Hydrology and
Earth System
Sciences

Development of a monthly to seasonal forecast framework tailored to inland waterway transport in central Europe

Dennis Meißner[1], **Bastian Klein**[1], **and Monica Ionita**[2,3]

[1]Federal Institute of Hydrology, 56068 Koblenz, Germany
[2]Alfred-Wegener-Institute, 27568 Bremerhaven, Germany
[3]MARUM, Bremen University, Bremen, Germany

Correspondence: Dennis Meißner (meissner@bafg.de)

Abstract. Traditionally, navigation-related forecasts in central Europe cover short- to medium-range lead times linked to the travel times of vessels to pass the main waterway bottlenecks leaving the loading ports. Without doubt, this aspect is still essential for navigational users, but in light of the growing political intention to use the free capacity of the inland waterway transport in Europe, additional lead time supporting strategic decisions is more and more in demand. However, no such predictions offering extended lead times of several weeks up to several months currently exist for considerable parts of the European waterway network.

This paper describes the set-up of a monthly to seasonal forecasting system for the German stretches of the international waterways of the Rhine, Danube and Elbe rivers. Two competitive forecast approaches have been implemented: the dynamical set-up forces a hydrological model with post-processed outputs from ECMWF general circulation model System 4, whereas the statistical approach is based on the empirical relationship ("teleconnection") of global oceanic, climate and regional hydro-meteorological data with river flows. The performance of both forecast methods is evaluated in relation to the climatological forecast (ensemble of historical streamflow) and the well-known ensemble streamflow prediction approach (ESP, ensemble based on historical meteorology) using common performance indicators (correlation coefficient; mean absolute error, skill score; mean squared error, skill score; and continuous ranked probability, skill score) and an impact-based evaluation quantifying the potential economic gain.

The following four key findings result from this study: (1) as former studies for other regions of central Europe indicate, the accuracy and/or skill of the meteorological forcing used has a larger effect than the quality of initial hydrological conditions for relevant stations along the German waterways. (2) Despite the predictive limitations on longer lead times in central Europe, this study reveals the existence of a valuable predictability of streamflow on monthly up to seasonal timescales along the Rhine, upper Danube and Elbe waterways, and the Elbe achieves the highest skill and economic value. (3) The more physically based and the statistical approach are able to improve the predictive skills and economic value compared to climatology and the ESP approach. The specific forecast skill highly depends on the forecast location, the lead time and the season. (4) Currently, the statistical approach seems to be most skilful for the three waterways investigated. The lagged relationship between the monthly and/or seasonal streamflow and the climatic and/or oceanic variables vary between 1 month (e.g. local precipitation, temperature and soil moisture) up to 6 months (e.g. sea surface temperature).

Besides focusing on improving the forecast methodology, especially by combining the individual approaches, the focus is on developing useful forecast products on monthly to seasonal timescales for waterway transport and to operationalize the related forecasting service.
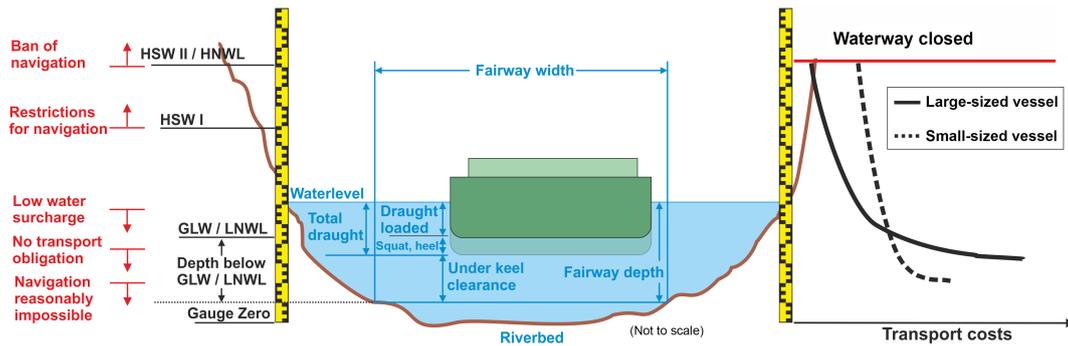
Figure 1. Diagram of the interaction between hydrologic conditions (water level), waterway parameters (fairway depths), specific navigation thresholds and transport costs.

1 Introduction

Competitive transport systems are vital for Europe's prosperity and its economic growth and the demand, especially for freight transport, will significantly increase within the coming years. Besides safeguarding a high degree of efficiency, accessibility and safety, today's European and national transport policy is clearly oriented towards sustainable and energy-efficient transport systems. In this regard, the importance of inland waterway transport (IWT) will further increase, because it is an environmentally friendly and safe mode of transport, which still has plenty of spare capacity (European Commission, 2001). But IWT also faces a few drawbacks. Besides the low transport velocity and the comparatively low network density, the natural variability of the fairway conditions and the availability along substantial parts of the European waterway network pose the main weaknesses. In this respect, low streamflows, floods and river ice are the significant hydrological hazards concerning IWT in central Europe. The relevance of these impacts depends on the geographical position, the season and the characteristic of the waterway (Nilson et al., 2012; Meißner and Klein, 2016). The main vulnerability of IWT with regard to hydrological impacts, results from (long-lasting) droughts leading to low streamflow and low water levels along the free-flowing waterways, which represent a substantial share of Europe's inland waterway network. Although there is no low-water threshold beyond which navigation is prohibited, low water levels and the corresponding reduced water depths are a limiting factor. During low-flow situations, additional (small) ships are needed to handle the same amount of cargo as during periods with mean water-level conditions. This causes increasing transport costs (see Fig. 1). Especially in the case of extreme or long-lasting low-flow periods, the available small-ship volume is limited compared to the transport demand and goods have to be shifted to other modes of transport, that is if this is technically possible and if transport volume is available. Besides determining the maximum cargo-carrying capacity the water level affects energy consumption and time of travel, altogether being reflected in the transport costs. Figure 1 depicts the close correlations between water level, fairway and vessel parameters and transportation costs. Furthermore, low-flow situations increase the danger of ship grounding and ship-to-ship collisions due to a reduced depth and width of the affected fairways.

Originally, navigation-related forecasts have been developed in order to primarily support the individual skipper, who aims at maximizing the load of an upcoming trip. Therefore current lead times of water level forecasts for the central European waterways range from one to several days complying with the travel time of the vessels to pass the main bottlenecks of a waterway leaving the loading port. Such forecasts allowing short-term operational decision-making remain vital to the waterway transport sector, but there is an increasing demand for additional forecast information going beyond this short- to medium-range (Klein and Meißner, 2016; Meißner and Klein, 2017). Extended forecast lead times offer the possibility to sustainably increase IWT efficiency and to support medium- to long-term waterway management. In the context of the EU-founded project EUPORIAS (European Provision Of Regional Impacts Assessments on Seasonal and Decadal Timescales; www.euporias.eu) the specific vulnerability of the waterway transportation sector on the Rhine was analysed and a prototypical service, called SOS RHINE, was designed with the collaboration of the German and the Dutch meteorological services and the German Federal Institute of Hydrology (Funk et al., 2015). However, so far, no such forecasts exist for the main parts of the trans-European waterway network – primarily due to the limited skill of long-term hydro-meteorological forecasts in northern extratropical regions when compared to other parts of the world (Ionita et al., 2008; Domeisen et al., 2015). Associated with this, there is still widespread scepticism about whether seasonal predictions can be trustworthy for decision-making in practice.

Despite the predictive limitations, several studies prove skill for seasonal river flow forecasts in different parts of Europe (Wilby et al., 2004; Gámiz-Fortis et al., 2008; Ionita et al., 2012, 2014). In recent years, multiple region-specific

forecast methods and systems have been developed to predict hydrological variables several weeks and months in advance to anticipate water availability in Europe (Olsson et al., 2016; Svensson, 2016; Demirel et al., 2015; Gelfan et al., 2015; Fundel et al., 2013, Jörg-Hess et al., 2015). The most common objectives behind these systems are water supply, reservoir operations and hydropower. The European Flood Awareness System (EFAS; Bartholmes et al., 2009) also publishes a seasonal hydrological outlook with a lead time of up to 2 months since 2016. This outlook offers Europe-wide homogenous information on surplus or deficit water resources as weekly averages aggregated for major hydrological units. Nevertheless the generation of early warning flood information up to 2 weeks ahead for regional, national or European authorities remains the focal point of EFAS, but the seasonal products supplement this continental hydrological information service. Regarding the underlying methodologies applied in seasonal hydrological forecasting, two categories, aside from using observed climatology, are usually distinguished: dynamical and statistical approaches (e.g. Crochemore et al., 2016, and references therein). Although the models of the second category rely on the statistical relationship between various observed predictors and the hydrological predictand (Demirel et al., 2013), the methods of the first category use seasonal meteorological forecasts and hydro-meteorological observations to drive hydrological models. Ensemble streamflow prediction (ESP) is probably the most famous dynamical approach applied already for many years in research contexts as well as in operational applications (Day, 1985; Wood et al., 2002; see also Sect. 3.3). The choice between the different approaches is usually based on the purpose and the region of the forecast as well as the availability of models and data, but sometimes the principles and philosophy within the executing institutions also influence such a decision. In addition, to use the two aforementioned methods alongside some studies show that within so-called hybrid or mixed approaches statistical and dynamical methods could complement each other leading to an increased forecast performance (Robertson et al., 2013).

This paper describes the set-up and performance of a monthly to seasonal forecasting framework for the major waterways crossing Germany, namely the Rhine, Danube and Elbe. The work was initiated by the Federal Institute of Hydrology, which is in charge of developing, maintaining and operating the navigation-related forecasting systems for the German waterways, realizing the high demand of long-term hydrological forecasts by the transport sector. The paper is structured as follows: first we describe the study area focusing on the hydrological characteristics relevant for IWT. This is followed by the specification of the forecasting methods and the underlying data implemented in the framework. After defining the methods and metrics to evaluate forecast skill and value, Sect. 4 shows the intercomparison of the different approaches. In Sect. 5 the main conclusions are discussed and an outlook on the forthcoming steps to an operational monthly to seasonal forecasting service tailored to IWT is given.

2   Study domain – the German waterways

The European inland waterways offer a more than 40 000 km network of canals, rivers and lakes connecting cities and industrial regions across the continent. The German inland waterway network – an integral part of the trans-European waterway system – comprises about 7350 km, of which approximately 75 % are rivers and 25 % canals. The major inland waterways with regard to freight transport are the Rhine (with its tributaries Neckar, Main, Moselle and Saar) and the Danube, as well as parts of the Elbe and some canals interconnecting the natural waterways. About two-thirds of the German waterways are of international relevance, whereupon the importance of the Rhine is outstanding: with almost 200 million tons transported along the Rhine per year (approximately two-thirds of the European IWT volume). The Rhine is not solely Germany's but also Europe's most important inland waterway (CCNR, 2016). Approximately one-third of the rivers in Germany used as waterways are free-flowing, so they are particularly affected by low flows as the dominating hydro-meteorological impact on IWT. Therefore this study focusses on the free-flowing stretches of the international waterways of the Rhine, Danube and Elbe (see Fig. 2).

The River Rhine, with a total length of 1230 km, drains an area of approx. 200 000 km^2 with a mean flow rate of approx. 2500 m^3 s^{-1} at its mouth in the North Sea. It is shippable for large vessels between Rotterdam and Basel on a length of about 800 km. Although the main shippable tributaries of the River Rhine are impounded, offering a guaranteed fairway depth, the Rhine itself is a free-flowing waterway between Iffezheim–Karlsruhe and the beginning of the delta near Pannerdensche Kop in the Netherlands (approximately 500 km). The flow regime of the River Rhine (see right panel of Fig. 2) shifts in downstream direction from a snow-dominated regime (nival, e.g. gauge Maxau) induced by the Alps to a complex flow regime in the middle (gauge Kaub) and lower Rhine stretch (gauge Ruhrort) due to the increasing influence of the rainfall-dominated (pluvial) flow regimes of the major tributaries (Neckar, Main, Moselle). Low flows, leading to restrictions for waterway transport, typically occur in the River Rhine in late summer and autumn due to high evaporation and low melt water input from the Alpine region (Funk et al., 2015). The River Danube, with a total length of 2826 km, drains an area of 817 000 km^2 with a mean flow rate of approximately 6500 m^3 s^{-1}. It is shippable on a length of 2415 km between Kelheim and the Black Sea. The German part of the waterway (ca. 220 km) is impounded offering a minimum fairway depth of 2.70 up to 2.90 m, except for a 70 km section between Straubing and Vilshofen an der Donau. The flow regime in this critical stretch for
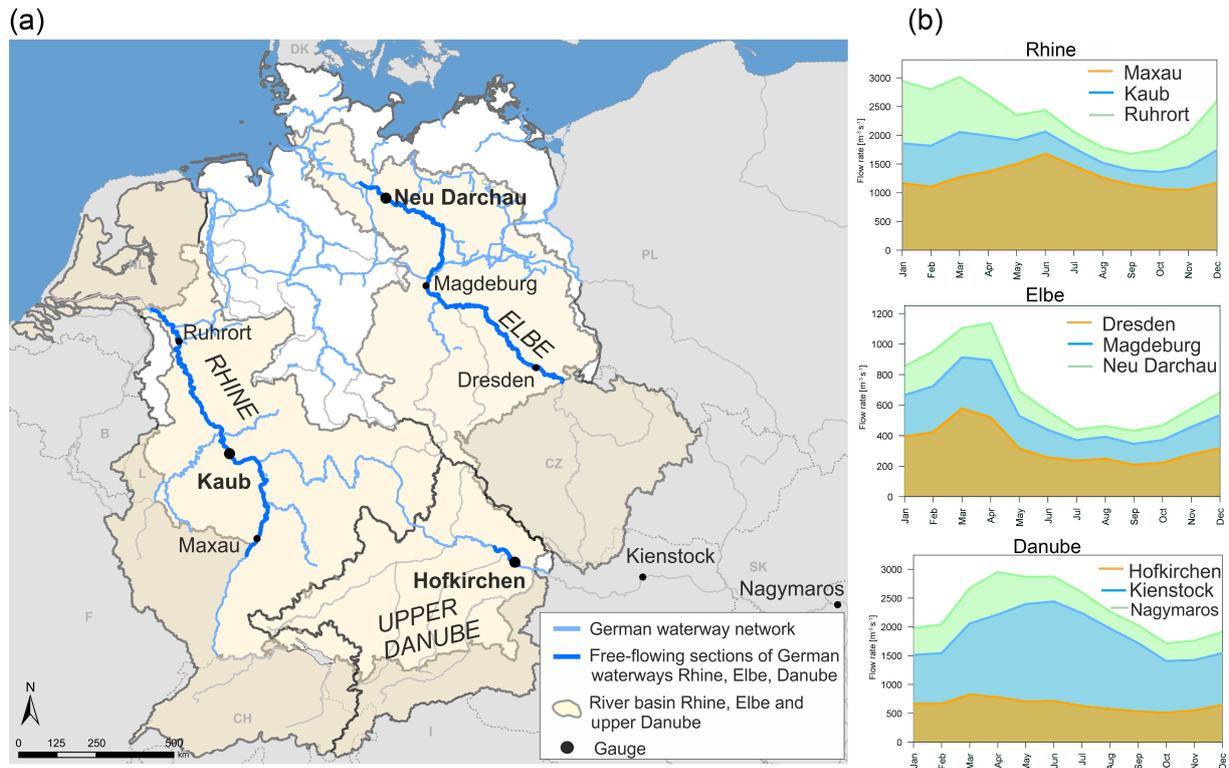
Figure 2. Map showing the German stretches of the international waterways the Rhine, Danube and Elbe **(a)**, for relevant gauges (black dots) the long-term monthly mean flow rates (1981–2000) are visualized **(b)**.

waterway transport is pluvio-nival with a complex broad-peaked runoff shape resulting from an overlapping of rainfall and snowmelt influences. Autumn is the typical low-flow season, often extended to the winter months. Further downstream after the Alpine river Inn enters the Danube, the flow regime changes to a nival regime (see gauges Kienstock and Nagymaros). The Elbe, with a total length 1090 km, drains an area of approx. 150 000 km² with a mean flow rate of approximately 860 m³ s⁻¹. About 930 km are shippable between Pardubice in the Czech Republic and the mouth in the North Sea at Cuxhaven, Germany. The stretch upstream of Geesthacht up to the German–Czech border (nearly 600 km) is free-flowing. The Elbe between Dresden and Neu Darchau shows pronounced pluvial runoff regimes with maximum flows in late winter to spring and lowest flow values in summer and autumn. Compared to the Rhine and Danube, the low-flow season relevant for waterway transport already starts in early summer and lasts until autumn. Figure 2b visualizes the flow regime of the three waterways represented by the long-term monthly mean flow rate at selected gauges (period 1981–2010).

