859

Mishra, A. K. and Desai, V. R.: Drought forecasting using feed-forward recursive neural network, Ecol. Model., 198, 127–138, doi:10.1016/j.ecolmodel.2006.04.017, 2006. 6859

Mishra, A. K. and Singh, V. P.: A review of drought concepts, J. Hydrol., 391, 202–216, doi:10.1016/j.jhydrol.2010.07.012, 2010. 6863

Moreira, E., Coelho, C., Paulo, a., Pereira, L., and Mexia, J.: SPI-based drought category prediction using loglinear models, J. Hydrol., 354, 116–130, doi:10.1016/j.jhydrol.2008.03.002, 2008. 6859

Morid, S., Smakhtin, V., and Bagherzadeh, K.: Drought forecasting using artificial neural networks and time series of drought indices, Int. J. Climatol., 27, 2103–2111, doi:10.1002/joc.1498, 2007. 6859

Murphy, A. H.: The Value of Climatological, Categorical and Probabilistic Forecasts in the Cost-Loss Ratio Situation, Mon. Weather Rev., 105, 803–816, doi:10.1175/1520-0493(1977)105<0803:TVOCCA>2.0.CO;2, 1977. 6864

Nalbantis, I. and Tsakiris, G.: Assessment of Hydrological Drought Revisited, Water Resour. Manag., 23, 881–897, doi:10.1007/s11269-008-9305-1, 2008. 6863

Özger, M., Mishra, A. K., and Singh, V. P.: Long Lead Time Drought Forecasting Using a Wavelet and Fuzzy Logic Combination Model: A Case Study in Texas, J. Hydrometeorol., 13, 284–297, doi:10.1175/JHM-D-10-05007.1, 2012. 6859

Richardson, D. S.: Skill and relative economic value of the ECMWF ensemble prediction system, Q. J. Roy. Meteor. Soc., 126, 649–667, doi:10.1002/qj.49712656313, 2000. 6864

Roulin, E.: Skill and relative economic value of medium-range hydrological ensemble predictions, Hydrol. Earth Syst. Sci., 11, 725–737, doi:10.5194/hess-11-725-2007, 2007. 6864

Schär, C., Vidale, P. L., Lüthi, D., Frei, C., Häberli, C., Liniger, M. A., and Appenzeller, C.: The role of increasing temperature variability in European summer heatwaves., Nature, 427, 332–336, doi:10.1038/nature02300, 2004. 6868

Shukla, S. and Lettenmaier, D. P.: Seasonal hydrologic prediction in the United States: understanding the role of initial hydrologic conditions and seasonal climate forecast skill, Hydrol. Earth Syst. Sci., 15, 3529–3538, doi:10.5194/hess-15-3529-2011, 2011. 6870

Smakhtin, V.: Low flow hydrology: a review, J. Hydrol., 240, 147–186, doi:10.1016/S0022-1694(00)00340-1, 2001. 6863

Steinemann, A. C.: Using Climate Forecasts for Drought Management, J. Appl. Meteorol. Clim., 45, 1353–1361, doi:10.1175/JAM2401.1, 2006. 6859

Tadesse, T., Wilhite, D. A., Hayes, M. J., Harms, S. K., and Goddard, S.: Discovering Associations between Climatic and Oceanic Parameters to Monitor Drought in Nebraska Using Data-Mining Techniques, J. Climate, 18, 1541–1550, doi:10.1175/JCLI3346.1, 2005. 6859

Talagrand, O., Vautard, R., and Strauss, B.: Evaluation of probabilistic prediction systems, in: Proc. ECMWF Workshop on Predictability, 1–25, ECMWF, Reading, Berkshire RG2 9AX, United Kingdom, 1997. 6868

Tallaksen, L. M., Madsen, H., and Clausen, B.: On the definition and modelling of streamflow drought duration and deficit volume, Hydrolog. Sci. J., 42, 15–33, doi:10.1080/02626669709492003, 1997. 6863

van Ogtrop, F. F., Vervoort, R. W., Heller, G. Z., Stasinopoulos, D. M., and Rigby, R. A.: Long-range forecasting of intermittent streamflow, Hydrol. Earth Syst. Sci., 15, 3343–3354, doi:10.5194/hess-15-3343-2011, 2011. 6859

