ERROR CORRECTING AND COMBINING MULTI MODEL FLOOD FORECASTING SYSTEMS

Konrad, Bogner[1]
Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Zürcherstrasse 111
CH-8903 Birmensdorf
konrad.bogner@wsl.ch

Katharina, Liechti, Massimiliano, Zappa
Swiss Federal Institute for Forest, Snow and Landscape Research WSL
kaethi.liechti@wsl.ch, massimiliano.zappa@wsl.ch

Phone: +41 (04) 739 2495

KEY WORDS

Flood forecast; Model combination; Verification; Predictive Uncertainties

ABSTRACT

The main objective of a flood forecasting system is the reliable and sharp estimation of the predictive uncertainty, which contains all information available about the forecast variable given the history of observed and predicted values. In order to increase the quality of forecasts post-processing methods have been developed for the removal of bias and dispersion errors and to derive predictive uncertainties. Therefore various statistical methodologies have been tested and applied to hydrological and hydraulic model simulations/forecasts. The possible improvements of these error correction methods have been analysed by the use the Continuous Ranked Probability Score (CRPS). Nowadays a great variety of different Numerical Weather Predictions (NWPs are available with different resolutions and forecast horizons and encompassing deterministic forecasts and Ensemble Prediction Systems (EPSs). A proper way to deal with such a large amount of information is to estimate a total predictive probability density function (pdf) combining all possible future stream-flow values. However, in order to calibrate these multi model forecasts, optimal weights for the combination of the forecasts should be estimated according to their previous forecast quality. Therefore a novel approach called Beta-transformed Linear Pool (BLP) has been tested and analysed at the Sihl river in Zurich (CH), where a flood warning system has been implemented combining meteorological, hydrological and hydraulic model outputs in order to assist decision makers in real-time. First results indicate that the proper combination of forecasts is preferable to the approach of choosing one best model a priori in view of reliability and sharpness of the forecast system.

1. INTRODUCTION

Accurate forecasts of floods are of enormous importance for all kind of decision makers in order to reduce losses, prevent costs and to save lives. Although huge progress has been achieved in this research area within the last decades, the forecasts are prone to a chain of errors ranging from hydro-meteorological measurements to water level predictions. Thus, besides the accuracy of the timing and the peak of the predicted water level, the precise and reliable estimation of the predictive uncertainties is equally crucial. Lorenz outlined already in the beginning of the 60’s, that there are upper limits of predictability caused by the intrinsic chaotic behaviour of atmospheric phenomena (Lorenz, 1963). However, for operational purposes it was not possible to cope with uncertainties in real-time until recently, mainly because of computational restrictions. The most convenient way to deal with these uncertainties is to run a probabilistic weather prediction based on so-called Ensemble Prediction Systems of NWP forecasts (Gneiting, et al., 2005). An ensemble consists of multiple runs of NWP

[1] Corresponding author

This document is the accepted manuscript version of the following article: Bogner, K., Liechti, K., & Zappa, M. (2018). Error correcting and combining multi-model flood forecasting systems. In P. Gourbesville, J. Cunge, & G. Caignaert (Eds.), Springer Water. Advances in hydroinformatics. SimHydro 2017 - choosing the right model in applied hydraulics (pp. 569-578). https://doi.org/10.1007/978-981-10-7218-5_40

models differing in the initial conditions and/or the model formulation with respect to the parameterized numerical representation of the atmosphere, thereby addressing the major sources of the forecast uncertainty.
In the beginning of the 90's the computational power reached a level, which allowed to run ensembles of weather forecasts with the goal to mimic the chaotic behaviour of nature by generating multiple possible traces of forecasts. Therefore different ensemble forecast systems have been developed at various forecast centres worldwide, for example by introducing small perturbations of the initial conditions representing uncertainty of model parameters and measurements. Theoretically these systems permit a statistical analysis of possible future states (see for example Molteni, et al. (1996)), however, practically the limited amount of ensemble members and systematic errors of the NWP's and the coupled hydraulic/hydrological models hinders a direct probabilistic interpretation (Bröcker & Smith, 2008), thus post-processing methods are required in order to derive predictive distributions properly. The herein proposed post-processing includes the error correction of historical model simulations and future predictions and the derivation of predictive uncertainties. Thus the procedures deal only with the updating of the model output without interfering with the meteorological or hydrological or hydraulic model itself. In this study an ensemble consisting of 16 members has been applied and verified in combination with a high resolution deterministic forecast. In order to estimate a predictive probability density function for each lead-time the different prediction systems and the different error correction methods will be optimally combined using a Beta transformation (Gneiting & Ranjan, 2013). The applied methods will be briefly summarized in the next section, the data and the results will be presented in section 3, followed by a discussion and conclusion section.

