Water Resour Manage (2018) 32:969–984
https://doi.org/10.1007/s11269-017-1849-5

Skill of Hydrological Extended Range Forecasts for Water Resources Management in Switzerland

Konrad Bogner[1] · **Katharina Liechti**[1] · **Luzi Bernhard**[1] · **Samuel Monhart**[1] · **Massimiliano Zappa**[1]

Received: 9 January 2017 / Accepted: 25 October 2017 /
Published online: 15 November 2017

Abstract There is a growing need for reliable medium to extended range hydrological forecasts in water and environmental management (e.g. hydro-power and agricultural production). The objective of this paper is a first assessment of the skill of hydrological forecasts based on Numerical Weather Predictions (NWPs) in comparison to the skill of forecasts based on climatology for monthly forecasts with daily resolutions and to identify possibilities of improvement by post-processing the hydrological forecasts. Various hydrological relevant model variables, such as the surface and subsurface runoff and the soil water content, will be analysed for entire Switzerland. The spatially aggregated predictions of these variables are compared to daily simulations and to long-term daily averages of simulations driven by meteorological observations (i.e. climatology). Besides this comparison of forecasts with simulations for model variables without direct measurements available, the skill of the monthly stream-flow forecasts is estimated at four catchments with discharge measurements. Additionally post-processing methods have been applied to remove bias and dispersion errors and to estimate the predictive uncertainty of the stream-flow. Some results of various verification measures like variants of the Geometric Mean for ratios of spatial aggregates and the Continuous Rank Probability Skill Score (CRPSS) will be shown. Apart from the indication of a strong diversity of upper limits of the forecast skill depending on catchment characteristics, the results of NWPs are generally superior to climatological predictions and could be applied gainfully for various kinds of long-term water management planning.

Keywords Ensemble forecasts · Extended-range · Skill score

✉ Konrad Bogner
konrad.bogner@wsl.ch

1 Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland

1 Introduction

The quality of the Numerical Weather Prediction (NWP) systems is steadily increasing and the lead-times for forecasts with reliable accuracy have been moved forward from just a few days to 10 and more days. Also for the extended-range, i.e. forecasts beyond one week up to monthly and seasonal forecast horizons, the predictability has improved significantly (Vitart 2014). However, since the beginning of the 60's it is already known that there are upper limits of predictability, caused by the intrinsic chaotic behaviour of atmospheric phenomena (Lorenz 1963). In the beginning of the 90's, the computational power reached a level which allowed ensembles of forecasts to be run with the goal to mimic the chaotic behaviour of nature by generating multiple possible traces of forecasts (Tracton and Kalnay 1993; Buizza et al. 1999). 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 (Toth and Kalnay 1997; Buizza and Palmer 1995; Houtekamer et al. 1996). This shift from deterministic to probabilistic, ensemble based forecasts helped to extend the predictive skill limit beyond 2 weeks (Buizza and Leutbecher 2015). The main objective and the novelty of this paper is to highlight the advantages of using NWP based ensemble forecasts for total Switzerland for different hydro-meteorological variables and to demonstrate possibilities of improvements at selected catchments. Therefore the monthly ENSemble prediction systems (ENS) from the European Centre of Medium Range Weather Forecasts (ECMWF), which issues 51 forecasts twice a week for the next 32 days, has been applied for predicting possible future water related system states. Ideally such systems permit a statistical analysis of possible future states (see for example et al. 1996), however, the limited amount of ensemble members and systematic errors of the NWP's and the hydrological models hinders a direct probabilistic interpretation (Bröcker and Smith 2008). Thus post-processing methods are required in order to derive predictive distributions properly. Until recently most extended-range forecasts of hydrological variables have been produced by applying statistical methodologies, for example by the use of ARIMA models or neural networks for the generation of monthly mean streamflows (e.g. Noakes et al. 1985; Wang et al. 2009; Yarar 2014; Zhang et al 2015). However, the application of monthly ensemble weather forecasts for the generation of daily hydrological predictions and the exploitation of its probabilistic information is rather limited until now. For example Pattanaik and Das (2015) demonstrated the usefulness of monthly ENS at an Indian catchment analysing one single flood event. In Orth and Seneviratne (2013) monthly reforecasts consisting of 5 ensemble members have been evaluated at 22 near-natural catchments in Switzerland looking at the ensemble mean alone. In the work of Fundel et al. (2013) monthly forecasts have been analysed regarding the low-flow conditions for one specific catchment in Switzerland. In Joerg-Hess et al. (2015) the 5 ensemble members of the monthly reforecasts were evaluated at several catchments in Switzerland regarding the enhancement of the predictability by incorporating daily snow depth measurements. The results of these studies have lead to the development of the Swiss drought information platform http://www.drought.ch, which produces model and measurement based information for drought specific actual and predicted environmental indicators and variables (Zappa et al. 2014). The next step will be the operational incorporation of extended-range forecasts into the drought information platform. Thus, the first time the quality of monthly forecast for entire Switzerland has been analysed in this study in order to highlight the potentially added value. Due to the scarcity of measurements of most of the investigated hydrological variables, spatial aggregates and averages driven by NWP based forecasts will be compared with climatology. This comparison allows the estimation of the upper limits of the monthly predictions, i.e. up to which lead-time horizon the NWP has a gain of information of the driving signal superior to long-term statistics. The NWP and the climatological based forecast will be tested against a reference simulation of hydrological variables driven by meteorological observations. As a verification measure the Geometric Mean Relative Absolute Error (GMRAE), recommended for example by Armstrong and Collopy (1992) and Fildes (1992), will be used with respect to the accuracy of the mean of the ensemble forecast. In order to evaluate the sharpness and the reliability of the ENS in comparison to climatology, the Continuous Rank Probability Skill Score (CRPSS) will be applied (Hersbach 2000; Gneiting and Raftery 2007).

At four selected catchments with available gauging stations the forecast will be compared directly with observations scaled by climatology in order to evaluate the effect of post-processing and possible improvements in applying ensemble weather forecasts rather than deterministic ones (e.g. the mean of the ENS). The four catchments will be verified using the CRPSS. In the next section the different case studies and data sets will be described. In section three and four technical details about the applied methods for calibrating and verifying the forecasts will be given. Finally, after a description of the results and the discussion of its applicability in Switzerland, the conclusion is given.