Vitart, F., Bonet, A., Alonso Balmaseda, M., Balsamo, G., Bidlot, J.-R., Buizza, R., Fuentes, M., Hofstadler, A., Molteni, F., and Palmer, T. N.: Merging VarEPS with the monthly forecasting system: a first step towards seamless prediction, ECMWF Newsletter, 115, 35–44, 2008a. 6861

Vitart, F., Buizza, R., Alonso Balmaseda, M., Balsamo, G., Bidlot, J.-R., Bonet, A., Fuentes, M., Hofstadler, A., Molteni, F., and Palmer, T. N.: The new VarEPS-monthly forecasting system: A first step towards seamless prediction, Q. J. Roy. Meteor. Soc., 134, 1789–1799, doi:10.1002/qj.322, 2008b. 6860, 6861

Viviroli, D. and Gurtz, J.: The Hydrological Modelling System Part II, Physical Model Description, Tech. Rep., Institute of Geography, University of Berne, Geographica Bernensia P40, Berne, 2007. 6862

Viviroli, D., Zappa, M., Gurtz, J., and Weingartner, R.: An introduction to the hydrological modelling system PREVAH and its pre- and post-processing-tools, Environ. Model. Softw., 24, 1209–1222, doi:10.1016/j.envsoft.2009.04.001, 2009. 6860, 6862

Webster, P. J., Jian, J., Hopson, T. M., Hoyos, C. D., Agudelo, P. A., Chang, H.-R., Curry, J. A., Grossman, R. L., Palmer, T. N., and Subbiah, A. R.: Extended-Range Probabilistic Forecasts of Ganges and Brahmaputra Floods in Bangladesh, B. Am. Meteor. Soc., 91, 1493–1514, doi:10.1175/2010BAMS2911.1, 2010. 6860

Weigel, A. P. and Mason, S. J.: The Generalized Discrimination Score for Ensemble Forecasts, Mon. Weather Rev., 135, 3069–3074, doi:10.1175/MWR-D-10-05069.1, 2011. 6865

Weigel, A. P., Liniger, M. A., and Appenzeller, C.: The Discrete Brier and Ranked Probability Skill Scores, Mon. Weather Rev., 135, 118, doi:10.1175/MWR3280.1, 2007. 6872

Wood, A. W., Maurer, E. P., Kumar, A., and Lettenmaier, D. P.: Long-range experimental hydrologic forecasting for the eastern United States, J. Geophys. Res., 107, 1–15, doi:10.1029/2001JD000659, 2002. 6859

Yoo, J., Kwon, H.-H., Kim, T.-W., and Ahn, J.-H.: Drought frequency analysis using cluster analysis and bivariate probability distribution, J. Hydrol., 420–421, 102–111, doi:10.1016/j.jhydrol.2011.11.046, 2011. 6863

Zaidman, M. D., Rees, H. G., and Young, A. R.: Spatio-temporal development of streamflow droughts in north-west Europe, Hydrol. Earth Syst. Sci., 6, 733–751, doi:10.5194/hess-6-733-2002, 2002. 6863

Zappa, M. and Kan, C.: Extreme heat and runoff extremes in the Swiss Alps, Nat. Hazards Earth Syst. Sci., 7, 375–389, doi:10.5194/nhess-7-375-2007, 2007. 6862, 6868

Zappa, M., Jaun, S., Germann, U., Walser, A., and Fundel, F.: Superposition of three sources of uncertainties in operational flood forecasting chains, Atmos. Res., 100, 246–262, doi:10.1016/j.atmosres.2010.12.005, 2011. 6870