2. METHODS

One of the simplest error correcting and model-output updating methods, which involves the forecasting of the errors of the hydrological simulations, is based on AutoRegressive (AR) time series models. The mathematical formulation of the AR updating procedure is a special limiting case of the linear Auto-Regressive eXogenous Input ARX model, also known as linear Transfer Function (TF) model (Shamseldin & O'Connor, 1999). That means that the AR approach predicts future differences between the forecast and the observation given previous errors between simulations and observations. In the ARX approach the AR model is fitted to the observations taken the simulations, resp. forecasts as exogenous model input.
Modifications of these models try to analyse and reproduce the underlying processes through decomposition into sub-processes with different time horizons by the use of Wavelet Transformations (WT). The decomposed time-series represent different scales of error and can be incorporated into a Vector ARX model (Zivot & Wang, 2006). An example of such a model called waveVARX for flood forecasting purposes is shown in (Bogner & Pappenberger, 2011). Similar to this decomposition approach knowledge extraction methods based on Neural Networks have been proposed by Jain & Kumar (2009). Other statistical approaches often applied in hydrological forecasting are Neural Networks (see for example Kişi (2007), Rezaeianzadeh, et al. (2013) and Quantile Regression (QR) models (e.g. Koenker (2005), Weerts, et al. (2011)). Recently methods have been proposed for combining QR models with Neural Networks in order to capture possible estimation problems stemming from non-linearities (Cannon, 2011). In this paper various approaches combining WT and Quantile Regression methods based on Neural Networks (Wave-QRNN, or simply QRNN later on) are applied.
Decisions related to uncertain future events need careful balancing out costs and the expected benefit. Therefore decision making requires the quantification of the total uncertainty about a hydrologic predictand (such as river stage, discharge, or run-off volume) in terms of a probability distribution, conditional on all available information and knowledge (Krzysztofowicz, 1999). This means that in order to estimate the expected benefit, it is necessary to assess the probability density of the future occurrence as a measure of the predictive uncertainty (PU). In Todini (2009) this concept of the PU is explained thoroughly and its application in flood forecasting systems is outlined in detail (see also Reggiani & Weerts (2008)).
The QRNN method results in direct estimates of the inverse cumulative density function (i.e. the quantile function), which in turn allows the derivation of the predictive uncertainty (see for example Weerts, et al. (2011), López López, et al. (2014), Dogulu, et al. (2015)), where the application of the QR in order to estimate Predictive Uncertainties (PU's) is outlined. If the number of estimated quantiles τ within the domain $\{0<\tau<1\}$ is sufficiently large the resulting distribution could be considered as continuous. In Quiñonero-Candela, et al. (2006) the cdf, respectively pdf is constructed by combining step-interpolation of probability densities for specified τ-quantiles with exponential lower and upper tails. In this study the pdf is constructed by monotone re-arranging the τ-quantiles and estimating a log-normal distribution to these quantiles for each lead-time Δt.
In the case of the COSMO-LEPS with 16 ensemble members the result will consist of 16 pdfs. Since the members are exchangeable, i.e. there is no memory between consecutive forecasts, these pdfs have to be aggregated into one single pdf. Two different ways of density aggregations have been tested for deriving the