In order to analyse the performance of the different forecast approaches implemented in the forecast framework, as described in the following sections, one gauge at the Rhine, Danube and Elbe has been selected, which is of special relevance for navigation along the particular waterway (black

dots, bold labels in Fig. 2). Gauge Kaub is an especially prominent station: up to 4500 requests per day on the current short-term forecasts, which are published via the river information system ELWIS (www.elwis.de), are recorded during low flows. Table 1 gives an overview of characterizing statistics of the three selected gauges. The mean flow has been calculated as the arithmetic mean of the daily flows within the reference period (listed in Table 1), while the mean low flow is calculated as the arithmetic mean of the lowest daily flows of each year within the particular reference period.

3 Forecast framework

3.1 Input data and hydrological model

Various sorts of input data from different providers have been integrated into the monthly to seasonal forecasting framework presented in this study. The basic requirement is that all data sources are operationally available, that means the data are continuously updated near real time. The input data selected could be grouped as hydrological measurements, climate and reanalysis data or seasonal meteorological forecasts.

Measured streamflow and water-level data (daily mean values) at the gauges relevant for navigation were provided by the Federal Waterway and Shipping Administration

Table 1. Catchment size, annual mean and mean low flow at selected gauges for the Rhine, Danube and Elbe.

Gauge	River	Area (km^2)	Mean flow (m^3 s^{-1})	Mean low flow (m^3 s^{-1})
Kaub	Rhine	103 488	1659 (1911–2011)	799 (1911–2011)
Hofkirchen	Danube	47 518	638 (1901–2008)	301 (1901–2008)
Neu Darchau	Elbe	131 950	710 (1890–2006)	290 (1890–2006)

(WSV) for the period 1951 to 2015. The data were used to validate model and forecast performance as well as to calculate climatological forecasts. The latter are generated by averaging the measured values for the same month(s) of the year as the one(s) forecasted.

The precipitation and temperature data, used to force the hydrological model in simulation mode up to the initialization of the particular forecast, is taken from the E-OBS dataset, version 13.1 (Haylock et al., 2008). The downward surface solar radiation is extracted from the ERA-Interim reanalysis (Dee et al., 2011) for the period 1979–2015. In order to preserve the statistical properties of the 5 km by 5 km HYRAS dataset (Rauthe et al., 2013), which was used for calibrating the hydrological model LASIM-ME, the data from E-OBS and ERA-Interim had to be downscaled and bias-corrected. A large number of different bias correction methods are available ranging from simple methods, such as linear scaling (see e.g. Lenderink et al., 2007), linear scaling with additional correction of the standard deviation (Leander and Buishand, 2007), to more complex methods such as the distribution-based correction method "quantile mapping" (see e.g. Piani et al., 2010). We applied these different methods to bias correct E-OBS and ERA-Interim and compared the goodness-of-fit of the flow simulation using the downscaled and bias-corrected meteorological input. As the goodness-of-fit measures such as the Nash–Sutcliffe efficiency and correlation were very similar, we decided to use the most simple bias correction method linear scaling. Monthly linear scaling with reference to the HYRAS dataset was applied on a coarse grid (25 km × 25 km). The monthly scaling factors and monthly additive terms for precipitation and temperature, respectively, have been derived for the period 1951–2000. Subsequently the processed E-OBS data were downscaled to the required 5 km by 5 km model grid by taking into account the long-term monthly ratio between HYRAS at the two different resolutions (5 km, 25 km) for precipitation and by assuming a constant lapse rate of 0.48° per 100 m between the grid cell heights for temperature. The ERA-Interim downward surface solar radiation was processed in a similar way as the E-OBS precipitation. As high-resolution global radiation reference data for bias correction and downscaling, the surface solar irradiance (SIS) of EURO4M (DWD, 2013) for the period 1991–2010 was used. In addition, to initialize the hydrological model before starting a forecast, the aforementioned input data were used to

run the continuous model simulation and the ESP forecasts (see Sect. 3.2).

For the statistical approach, different meteorological, climatological and oceanic data products have been selected as predictors. These datasets or reanalysis products are listed in Table 2.

As seasonal meteorological forecast used in the dynamical forecast approach, we used the reforecast dataset from ECMWF's Seasonal Forecast System 4 (S4 hereafter) for the period 1981–2014. For the period 1981–2011 the ensemble size varies between 15 members (initialization months January, March, April, June, July, September, October, December) and 51 members (for the remaining months). Since 2012, the ensemble size is 51 members throughout the year. Before feeding the hydrological model, the output from S4 (daily total precipitation and air temperature), interpolated to a 50 km × 50 km grid (multiple of the 5 km × 5 km model grid), was bias-corrected with the meteorological observation dataset used for the baseline simulation. Again several bias correction and post-processing methods of different complexities for ensemble forecasts are available (see e.g. Crochemore et al., 2016; Zhao et al., 2017). After the experiences of the bias correction of E-OBS and ERA-Interim, we decided to stick to the most simple bias correction method linear scaling, successfully applied for bias correction of seasonal forecasts (Crochemore et al., 2016). We corrected daily values of the different parameters on a monthly basis, which means each daily value of the same month is corrected by the same scaling. In future applications, different bias-correction and post-processing methods will be applied and analysed. As meteorological seasonal forecasts tend to drift towards the climate model from which they are issued with increasing lead time, giving rise to model bias, separate bias correction factors have been estimated for each forecast initialization date (calendar month) and monthly lead time (month 1 to month 6). In the final step the corrected precipitation and temperature are downscaled to the 5 km × 5 km model grid.

The hydrological model used in this forecasting environment is based on the model software LARSIM (Large Area Runoff SImulation Model). LARSIM is a deterministic distributed conceptual hydrological model for the simulation and forecasting of the terrestrial water cycle and flow in rivers. It has been originally developed by Ludwig and Bremicker (2006) and is currently maintained and developed by a transnational developer community of several forecasting centres from Germany and Switzerland. The spatial dis-

Table 2. Climate and oceanic data sources used in the statistical forecast approach.

Dataset (Provider)	Variables	Spatial coverage	Spatial resolution	Reference
E-OBS v13.1 (ECA&D Project, KNMI)	precipitation temperature	Europe	$0.25° \times 0.25°$	Haylock et al. (2008)
CDC (DWD)	precipitation temperature	Germany	areal mean values $(2500–48\,000\,km^2)$	–
ERSST v4 (NOAA)	sea surface temperature	global	$2° \times 2°$	Huang et al. (2010)
NCEP/NCAR Reanalysis 1 (NOAA NCEP-NCAR)	volumetric soil moisture	global	$2.5° \times 2.5°$	Kistler et al. (2001)
CDAS-1 MONTHLY (NOAA NCEP-NCAR)	soil moisture geopotential height (700 mbar) relative humidity, specific humidity zonal and meridional wind (700 mbar) mean sea level pressure	global	$2.5° \times 2.5°$	Kalnay et al. (1996)

cretization of the model can be grid-based subareas or subareas according to hydrologic sub catchments. Hydrological processes are modelled for each single land use category or alternatively for each land use and soil type combination in a subarea (hydrological response unit, HRU). Due to the strong altitude dependence of temperature HRU could be further subdivided into elevation zones for the simulation of the snow processes. To avoid unnaturally high snow accumulation and to consider the snow mass transport mechanism, the new LARSIM snow mass transport option could be activated. Using this option, snow can only accumulate up to a gradient-dependent threshold and exceeding snow is simply passed to the next model element downhill. These options (subdivision in elevation zones and snow mass transport) in LARSIM are currently used for the Rhine basin. For the Elbe and Danube they will be implemented in the near-future. The LARSIM model used in this context is called LARSIM-ME (ME – *Mittel Europa* or central Europe). LARSIM-ME covers the catchments of the rivers Rhine, Elbe, Weser–Ems, Odra and Danube up to gauge Nagymaros in Hungary. The total catchment size, simulated by the model, is approximately $800\,000\,km^2$. The spatial resolution is $5\,km \times 5\,km$ and the computational time step is daily. To estimate the model parameters of such a large model domain, a regionalization approach based on clustering was applied. In total nine clusters with similar flow characteristics have been identified for the model domain based on the following steps:

1. Definition of statistical flow values for 132 headwater catchments, relatively free of anthropogenic influences. As statistical values we chose

   – mean flow in relation to basin area,

   – high flow with a 2-year recurrence frequency in relation to the mean flow,

   – monthly mean low flow in relation to the mean flow,

   – mean flow in winter in relation to the mean flow in summer.

2. Identification of nine clusters using the k-means clustering algorithm,

3. Selection of seven geographical factors (e.g. height, slope, areal share of forest, areal share of field, areal share of unconsolidated rock, mean permeability of upper and lower soil) in order to characterize the clusters,

4. Rule-based mapping of all subbasins to the clusters.

Subsequently a subset of 72 catchments, that are relatively free from anthropogenic effects (catchment sizes ranging from 200 to $2100\,km^2$) and evenly distributed over the model domain, were calibrated manually together with the nine clusters using the HYRAS dataset as forcings. The manual calibration for the period 1998 to 2006 (validation 1976–2006) followed the guideline to calibrate LARSIM water balance models (Haag et al., 2016), which recommends the relevant parameters to be calibrated, the parameter range and a calibration procedure. The manual calibration strategy applied involved the following steps: (1) adjustment of the parameters of the snow modules with a focus on the water balance and floods caused by snow melting, (2) adjustment of base flow storage relevant parameters to reproduce discharges at low-flow conditions, (3) adjustment of the parameters relevant for interflow to reproduce mean flow conditions, (4) adjustment of relevant parameters to reproduce flood hydrographs and (5) final validation and fine adjustment of all parameters. By using this well-proven process-based calibration strategy for LARSIM models conducted by experienced hydrologists instead of a non-process based automatic calibration procedure, we expected to reduce the degree of parameter uncertainty in the clusters due to the problem of parameter equifinality (Beven, 1996). But the fact that different sets of model parameters reproduce equally good output signals remains an issue in any hydrological model calibration
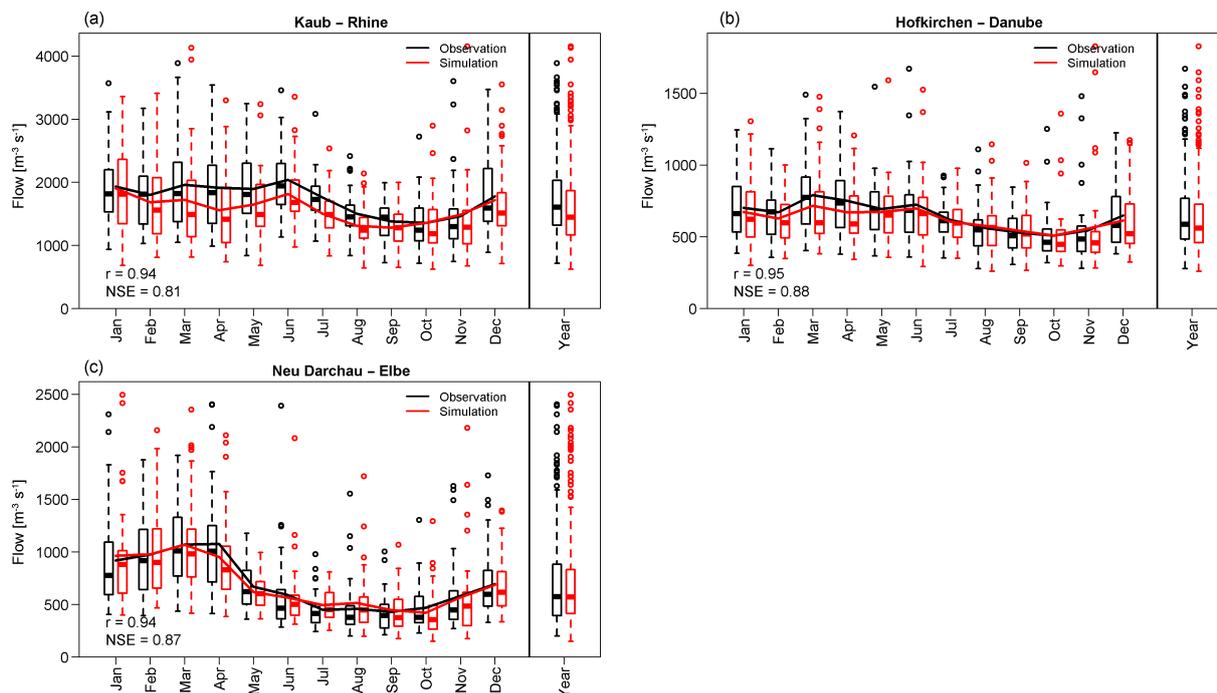
Figure 3. Comparison of the simulated (red) and observed (black) long-term (1981–2015) monthly mean streamflow and the distribution of the simulated and observed monthly mean streamflow within the different months and for the whole year illustrated as box-and-whisker plots of the gauges Kaub–Rhine **(a)**, Hofkirchen–Danube **(b)** and Neu Darchau–Elbe **(c)**.

and it is a significant aspect in particular when applying those models for predictions. Afterwards, the parameter means and parameter spans have been derived for the clusters and transferred to the respective clusters in the whole model domain. As a next step, a fine calibration of the model parameters within the parameter spans of the clusters have been conducted for selected parts or gauges of the upper Danube, Elbe and River Rhine. The fine calibration for the Elbe, Danube and Rhine was based on the same period as the input dataset (HYRAS). Special attention was given to anthropogenic effects dominating the flow behaviour in several catchments (dams, regulated lakes, water transfers) and the most relevant structures have been implemented explicitly. In order to retain the consistent spatial parameter distribution, the parameters have been optimized within the identified parameter spread of the specific cluster of the initial calibration. The standard deviation of the parameter values has been used as indicator to identify potential room for parameter optimization.