2 Case Study and Data

In this paper the monthly ENS forecasts from 2012 - 2016 will be used as hydrological model input consisting of 51 members and with a spatial resolution of about 50km. Within that period the ENS has been updated several times (Cycle 38r1 to Cycle 41r2), which may cause inconsistencies in the analysis. However, given the limited amount of data and the impossibility of accessing reforecasts for the 51 member ensemble products, these effects caused by ENS updates have had to be ignored. The chronology of model updates can be found at http://www.ecmwf.int. In order to make the forecasts applicable for hydrological modelling purposes with much higher resolution (e.g. in this case study 200m to 500m), bilinear interpolation and temperature lapse rates are used to adjust this resolution gap (see Addor et al. 2011 for details). Other more sophisticated downscaling processes are under investigation at the moment, however the objective of this study is the evaluation of the hydrological forecast quality using raw and uncorrected ENS forecasts and the quality of the meteorological forecast itself will be evaluated indirectly only.

The analysis of the resulting hydrological forecasts will be carried out for entire Switzerland divided into 307 sub-catchments with sizes of approximately 150 km^2, which are aggregated to 57 regions with an average size of 1000 km^2. The delineation of the regions is according to the model setup used for the Swiss drought information platform and the information provided by the Swiss Federal Office for the Environment (FOEN). Four gauged catchments have been selected with sizes ranging from $\sim$120 to $\sim$1700 km^2 (Table 1; Fig. 1) which are representative for different types of water resources management (hydro-power, agriculture) and are sensitive to drought and floods.

The core of the forecast system is the hydrological model PREVAH (Viviroli et al. 2009), which will be run in reference mode taking meteorological observations as input for predicting past and initial conditions and in forecast mode taking the meteorological ENS forecast as driving forces for predicting future hydrological states. PREVAH is a conceptual, process-oriented hydrological modelling system and has been developed especially for mountainous environments using a HBV-based runoff generation module (Bergström and Forsman 1973). Within this study the fully distributed version of PREVAH is used, which

Table 1 Geomorphologic characteristics (area, altitude, slope) of the Thur, Emme, Broye and Verzasca catchment

Catchment	Thur	Emme	Broye	Verzasca
PREVAH ID	Thu200	EmE200	Bry200	VAG200
No. (see Fig. 1)	1	2	3	4
Area [km^2]	1696	124	392	186
Min. Elev. [m a.s.l.]	356	745	450	530
Max. Elev. [m a.s.l.]	2415	2170	1495	2120
Average Elev. [m a.s.l.]	770	1189	710	1672
Mean Channel Slope [%]	0.7	1.8	0.9	5.6

allows an explicit incorporation of the meteorological forecast grids. The analysis of the quality of the ENS for entire Switzerland will be run with a 500m grid version, whereas the four catchments will be run with 200m resolution. The application at the scale of Switzerland is designed to represent potential of the forecast system for ungauged areas. The setup of the four gauged catchments mimics the one presented in Speich et al. (2015), Orth et al. (2015), Joerg-Hess et al. (2015).

In Fig. 1 at the top, the 307 catchments (black lines) and the aggregated 57 regions (orange delineation) are shown and the grid points of the ENS are given in blue dots.

2.1 Simulated and observed variables

For clarity reasons the analysis will be restricted to the most relevant variables for agricultural management purposes and for hydro-power production (for more details about the variables and their abbreviation please refer to Viviroli et al. 2009):

- Soil Moisture Storage: SSM [mm]
- Total Runoff: RGS, comprises the quick (surface) runoff R0, the delayed R1 and the slow runoff (baseflow) R2 [mm]
- Total Baseflow: R2 [mm]
- Areal catchment precipitation: P [mm]
- Cumulative weekly discharge [m^3]

The four catchments shown in Fig. 1 represent different hydrological regimes with maximum discharges in Winter (Broye, 3) or Spring (Thur and Emme, 1 and 2) on the north side of the Alps and with two peaks (late Spring and Autumn) on the south side (Verzasca, 4). The water of the latter catchment is collected by an artificial dam and is used for hydro-power generation, whereas the management of the selected catchments on the north side is dominated by agricultural purposes. Whilst the Verzasca (Liechti et al. 2013) and the Emme are rather prone to serious flooding, the discharge of the Broye and of the Thur is more sensitive to hydrological drought (Fundel et al. 2013).

3 Error Correction

The comparison between model simulation and forecasts evaluates the forecast quality conditioned on climatology disregarding the errors introduced by the hydrological measurements and models. This approach will be applied for evaluating the forecasts for Switzerland