density of the total ensemble. One method is based on averaging the quantiles of the 16 ensemble members directly and the other one is calculated by averaging the probabilities derived from the approximated pdfs similar to the work of Kenneth C. Lichtendahl, et al. (2013), which will be called QRNN-q-averaged., resp. QRNN-p-averaged.
For the ARX based models the PU is estimated for each lead time by assuming that the pdf of the 16 ensemble members could be approximated with a normal distribution applying a Normal Quantile Transformation . Thus the uncertainty stemming from the model and the uncertainty from the forecast can be integrated into the total PU as outlined in the work of Krzysztofowicz & Kelly (2000).
There exists a long list of publications demonstrating the advantages of combining different forecast systems with different complexity of aggregation (Bates & Granger, 2001; Wallis, 2011). However, any non-trivially weighted average of distinct and calibrated probability forecasts will be uncalibrated and lack sharpness (Ranjan & Gneiting, 2010). The aggregation method introduced by Ranjan & Gneiting (2010) and Gneiting & Ranjan (2013) considers the Beta transformed Linear Pool (BLP) for a set of predictive cdfs $F_1, \dots, F_M$ as

$$F(y|\theta) = B_{\alpha,\beta}\left(\sum_{m=1}^{M} \omega_m F_m(y)\right)$$

for $y \in \mathbb{R}$, where $\theta = (\alpha, \beta, \omega)$, $B_{\alpha,\beta}$ denotes the cdf of the standard Beta distribution with parameters $\alpha > 0$ and $\beta > 0$ and density function $b_{\alpha,\beta}(x)$, proportional to $x^{\alpha-1}(1-x)^{\beta-1}$ on the unit interval and ω belongs to the unit simplex in $\mathbb{R}^M$: $\Delta_M = \{\omega = (\omega_1, \dots, \omega_M) \in [0,1]^M : \sum_{m=1}^{M} = 1\}$.
Thus $B_{\alpha,\beta}$ can be interpreted as a parametric calibration function for combining $F_1, \dots, F_M$ with mixture weights $\omega \in \Delta_M$, which assign relative importance to the individual predictive distributions. This BLP approach has been applied now to combine the different forecast systems.
In order to compare and quantify the quality of the different forecast systems and its combination, the Continuous Ranked Probability Score (CRPS) will be used as verification measure of the forecast skill.
The CRPS addresses both important forecast properties, the sharpness and the reliability, and is defined as the integral of the Brier score at all possible threshold values for the continuous predictand (Hersbach, 2000). Thus, it compares the forecast probability distribution with the observation and both are represented as cumulative distribution functions (cdfs). Therefore an ensemble of predictions can be converted into a piecewise constant cdf with jumps at the different ensemble members, and the observation is a Heaviside distribution with a single step from 0 to 1 at the observed value of the variable. In the case of QR models the cdf is derived with quantile estimates. If F is the predictive cdf and y is the verifying observation, the CRPS is defined as:

$$CRPS(F, y) = \int_{-\infty}^{\infty} [F(t) - H(t-y)]^2 dt$$

where $H(t-y)$ denotes the Heaviside function. It is important to note that the CRPS generalizes the Mean Absolute Error (MAE), to which it reduces if F is a point forecast (Gneiting & Raftery, 2007).

3. DATA AND RESULTS

A detailed description of the operational forecast system for the Sihl river at the city of Zurich (CH) can be found in (Addor, et al., 2011) and only a short summary will be given below.
The stream-flow forecasts of the Sihl catchment are driven by the COSMO-Limited-area ensemble prediction system (LEPS, (Montani, et al., 2011)), which is nested into the ensemble prediction system of ECMWF (Molteni, et al., 1996; Buizza, et al., 2007). COSMO stands for the Consortium for Small-scale Modeling (`http://www.cosmo-model.org/`). The Sihl flood forecasting system was supplemented operationally with two deterministic numerical weather predictions versions of the COSMO produced at MeteoSwiss, the COSMO-2 and COSMO-7 (see Tab. 1), however COSMO-2 will not be used in the following analysis because of its limited forecast horizon.