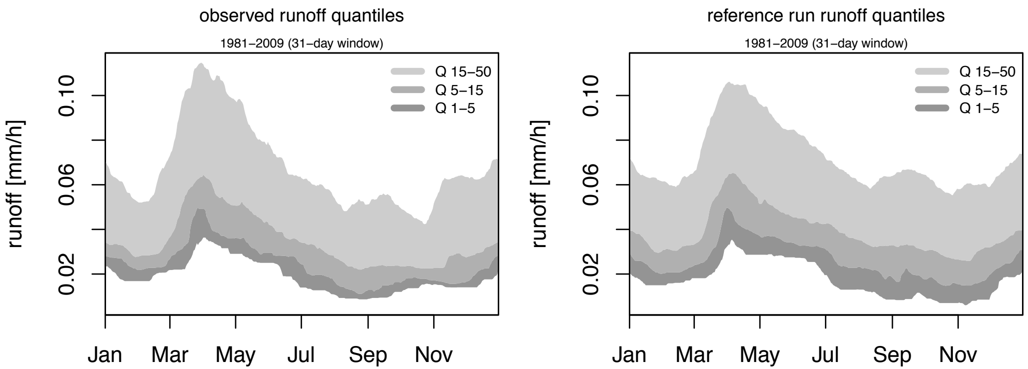

Fig. 1. Lower runoff quantiles for the Thur catchment from gauge measurements (left) and the hydrological reference run (right). The quantiles were calculated for daily mean runoff and non-exceedance probabilities of 1, 5, 15 and 50 %, utilizing a window of 31 days around the day of interest.

HESSD
9, 6857–6887, 2012

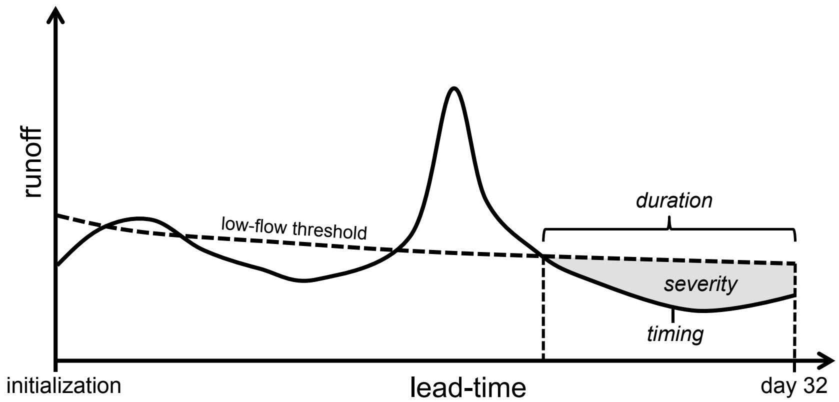

Fig. 2. Illustration of the low-flow indices considered in this study. The solid line is the observed or forecast runoff that is, in certain periods, above or below the low-flow threshold (dashed line). For each forecast member of each 32 day forecast and the corresponding observation, the longest consecutive period below the threshold (low-flow duration) is evaluated. The water deficit during this period (severity, shaded area) is the cumulative difference between the threshold and runoff. The quotient of severity and duration, called magnitude, is evaluated as well. Timing is defined as the moment when half of the event has happened.

HESSD
9, 6857–6887, 2012

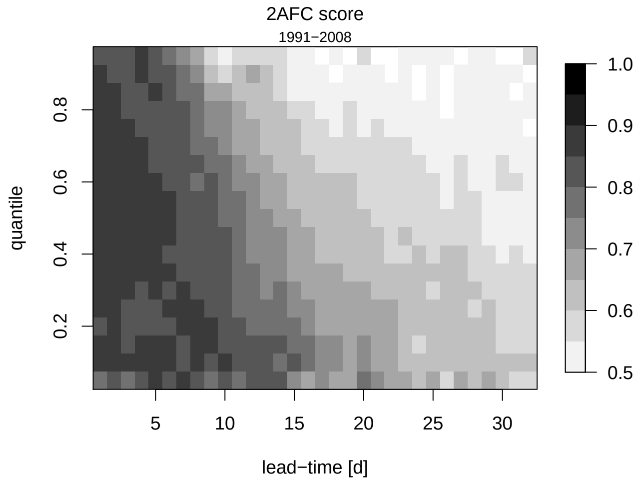

Fig. 3. 2AFC scores of probabilistic VarEPS/PREVAH daily runoff forecasts for different, quantile-based thresholds.