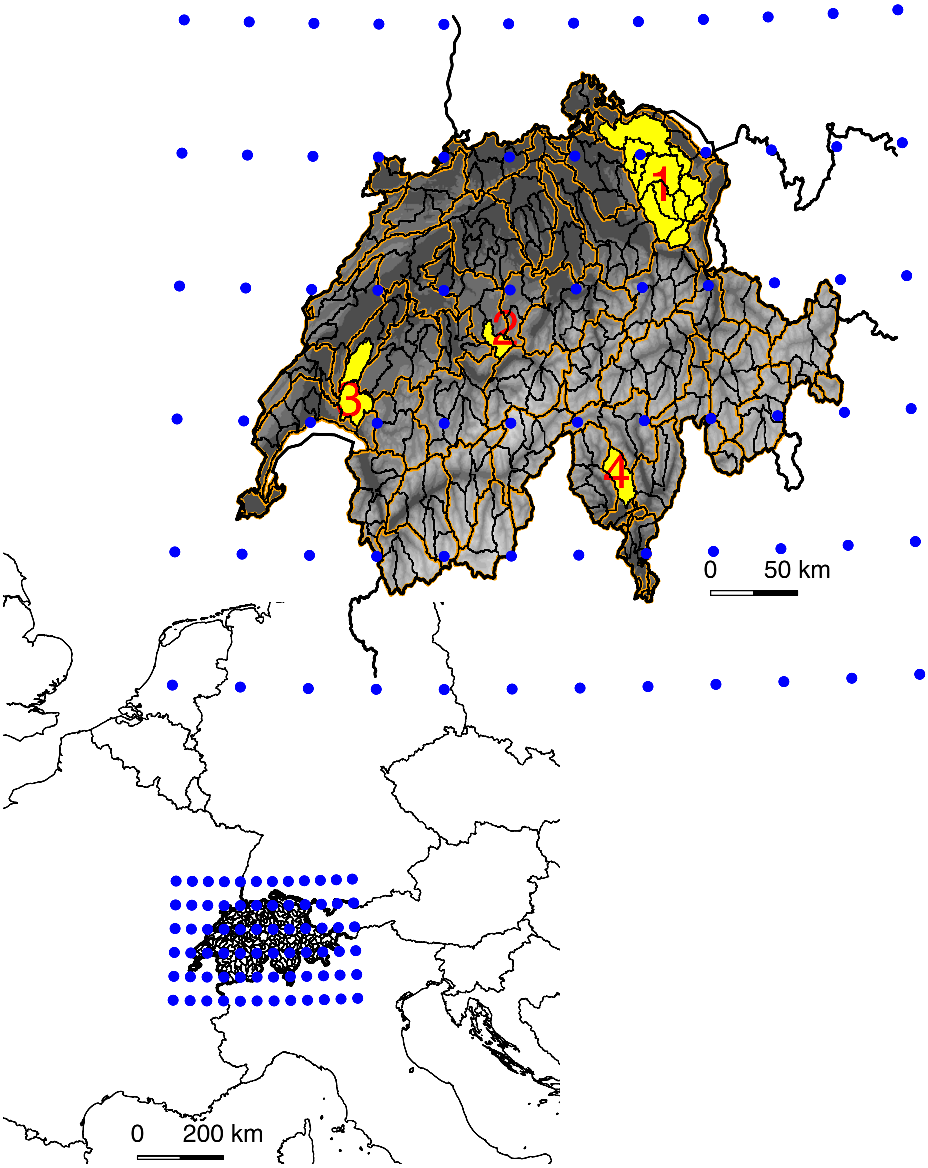

Fig. 1 Switzerland divided into 307 catchments (black lines) and the aggregated 57 Regions (orange border lines) and the four selected gauged catchments (yellow areas). Blue dots represent the grid points of the monthly forecast system ENS from ECMWF covering Switzerland. Swiss GIS elements reproduced with the authorisation of swisstopo (JA100118)

separated in 57 regions without direct measurements. However, at the four catchments with measurements available, the forecasts will be compared with the observed stream-flow values directly. Thus, the resulting difference between the model forecast and the observation comprises all kinds of hydrological and meteorological measurement, modelling and forecast errors. In order to reduce these errors, different post-processing methods have been developed, with an emphasis on ensemble forecasts in recent years (see for example van Andel et al. 2013).

Most often AutoRegressive (AR) models have been fitted in hydrology to series of differences between observations and predictions because of their effectiveness and simplicity. VectorAutoRegressive (VAR) models (for example Gilbert 1995; Zivot and Wang 2006), which have been developed and used in economic research fields for describing the evolution of multiple variables at the same time depending on possibly different lag-times for each variable, build a generalisation of the AR models and could be a promising alternative for hydrological purposes as well. Especially if the problem of scale dependency occurring in hydro-meteorological processes is taken into consideration, where errors manifest at very different temporal resolutions ranging from very short time-scales (minutes or hours) to long lasting processes (days, weeks, months and beyond). This problem could be solved by decomposing the time-series of simulations/forecasts into different levels of resolution applying for example wavelet transformations. In Bogner and Pappenberger (2011) such a 'waveVARX' model has been described in detail, where the decomposed stream-flow observations form the VAR model and the decomposed predictions (simulations and forecasts) compose the exogenous input.

The method of Quantile Regression (QR) developed by Koenker and Bassett (1978) has been applied successfully for hydrological post-processing purposes (e.g. Weerts et al. 2011; López López et al. 2014; Dogulu et al. 2015). Recently a modified QR model called QRNN (Taylor 2000; Cannon 2011), combining QR and Neural Networks, has been tested at some catchments in Switzerland (Bogner et al. 2016) and will be applied in this study as well. Since the results of the QRNN method will be an approximated cdf for each ensemble member, the final overall cdf can be estimated by two different methods, one based on direct quantile averaging and one calculated by averaging the probabilities derived from approximated pdfs similar to the work of Kenneth et al. (2013), which will be called QRNN-q-ave., resp. QRNN-p-ave. (see also Bogner et al. 2016). The predictors in the QRNN model are the different levels of wavelet decomposed simulations/forecasts plus an additional series of past observations available up to the initialisation time. This initialisation time represents the start of predicting future stream-flows given meteorological forecasts, resp. given meteorological measurements in case of model simulation.

It should be noted that in the QRNN model the decomposed simulations/forecasts are the main predictors and a series with past (lagged) observations is given as additional (exogenous) information. In the waveVARX model it is the other way round and the decomposed and lagged observations are the predictors and main drivers of the error correction model and the decomposed simulations/forecasts represent the exogenous model input.

4 Verification

4.1 Deterministic single value verification

For estimating the quality of deterministic and probabilistic forecasts a lot of different verification measures have been developed, especially in the field of atmospheric sciences (see for example Jolliffe and Stephenson 2011) and for hydro- meteorological applications (for example Brown and Seo 2010). The Mean Absolute Percentage Error (MAPE) is one of the most popular measures of the deterministic forecast accuracy and is recommended in most textbooks. The disadvantage of MAPE is, that it yields extremely large percentage errors for values close to zero, resp. infinite MAPEs for zero values.

Variants of the MAPE are the Mean Relative Absolute Error (MRAE) and the Geometric Mean Relative Absolute Error (GMRAE), which belong to the category of scale independent measures. Both involve the division of each error by the error obtained using some benchmark method of forecasting. Because the GMRAE is based on relative errors it is less scale sensitive than the MAPE and allows the comparison of forecasts belonging to different scales.

The Relative Absolute Error RAE and the GMRAE are defined as:

$$\mathrm{RAE}_t = \left|\frac{Y_t - F_t}{Y_t - F_t^*}\right|; \quad \mathrm{GMRAE} = \sqrt{\prod_{t=1}^{N} \mathrm{RAE}_t}, \tag{1}$$

where Y_t and F_t denote the actual and forecast values at time t and F_t^* is a benchmark forecast (e.g. climatology).

The GMRAE will be applied to test the accuracy of the ensemble mean ($\overline{\mathrm{ENS}}$) in comparison to the climatology. Thus, the uncertainty of the forecast system given by the spread of the ensemble is not taken into account, which will be accomplished with the Continuous Ranked Probability Score (CRPS).

4.2 Probabilistic continuous verification

The CRPS addresses two important forecast properties, the sharpness (measured by the spread of the forecast probability density function (pdf)) and the reliability (i.e. the matching of the forecast probabilities and the observed frequencies), and is defined as the integral of the Brier score at all possible threshold values for the continuous predictand (Hersbach 2000; Gneiting and Raftery 2007). The Continuous Rank Probability Skill Score (CRPSS) is a dimensionless indicator of the skill comparing the CRPS of the forecast with the CRPS of a benchmark (see for example Bradley and Schwartz 2011). If the benchmark is a deterministic forecast, e.g. based on climatology, the CRPS reduces to the Mean Absolute Error (MAE) and the CRPSS will be defined as:

$$\mathrm{CRPSS}(F, F^*, y) = 1 - \frac{\mathrm{CRPS}(F, y)}{\mathrm{MAE}(F^*, y)} \tag{2}$$

The CRPSS ranges between 1 (for perfect predictions) to $-\infty$, however only values > 0 indicate positive skill. These measures will be used for the analysis of the ENS forecast system and of the predictive densities derived with error correction models (see also Bogner et al. 2016).