Table 1: Numerical Weather Prediction Systems

System	Spatial Resolution km^2	Forecast horizon Hours	Ensemble Members	Update cycle hours
COSMO-2	2.2 x 2.2	24	-	3
COSMO-7	6.6 x 6.6	72	-	8
COSMO-LEPS	7 x 7	132	16	24

The operational forecast for the Sihl is running hourly, thus a set of 132 parameters for the ARX based and QRNN models should be estimated, i.e. for each hour of the forecast horizon of the COSMO-LEPS. Since we combine the COSMO-LEPS with the COSMO-7 model only the overlapping period of the first 72 hours will be analysed.

To calibrate the post-processing models at the Sihl a period from 2009 to 2014 was available, where the first half was used for estimating the ARX based model and QRNN parameters and the second half was used for validation. To verify the operational forecast system (i.e. the hindcast) itself a period from 2011 to 2015 was analysed.
The model has been running quasi-operational with the COSMO-LEPS forecasts (hindcast) for approximately five years (2011-2015). There is a temporal overlap of the model validation and the hindcast period of four years (2011-2014), however the meteorological data sets are different (observed meteorological data for the validation period, resp. COSMO-LEPS forecast data during the hindcast period), thus the resulting stream-flow series show differences as well. The forecast time resolution is hourly, however the forecasts are updated only once per day, when the new 12:00 o´clock run of the COSMO-LEPS forecast becomes available.

These limited-area atmospheric forecasts are taken as input for the hydrological model. The stream-flows are estimated by the use of the conceptual hydrological model PREVAH (Precipitation-Runoff-EVApotranspiration HRU Model). Originally PREVAH was based on hydrologic response units (HRU), i.e. clusters of raster grids of similar hydrological properties (Gurtz, et al., 1999). This HRU version is used for the Sihl catchment. Because of the elongated shape of the basin, proper flood wave propagation is essential. Therefore PREVAH is coupled with a hydraulic model called FLORIS, a commercial 1-D simulation program developed in the 1990s by the Laboratory of Hydraulics, Hydrology and Glaciology (VAW) of the ETH Zurich. Further information about PREVAH structure, physics, tuneable parameters and tools can be found in Viviroli, et al. (2009). The catchment and model structure of the forecast chain are shown in Fig.1 and 2.