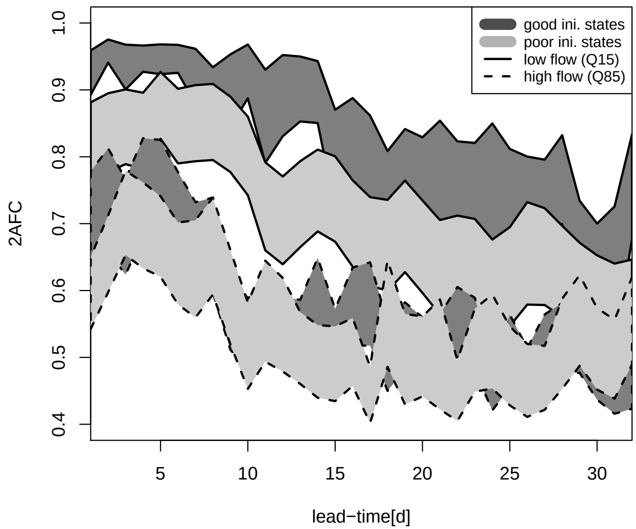

Fig. 4. 2AFC scores of probabilistic VarEPS/PREVAH daily runoff forecasts for different lead-time ranges, different exceedance thresholds and different initial states qualities. The polygons show the 90 % confidence interval, found by resampling 1000 times.

HESSD
9, 6857–6887, 2012

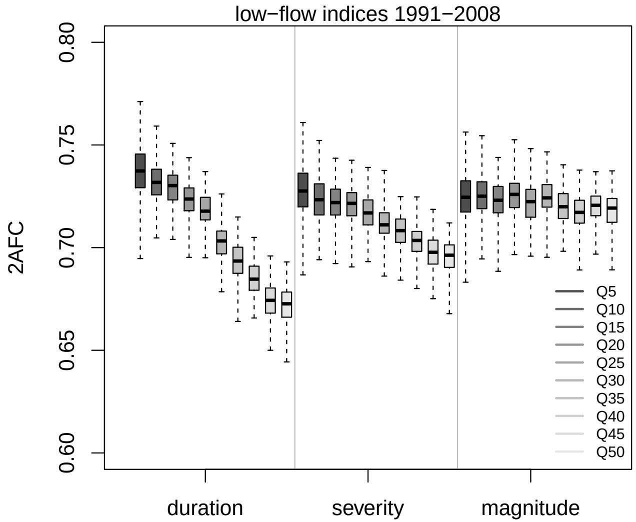

Fig. 5. Overall performance of duration, severity and magnitude forecasts subject to the chosen low-flow threshold for the Thur catchment. The boxes indicate the range, inter-quartile range and median of the 2AFC score, found by resampling the forecast in the verification period 1000 times.