5 Results

This section will be separated into results regarding the spatial aggregates of the 57 regions and the four selected catchments.

5.1 Regional analysis

For the regional analysis the hydrologically relevant variables, namely the areal precipitation (P), total runoff (RGS), baseflow (R2) and soil moisture storage (SSM) have been aggregated to 57 regions and the raw ensemble forecasts have been compared to climatology. At first the GMRAE has been estimated for each lead-time (from 1 to 32 days) separately looking at the mean of ENS ($\overline{\text{ENS}}$) only. In Fig. 2 the four variables are shown for the lead-times of 7, 14, 21 days (three columns). The lower (brighter) the GMRAE, the better is the skill of the forecast in comparison to climatology. If the GMRAE is ≥ 1 (dark orange), the forecast has the same or less skill as climatology, that means the NWP based forecast is getting worthless. In Fig. 3 the GMRAE of all regions are pooled for each lead-time from 1 to 32 days. The blue boxes represent the interquartile range of all available catchments, the median is given as red bar and the black vertical lines show the lower and upper range (values $\leq 25\%$, resp. $\geq 75\%$). The green line indicates the percentage of all 307 sub-catchments, which have a GMRAE less than 1.

Besides the GMRAE, the CRPSS has been used for verifying the regional forecasts as well. Therefore instead of observations, which are not available for these regional averaged catchments, the model simulation is taken as y in Eq. 2 using meteorological observations as model input. In order to estimate the gain of informations using ENS, the climatology is taken as a benchmark. Thus, a CRPSS greater than zero indicates an improvement over the climatology. The results are shown in Fig. 4, for the same variables and lead-times and with the same colour scheme as in the GRMAE analysis (the brighter the better). In Fig. 5 the CRPSS is shown for all model variables, whereas in Fig. 6 the CRPS and the MAE for

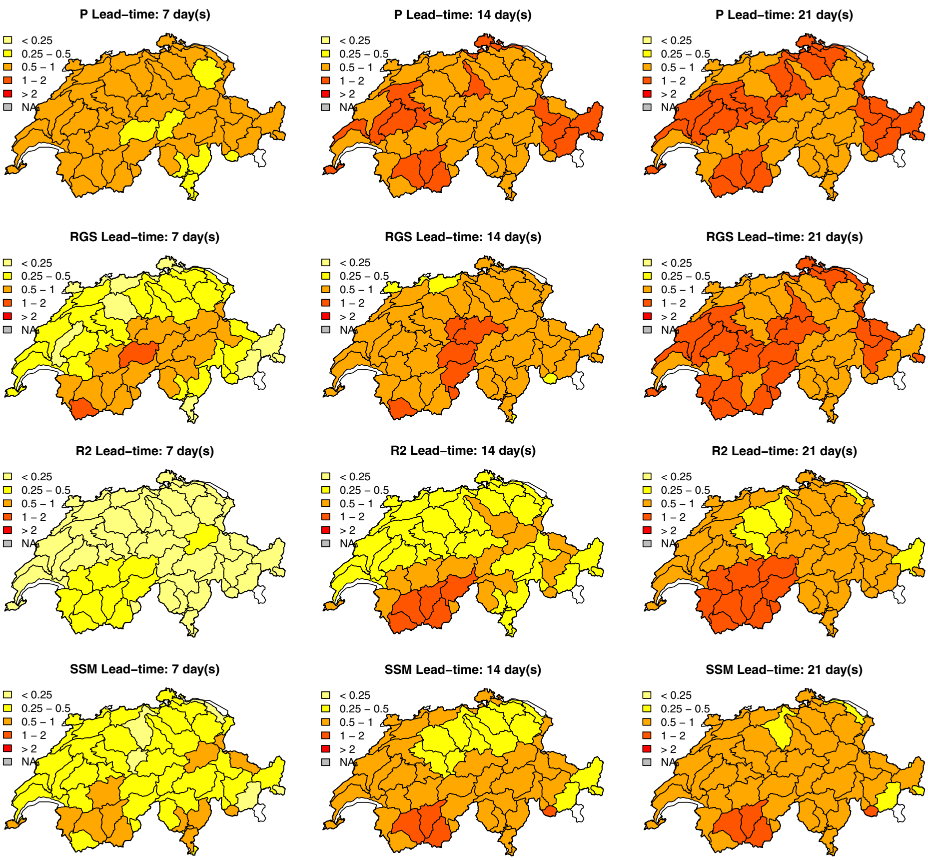

Fig. 2 Spatial GMRAE of the areal catchment precipitation P, total runoff RGS, total baseflow R2, soil moisture storage SSM (rows from top to the bottom) for lead-times of 7, 14 and 21 days (left, middle, right)

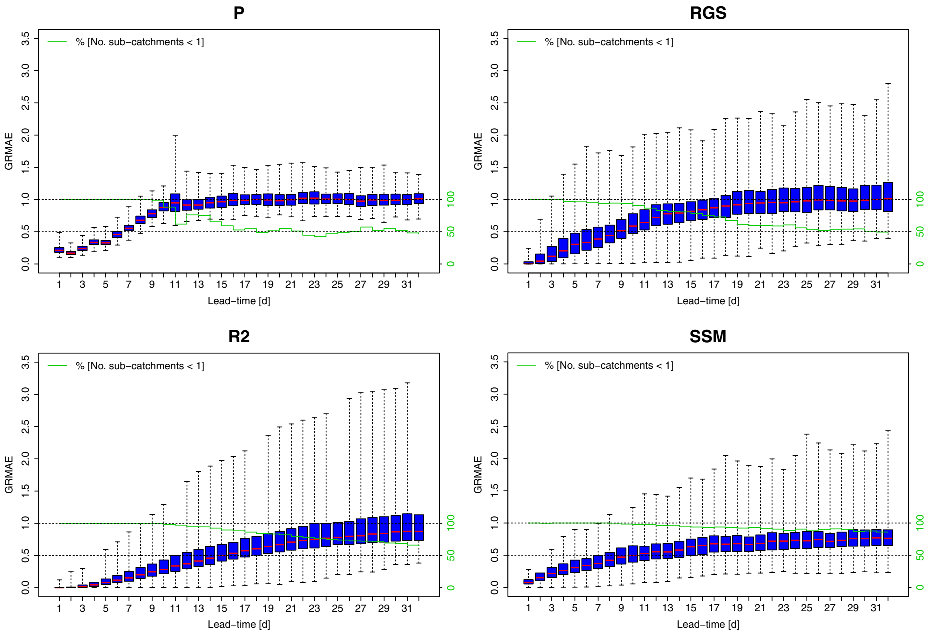

Fig. 3 GMRAE of all 307 sub-catchments pooled together and represented as boxes for lead-times from 1 to 32 days for the areal catchment precipitation P, total runoff RGS, total baseflow R2, soil moisture storage SSM. The green line indicates the percentage of all sub-catchments, which have a GMRAE less than 1

the precipitation P and runoff RGS are shown for all regions pooled together for lead-times from 1 to 32 days.