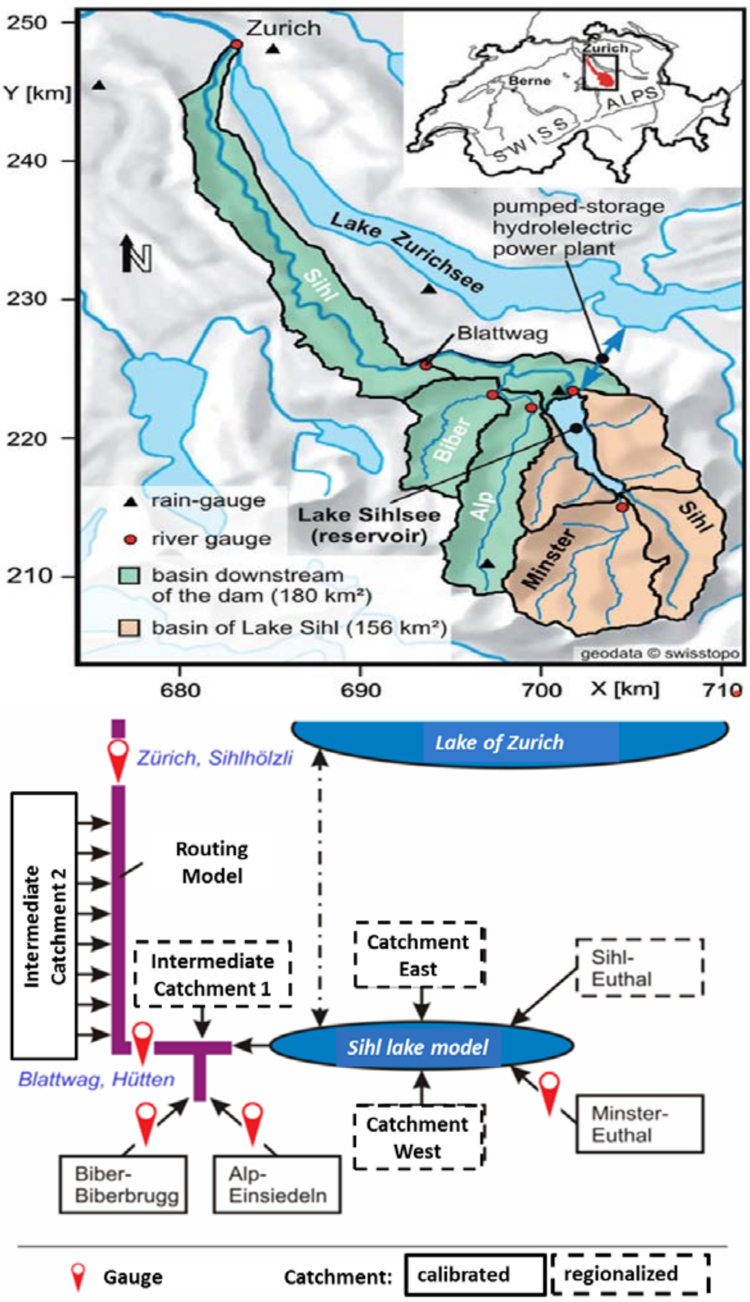

Figure 1: Catchment of the Sihl (top) and its modelling structure (below). The upper part of the catchment coloured in orange (46 %) belong to the accumulation area of the reservoir lake Sihlsee. The remaining catchment area consists of the two tributaries Alp and Biber and the narrow Sihl valley between the lake and Zurich City (source Addor et al. 2011).

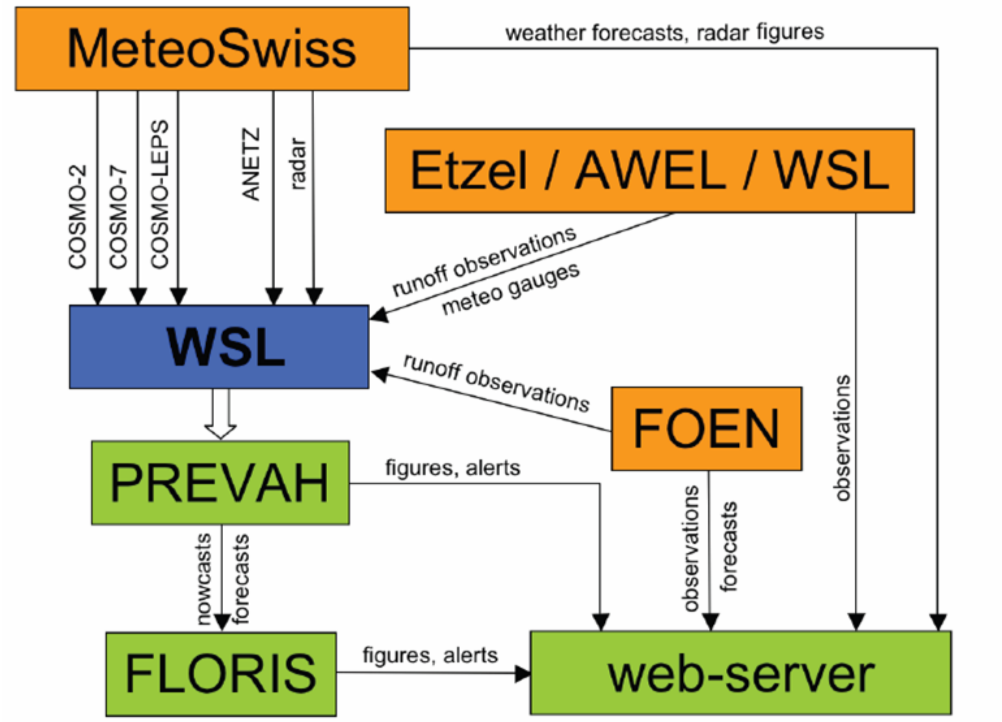

Figure 2: Operational forecast chain

In Fig. 3 the result of the CRPS analysis is shown for the hindcast period. The BLP method has been applied using a sliding window size of 7 days in order to optimize the weighting parameters. The set of predictive cdfs comprises the raw and the post-processed COSMO-LEPS forecasts (i.e. the QRNN-q.averaged, QRNN-p.averaged, waveVARX) and the post-processed COSMO7 forecast (QRNN-C7).