In the forecast framework, different meteorological inputs (E-OBS, ERA-Interim instead of the non-operational HYRAS dataset) have to be used. Therefore in Fig. 3 we show the output from the LARSIM model set-up for monthly to seasonal forecasting for the hindcast period from 1981 to 2015. Figure 3 illustrates the simulated and observed long-term monthly mean streamflow of the period 1981–2015, the distribution of the simulated and observed monthly mean

flows within the individual months of the year as box-and-whisker plots, as well as the correlation r and Nash–Sutcliffe efficiency (NSE). The boxes of the box-and-whisker plots represent the 25–75 % inter-quantile range with the median as a band inside the boxes. The whiskers represent 1.5 times the inter-quantile range and values beyond are plotted as single data points. All gauges show an NSE > 0.8 and r > 0.9 for monthly mean values. On the original daily time step, the gauges show an NSE > 0.7 and r > 0.85. Especially for the Elbe, the model performance on a daily basis is somewhat lower, mainly due to shortcomings in the river routing. The climatological seasonality of streamflow is well reproduced at the gauges Hofkirchen–Danube and Neu Darchau–Elbe. At gauge Kaub–Rhine, especially in the spring months effected by snow melt, LARSIM-ME underestimates the flow. This underestimation is an indicator for problems in modelling snow processes, which will be a focus in the future model developments of LARSIM-ME.

Another effect is anthropogenic influences, especially regulated dams, lakes and water transfers. As it is quite challenging to account for such impacts in a large-scale hydrological model like LARSIM-ME, just the major barrages and water transfers together with their regulation rules have been implemented so far. This is not due to limitations in the model functionality but due to the difficulties in getting the required information. Therefore some of the problems the LARSIM

models show to reproduce the flow behaviour result from missing or incomplete reproduction of anthropogenic effects.

3.2 Forecast parameter and forecast benchmark

In preparation for the development of the forecast framework, three forecast parameters have been selected in agreement with the users. For the monthly forecast, the monthly mean flow (MoMQ) and the lowest arithmetic mean of flow on seven consecutive days within a month (MoNM$_7$Q) were chosen. For the seasonal forecast, the tri-monthly mean flow (3MoMQ; the average of three months) is predicted. MoMQ and 3MoMQ are quite common forecast parameters on longer lead times (e.g. Yossef et al., 2013; Svensson, 2016; Tucci et al., 2003), NM$_7$Q is a variable primarily used as low-flow indicator in the context of hydrological monitoring or ecological purposes (Richter et al., 1998; Marke, 2008). Low NM$_7$Q values imply the existence of a long-lasting drought period (Klein and Meißner, 2016). Furthermore, the NM$_7$Q is a robust indicator, because it is insensitive to distorting singularities like short-term fluctuations due to natural or anthropogenic errors. For decision-making in the IWT sector on monthly or seasonal timescales a monthly or tri-monthly resolution is sufficient. Users do not estimate the load of a specific ship on a specific trip, as it is the case on short- to medium-ranges timescales, but typical decisions (e.g. how to compose an optimal fleet for the coming month) require information on "average" or the "worst" conditions characterized e.g. by the monthly mean flow or the lowest arithmetic mean of seven consecutive daily values within a month.

Traditionally, for short- to medium-range forecasts, the water level is the parameter of main interest for navigation purposes as it determines via the shape of the river bed the available water depths and therefore the possible vessel draught. As in many river stretches the shape of the riverbed changes over time due to morphological dynamics and this is problematic to compare water levels over long periods of time. To overcome this issue, we decided at this stage of development to analyse discharges rather than water levels.

As reference data, in order to evaluate the forecast quality of each forecast approach, we used two datasets: (1) observed discharges at the forecasting gauges representing the real-world situation and (2) simulated discharges at the forecasting gauges generated by the hydrological model LARSIM-ME forced with observed meteorology (see Sect. 3.1). This reference, also called pseudo-observations, was used in some of the analysis on the predictability in order to mask the error coming out of the hydrological model itself (Shukla and Lettenmaier, 2011; van Dijk et al., 2013; Wood et al., 2016).

Forecast benchmarks are used to demonstrate and quantify the added skill and value of the analysed forecast approaches. On the one hand the benchmarks have been selected on current practice (climatological forecast) and on the other hand a standard method requiring extensive input

data was chosen. As Fig. 2 shows, the discharges along the waterways are subject to seasonal variability dependent on the flow regime. Therefore, only flows of the same month in each year have been included into the respective climatological forecast, e.g. the climatological forecast for January is based on the measured flows of the first 31 days within each year. As standard seasonal forecasting method, we applied the already mentioned ESP approach. The set-up of ESP is relatively simple although a hydrological model of the basin of interest is required. Each forecast run is initialized with the best estimated initial hydrological conditions, which is based on forcing the hydrological model with measured meteorological inputs. Potential improvements might be achieved by data assimilation techniques (see e.g. Yossef et al., 2013). Based on this initialization, from which the predictive skill of ESP originates, the hydrological model is forced with an ensemble of historical time series of observed meteorology from previous years (Wood et al., 2002). ESP does not require seasonal meteorological forecasts but it solely relies on the resampled meteorology, which is a limitation at the same time, because meteorology does not contribute to an improved forecast skill in relation to the climatological forecast. Nevertheless, ESP proved to be a robust forecast approach used in several operational applications for years.

In addition to ESP, Wood and Lettenmaier (2008) suggested a complementary approach for sensitivity analysis, called reverse ESP. In contrast to an ESP forecast, the hydrological model in a reverse ESP run is initialized with an ensemble of initial conditions based on climatology. Along the forecast period the model is driven with the measured meteorology. As reverse ESP requires the "perfect" meteorology along the forecast lead time, it is not suitable as forecast approach in operational practice. Although the skill of ESP results from the initial hydrological conditions, reverse ESP obtains skill from the (perfect) meteorological forecast. That is why Wood and Lettenmaier (2008) suggested comparing the skill of ESP and reverse ESP as a function of lead time, season, basin etc. in order to determine which of the two main components of a seasonal hydrological forecast (hydrological memory or meteorological forcing) is dominating the particular forecast skill. Recently, this method was extended by Wood et al. (2016) to a method called VESPA (Variational Ensemble Streamflow Prediction Assessment), which is able to blend the two sources of seasonal forecast skill systematically.

3.3 Forecast approaches

In order to find an optimized (related to forecast quality, data and model requirements, computing time etc.) seasonal forecast procedure for navigation-related forecasting, multiple approaches representing the different philosophies (dynamical versus statistical) in seasonal hydrological forecasting have been implemented and tested under operational conditions. The above-mentioned dynamical method ESP is pri-

marily applied to analyse the different sources of predictability on seasonal timescales as well as to act as benchmark for the other methods. Furthermore, a dynamical approach similar to the one used for short- to medium-range forecasting was implemented, linking a hydrological model with seasonal meteorological predictions. The hydrological model LARSIM-ME was forced with measured data (up to the forecast starting date) and subsequently with the forecasts from ECMWF System 4. The hydrological model, as well as the processing of the meteorological inputs, is described in Sect. 3.1.

For the statistical approach, a methodology has been adopted, which was already successfully applied to predict seasonal streamflow anomalies at the Romanian Danube (Rimbu et al., 2005) as well as monthly to tri-monthly streamflow at the lower Elbe in Germany for specific events and seasons (Ionita et al., 2009, 2015). The basic idea is to use climate and hydro-meteorological variables (e.g. sea surface temperature, precipitation etc.) as predictors instead of climate indices (e.g. North Atlantic Oscillation, Southern Oscillation Index). Since the early days of seasonal hydrological forecasting, large-scale climatic patterns have been used as predictors for seasonal streamflow anomalies (Maurer and Lettenmaier, 2003; Wang et al., 2011). For Europe, the North Atlantic Oscillation (NAO) and the El Niño–Southern Oscillation (ENSO) indices are most commonly used as predictors of hydrological variables like streamflow (Rimbu et al., 2004; Trigo et al., 2004). Although these teleconnections are detectable, they are significantly less pronounced for Europe than for other continents like Africa or Australia. Furthermore, they are characterized by non-stationarity issues, which means that the strength of the correlation between the indices of these two phenomena and streamflow anomalies varies over time (Ionita et al., 2008). The climate and hydrometeorological variables used in the approach presented here have to fulfil a stability criterion for the correlation between predictor and predictand. The concept of predictor stability, wherein the so-called stability maps are a crucial tool, was introduced by Lohmann et al. (2005). In order to detect stable predictors the variability of the correlation between the streamflow at a specific location and the potential predictors are investigated within a 31-year moving window within the period 1948 to 2012. The correlation is considered to be stable for those spatial units where the current streamflow and previous months climate variables are significantly correlated at the 90 % or 80 % level for more than 80 % of the moving window. Based on the following three steps stability correlation maps are generated:

1. Each predictand (e.g. 3MoMQ for March–April–May at station Kaub/Rhine) is correlated with numerous potential predictors. Different lags (e.g. mean sea level pressure in February, mean sea level pressure in January) and regions of the same variable are regarded as independent predictor.

2. The correlation is computed in a moving window of 31 years within the period from 1948 to 2012 (window 1: 1948–1979, window 2: 1949–1980, ..., window 34: 1981–2012).

3. The correlation is considered to be stable for those grid points or regions where predictor and predictand are significantly correlated at the 90 % level and 80 % level, respectively, for more than 80 % of the 31-year window within the period 1948–2012. According to the level of significance, grid points or regions are coloured in red to yellow shades (stable positive correlation), shades of blue to green (stable negative correlation) or white (non-stable correlation) in order to create a stability correlation map for each potential predictor. Figure 4 illustrates this procedure: tri-monthly mean flow in spring (March–April–May) at Kaub/Rhine and mean sea level pressure (SLP) of previous winter (December–January–February) from region (a) are positively correlated for all 31-year windows of the period 1948–2012 and above the 90 % significance level for more than 80 % of the windows (left panel of Fig. 4). Therefore, these grid points are stable, correlated with streamflow and are represented on the stability map in red (right panel of Fig. 4). The 3MoMQ and SLP from the grid points in region (b) are negative and above the 90 % significance level correlated for more than 80 % of the moving windows. These grid points are represented on the stability correlation map as blue. Grid points significantly correlated for less than 80 % windows, like those in region (c) are regarded as "unstable" (white colour on the stability correlation map). In the left panel of Fig. 4, the correlation is plotted in the middle of each 31-year window. Therefore the first point represents the correlation between 3MoMQ and previous winter SLP from 1948 to 1979, while the last point represents the correlation from 1981 to 2012.

The final composition of the predictors for the forecasting model is established by stepwise regression of the stable predictors using the Akaike information criterion and the explained variance of forecast errors. Originally, this method was solely applied for European and global climate datasets like E-OBS or ERSST (see Table 2). In the context of this navigation-related forecast framework, regional to local precipitation and temperature datasets from the German Meteorological Service (DWD) also have been included, as well as the historical discharges at the gauges of interest. The consideration of the measured discharges, which are an aggregated proxy of the hydrological history and of the current conditions, especially led to an additional increase in forecast skill. To quantify the importance of measured discharges it has been excluded from the regression model in an experimental set-up. The last column of Fig. 11 shows the forecast results without using measured discharges as predictor for June (month with the highest MoNM$_7$Q), September and
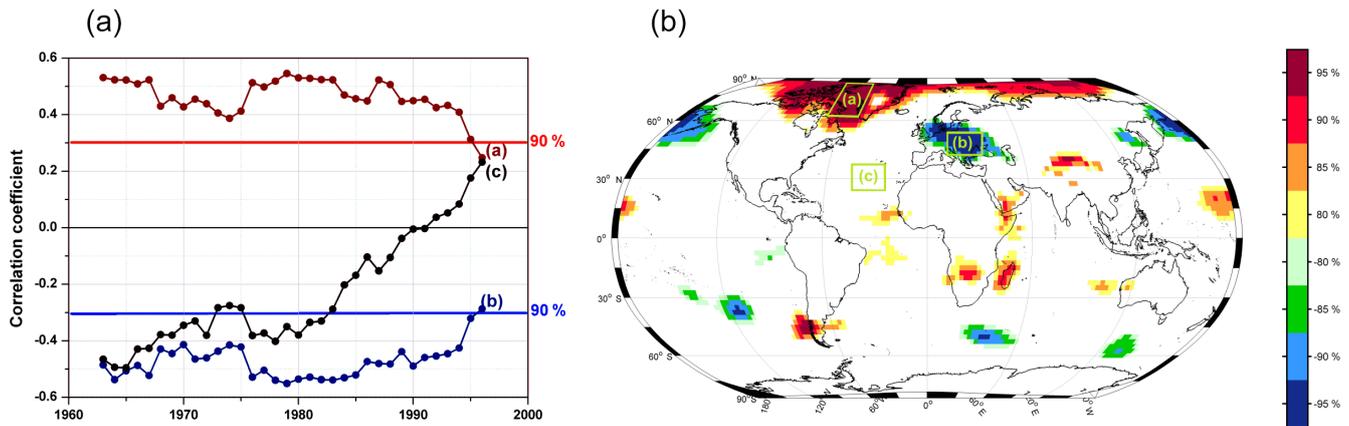
Figure 4. Example of a stability correlation map. **(a)** correlation between tri-monthly streamflow at Kaub/Rhine and SLP of the previous winter in a 31-year moving window for selected grid areas, **(b)** map showing the areas with the corresponding correlations.

November (a low-flow month, typically affecting inland navigation) MoNM$_7$Q at the station Kaub/Rhine. The selected scores clearly indicate an improved forecast skill when using measured discharge of previous months as predictor.