5.2 Selected catchments

The difference in the analysis of the four selected catchment (Broye, Emme, Thur, Verzasca) to the regional analysis is that the verifying variable is the observed stream-flow (in-flow) measured at gauging stations at the outlet of the catchment. Therefore the result outlines not only the quality of the forecast itself, but comprises the overall predictive quality including the hydrological model and measurement uncertainty. Thus, in case of probabilistic forecasts the total predictive uncertainty of the forecast system can be inferred. Additionally the availability of stream-flow measurements allows the application of post-processing methods in order to minimise the error between observations and model simulations, resp. forecasts. Thus, the following results compare the behaviour of the raw ensemble and post-processed forecasts with respect to climatology. In Fig. 7 the CRPSS is shown for each of the four catchments and for all lead-times.

6 Discussion

6.1 Regional analysis

For the sake of clarity only the results for lead-times of 7, 14 and 21 days are shown in the Figs. 2 and 4, but these selected examples illustrate quite well the spatial development

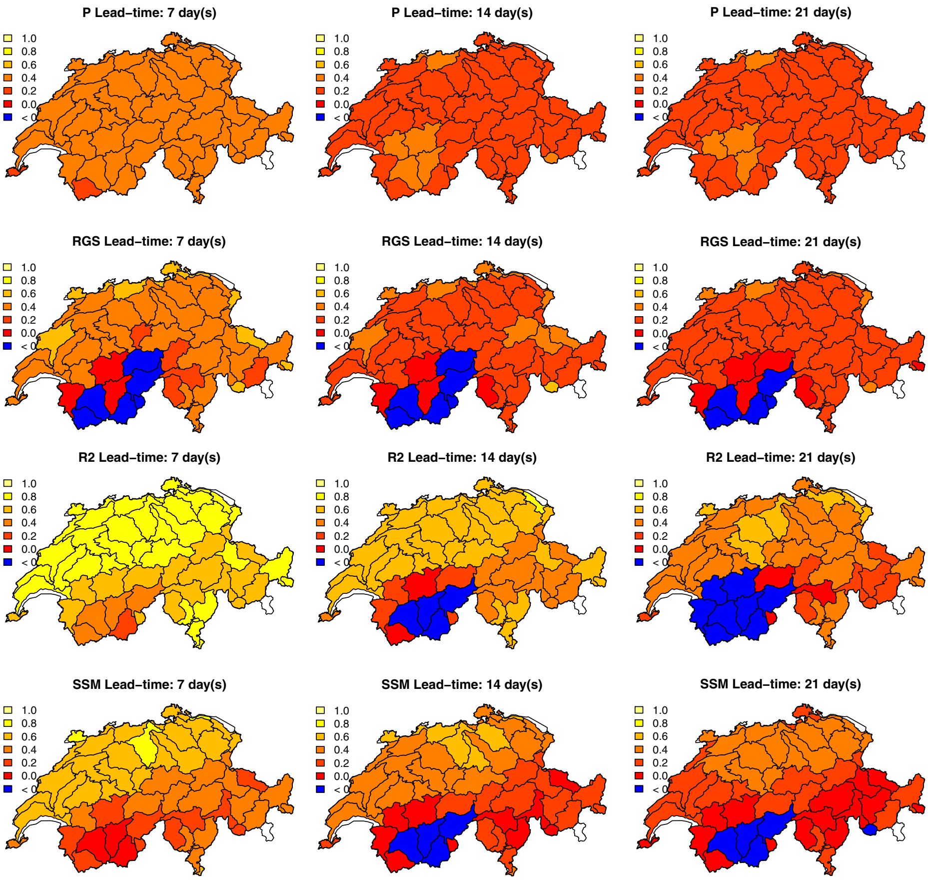

Fig. 4 Spatial CRPSS of Precipitation P, RGS, R2, SSM (rows from top to the bottom) for lead-times of 7, 14 and 21 days (left, middle, right)

of the GMRAE and CRPSS. The general temporal evolution of the forecast quality for the forecast horizon from 1 to 32 days is shown in the Figs. 3 and 5.

In the first row on the top of Fig. 2 the GMRAE of the average catchment precipitation P used to force PREVAH is shown for the 57 regions. At the lead-time of 7 days all regions show gain in the forecast skill of the $\overline{\text{ENS}}$ over climatology. At a lead-time of 14 days still the majority of the regions show the superiority of the $\overline{\text{ENS}}$ forecast, and only at a lead-time of 21 days approximately half of the regions have GMRAE values greater than 1. The total runoff (RGS, second row in Fig. 2) shows at the lead-time of 7 days for the majority of the regions a significant improvement in the skill of the $\overline{\text{ENS}}$ forecasts (GMRAE values less than 0.5). However, central Switzerland with mostly steep, alpine catchments and fast response times shows GMRAE values between 0.5 and 1. Lead-times of 14 and 21 days show a similar pattern to the precipitation results. The variables baseflow R2 and soil moisture storage SSM describing the sub-surface processes (row three and row four in Fig. 2) show a physically reasonable delay of almost one week in the GMRAE, i.e. the pattern of R2 and SSM at lead-time 14 is comparable to lead-time 7 of RGS, and the pattern of

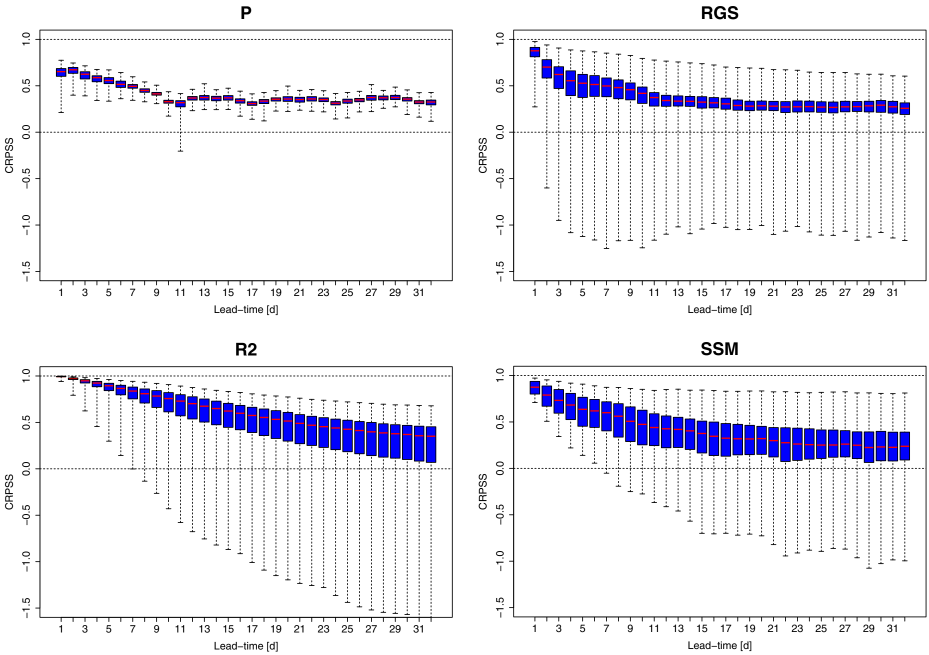

Fig. 5 CRPSS of all 307 sub-catchments pooled together and represented as boxes for lead-times from 1 to 32 days for the areal catchment precipitation P, total runoff RGS, total baseflow R2, soil moisture storage SSM