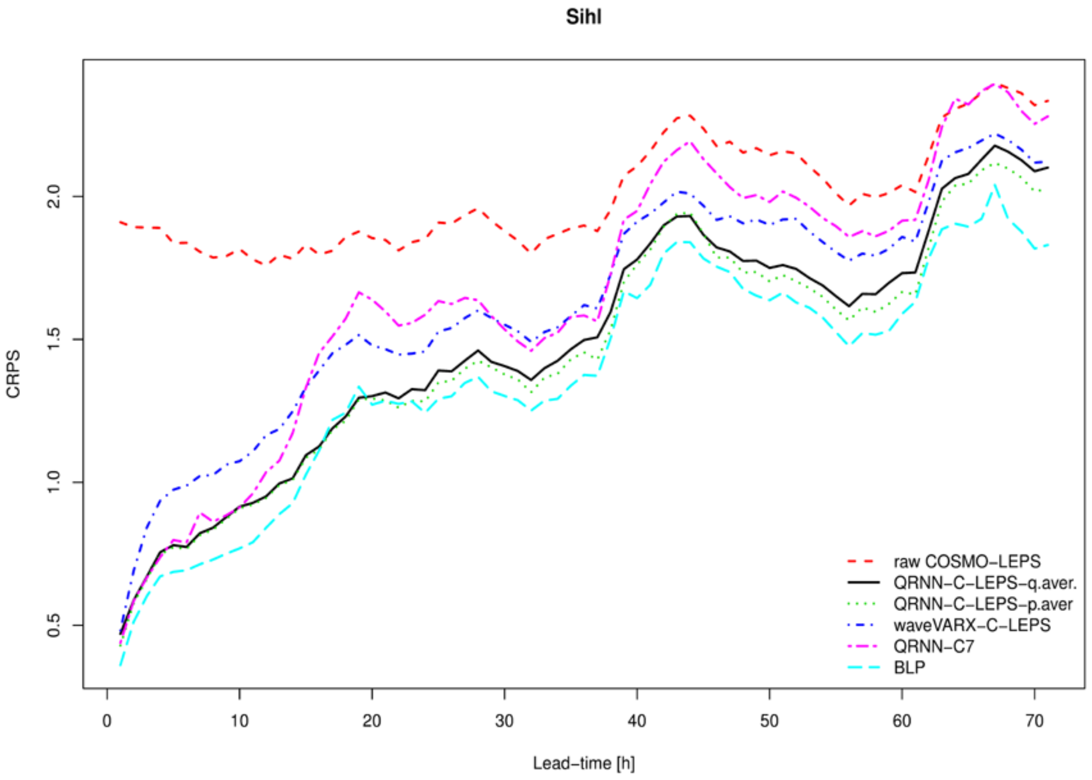

Figure 3: Continuous Ranked Probability Score (CRPS) of the raw COSMO-LEPS (C-LEPS) forecast, post-processed C-LEPS and COSMO7 (C7) and the Beta transformed Linear Pool (BLP) combined forecast using a sliding window of 7 days.

4. DISCUSSION

The analysed forecast chain comprises all kind of uncertainties stemming from the measurements, modelling and forecasting of meteorological, hydrological and hydraulic processes. The modelling uncertainty has been estimated by the QRNN and the waveVARX approach using historical pairs of stream-flow (or water-level) measurements and simulations. This part of the post-processing has been calculated off-line. The forecast uncertainty needs to be estimated in real-time. A first indicator of the forecast uncertainty is given by the spread of the ensemble forecast. However, in the case of having different forecast systems and error correction methods available, a simple aggregation (averaging) will most probably result in an uncalibrated total predictive uncertainty (Gneiting & Ranjan, 2013). Thus the weighting procedure BLP has been tested in order to optimally combine the different available cdfs. All analysis have been based on the results of the CRPS, which comprises the two important quality characteristics, reliability and sharpness, and is negatively oriented (i.e. the lower the better). It is interesting to see in Fig. 3 that the BLP shows the best results for all lead-times. The COSMO-LEPS based forecasts corrected with the QRNN method are slightly worse, followed by the waveVARX method. The QRNN - COSMO7 based forecast shows only for the first ten hours of lead-time good results comparable to the QRNN-COSMO-LEPS forecasts, but decreases significantly afterwards and reaches finally the quality of the uncorrected COSMO-LEPS forecast. This raw ensemble forecast shows the worst results regarding the CRPS for the whole forecast period.