3.4 Forecast evaluation – skill and value

The skill of the forecasts was assessed in terms of the correlation coefficient (CC), the mean absolute error (MAE) and the mean squared error (MSE) as deterministic measures. In order to evaluate the skill improvement with respect to the reference forecast (climatology, ESP) as well as to be able to compare the skill amongst the different waterways, the corresponding skill scores (SS) have been additionally used for evaluation (MAE-SS, MSE-SS). Although the perfect score is 0 for MAE and MSE, an optimal forecast produces a CC, MAE-SS and MSE-SS of 1. The skill scores are a function of the forecast as well as the reference forecast and the observations. The skill scores are positive (negative) if the forecast skill is higher (lower) than the one of the reference forecasts. A skill score might be interpreted as the percentage improvement with regard to the reference forecast by multiplying the skill score by 100. As probabilistic measure of forecast skill, we applied the continuous ranked probability score (CRPS) and the respective CRPSS as its corresponding skill score (Hersbach, 2000). CRPS and CRPSS are appropriate indicators of the overall performance of probabilistic forecast systems comparable to the MAE and MAE-SS in the case of a deterministic forecast.

For the ensemble-based forecasts (ESP, dynamic approach based ECMWF S4) the deterministic metrics MAE and MSE have been determined relative to the observation for each ensemble member separately and afterwards the average of the single values was calculated. The CC is calculated from the ensemble mean. The forecasts of interest as well as the climatology are thereby treated as real ensembles instead of single realizations (e.g. the ensemble mean or median).

For the climatological forecast, as well as for the ESP approach, we considered a subsample of observations covering the study period from 1981 to 2014. In each case, we used the historical values of the same days in each year along the forecast length. Additionally, we followed the leave-one-out cross validation procedure by excluding the values of the validation year from the measurements when generating the respective climatological forecast as well as the meteorological input to the hydrological model in the case of ESP.

Besides the verification of forecast skill, we aim at evaluating the economic value of the forecast too. The value of a forecast arises by its ability to improve decisions made by the forecast users (Murphy, 1993). Forecast value and skill are not necessarily the same and their relationship in real-world applications could be quite complex as the analysis of forecast value always has to consider the specific user context. A feasible approach successfully applied in several applications before, most often in meteorological (Richardson, 2000, 2011; Wilks, 2001) and seldom in hydrological contexts (Roulin, 2007; Fundel and Zappa, 2011), is the concept of the relative economic value. In order to evaluate the potential economic gain of a forecast dependent on the user-specific cost/loss ratio, the original continuous hydrological forecasts are converted to categorical forecasts and subsequently combined with a relatively simple static cost/loss model. According to this model, costs (C) will incur whenever the forecast indicates an event, because the user will take preventive actions. It is assumed that these preventive actions offer a total protection, so that the investment prevents any losses (L). Losses will only incur, if the forecast misses an event. If no event is forecasted and it actually did not happen, neither costs nor losses incur. The decision strategy behind the relative economic value assumes that the user aims at a long-term economic optimum, the users decisions solely depend on economic reasons and that the users actions are risk-neutral. In order to compare different forecasting systems based on their economic value Richardson (2000) sug-

gested to calculate a relative score, showing the added value of a forecast compared to the climatological forecast. The relative economic value score V is defined as the difference of the long-term average expected expenses (EE) of the climatological forecast and the forecast of interest in relation to the difference of the climatological and a perfect forecast:

$$V = \frac{\text{EE}_{\text{clim}} - \text{EE}_{\text{forecast}}}{\text{EE}_{\text{clim}} - \text{EE}_{\text{perfect}}}. \tag{1}$$

A perfect forecast prevents losses (all events are predicted correctly) and costs only incur in case of an event which occurs with climatological recurrence interval P_i. Therefore $\text{EE}_{\text{perfect}} = P_i \cdot C$. The best strategy to act, based on a climatological forecast, is to find the optimum of the two options "always protect" or "never protect" by minimizing the related expenses $\text{EE}_{\text{clim}} = \min(C, P_i \cdot L)$. In the long-term average, a user with a specific cost/loss ratio below the climatological recurrence interval P_i of the event will always protect, otherwise it is economically advantageous to accept the losses of the event. All of the aforementioned assumptions lead to the definition by Richardson (2000):

$$V = \tag{2}$$
$$\frac{\min\left(\frac{C}{L}, P_i\right) - \text{POFD} \cdot \frac{C}{L} - (1 - P_i) + \text{POD} \cdot P_i \cdot \left(1 - \frac{C}{L}\right) - P_i}{\min\left(\frac{C}{L}, P_i\right) - P_i \cdot \frac{C}{L}},$$

with

- POD is the probability of detection or hit rate (fraction of observed events that is forecast correctly)

- POFD is the probability of false detection or false alarm rate (fraction of false alarms conditioned on observations)

- C/L is the cost/loss ratio

- P_i is the recurrence interval of the disastrous event.

The maximum value score is 1 for perfect forecasts, while $V = 0$ indicates no added value of the forecast under investigation compared to the optimal use of a climatological forecast. In case of relative economic values below zero, the user should refuse the forecast and better use climatology. The maximum relative value score V_{max} is reached for $P_i = C/L$. V_{max} corresponds to the difference in hit and false alarm rates, which is also known as the Pierce skill score or Kuipers skill score (Manzato, 2007). The economic value depends on the quality of the forecast (expressed via the hit rate and the false alarm rate); the definition of the event, expressed through the climatological frequency; and on the individual user represented by the cost/loss ratio. As different forecast users with different decision problems will gain different levels of economic value from an optimal use of the same forecasts, the relative value score is often expressed graphically as a function of the cost/loss ratio. A suitable

numerical score, aside from V_{max}, is the area below the economic value function (Figs. 9 and 10), with an optimum value of 1.

4 Results

4.1 Sources of predictability

In the course of designing the forecasting framework, we conducted the typical ESP and reverse ESP experiments, as various predictability studies did before (Wood and Lettenmaier, 2008). Being aware that this experiment is just able to represent the two (in most cases unrealistic) endpoints of forecast uncertainty (zero and perfect information about future forcings and initial conditions, see Sect. 3.2), it is a feasible and pragmatic way to gain more insight into the relative role of the two main sources of predictability as a function of forecast location, lead time and initialization month. Figure 5 visualizes the MSE-SS for ESP and reverse ESP in relation to the simulated climatology at the gauges Kaub/Rhine, Hofkirchen/Danube and Neu Darchau/Elbe for forecast months 1 to 6. Eight initialization months (January, March, May, July, August, September, October, December) have been selected in order to clearly arrange the graphs, while focussing on the typical low-flow period (July–October).

Figure 5 clearly indicates that the differences in flow regime, climate region and catchment characteristics at the waterways (see Sect. 2) sustainably affect the relative importance of the predictability sources within the seasonal cycle. However, despite all differences amongst the gauges or waterways, the overall conclusion of the ESP and reverse-ESP experiments is that for the majority of initialization months and lead times, the mean squared forecast error is dominated by the meteorological forcing. In many cases, already in the first forecast month, future weather is the leading source of forecast skill as the MSE-SS of the ESP drops below the corresponding reverse ESP values. Nevertheless, in some months (e.g. July and August at gauge Kaub), the initial hydrological conditions noticeably influence forecast skill, at least for the first forecast month; in rare cases, also for the subsequent months (e.g. forecast at gauge Neu Darchau initialized in December). Although a sound estimation of initial hydrological conditions are essential (see also Demirel et al., 2013; Fundel et al., 2013), it could be concluded that solely relying on the "hydrological memory" as source of predictability won't be sufficient to produce skilful streamflow forecasts with lead times beyond 1 month for the German waterways.

4.2 Long-term evaluation

Following the history of the stepwise set-up of the forecast framework for the German waterways, we first compare the dynamical forecast approach based on ECMWF S4 forecasts
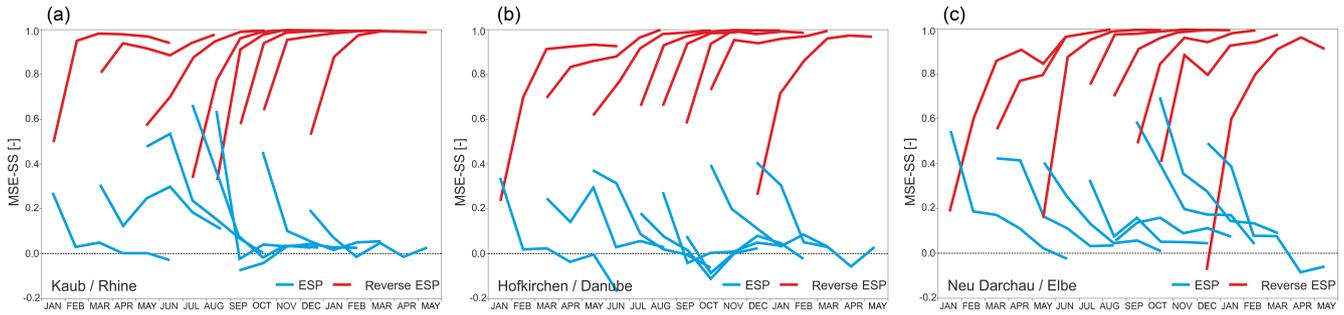
Figure 5. MSE-SS of the ESP (blue) and reverse ESP (red) forecast (reference: simulated to the MSE from using a climatology) for the gauges Kaub/Rhine **(a)**, Hofkirchen/Danube **(b)** and Neu Darchau/Elbe **(c)** over a lead time of 6 months (verification period 1981–2014).
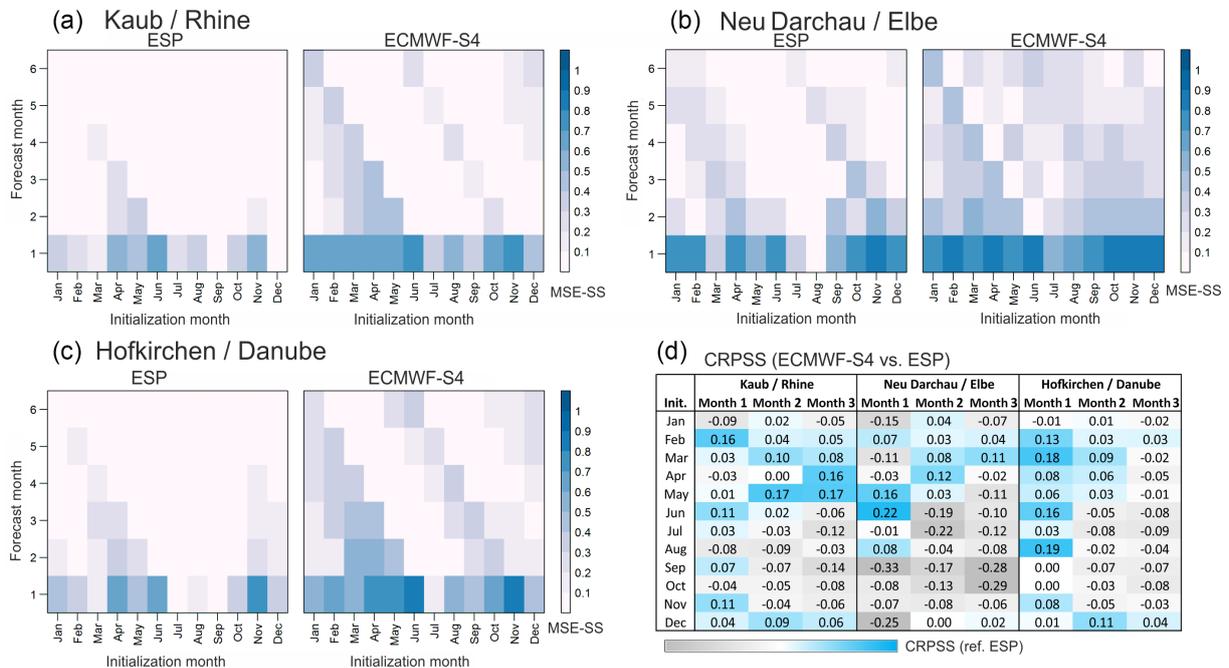

Figure 6. MSE-SS for ESP and S4-driven forecast for each initialization and forecast month at the stations Kaub/Rhine **(a)**, Neu Darchau/Elbe **(b)**, Hofkirchen/Danube **(c)** and CRPSS values of S4-driven forecast related to ESP for forecast months 1 to 3 (verification period 1981–2014) **(d)**.

with the ESP approach for the period 1981 to 2014. Although it is well-known that central Europe is a region offering limited skill for seasonal meteorological forecasts, in particular for precipitation as the most-important input to hydrological models, the question was if the information of these forecasts could provide some additional information to the seasonal hydrological forecasts. Figure 6 displays the MSE-SS between the ESP-based as well as the S4-based forecasts and observed climatology for the gauges Kaub, Neu Darchau and Hofkirchen as a function of forecast month (1–6) and initialization month (January–December). Dark coloured pixels indicate high forecast skill compared to climatology. It is obvious that for both approaches and all stations, the skill significantly diminishes with increasing lead time, but that the use of the S4 forecasts leads to additional skill for the majority

of lead times and initialization months at all gauges. Overall, the forecast skill for the Elbe (gauge Neu Darchau) is higher than for the Rhine (gauge Kaub) and Danube (gauge Hofkirchen). The skill scores at all stations show a noticeable pattern indicating higher scores in spring and late autumn, which might be induced snow melt (spring) and snow accumulation (autumn). This characteristic remains visible, in some cases becomes even more obvious, for the S4-based forecasts (e.g. for the October to November forecast at for the Rhine and Danube).

The increase in forecast skill is proved by the CRPSS shown for the first three months of lead time at Kaub, Neu Darchau and Hofkirchen in the table in Fig. 6d. As reference forecast, we selected ESP so that a positive CRPSS directly implies an improved forecast skill by using S4 inputs.