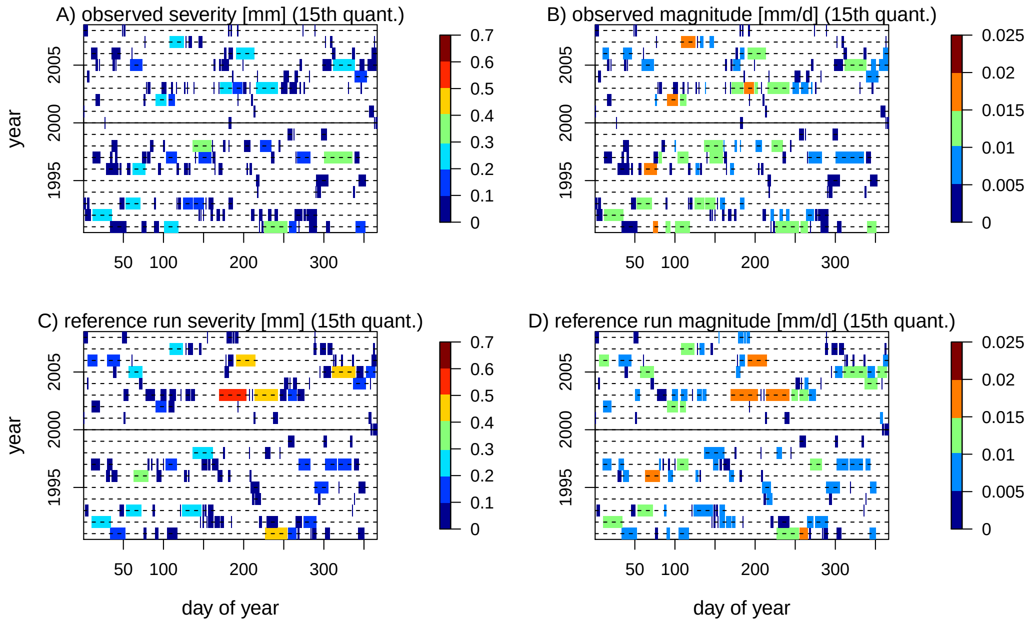

Fig. 6. Periods of low-flow occurrence, with severity (**A** and **C**) and magnitude (**B** and **D**) indicated by colors. A and B are derived from the observed runoff while C and D are derived from the reference run

HESSD
9, 6857–6887, 2012

Predictions of drought parameters

F. Fundel et al.

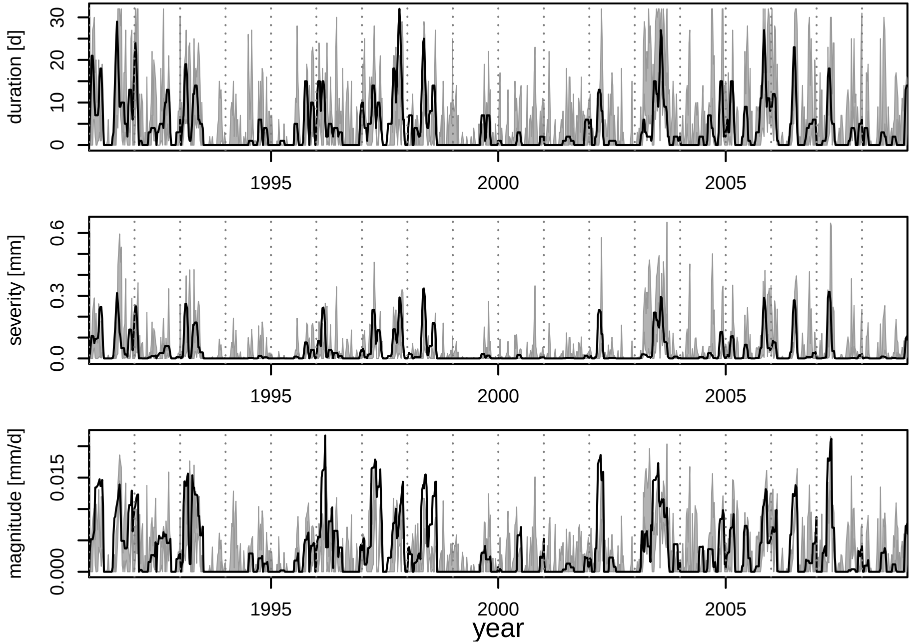

Fig. 7. Weekly issued forecasts of low-flow duration, severity and magnitude from the 32 day lead-time VarEPS ensemble reforecasts (ensemble range shaded in gray), and corresponding observations (black line). The threshold to characterize a low-flow event is the monthly varying 15th quantile.