5. CONCLUSION

Contrary to the approach of choosing the right model beforehand this analysis of the forecast chain for the Sihl river at Zurich (CH) demonstrates that the optimal combination of all available models could lead to significant improvements in the forecast quality. The application of the Beta transformed Linear Pool outperformed all single model approaches, even the best ones. Therefore the different post-processing methods based on Quantile Regression Neural Networks and Vector Autoregressive Models in the wavelet domain are combined with the uncorrected forecast. This combination method is very promising and attractive, since it avoids the necessity to choose the best model a priori. Especially in the case of flood forecasting each extreme event will be caused by a different weather situation and the best model to capture this situation will vary between events or even within different lead-times of one single event. Since parameters of the post-processing methods are estimated off-line, the computational costs for the BLP in real-time are negligible. These first results show the importance of forecast combination in order to improve the reliability and sharpness quality of a flood forecasting system in Switzerland.

ACKNOWLEDGMENT

The real-time operational system for the Sihl basin is financed by the Office of Waste, Water, Energy and Air of the Canton of Zurich. This study was conducted in the framework of the Swiss Competence Center for Energy Research - Supply of Electricity (SCCER-SoE) with funding from the Commission for Technology and Innovation CTI (grant 2013.0288). MeteoSwiss is greatly acknowledged for providing all used meteorological data. The Swiss Federal Office for Environment (FOEN) provided the observed discharge data. We would like to thank you also J. Schwanbeck (University of Bern) for providing Figure 1.

REFERENCES AND CITATIONS

Addor, N., Jaun, S., Fundel, F. & Zappa, M. (2011). An operational hydrological ensemble prediction system for the city of Zurich (Switzerland): Skill, case studies and scenarios. *Hydrology and Earth System Sciences*,**15**, 2327-2347.

Bates, J. M. & Granger, C. W. J. (2001). The combination of forecasts. In: *Essays in econometrics.* New York, NY, USA: Cambridge University Press, 391-410.

Bogner, K. & Pappenberger, F. (2011). Multiscale error analysis, correction, and predictive uncertainty estimation in a flood forecasting system. *Water Resources Research*, **47**.

Bröcker, J. & Smith, L. A. (2008). From ensemble forecasts to predictive distribution functions. *Tellus A*, **60**, 663-678.

Buizza, R. et al. (2007). The new ECMWF VAREPS (Variable Resolution Ensemble Prediction System). *Quarterly Journal of the Royal Meteorological Society*, **133**, 681-695.

Cannon, A. J., (2011). Quantile regression neural networks: Implementation in R and application to precipitation downscaling. *Computers \& Geosciences,* **37**, 1277-1284.

Dogulu, N. et al. (2015). Estimation of predictive hydrologic uncertainty using the quantile regression and UNEEC methods and their comparison on contrasting catchments. *Hydrology and Earth System Sciences*, **19**, 3181-3201.

Gneiting, T. & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, **102**, 359-378.

Gneiting, T., Raftery, A. E., Westveld III, A. H. & Goldman, T. (2005). Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Monthly Weather Review*, **133**, 1098-1118.

Gneiting, T. & Ranjan, R. (2013). Combining predictive distributions. *Electron. J. Statist.*, **7**, 1747-1782.

Gurtz, J., Baltensweiler, A. & Lang, H. (1999). Spatially distributed hydrotope-based modelling of evapotranspiration and runoff in mountainous basins. *Hydrological Processes,* **13**, 2751-2768.
Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting,* **15**, 559-570.