Initialization month	Category	Kaub / Rhine			Neu Darchau / Elbe			Hofkirchen / Danube		
		ESP	S-4	MLR	ESP	S-4	MLR	ESP	S-4	MLR
January	CC	0.673	0.682	0.939	0.860	0.783	0.936	0.679	0.703	0.953
	MSE-SS	0.41	0.54	0.89	0.70	0.73	0.88	0.49	0.63	0.91
	MAE-SS	0.23	0.34	0.65	0.47	0.54	0.62	0.29	0.44	0.68
	CRPSS	0.25	0.17		0.48	0.400		0.29	0.28	
February	CC	0.640	0.789	0.856	0.728	0.790	0.901	0.622	0.766	0.864
	MSE-SS	0.36	0.63	0.74	0.57	0.71	0.82	0.46	0.71	0.76
	MAE-SS	0.24	0.41	0.43	0.39	0.51	0.61	0.36	0.51	0.50
	CRPSS	0.19	0.32		0.37	0.41		0.29	0.39	
March	CC	0.654	0.832	0.912	0.861	0.842	0.940	0.620	0.793	0.830
	MSE-SS	0.10	0.46	0.84	0.55	0.72	0.88	0.38	0.70	0.70
	MAE-SS	0.08	0.27	0.64	0.39	0.51	0.67	0.26	0.51	0.46
	CRPSS	-0.02	0.00		0.47	0.41		0.23	0.37	
April	CC	0.726	0.796	0.903	0.734	0.767	0.836	0.768	0.828	0.860
	MSE-SS	0.36	0.52	0.82	0.58	0.67	0.71	0.59	0.74	0.75
	MAE-SS	0.23	0.34	0.60	0.39	0.46	0.45	0.40	0.53	0.52
	CRPSS	-0.05	-0.08		0.24	0.22		0.26	0.32	
May	CC	0.839	0.890	0.966	0.519	0.712	0.882	0.766	0.786	0.953
	MSE-SS	0.50	0.67	0.93	0.35	0.63	0.78	0.63	0.72	0.91
	MAE-SS	0.29	0.40	0.76	0.22	0.42	0.55	0.42	0.51	0.71
	CRPSS	0.02	0.03		0.16	0.29		0.37	0.41	
June	CC	0.857	0.875	0.872	0.721	0.881	0.840	0.787	0.823	0.884
	MSE-SS	0.52	0.67	0.77	0.61	0.79	0.72	0.60	0.66	0.79
	MAE-SS	0.32	0.45	0.54	0.36	0.52	0.36	0.38	0.46	0.54
	CRPSS	0.28	0.36		0.29	0.45		0.39	0.48	

Initialization month	Category	Kaub / Rhine			Neu Darchau / Elbe			Hofkirchen / Danube		
		ESP	S-4	MLR	ESP	S-4	MLR	ESP	S-4	MLR
July	CC	0.677	0.672	0.955	0.609	0.604	0.832	0.378	0.412	0.518
	MSE-SS	0.15	0.27	0.91	0.18	0.50	0.69	0.01	0.18	0.15
	MAE-SS	-0.06	0.05	0.60	0.10	0.29	0.43	-0.03	0.08	0.00
	CRPSS	-0.28	-0.241		0.18	0.172		-0.02	0.02	
August	CC	0.618	0.695	0.831	0.770	0.816	0.990	0.446	0.631	0.872
	MSE-SS	0.14	0.25	0.71	0.49	0.69	0.98	0.18	0.44	0.77
	MAE-SS	0.03	0.10	0.43	0.33	0.44	0.86	0.11	0.30	0.48
	CRPSS	-0.26	-0.36		0.36	0.41		0.13	0.30	
September	CC	0.625	0.735	0.890	0.868	0.777	0.966	0.585	0.669	0.780
	MSE-SS	0.20	0.46	0.79	0.63	0.63	0.94	0.03	0.38	0.63
	MAE-SS	0.10	0.27	0.57	0.40	0.37	0.75	0.10	0.24	0.40
	CRPSS	0.06	0.12		0.39	0.19		0.18	0.18	
October	CC	0.678	0.713	0.864	0.842	0.860	0.980	0.651	0.634	0.871
	MSE-SS	0.49	0.60	0.76	0.47	0.53	0.96	0.53	0.59	0.76
	MAE-SS	0.31	0.39	0.52	0.26	0.28	0.81	0.28	0.37	0.47
	CRPSS	0.30	0.27		0.00	-0.08		0.19	0.19	
November	CC	0.653	0.802	0.908	0.866	0.866	0.952	0.763	0.870	0.918
	MSE-SS	0.47	0.70	0.82	0.72	0.76	0.91	0.64	0.80	0.85
	MAE-SS	0.33	0.48	0.52	0.43	0.47	0.59	0.43	0.54	0.48
	CRPSS	0.30	0.37		0.25	0.19		0.38	0.43	
December	CC	0.721	0.677	0.879	0.925	0.928	0.971	0.725	0.681	0.813
	MSE-SS	0.46	0.54	0.77	0.78	0.84	0.94	0.48	0.52	0.67
	MAE-SS	0.31	0.43	0.46	0.54	0.56	0.73	0.34	0.46	0.38
	CRPSS	0.27	0.30		0.51	0.39		0.30	0.30	

Figure 7. Comparison of monthly forecast skill of ESP, S4-driven and statistical forecasts at the stations Kaub, Neu Darchau, Hofkirchen predicting lowest 7-day mean flow $MoNM_7Q$ of the current month (verification period 1981–2014).

Initialization month	Category	Kaub / Rhine			Neu Darchau / Elbe			Hofkirchen / Danube		
		ESP	S-4	MLR	ESP	S-4	MLR	ESP	S-4	MLR
March	CC	0.404	0.632	0.945	0.651	0.737	0.913	0.476	0.657	0.914
	MSE-SS	0.12	0.48	0.90	0.40	0.65	0.84	0.29	0.61	0.84
	MAE-SS	0.08	0.33	0.68	0.24	0.41	0.61	0.18	0.40	0.63
	CRPSS	0.00	0.04		0.24	0.245		0.13	0.20	
June	CC	0.647	0.643	0.921	0.772	0.778	0.857	0.625	0.648	0.776
	MSE-SS	0.26	0.29	0.86	0.58	0.64	0.74	0.42	0.34	0.61
	MAE-SS	0.13	0.12	0.63	0.33	0.39	0.51	0.25	0.21	0.36
	CRPSS	0.03	-0.10		0.27	0.34		0.26	0.23	
September	CC	0.175	0.286	0.846	0.752	0.717	0.938	0.127	0.230	0.849
	MSE-SS	-0.37	0.10	0.72	0.37	0.56	0.89	-0.33	0.31	0.72
	MAE-SS	-0.10	0.08	0.40	0.24	0.37	0.65	-0.11	0.18	0.40
	CRPSS	0.02	-0.14		0.29	0.14		-0.01	-0.03	
December	CC	0.451	0.528	0.861	0.724	0.751	0.948	0.525	0.543	0.809
	MSE-SS	0.04	0.19	0.76	0.52	0.63	0.90	0.28	0.29	0.67
	MAE-SS	0.01	0.08	0.52	0.25	0.36	0.68	0.14	0.17	0.40
	CRPSS	0.07	-0.04		0.27	0.27		0.12	0.02	

Figure 8. Comparison of forecast skill of ESP, ECMWF-driven and statistical forecasts at the stations Kaub, Neu Darchau, Hofkirchen predicting mean flow of the next 3 months 3MoMQ initialized in March, June, September and December (verification period 1981–2014).

The first forecast months especially show an improved skill. Unfortunately, the improvements are less pronounced in the typical low-flow season (July–October) particularly relevant for IWT.

As second forecast approach, the multiple linear regression model (MLR) based on the stability analysis (Sect. 3.3) was implemented. In Fig. 7 all three forecast approaches currently implemented in the forecast framework are compared by different skill metrics (see Sect. 3.4) for the period 1981 to 2014 for the first forecast month and initializations. The monthly $MoNM_7Q$ was chosen as forecast variable because it shows slightly more robust forecast results as the monthly MoMQ. From Fig. 7 it is evident that the statistical approach is able to further improve forecast quality of the dynamical approach. Although, forecast skill is rather fluctuating for ESP and S4, it is significantly more stable for the statistical approach. There is still a decrease in forecast skill within the typical low-flow period in late summer to autumn, but still on a proper level.

The results for the seasonal forecasts (tri-monthly mean discharge of the upcoming meteorological season) shown in Fig. 8 confirm the findings described above. While forecast skill (see Sect. 3.4) based on ESP and S4 at the Rhine and especially the Danube is comparatively low when compared to climatology, the skill for the Elbe turns out to be significantly higher. The statistical approach leads to a sustainable increase in forecast skill at all waterways, even the skill for the Elbe could be further increased. Based on the skill metrics applied, the statistical approach shows the best results overall.

The inter-comparison of the forecast approaches, based on the relative economic value, was conducted for $MoNM_7Q$. In order to calculate the relative economic value, three discharge thresholds have been selected to generate the categorical forecast: the median of the observed NM_7Q of the partic-
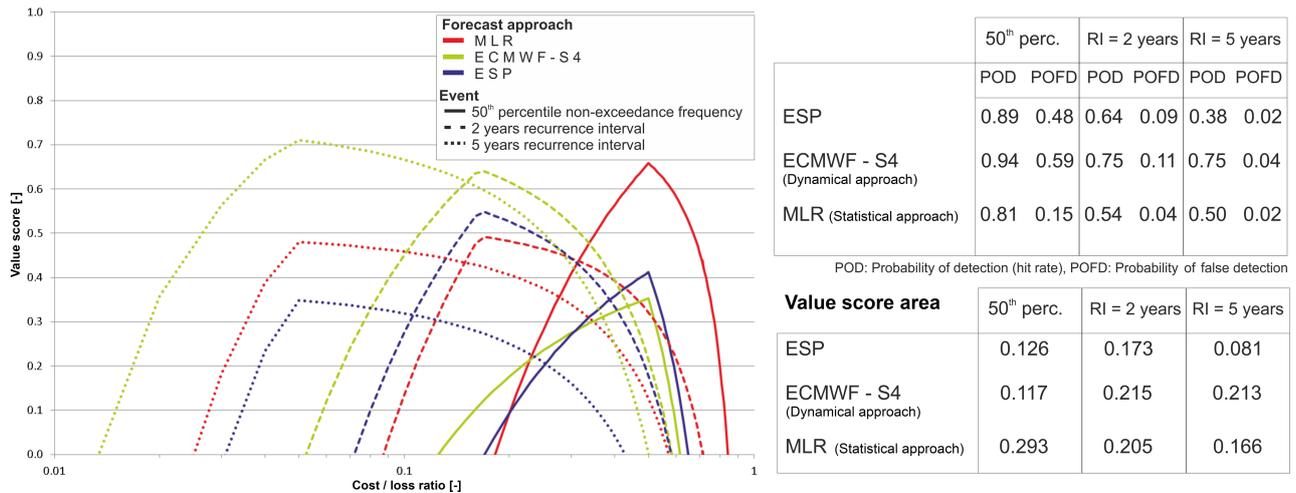
		50th perc.		RI = 2 years		RI = 5 years	
		POD	POFD	POD	POFD	POD	POFD
ESP		0.89	0.48	0.64	0.09	0.38	0.02
ECMWF - S4 (Dynamical approach)		0.94	0.59	0.75	0.11	0.75	0.04
MLR (Statistical approach)		0.81	0.15	0.54	0.04	0.50	0.02

POD: Probability of detection (hit rate), POFD: Probability of false detection

Value score area	50th perc.	RI = 2 years	RI = 5 years
ESP	0.126	0.173	0.081
ECMWF - S4 (Dynamical approach)	0.117	0.215	0.213
MLR (Statistical approach)	0.293	0.205	0.166

Figure 9. Economic value score plotted against cost/loss ratio for three forecast approaches predicting the lowest 7-day mean MoNM$_7$Q of the current month at the gauge Kaub/Rhine for three different event thresholds within the typical low-flow season of July to November (verification period 1981–2014).

ular month the forecast is issued for (period 1951–2014), the NM$_7$Q with a recurrence interval (RI) of 2 and 5 years. The recurrence intervals have been calculated based on the time series 1961–2015 using a Weibull-3 distribution fitted by a method of moments. The relative economic value was examined for the non-exceedance of the aforementioned thresholds. In Fig. 9 the relative economic value is shown for the station Kaub (Rhine) and the different forecast approaches as a function of the cost/loss ratio.

The value was calculated for the relevant low-flow season between July and November usually affecting IWT along the River Rhine within the period 1981–2014. Overall the three approaches provide positive economic values for a wide range of cost/loss situations, but the economic value considerably varies amongst the forecast methods as well as the selected events. For the statistical approach the economic values decrease with decreasing return period, while the S4-driven dynamical approach achieves stable or even higher economic values for more extreme low-flow events. The S4-driven forecasts show the highest POD for all events, but especially for the highest threshold (50th percentile), the POFD is quite high (0.59) too. In this case, the median do not seem to be the optimal representative for the ensemble and choosing another quantile might lead to better economic values. Also, the ESP forecast suffer from a comparatively high POFD. For the more rare low-flow events, the POFD of the S4-driven forecast significantly drops and is relatively close to the one from ESP and MLR forecasts, while the POD stays the best. For all events and cost/loss situations the benchmark approach (ESP) could be improved at least by one of the two alternative approaches.

An inter-comparison of the economic value between the three waterways based on the statistical forecast approach is shown in Fig. 10. Three different thresholds based on the

50th, 25th and 10th percentiles have been selected. For the event occurring most often (50th percentile threshold), the forecasts for the Elbe river produce the highest economic value, which corresponds to the best forecast skill achieved for Neu Darchau when compared to Kaub and Hofkirchen (see Fig. 7). Overall, the forecasts for the Danube provide the lowest relative economic values for all selected thresholds. Nevertheless, the values are still positive (that means added value compared to the currently used climatological forecast) for a wide range of cost/loss situations. For more extreme low-flow events, the economic values for the Rhine and Elbe get closer and for the 10th percentile threshold, the forecast at the Rhine reaches slightly higher relative economic values than the ones at the Elbe (see also the value score area).