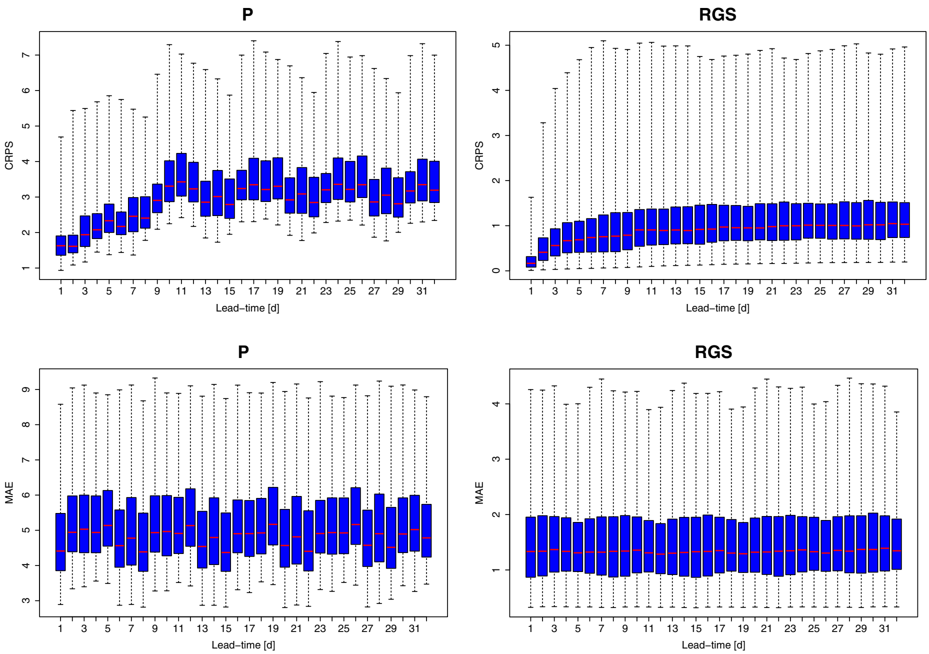

Fig. 6 CRPS (top) and MAE (bottom) of all 307 sub-catchments pooled together and represented as boxes for lead-times from 1 to 32 days for the areal catchment precipitation P and runoff RGS

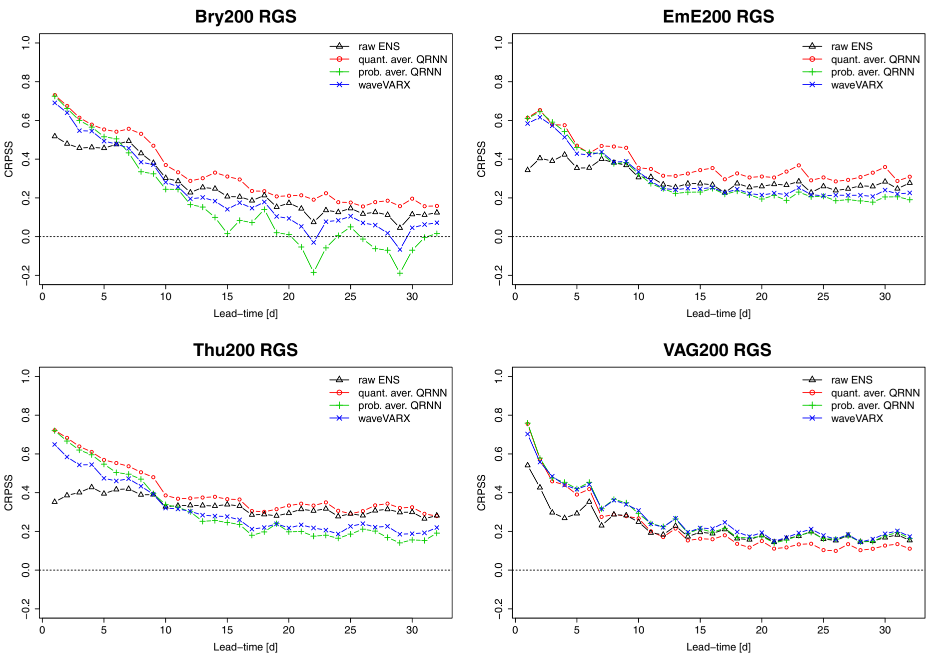

Fig. 7 CRPSS of daily stream-flow series $[m^3/s]$ for the Broye (Bry200), Emme (EmE200), Thur (Thu200) and the Verzasca (VAG200) catchments. Each panel compares the raw Ensemble (without post-processing, black line) with the QRNN method (quantile averaging in red; probability averaging in green) and the waveVARX method (blue line)

R2 and SSM at lead-time of 21 days looks similar to day 14 of RGS. In Fig. 3 the lead-time dependency of the GMRAE is shown for all regions pooled together. Obviously the skill of the precipitation P (top left) converges to 1 within the first 10 days and at a lead-time of about 16 days the median of all regional forecasts is approximately equal to 1. It is interesting to see that for lead-times from 17 onwards, this relation keeps constant with less than 50% of the sub-catchments having a GMRAE greater than 1. In general these temporal analysis highlight the dependency of the information gain on the reaction time of the investigated model variable. The slower the hydrological variable is reacting (i.e. the more persistent the water stays in the system), the longer the skill of the NWP forecast will be superior to climatology. This similarity between memory and predictability has also been found by Orth and Seneviratne (2013).