HESSD
9, 6857–6887, 2012

Predictions of drought parameters

F. Fundel et al.

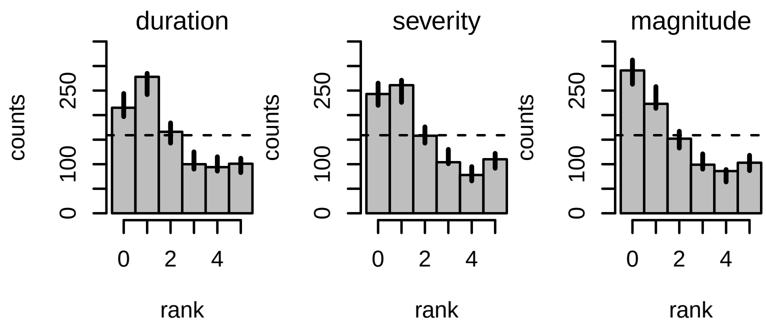

Fig. 8. Rank histograms with 90 % confidence intervals of forecast duration, severity and magnitude, showing the number of times the observation falls within one of the bins spanned by the ensemble. A well-calibrated ensemble prediction system would match the dashed line.

HESSD
9, 6857–6887, 2012

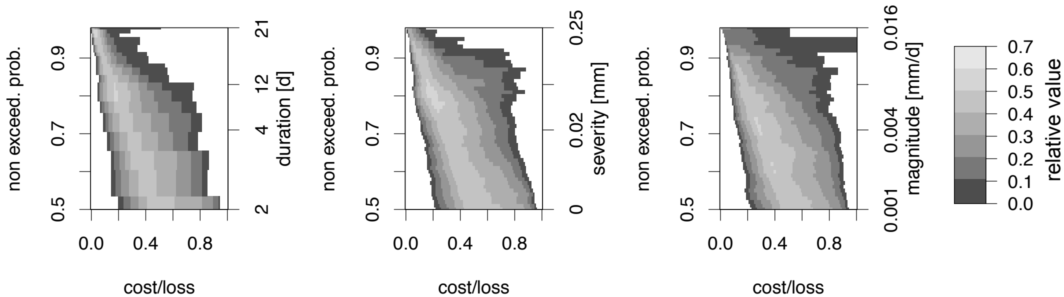

Fig. 9. Relative economic value of Thur low-flow ensemble forecasts of duration, magnitude and severity (columns) using the 15th seasonally varying quantile of runoff as the low-flow detection threshold. The economic value is shaded in gray for forecast users with different cost-loss ratios (x-axis) and for different event thresholds. The left y-axis shows the non exceedance probability and the right y-axis the quantiles for the 50th, 70th, 80th and 95th non exceedance probability.

HESSD
9, 6857–6887, 2012

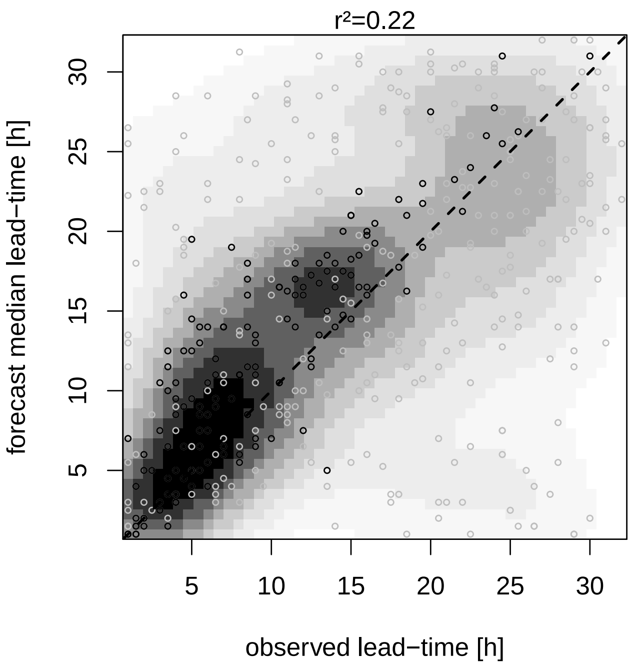

Fig. 10. Observed vs. forecast lead-times (median) of low-flow events. The center between the onset and offset of an event was chosen as the lead-time. The background field is the estimated 2-D density (dimensionless). The black dots mark the events that overlap by at least 1 day with the observed event.

HESSD
9, 6857–6887, 2012