As the most recent step in setting up the monthly to seasonal forecast framework, we tested the combination of the statistical approach with the ESP method. This combined method does not require seasonal meteorological forecasts (and its respective post-processing to force the hydrological model), but it might benefit from the ability of the hydrological model to emulate the initial hydrological conditions prior to the forecast. Therefore, we added the ESP benchmark forecasts for NM$_7$Q at the gauge Kaub/Rhine into the corresponding final statistical forecast model. Figure 11 contains the selected skill measures comparing the basic statistical model (MLR) with the one extended by ESP results as additional predictor (MLR + ESP) for three initialization months: June as the month with the highest ESP skill, September as one of the months showing the lowest skill of the ESP forecasts and November, which shows an intermediate ESP skill (see Fig. 7). The forecasts of NM$_7$Q in June and November significantly benefit from the ESP forecasts as all measures indicate. For the September forecasts, where ESP shows a relative low skill, the combined approach gives comparable
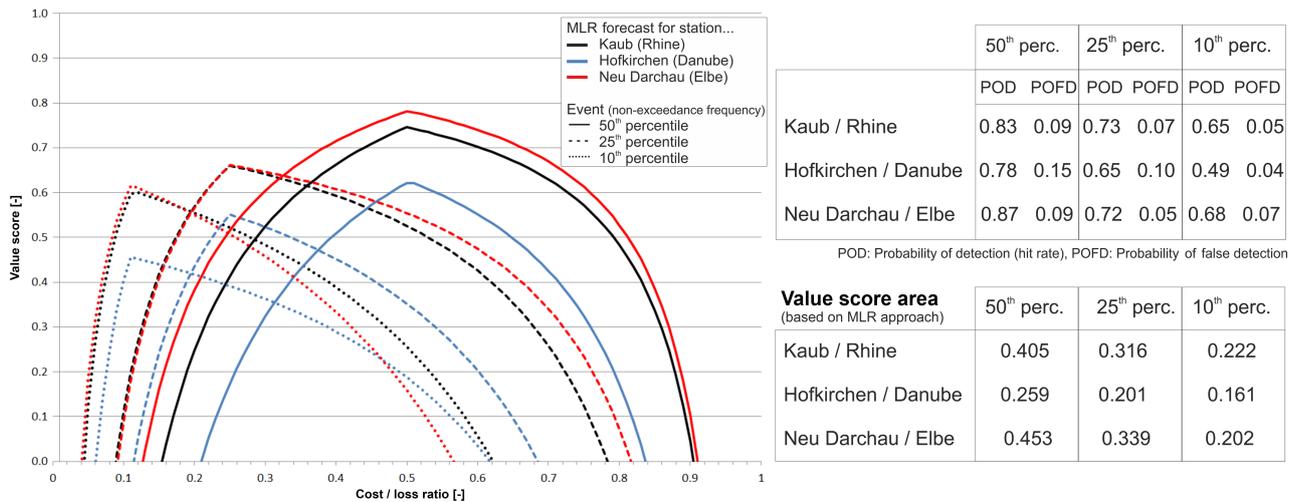
	50th perc.		25th perc.		10th perc.	
	POD	POFD	POD	POFD	POD	POFD
Kaub / Rhine	0.83	0.09	0.73	0.07	0.65	0.05
Hofkirchen / Danube	0.78	0.15	0.65	0.10	0.49	0.04
Neu Darchau / Elbe	0.87	0.09	0.72	0.05	0.68	0.07

POD: Probability of detection (hit rate), POFD: Probability of false detection

Value score area (based on MLR approach)	50th perc.	25th perc.	10th perc.
Kaub / Rhine	0.405	0.316	0.222
Hofkirchen / Danube	0.259	0.201	0.161
Neu Darchau / Elbe	0.453	0.339	0.202

Figure 10. Economic value score plotted against cost/loss ratio for the forecast of the lowest 7-day mean flow MoNM$_7$Q of the current month at the gauges Kaub/Rhine, Hofkirchen/Danube and Neu Darchau/Elbe based on the statistical approach (verification period 1951–2014).

results as the basic statistical approach. So, adding (low skill) ESP results as predictor does not improve forecast skill, as expected but neither does it deteriorate the skill.

4.3 Evaluation for a significant low-flow event

In 2015 a long-lasting drought hit Europe, especially affecting its central and eastern part, where it was one of the worst drought events since the major droughts of 1976 and 2003 (van Lanen et al., 2016; Ionita et al., 2017; Laaha et al., 2017). Large-scale deficits in precipitation in combination with high evapotranspiration losses led to deficit in soil moisture and subsequently manifested itself as a long-lasting hydrological drought, with low water levels and deficits in streamflow along several major European rivers. The 2015 drought showed numerous socio-economic impacts, like constraints in drinking water supply, energy production and agriculture. Also IWT was significantly impacted, especially in the second half of 2015, notably in France, the Netherlands, Germany and eastern Europe. In Germany, load losses on the Rhine, Danube, Elbe, Odra and Weser rivers and in Russia on the Don River were up to 50 % (van Lanen et al., 2016).

At the beginning of 2015, the aforementioned forecast framework was in place, at least for off-line use to support advisory activities. The forecasts could be issued within the first days of the particular month or the particular seasons as soon as all input data and predictands became available. The monthly MoNM$_7$Q for the year 2015 at the station Kaub/Rhine is plotted in Fig. 12, together with the climatology (mean, selected percentiles) and the forecasts (mean, uncertainty range between the 5th and 95th percentiles) based on the ESP method, the dynamical approach (S4) and the statistical approach (MLR). Within the first half of 2015, the already ongoing meteorological drought was not yet visible along the Rhine and the observed values were quite close to

Initialization month	Score	Kaub / Rhine				
		ESP	S-4	MLR	MLR + ESP	MLR excl. Q
June	CC	0.857	0.875	0.872	0.913	0.751
	MSE-SS	0.52	0.67	0.77	0.84	0.58
	MAE-SS	0.32	0.45	0.54	0.62	0.33
	CRPSS	0.28	0.36			
September	CC	0.625	0.735	0.890	0.889	0.830
	MSE-SS	0.20	0.46	0.79	0.79	0.69
	MAE-SS	0.10	0.27	0.57	0.57	0.45
	CRPSS	0.06	0.12			
November	CC	0.653	0.802	0.908	0.917	0.865
	MSE-SS	0.47	0.70	0.82	0.84	0.75
	MAE-SS	0.33	0.48	0.52	0.54	0.49
	CRPSS	0.30	0.37			

Figure 11. Comparison of forecast skill of the statistical forecast approach using ESP results as additional predictor for June and September forecasts at station Kaub/Rhine (predictand lowest 7-day mean flow MoNM$_7$Q of the current month, verification period 1981–2014).

the climatological mean (except for January). As expected, in this period the use of a climatological forecast produced good results. From July onwards, the flow dropped significantly for the rest of the year. This change was predicted most accurately by the statistical approach. In the subsequent months, all methods provided meaningful forecasts, while ESP and the dynamical approach are slightly advantaged compared to the statistical method in this particular situation. The statistically based forecasts tended to underestimate the MoNM$_7$Q, with the September forecast being significantly too low. Although the range of the 5th to 95th percentiles significantly differs in the first half of 2015, with the ESP-based forecasts showing the widest range, the ranges converge in the second very dry half of the year 2015 and all approaches show similar ranges.

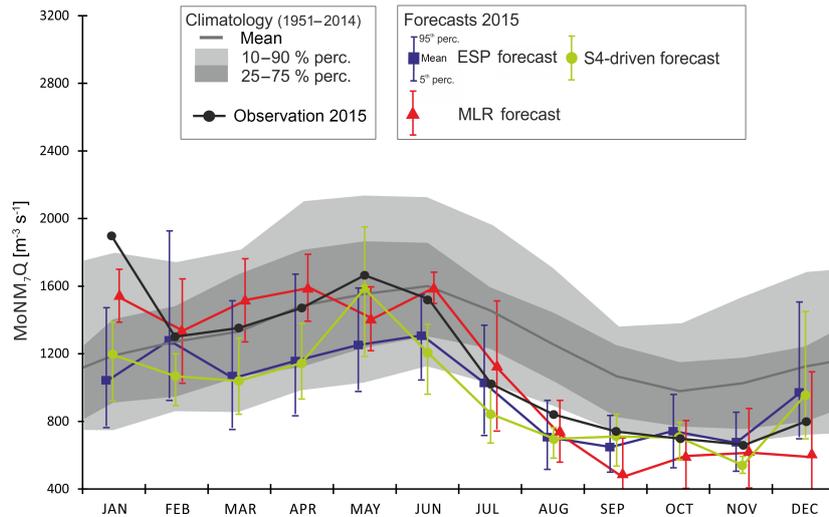Regarding the tri-monthly forecast issued at the beginning of the particular meteorological season, the statistical

Figure 12. Comparison between the three navigation-related forecast approaches for the lowest 7-day mean $MoNM_7Q$ of the current month in the year 2015 at gauge Kaub/Rhine in relation to the observed values and the climatology (1951–2014).
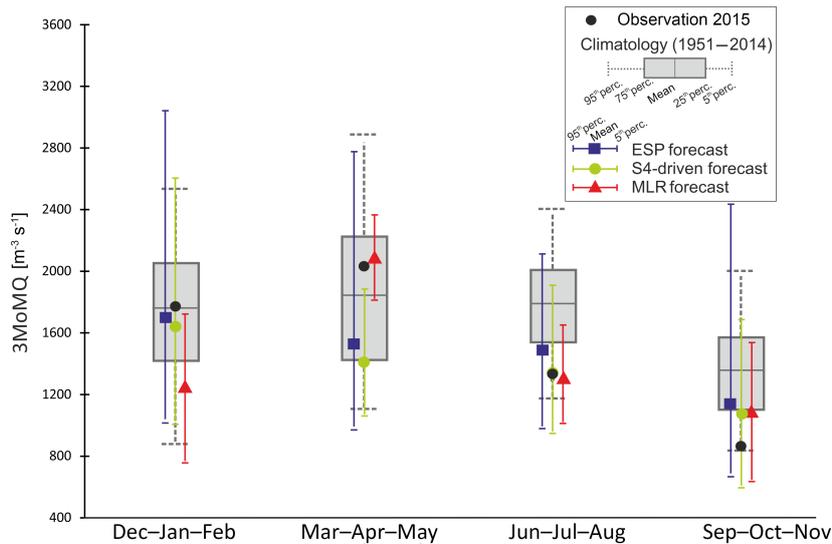

Figure 13. Comparison between the three navigation-related forecast approaches for the monthly mean flow of the next three months 3MoMQ initialized in December 2014, March 2015, June 2015, September 2015 (meteorological seasons 2015) at gauge Kaub/Rhine in relation to the observed values and the climatology (1951–2014).

approach significantly outperformed the other methods in spring (Fig. 13). For summer (JJA) and autumn (SON) the forecasts of all approaches are relatively similar. But the statistical approach and the S4-driven forecast slightly outperformed the ESP-based results. For the autumn season, all approaches overestimated the observed value (up to 25 %), but they at least indicate below-average conditions, which is already a valuable information for the navigational users. Regarding the range of the 5th to 95th percentiles, the statistical approach produces the sharpest forecast with the narrowest ranges. Except for the winter forecast (DJF), the observed values fall in the range of the statistical approach. Although

all approaches overestimated the flow in autumn, the measured values still fall into the 5th to 95th percentiles range of all approaches.

The performance of the seasonal forecasts for the waterways Elbe and Danube is shown in Fig. 14. These results approve of the long-term evaluation. The seasonal forecast results for the Elbe were markedly good, particularly for the statistical and the S4-driven approach producing nearly perfect predictions of MAM and SON flows. For JJA the flow was overestimated by all methods, but the forecasts consistently indicated below the long-term average conditions. The range between the 5th and 95th percentile are similar
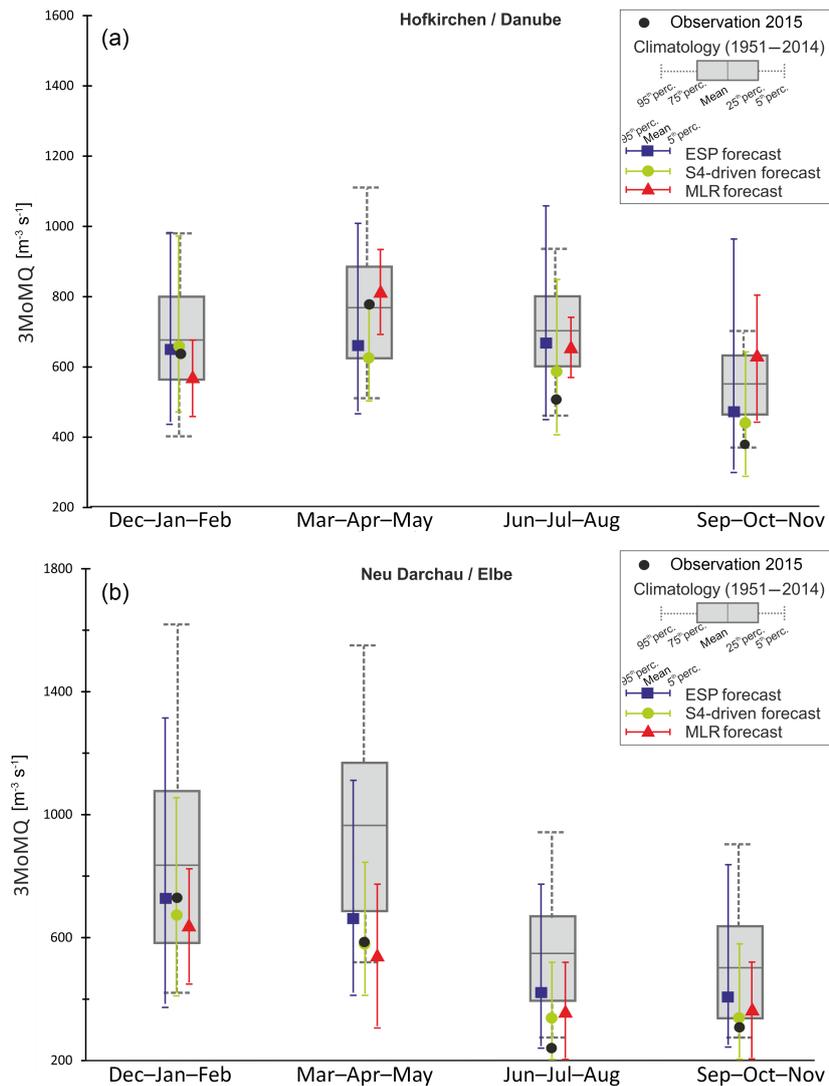
Figure 14. Comparison between the statistical forecast approaches for the monthly mean flow of the next 3 months 3MoMQ initialized in December 2014, March 2015, June 2015, September 2015 at gauges Hofkirchen/Danube **(a)** and Neu Darchau/Elbe **(b)** in relation to the observed values and the climatology (1951–2014).

for the statistical and the S4-driven forecast, while the ESP-based forecast is significantly less sharp. For the 2015 event at the Danube, gauge Hofkirchen, the S4-driven dynamical approach produced the best forecast results, except for the spring season. Although the dynamical approach still overestimated the flows in summer and autumn, it was a good indicator for the significant low-flow situation observed. The measured value was covered by the 5th to 95th percentile range at least. The statistical approach especially failed to predict these low values. Even the 5th to 95th percentile range did not cover the measured values in summer and autumn.

5 Discussion

The increasing number of seasonal hydrological forecasting systems (see e.g. Zappa et al., 2014; Olsson et al., 2016; Bell et al., 2017) proves the need as well as the feasibility of long-range predictions over Europe despite the generally limited hydro-meteorological predictability when compared to other continents. Furthermore, the European Flood Awareness System EFAS provides a continental-wide seasonal hydrological outlook with a lead time of 2 months of weekly river flow anomalies and its probability for large European regions since 2016. The work presented in this paper fit into this progress and the results prove possibilities and limitations with regard to monthly to seasonal forecasting along Germany's major inland waterways. Our findings sup-

Table 3. Statistics for the forecast models based on NCEP and GDM soil moisture data for forecasted lowest 7-day mean flow $MoNM_7Q$ of June at gauge Kaub/Rhine (period 1954–2014).