The results of the CRPSS (Figs. 4 and 5) underpin these outcomes from the GMRAE analysis, although the methodologies are quite different. The GMRAE looks at the $\overline{ENS}$ alone and in such a way it represents a deterministic forecast verification method, whereas the CRPSS takes the whole ensemble under consideration and thus it verifies a probabilistic forecast. Both methods stress some problems of forecasting the sub-surface process variables R2 and SSM (third and fourth row in Fig. 4) at longer lead-times (beyond one week) in the south-western part of Switzerland, which are high alpine regions with thin and highly variable soil layers. Also the runoff RGS (second row) itself shows some deficits in these areas, which are most probably caused by runoffs driven by glacier melting processes. However, the skill of the precipitation (first row) remains significantly positive (CRPSS about 0.4) for that region for lead-times of 14 and 21 days. The reason for that phenomena could be the long lasting dry spells, which are predominant in such inner alpine regions. It is interesting to see that the analysis of the GMRAE looking at the $\overline{\text{ENS}}$ alone does not reveal this characteristic feature.

Similar to the results of the GMRAE (Fig. 3), the results in Fig. 5 of the CRPSS highlight the dependency of the information gain on the reaction time of the investigated model variable. However, it is also interesting to see how the median of the GMRAE converges to 1, whereas the median of the CRPSS converges to values greater than 0.2. This result stresses the importance of using the whole ENS (as for the CRPSS) instead of using the mean of the ENS (as in case of the GMRAE). Only when all 51 members of the ENS are used, the superiority of the application of forecasts based on NWPs over climatology for the whole range of lead-times will be revealed. The explanation why the CRPSS of the precipitation P shows such a reduction in the interquartile range is given in Eq. 2. Since the variability of the MAE between the model simulation and the climatology will be quite large between different catchments in comparison to the CRPS based on the ENS and the model simulation (see Fig. 6), the CRPSS will become small. The MAE itself stays more or less constant, since the quality of the model simulation is calculated with observed meteorological data, thus it is independent from the forecast lead-time. A similar, but not that prominent, effect of dumped variability is occurring for the runoff at lead-times greater than 10 days.

Both results, GMRAE and CRPSS, of the regional analysis are quite promising for the application of $\overline{\text{ENS}}$ and ENS in the case of agricultural management purposes, since they will benefit particularly from forecasts of water storage in the soil with longer lead-times, which are given by the variables R2 and SSM.

6.2 Selected catchments

The CRPSS for the four catchment indicates similar improvements regarding climatology (CRPSS values greater than zero), but with a different favourable post-processing methods. Only for the Broye catchment is the CRPSS of the QRNN-probability averaging method below 0 for lead-times greater than 20 days. For the Broye, Emme and Thur catchment the QRNN quantile averaging method is superior to the other post-processing methods for all lead-times. At the Verzasca catchment the probability averaging method and waveVARX method show both equally better results. However, at the Thur and the Verzasca catchment the improvements due to post-processing degrades to the uncorrected raw ENS after a lead-time of about 15 days. Nonetheless, it should be stressed that any slight improvement in the forecast skill could manifest in significant economical gains, especially in the case of flood forecasting or decisions concerning management of hydro-power.

All the tested post-processing methods calibrate a correction approach to the past errors between model simulations and observations at gauging stations and apply the resulting parameter fits to the forecasts treating these predictions as future simulations. Thus, the gained improvement will decrease with increased lead-time, when the error caused by the meteorological forecasts will more and more predominate the hydrological model error.

7 Conclusions

This study analyses the gain of applying NWP based extended-range forecasts. The spatial and temporal aggregation of hydrological relevant variables clearly demonstrates the benefits of NWP based forecasts and their advantage over climatology for lead-times up to 32 days. The analysis of the geometric mean relative absolute error (GMRAE) shows that these findings are valid for the mean of the ensemble forecast of sub-surface processes, whereas these gains in skill disappear for precipitation and runoff forecasts after two weeks. A more detailed evaluation of all 51 ensemble members based on the continuous ranked probability score (CRPS) allows the verification of the reliability and the sharpness, two important properties of probabilistic forecasts. The results of this skill score highlight the necessity to look at the whole ensemble in order to draw exhaustive conclusions about the forecast quality. In general both verification measures outline the profit of NWPs for water management purposes especially in the field of agriculture, where sub-surface water flows and storage are highly relevant. These processes show improved forecast skills in comparison to climatology for the whole forecast period because of the persistence and memory effects of the water in the soil. But also the management of hydro-power plants could gain from extended-range stream-flow forecasts allowing a greater flexibility in decisions regarding the regulation of reservoir in- and out-flows and residual flows. Hence, public information systems like the Swiss drought information platform www.drought.ch will definitely benefit from implementing extended-range forecasts. Apparently there are limitations in deriving skilful flood forecasts in order to answer questions about exact timings and precise magnitudes of peak flows at gauging stations for lead-time horizons greater than two weeks. Also the benefits of the applied error correction methods will vanish after such long forecast horizons. Further analysis of possible improvements applying sophisticated post-processing methods to the meteorological forecasts directly, before they are fed into the hydrological forecast system, are under investigation.

Acknowledgements The application www.drought.ch is a product of the DROUGHT-CH project financed by Swiss National Research Programme on Sustainable Water Management (NRP 61). The operational demonstration of www.drought.ch has been financed by the Swiss Federal Office for Environment and supported by WSL and MeteoSwiss. Konrad Bogner's contribution is part of the Swiss Competence Centre for Energy Research - Supply of Electricity (SCCER-SoE) and is funded by the Commission for Technology and Innovation (CTI). Samuel Monhart's contribution is financed by the NRP 70 - Energy Turnaround project (Grant No. 407040_153929). The authors would like to thank especially Vanessa Round for proofreading.

References

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. Hydrol Earth Syst Sci 15(7):2327–2347

Armstrong J, Collopy F (1992) Error measures for generalizing about forecasting methods: Empirical comparisons. Int J Forecast 8(1):69–80

Bergström S, Forsman A (1973) Development of a conceptual deterministic rainfall-runoff model. Hydrol Res 4(3):147–170

Bogner K, Pappenberger F (2011) Multiscale error analysis, correction, and predictive uncertainty estimation in a flood forecasting system. Water Resour Res 47(7):W07524. https://doi.org/10.1029/2010WR009137

Bogner K, Liechti K, Zappa M (2016) Post-processing of stream flows in switzerland with an emphasis on low flows and floods. Water 8(4):115. https://doi.org/10.3390/w8040115

Bradley A, Schwartz S (2011) Summary verification measures and their interpretation for ensemble forecasts. Mon Weather Rev 139(9):3075–3089

Bröcker J, Smith LA (2008) From ensemble forecasts to predictive distribution functions. Tellus A 60(4):663–678. https://doi.org/10.1111/j.1600-0870.2008.00333.x

Brown J, Seo DJ (2010) A nonparametric postprocessor for bias correction of hydrometeorological and hydrologic ensemble forecasts. J Hydrometeorol 11(3):642–665. https://doi.org/10.1175/2009JHM1188.1