Jain, A. & Kumar, S. (2009). Dissection of trained neural network hydrologic models for knowledge extraction. *Water Resources Research,* **45**.

Kenneth C. Lichtendahl, J., Grushka-Cockayne, Y. & Winkler, R. L. (2013). Is It Better to Average Probabilities or Quantiles?. *Management Science,* **59**, 1594-1611.

Kişi, Ö. (2007). Streamflow forecasting using different artificial neural network algorithms. *Journal of Hydrologic Engineering,* **12**, 532-539.

Koenker, R. (2005). *Quantile Regression.,* Cambridge University Press.
Krzysztofowicz, R. (1999). Bayesian Forecasting via Deterministic Model. *Risk Analysis,* **19**, 739-749.

Krzysztofowicz, R. & Kelly, K. S. (2000). Hydrologic uncertainty processor for probabilistic river stage forecasting. *Water Resources Research,* **36**, 3265-3277.

López López, P., Verkade, J. S., Weerts, A. H. & Solomatine, D. P. (2014). Alternative configurations of quantile regression for estimating predictive uncertainty in water level forecasts for the upper Severn River: a comparison. *Hydrology and Earth System Sciences,* Volume 18, 3411-3428.

Lorenz, E. N. (1963). Deterministic Nonperiodic Flow. *J. Atmos. Sci.,* **20**, 130-141.

Molteni, F., Buizza, R., Palmer, T. N. & Petroliagis, T. (1996). The ECMWF Ensemble Prediction System: Methodology and validation. *Quarterly Journal of the Royal Meteorological Society,* **122**, 73-119.

Montani, A., Cesari, D., Marsigli, C. & Paccagnella, T. (2011). Seven years of activity in the field of mesoscale ensemble forecasting by the COSMO-LEPS system: main achievements and open challenges. *Tellus Series A-Dynamic Meteorology and Ocenaography,* **63**, 605-624.

Quiñonero-Candela, J. et al. (2006). Evaluating Predictive Uncertainty Challenge. In: J. Quiñonero-Candela, I. Dagan, B. Magnini & F. dÁlché Buc, eds. *Machine Learning Challenges. Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Tectual Entailment.*, Springer Berlin Heidelberg, 1-27.

Ranjan, R. & Gneiting, T. (2010). Combining probability forecasts. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* **72**, 71-91.

Reggiani, P. & Weerts, A. H. (2008). A Bayesian approach to decision-making under uncertainty: An application to real-time forecasting in the river Rhine. *Journal of Hydrology,* **356**, 56-69.

Rezaeianzadeh, M. et al. (2013). Flood flow forecasting using ANN, ANFIS and regression models. *Neural Computing and Applications,* **25**, 25-37.

Shamseldin, A. Y. & O'Connor, K. M. (1999). A real-time combination method for the outputs of different rainfall-runoff models. *Hydrological Sciences Journal,* **44**, 895-912.

Todini, E. (2009). Predictive uncertainty assessment in real time flood forecasting. In: P. C. Baveye, M. Laba & J. Mysiak, eds. *Uncertainties in Environmental Modelling and Consequences for Policy Making.* Springer Netherlands, 205-228.

Viviroli, D., Zappa, M., Gurtz, J. & Weingartner, R. (2009). An introduction to the hydrological modelling system PREVAH and its pre- and post-processing-tools. *Environmental Modelling and Software,* **24**, 1209-1222.

Wallis, K. F. (2011). Combining forecasts - forty years later. *Applied Financial Economics*, **21**, 33-41.
Weerts, A. H., Winsemius, H. C. & Verkade, J. S. (2011). Estimation of predictive hydrological uncertainty using quantile regression: examples from the National Flood Forecasting System (England and Wales). *Hydrology and Earth System Sciences,* **15**, 255-265.

Xiong, L. & O'Connor, K. M. (2002). Comparison of four updating models for real-time river flow forecasting. *Hydrological Sciences Journal,* **47**, 621-640.

Zivot, E. & Wang, J. (2006). Vector Autoregressive Models for Multivariate Time Series. In: *Modeling Financial Time Series with S-PLUS* © ,Springer New York, 385-429.