Skill measure	NCEP	GDM	NCEP	GDM
	Soil data – May		Soil data – April	
Pearson correlation coefficient (–)	0.43	0.71	0.50	0.60
Coefficient of determination (–)	0.19	0.51	0.25	0.36
p value (–)	3.826×10^{-4}	6.523×10^{-11}	4.120×10^{-5}	2.248×10^{-7}
Residual standard error ($m^3\,s^{-1}$)	373	290	361	332

port previous studies (e.g. by Yossef et al., 2013) which indicate that meteorological forcings dominate initial hydrological conditions for many European rivers beyond lead times of 1–2 months, within specific seasons already for shorter lead times. Furthermore, the study reveals that for setting-up a specific forecasting system tailored to particular user needs the heterogeneity within Europe requires a basin-wise as well as a use-oriented consideration of forecast sensitivities and forecast approaches. Already the differences between the Rhine, upper Danube and Elbe are significant in order to identify skilful initialization months and corresponding forecast approaches (see Sect. 4.1 and 4.2). Differences within Europe become even more apparent when comparing our findings for the German waterways with related studies in other parts of the continent. For example Olsson et al. (2016) analysed a quite similar framework of competing forecast approaches for predicting the spring-flood volume relevant for hydro power generation in a Swedish river basin. Here, none of the approaches tested (conditional ESP approach, dynamical approach using the HBV model forced with ECMWF seasonal meteorological forecasts, a statistical approach based on the correlation of large-scale circulation variables) consistently outperformed the ESP-based method currently applied operationally. This proves the importance of initial hydrological conditions, primarily snow pack, for this forecasting service. In contrast, along the German waterways the dynamical and the statistical approach showed significantly higher skill (and economic value) for the majority of initialization months and lead times than the ESP forecasts, which mainly reveals the different effect of initial hydrological conditions for long-range forecasting in both regions. For parts of Europe, like Scandinavia (Olsson et al., 2016) or the Alpine region (Fundel et al., 2013; Jörg-Hess et al., 2015), snow pack accumulated over the season prior to the initialization of a forecast could significantly dominate the predicted flow for several months, at least within specific seasons. Even for the central European rivers initial hydrological conditions could affect forecast skill on monthly to seasonal timescales (see Sect. 4.2), but compared to the aforementioned regions the signal is noticeably masked by the meteorological impact within the forecast period. Furthermore over the course of the year other elements of initial hydrological conditions could have a stronger influence on

flows than snow. Soil moisture and anthropogenic storages, along the Rhine, upper Danube and Elbe (mainly dam management and lake regulation), especially act as buffers to the meteorological input. But in order to benefit from this "hydrological memory" in forecasting, the relevant processes or information have to be included adequately in the specific approaches.

Although data on soil moisture is already used as predictor for the statistical approach applied in this study (see Table 2), no explicit information on snow pack or on any kind of anthropogenic storage or water transfer between catchments is used so far. Although snow pack is at least implicitly considered via precipitation and temperature, no information on filling levels or released volumes of lakes etc. are used as input to the statistical model. For snow pack, potentially useful data products are operationally available (Alverado Montero et al., 2016), this might be more difficult for lake or dam-related data due to legal restrictions often imposed by hydropower companies. In order to evaluate and to demonstrate the additional predictive skill of improved data products representing initial hydrological conditions, e.g. the soil moisture, we modified the statistical approach. The results presented in the preceding sections are based on the global reanalysis product provided by NCEP (see Table 2) with a spatial resolution of 2.5°, which is relatively coarse for most of the central European catchments with their spatial heterogeneity. Therefore we tested an alternative soil moisture information for Germany with a resolution of 4 km by 4 km, which is provided operationally by the German drought monitor (GDM, www.ufz.de/droughtmonitor) since 2014 (Zink et al., 2016). In order to evaluate the sensitivity of the forecast skill to the different soil moisture data, we evaluated the results for the forecasts of $MoNM_7Q$ at gauge Kaub/Rhine solely based on the two different datasets (Table 3). We have tested the model with 1 and 2 month lags, resulting in four forecast models using the April and May soil moisture data of the respective source. As the metrics indicate, using the soil moisture data from the GDM could further increase the skill of the forecast considerably.

For the dynamical approaches applied in this study, snow pack and soil moisture are modelled by the hydrological model LARSIM-ME. Anthropogenic storages and water transfers, at least the most relevant ones, are implemented

in the hydrological model for the Rhine, Elbe and upper Danube too. The hydrological processes like snow accumulation/snow melting are modelled in a simplified manner and the management rules had to be idealized. Furthermore the real-world regulations of any measure usually differ from theoretical rules, and so in this regard two important aspects to better exploit forecast skill hidden in the hydrological system exist. On the one hand the implementation of anthropogenic storages and water transfers in the hydrological model should be completed, together with a better definition of their regulation. Additionally the use of operationally available data products representing snow pack, soil moisture, lake levels etc. via data assimilation techniques is an important issue to improve the internal states of the hydrological model. The amount and quality of useful data to be assimilated is continuously increasing, especially pushed by satellite missions in recent years (Jörg-Hess et al., 2015; Alverado Montero et al., 2016).

Another aspect towards an improved forecast skill, which is briefly touched on by our study and which poses a similarity to the findings by Olsson et al. (2016), is the combination of different forecast approaches. Olsson et al. (2016) tested two pragmatic ways of combining multiple forecasts in a post-processing step. This finding corresponds to a first attempt in our study testing the combination of the statistical approach with the ESP method, but in a more direct way of coupling. As shown in Sect. 4.2 the forecast skill could noticeably increase by merging both approaches, but the added value depends on the skill of the ESP approach. An advantage of the combination procedure we used is that by merging both forecasts, forecast skill does not deteriorate even in the case of low ESP skill. Therefore, we see different aspects and potential ways to further improve forecast skill even within central Europe. Additionally, the dynamical approaches using seasonal meteorological forecasts may benefit from future improvements achieved in the area of climate and seasonal meteorological forecast skill, potentially even with a disproportionate impact (Wood et al., 2016).

6 Conclusions and prospects

This paper asses the skill and economic value of a monthly to seasonal forecasting system set-up for the main European waterways the Rhine, upper Danube and Elbe over a 35-year hindcast periods with a specific focus on the 2015 drought event. IWT along these free-flowing rivers is prone to (long-lasting) droughts leading to low streamflows and low water levels, which subsequently limit the maximum cargo-carrying capacity of vessels, increase their energy consumption and their time of travel. Monthly to seasonal hydrological forecasting services are one measure to cope with these impacts and to sustainably increase IWT efficiency as well as to support medium- to long-range waterway management. Despite the overall limited hydro-meteorological predictabil-

ity in central Europe, the results of the different forecast approaches tested reveal the existence of a valuable predictability of streamflow on monthly and to some extent even up to seasonal timescales along the major waterways in Germany. We found that the skill of the meteorological forcing has a larger effect than the quality of initial hydrological conditions on seasonal forecast quality for relevant stations along the German waterways. Just for a few initialization months, the initial hydrological conditions could noticeably affect forecast skill for the first forecast month. Nevertheless a good estimation of the initial hydrological conditions (especially soil moisture, snow pack, groundwater storage, but also human activities like regulated lakes/dams) forms the basis of any monthly to seasonal hydrological forecast. Therefore, for future development we have in mind the improved integration of such kinds of information off-line in the course of setting-up hydrological models as well as in real-time by data assimilation techniques.

The more physically based and the statistical approach are able to improve the predictive skills and economic value compared to the climatology and the ESP approaches. The forecast intercomparison showed that the specific forecast skill along the German waterways highly depends on the forecast location, the lead time and the season. Overall, the statistical approach currently seems to be most skilful for the three waterways investigated. The lagged relationship between the monthly or seasonal streamflow and the climatic or oceanic variables vary between 1 month (e.g. local precipitation and temperature, soil moisture) up to 6 months (e.g. sea surface temperature). However, we also observed that in some situations the statistical approach forecasts extremely low streamflows, while the more physically based approaches respond more moderately. This effect needs further investigation.

Based on the present study two additional aspects will become a focal point in the near-future: (i) besides improving the individual forecast methods to tap their full potential (e.g. increase of input data resolution, enhancement of the hydrological model) an optimal combination, in terms of hybrid or hierarchical procedures, will be investigated systematically. (ii) In close cooperation with the users, monthly to seasonal forecast products have to be designed in order to push the usage of such forecast information in decision-making processes and in order to communicate the value and uncertainty associated with such forecasts in a transparent way.

As such, the results shown in this paper represent the basis for setting up such an operational service, which will sustainably extend the existing forecast portfolio for waterway users and meet the growing needs in order to increase the modal share of water-borne transportation.

The climate and oceanic data used in the statistical forecast approach are publicly available. Gridded precipitation and temperature data are extracted from the E-OBS version 8 dataset (Haylock et al., 2008) obtained via http://www.ecad.eu/download/ensembles/ensembles.php.

Precipitation and temperature data, at country level and for each German federal state, are extracted from the ftp server of the German Meteorological Service DWD: ftp://ftp-cdc.dwd.de/pub/CDC/.

The Extended Reconstructed Sea Surface Temperature dataset (ERSST v4; Huang et al., 2014) was downloaded from https://www.ncdc.noaa.gov/data-access/marineocean-data/extended-reconstructed-sea-surface-temperature-ersst-v4.

The Volumetric Soil Moisture data are based on the NCEP/NCAR reanalysis product R-1 (Kistler et al., 2001), which was downloaded from https://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.derived.surfaceflux.html.

Datasets for soil moisture, geopotential height, sea level pressure and relative humidity are provided by the National Centre for Atmospheric Research (NCAR, Kalnay et al., 1996) via http://iridl.ldeo.columbia.edu/SOURCES/.NOAA/.NCEP-NCAR/.CDAS-1/.MONTHLY/.

In addition to the E-OBS dataset, the ERA-Interim dataset was used (Dee et al., 2011). It is a public dataset offered by ECMWF: http://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/.

The HYRAS dataset, used to calibrate the hydrological model LARSIM, is not currently publicly available. It is managed by the German Meteorological service (DWD: hydromet@dwd.de).

The seasonal meteorological forecasts from ECMWF (Molteni et al., 2011) are not public, but depending on who you are, different ways and licenses in order to access the data are offered (http://www.ecmwf.int/en/forecasts/accessing-forecasts).

and the data providers in the ECA&D project (http://www.ecad.eu). Surface solar radiation downwards from the ERA-Interim data were provided courtesy of ECMWF. Streamflow and water-level data were provided by the Federal Waterways and Shipping Administration (WSV), the Federal Institute of Hydrology (BfG; gauges Maxau, Kaub, Ruhrort, Hofkirchen, Dresden, Magdeburg-Strombrücke, Neu Darchau), ViaDonau (gauges Kienstock and Wildungsmauer), and the Ministry of Interior, the General Directorate of Water Management, Republic of Hungary (gauge Nagymaros).

References

Alverado Montero, R., Schwanenberg, D., Krahe, P., Lisniak, D., Sensoy, A., Sorman, A. A., and Akkol, B.: Moving horizon estimation for assimilating H-SAF remote sensing data into the HBV hydrological model, Adv. Water Resour., 92, 248–257, https://doi.org/10.1016/j.advwatres.2016.04.011, 2016.

Bell, V. A., Davies, H. N., Kay, A. L., Brookshaw, A., and Scaife, A. A.: A national-scale seasonal hydrological forecast system: development and evaluation over Britain, Hydrol. Earth Syst. Sci., 21, 4681–4691, https://doi.org/10.5194/hess-21-4681-2017, 2017.

Beven, K. J.: Equifinality and Uncertainty in Geomorphological Modelling, in: The Scientific Nature of Geomorphology, edited by: Rhoads, B. L. and Thorn, C. E., Wiley, Chichester, UK, 289–313, 1996.

CCNR: Annual Report 2016 – Inland Navigation in Europe – Market Observation, Central Commission for the Navigation of the Rhine, Strasbourg, France, 2016.

Crochemore, L., Ramos, M.-H., and Pappenberger, F.: Bias correcting precipitation forecasts to improve the skill of seasonal streamflow forecasts, Hydrol. Earth Syst. Sci., 20, 3601–3618, https://doi.org/10.5194/hess-20-3601-2016, 2016.

Day, G. N.: Extended streamflow forecasting using NWSRFS, J. Water Resour. Plan. Man., 111, 157–170, 1985.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Holm, E. V., Isaksen, L., Kallberg, P., Kohler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J. N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. Roy. Meteorol. Soc., 137, 553–597, 2011.

Demirel, M. C., Booij, M. J., and Hoekstra, A. Y.: Identification of appropriate lags and temporal resolutions for low flow indicators in the River Rhine to forecast low flows with different lead times, Hydrol. Proc., 27, 2742–2758, https://doi.org/10.1002/hyp.9402, 2012.

Demirel, M. C., Booij, M. J., and Hoekstra, A. Y.: Effect of different uncertainty sources on the skill of 10 day ensemble low flow

forecasts for two hydrological models, Water Resour. Res., 49, 4035–4053, https://doi.org/10.1002/wrcr.20294, 2013.

Demirel, M. C., Booij, M. J., and Hoekstra, A. Y.: The skill of seasonal ensemble low-flow forecasts in the Moselle River for three different hydrological models, Hydrol. Earth Syst. Sci., 19, 275–291, https://doi.org/10.5194/hess-19-275-2015, 2015.

Domeisen, D. I. V., Butler, A. H., Fröhlich, K., Bittner, M., Müller, W. A., and Baehr, J.: Seasonal Predictability over Europe Arising from El Niño and Stratospheric Variability in the MPI-ESM Seasonal Prediction System, J. Climate, 28, 256–271, https://doi.org/10.1175/JCLI-D-14-00207.1, 2015.

DWD: Deliverable D.1.7 Gridded 20 yr dataset of surface solar radiation, cloud albedo, cloud fraction and surface albedo derived from MVIRI (SEVIRI) using the Heliosat method (0.05°), EU funded project EURO4M, EURO4M project report, available at http://www.euro4m.eu/downloads/D1.7_Gridded-20yr-dataset-of-SIS-SAL-CFC-CAL-derived-from-MVIRI-SEVIRI.pdf (last access: 11 December 2017), 2013.

European Commission: White paper – European transport policy for 2010: time to decide, COM(2001)370, Brussels, available at: https://ec.europa.eu/transport/sites/transport/files/themes/strategies/doc/2001_white_paper/lb_com_2001_0370_en.pdf (last access: 11 December 2017), 2001.

Fundel, F. and Zappa, M.: Hydrological Ensemble Forecasting in Mesoscale Catchments: Sensitivity to Initial Conditions and Value of Reforecasts, Water Resour. Res., 47, W09520, https://doi.org/10.1029/2010WR009996, 2011.