Buizza R, Leutbecher M (2015) The forecast skill horizon. Q J R Meteorol Soc 141(693):3366–3382. https://doi.org/10.1002/qj.2619

Buizza R, Palmer T (1995) The singular-vector structure of the atmospheric global circulation. J Atmos Sci 52(9):1434–1456

Buizza R, Milleer M, Palmer TN (1999) Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. Q J R Meteorol Soc 125(560):2887–2908

Cannon AJ (2011) Quantile regression neural networks: Implementation in R and application to precipitation downscaling. Comput Geosci 37(9):1277–1284. https://doi.org/10.1016/j.cageo.2010.07.005

Dogulu N, López López P, Solomatine DP, Weerts AH, Shrestha DL (2015) Estimation of predictive hydrologic uncertainty using the quantile regression and uneec methods and their comparison on contrasting catchments. Hydrol Earth Syst Sci 19(7):3181–3201. https://doi.org/10.5194/hess-19-3181-2015

Fildes R (1992) The evaluation of extrapolative forecasting methods. Int J Forecast 8(1):81–98

Fundel F, Jörg-Hess S, Zappa M (2013) Monthly hydrometeorological ensemble prediction of streamflow droughts and corresponding drought indices. Hydrol Earth Syst Sci 17(1):395–407

Gilbert P (1995) Combining VAR estimation and state space model reduction for simple good predictions. J Forecast Spec Issue VAR Modell 14:229–250

Gneiting T, Raftery A (2007) Strictly proper scoring rules, prediction, and estimation. J Amer Stat Assoc 102(477):359–378

Hersbach H (2000) Decomposition of the continuous ranked probability score for ensemble prediction systems. Weather Forecast 15(5):559–570

Houtekamer P, Lefaivre L, Derome J, Ritchie H, Mitchell H (1996) A system simulation approach to ensemble prediction. Mon Weather Rev 124(6):1225–1242

Joerg-Hess S, Griessinger N, Zappa M (2015) Probabilistic forecasts of snow water equivalent and runoff in mountainous areas. J Hydrometeorol 16(5):2169–2186. https://doi.org/10.1175/JHM-D-14-0193.1

Jolliffe I, Stephenson D (2011) Forecast verification: A practitioner's guide in atmospheric science. Wiley, New Jersey

Kenneth C, Lichtendahl J, Grushka-Cockayne Y, Winkler RL (2013) Is it better to average probabilities or quantiles? Manag Sci 59(7):1594–1611. https://doi.org/10.1287/mnsc.1120.1667

Koenker R, Bassett Jr G (1978) Regression quantiles. Econometrica: Journal of the Econometric Society 46(1):33–50

Liechti K, Zappa M, Fundel F, Germann U (2013) Probabilistic evaluation of ensemble discharge nowcasts in two nested alpine basins prone to flash floods. Hydrol Process 27(1):5–17. https://doi.org/10.1002/hyp.9458

López López P, Verkade JS, Weerts AH, Solomatine DP (2014) Alternative configurations of quantile regression for estimating predictive uncertainty in water level forecasts for the upper severn river: a comparison. Hydrol Earth Syst Sci 18(9):3411–3428. https://doi.org/10.5194/hess-18-3411-2014

Lorenz EN (1963) Deterministic Nonperiodic Flow. J Atmos Sci 20(2):130–141

Molteni F, Buizza R, Palmer TN, Petroliagis T (1996) The ecmwf ensemble prediction system: Methodology and validation. Q J R Meteorol Soc 122(529):73–119

Noakes DJ, McLeod AI, Hipel KW (1985) Forecasting monthly riverflow time series. Int J Forecast 1(2):179–190

Orth R, Seneviratne SI (2013) Predictability of soil moisture and streamflow on subseasonal timescales: A case study. J Geophys Res Atmos 118(19):10,963–10,979

Orth R, Staudinger M, Seneviratne SI, Seibert J, Zappa M (2015) Does model performance improve with complexity? A case study with three hydrological models. J Hydrol 523:147–159. https://doi.org/10.1016/j.jhydrol.2015.01.044

Pattanaik DR, Das AK (2015) Prospect of application of extended range forecast in water resource management: a case study over the mahanadi river basin. Nat Hazards 77(2):575–595

Speich MJR, Bernhard L, Teuling AJ, Zappa M (2015) Application of bivariate mapping for hydrological classification and analysis of temporal change and scale effects in Switzerland. J Hydrol 523:804–821. https://doi.org/10.1016/j.jhydrol.2015.01.086

Taylor JW (2000) A quantile regression neural network approach to estimating the conditional density of multiperiod returns. J Forecast 19(4):299–311

Toth Z, Kalnay E (1997) Ensemble forecasting at ncep and the breeding method. Mon Weather Rev 125(12):3297–3319

Tracton MS, Kalnay E (1993) Ensemble forecasting at NMC: practical aspects. Wea Forecast 8:379–398

van Andel SJ, Weerts A, Schaake J, Bogner K (2013) Post-processing hydrological ensemble predictions intercomparison experiment. Hydrol Process 27(1):158–161. https://doi.org/10.1002/hyp.9595

Vitart F (2014) Evolution of ECMWF sub-seasonal forecast skill scores. Q J Royal Meteorol Soc 140(683):1889–1899. https://doi.org/10.1002/qj.2256

Viviroli D, Zappa M, Gurtz J, Weingartner R (2009) An introduction to the hydrological modelling system prevah and its pre- and post-processing-tools. Environ Model Softw 24(10):1209–1222. https://doi.org/10.1016/j.envsoft.2009.04.001, cited By 89

Wang WC, Chau KW, Cheng CT, Qiu L (2009) A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. J Hydrol 374(3–4):294–306

Weerts A, Winsemius H, Verkade J (2011) Estimation of predictive hydrological uncertainty using quantile regression: Examples from the National Flood Forecasting System (England and Wales). Hydrol Earth Syst Sci 15(1):255–265

Yarar A (2014) A hybrid wavelet and neuro-fuzzy model for forecasting the monthly streamflow data. Water Resour Manag 28(2):553–565

Zappa M, Bernhard L, Spirig C, Pfaundler M, Stahl K, Kruse S, Seidl I, Stähli M (2014) A prototype platform for water resources monitoring and early recognition of critical droughts in Switzerland. IAHS Publ 364:492–498

Zhang X, Peng Y, Zhang C, Wang B (2015) Are hybrid models integrated with data preprocessing techniques suitable for monthly streamflow forecasting? Some experiment evidences. J Hydrol 530:137–152

Zivot E, Wang J (2006) Vector autoregressive models for multivariate time series. In: Modeling Financial Time Series with S-$PLUS^{©}$. Springer, New York, pp 385–429