Fundel, F., Jörg-Hess, S., and Zappa, M.: Monthly hydrometeorological ensemble prediction of streamflow droughts and corresponding drought indices, Hydrol. Earth Syst. Sci., 17, 395–407, https://doi.org/10.5194/hess-17-395-2013, 2013.

Funk, D., Pouget, L., Dubus, L., Falloon, P., Meißner, D., Klein, B., Palin, E., Viel, C., Foster, K., Lootvoet, M., Bosi, L., Creswick, J., Davis, M., and Jimenez, I.: White paper on sector specific vulnerabilities, Deliverable 11.2, EUPORIAS – European Provision Of Regional Impacts Assessments on Seasonal and Decadal Timescales, Grant Agreement 308291, EUPORIAS project report, available at: http://www.euporias.eu/system/files/D11.2_Final.pdf (last access: 11 December 2017), 2015.

Gámiz-Fortis, S., Pozo-Vázquez, D., Trigo, R. M., and Castro-Díez, Y.: Quantifying the predictability of winter river flow in Iberia, Part II: seasonal predictability, J. Climate, 21, 2503–2518, 2008.

Gelfan, A. N., Motovilov, Yu. G., and Moreido, V. M.: Ensemble seasonal forecast of extreme water inflow into a large reservoir, Proc. IAHS, 369, 115–120, https://doi.org/10.5194/piahs-369-115-2015, 2015.

Haag, I., Johst, M., Sieber, A., and Bremicker, M.: Guideline for the Calibration of LARSIM Water Balance Models for operational Application in Flood Forecasting–Calibration Guide, available at: http://hmdblog.rlp.de/luwg/larsim/, last access: 3 March 2016.

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., and New, M. A.: European daily high-resolution gridded dataset of surface temperature and precipitation, J. Geophys. Res.-Atmos., 113, D20119, https://doi.org/10.1029/2008JD010201, 2008.

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

Huang, B., Banzon, V. F., Freeman, E., Lawrimore, J., Liu, W., Peterson, T. C., Smith, T. M., Thorne, P. W., Woodruff, S. D., and Zhang, H.-M.: Extended Reconstructed Sea Surface Temperature version 4 (ERSST.v4): Part I, Upgrades and intercomparisons, J. Climate, 28, 911–930, https://doi.org/10.1175/JCLI-D-14-00006.1, 2014.

Leander, R. and Buishand, T.: Resampling of regional climate model output forthe simulation of extreme river flows, J. Hydrol., 332, 487–496, 2007.

Ionita, M., Lohmann, G., and Rimbu, N.: Prediction of Elbe discharge based on stable teleconnections with winter global temperature and precipitation, J. Climate, 21, 6215–6226, 2008.

Ionita, M., Lohmann, G., Rimbu, N., and Chelcea, S.: Interannual Variability of Rhine River Streamflow and Its Relationship with Large-Scale Anomaly Patterns in Spring and Autumn, J. Hydrometeorol., 13, 172–188, 2012.

Ionita, M., Dima, M., Lohmann, G., Scholz, P., and Rimbu, N.: Predicting the June 2013 European Flooding based on Precipitation, Soil Moisture and Sea Level Pressure, J. Hydrometeorol., 16, 598–614, https://doi.org/10.1175/JHM-D-14-0156.1, 2014.

Ionita, M., Tallaksen, L. M., Kingston, D. G., Stagge, J. H., Laaha, G., Van Lanen, H. A. J., Scholz, P., Chelcea, S. M., and Haslinger, K.: The European 2015 drought from a climatological perspective, Hydrol. Earth Syst. Sci., 21, 1397–1419, https://doi.org/10.5194/hess-21-1397-2017, 2017.

Jörg-Hess, S., Griessinger, N., and Zappa, M.: Probabilistic Forecasts of Snow Water Equivalent and Runoff in Mountainous Areas, J. Hydrometeorol., 16, 2169–2186, https://doi.org/10.1175/JHM-D-14-0193.1, 2015.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, B. Am. Meteorol. Soc., 77, 437–470, 1996.

Kistler, R., Collins, W., Saha, S., White, G., Woollen, J., Kalnay, E., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., van den Dool, H., Jenne, R., and Fiorino, M.: The NCEP-NCAR 50-Year Reanalysis: Monthly Means CD-ROM and Documentation, B. Am. Meteorol. Soc., 82, 247–267, https://doi.org/10.1175/1520-0477(2001)082<0247:TNNYRM>2.3.CO;2, 2001.

Klein, B. and Meißner, D.: Vulnerability of Inland Waterway Transport and Waterway Management on Hydro-meteorological Extremes, Deliverable 9.1, IMPREX (Improving Predictions of Hydrological Extremes), Grant Agreement Number 641811, IMPREX project report, available at: http://www.imprex.eu/system/files/generated/files/resource/d9-1-imprex-v2-0.pdf (last access: 11 December 2017), 2016.

Laaha, G., Gauster, T., Tallaksen, L. M., Vidal, J.-P., Stahl, K., Prudhomme, C., Heudorfer, B., Vlnas, R., Ionita, M., Van Lanen, H. A. J., Adler, M.-J., Caillouet, L., Delus, C., Fendekova, M., Gailliez, S., Hannaford, J., Kingston, D., Van Loon, A. F., Mediero, L., Osuch, M., Romanowicz, R., Sauquet, E., Stagge, J. H., and Wong, W. K.: The European 2015 drought from a hydrological perspective, Hydrol. Earth Syst. Sci., 21, 3001–3024, https://doi.org/10.5194/hess-21-3001-2017, 2017.

Lenderink, G., Buishand, A., and van Deursen, W.: Estimates of future discharges of the river Rhine using two scenario methodolo-

gies: direct versus delta approach, Hydrol. Earth Syst. Sci., 11, 1145–1159, https://doi.org/10.5194/hess-11-1145-2007, 2007.

Lohmann, G., Rimbu, N., and Dima, M.: Where can the Arctic oscillation be reconstructed? Towards a reconstruction of climate modes based on stable teleconnections, Clim. Past Discuss., https://doi.org/10.5194/cpd-1-17-2005, 2005.

Ludwig, K. and Bremicker, M.: The Water Balance Model LAR-SIM, Design, Content and Applications, edited by: Leibundgut, C., Demuth, S., and Lange, J., Freiburger Schriften zur Hydrologie, Institut für Hydrologie, Universität Freiburg im Breisgau, Freiburg, 141 pp., 2006.

Marke, T.: Development and Application of a Model Interface to couple Land Surface Models with Regional Climate Models for Climate Change Risk Assessment in the Upper Danube Watershed, Dissertation, LMU München, Fakultät für Geowissenschaften, 2008.

Maurer, E. P. and Lettenmaier, D. P.: Predictability of seasonal runoff in the Mississippi river basin, J. Geophys. Res., 108, 8607, https://doi.org/10.1029/2002JD002555, 2003.

Meißner, D. and Klein, B.: Probabilistic Shipping Forecast, in: "Handbook of Hydrometeorological Ensemble Forecasting", edited by: Duan, Q., Pappenberger, F., Thielen, J., Wood, A., Cloke, H. L., and Schaake, J. C., Springer, https://doi.org/10.1007/978-3-642-40457-3_58-1, 2016.

Molteni, F., Stockdale, T., Balmaseda, M., Balsamo, G., Buizza, R., Ferranti, L., Magnusson, L., Mogensen, K., Palmer, T., and Vitart, F.: The new ECMWF seasonal forecast system (System 4), ECMWF Tech. Memo., 656, 49 pp., available at: http://www.ecmwf.int/sites/default/files/elibrary/2011/11209-new-ecmwf-seasonal-forecast-system-system-4.pdf (last access: 29 August 2016), 2011.

Murphy, A. H.: What Is a Good Forecast? An Essay on the Nature of Goodness in Weather Forecasting, Weather Forecast., 8, 281–293, 1993.

Nilson, E., Lingemann, I., Klein, B., and Krahe, P.: Impact of hydrological change on navigation conditions, Deliverable 1.4 ECCONET – Effects of climate change on the inland waterway transport network, ECCONET project report, available at: http://www.tmleuven.be/project/ecconet/ECCONET_D1.4_final.pdf (last access: 11 December 2017), 2012.

Olsson, J., Uvo, C. B., Foster, K., and Yang, W.: Technical Note: Initial assessment of a multi-method approach to spring-flood forecasting in Sweden, Hydrol. Earth Syst. Sci., 20, 659–667, https://doi.org/10.5194/hess-20-659-2016, 2016.

Piani, C., Haerter, J., and Coppola, E.: Statistical bias correction for daily precipitation in regional climate models over Europe, Theor. Appl. Climatol., 2010, 187–192, https://doi.org/10.1007/s00704-009-0134-9, 2010.

Rauthe, M., Steiner, H., Riediger, U., Mazurkiewicz, A., and Gratzki, A.: A Central European precipitation climatology – Part I: Generation and validation of a high-resolution gridded daily data set (HYRAS), Meteorol. Z., 22, 235–256, 2013.

Richardson, D. S.: Skill and relative economic value of ECMWF ensemble prediction system, Q. J. Roy. Meteorol. Soc., 126, 649–667, 2000.

Richardson, D. S.: Economic Value and Skill, in: Forecast Verification: A Practitioner's Guide in Atmospheric Science, edited by: Joliffe, S. B., 2nd edn., John Wiley & Sons, Ltd., 167–184, 2012.

Richter, B. D., Baumgartner, J. V., Braun, D. P., and Powell, J.: A spatial assessment of hydrologic alteration within a river network, River Res. Appl., 14, 329–340, 1998.

Rimbu, N., Dima, M., Lohmann, G., and Stefan, S.: Impacts on the North Atlantic Oscillation and the El Niño–Southern Oscillation on Danube river flow variability, Geophys. Res. Lett., 31, L23203, https://doi.org/10.1029/2004GL020559, 2004.

Rimbu, N., Dima, M., Lohmann, G., and Musat, I.: Seasonal prediction of Danube flow variability based on stable teleconnection with sea surface temperature, Geophys. Res. Lett., 32, L21704, https://doi.org/10.1029/2005GL024241, 2005.

Robertson, D. E., Pokhrel, P., and Wang, Q. J.: Improving statistical forecasts of seasonal streamflows using hydrological model output, Hydrol. Earth Syst. Sci., 17, 579–593, https://doi.org/10.5194/hess-17-579-2013, 2013.

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

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, https://doi.org/10.5194/hess-15-3529-2011, 2011.

Svensson, C.: Seasonal river flow forecasts for the United Kingdom using persistence and historical analogues, Hydrol. Sci. J., 61, 19–35, https://doi.org/10.1080/02626667.2014.992788, 2016.

Trigo, R. M., Pozo-Vázquez, D., Osborn, T. J., Castro-Diez, Y., Gamiz-Fortis, S., and Esteban-Parra, M. J.: North Atlantic Oscillation influence on precipitation, river flow and water resources in the Iberian Peninsula, Int. J. Climatol., 24, 925–944, 2004.

Tucci, C. E. M., Clarke, R. T., Collischonn, W., da Silva Dias, P. L., and de Oliveira, G. S.: Long-term flow forecasts based on climate and hydrologic modeling: Uruguay River basin, Water Resour. Res., 39, 1181, https://doi.org/10.1029/2003WR002074, 2003.

Van Lanen, H. A. J., Laaha, G., Kingston, D. G., Gauster, T., Ionita, M., Vidal, J.-P., Vlnas, R., Tallaksen, L. M., Stahl, K., Hannaford, J., Delus, C., Fendekova, M., Mediero, L., Prudhomme, C., Rets, E., Romanowicz, R. J., Gailliez, S., Wong, W. K., Adler, M.-J., Blauhut, V., Caillouet, L., Chelcea, S., Frolova, N., Gudmundsson, L., Hanel, M., Haslinger, K., Kireeva, M., Osuch, M., Sauquet, E., Stagge, J. H., and Van Loon, A. F.: Hydrology needed to manage droughts: the 2015 European case, Hydrol. Process., 16, 1373–1381, https://doi.org/10.1002/hyp.10838, 2016.

Wang, E., Zhang, Y., Luo, J., Chiew, F. H. S., and Wang, Q. J.: Monthly and Seasonal Streamflow Forecasts Using Rainfall-Runoff Modeling and Historical Weather Data, Water Resour. Res., 47, W05516, https://doi.org/10.1029/2010WR009922, 2011.

Wilby, R. L., Wedgbrow, C. S., and Fox, H. R.: Seasonal predictability of the summer hydrometeorology of the River Thames, UK, J. Hydrol., 295, 1–16, 2004.

Wilks, D. S.: A skill score based on economic value for probability forecasts, Meteorol. Appl., 8, 209–219, 2001.

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.-Atmos., 107, 4429, https://doi.org/10.1029/2001JD000659, 2002.

Hydrol. Earth Syst. Sci., 21, 6401–6423, 2017

www.hydrol-earth-syst-sci.net/21/6401/2017/

Wood, A. W. and Lettenmaier, D. P.: An ensemble approach for attribution of hydro-logic prediction uncertainty, Geophys. Res. Lett., 35, L14401, https://doi.org/10.1029/2008GL034648, 2008.

Wood, A. W., Hopson, T., Newman, A., Brekke, L., Arnold, J., and Clark, M.: Quantifying Streamflow Forecast Skill Elasticity to Initial Condition and Climate Prediction Skill, J. Hydrometeorol., 17, 651–668, https://doi.org/10.1175/JHM-D-14-0213.1, 2016.

Yossef, N. C., Winsemius, H., Weerts, A., van Beek, R., and Bierkens, M. F. P.: Skill of a global seasonal streamflow forecasting system, relative roles of initial conditions and meteorological forcing, Water Resour. Res., 49, 4687–4699, https://doi.org/10.1002/wrcr.20350, 2013.

Zappa, M., Bernhard, L., Spirig, C., Pfaundler, M., Stahl, K., Kruse, S., Seidl, I., and Stähli, M.: A prototype platform for water resources monitoring and early recognition of critical droughts in Switzerland, Proc. IAHS, 364, 492–498, https://doi.org/10.5194/piahs-364-492-2014, 2014.

Zhao, T., Bennett, J. C., Wang, Q. J., Schepen, A., Wood, A. W., Robertson, D. E., and Ramos, M. H.: How Suitable is Quantile Mapping For Postprocessing GCM Precipitation Forecasts? J. Climate, 30, 3185–3196, 2017.

Zink, M., Samaniego, L., Kumar, R., Thober, S., Mai, J., Scäfer, D., and Marx, A.: The German drought monitor, Environ. Res. Lett., 11, 074002, https://doi.org/10.1088/1748-9326/11/7/074002, 2016.