Benchmarking an operational hydrological model for providing seasonal forecasts in Sweden

Marc Girons Lopez, Louise Crochemore, Ilias G. Pechlivanidis

Swedish Meteorological and Hydrological Institute, 601 76 Norrköping, Sweden

Correspondence to: Marc Girons Lopez (marc.girons@smhi.se)

Abstract. Probabilistic seasonal forecasts are important for many water-intensive activities requiring long-term planning. Among the different techniques used for seasonal forecasting, the Ensemble Streamflow Prediction (ESP) approach has long been employed due to the singular dependence on past meteorological records. The Swedish Meteorological and Hydrological Institute is currently extending the use of long-range forecasts within its operational warning service, which requires a thorough analysis of the suitability and applicability of different methods with the national S-HYPE hydrological model. To this end, we aim to evaluate the skill of ESP forecasts over 39,493 catchments in Sweden, understand their spatiotemporal patterns, and explore the main hydrological processes driving forecast skill. We found that ESP forecasts are generally skilful for most of the country up to 3 months into the future but that large spatiotemporal variations exist. Forecasts are most skilful during the winter months in northern Sweden, except for the highly-regulated hydropower-producing rivers. The relationships between forecast skill and 15 different hydrological signatures show that forecasts are most skilful for slowly-reacting, baseflow-dominated catchments and least skilful for flashy catchments. Finally, we show that forecast skill patterns can be spatially clustered in 7 unique regions with similar hydrological behaviour. Overall, these results contribute to identify in which areas, seasons, and how long into the future ESP hydrological forecasts provide an added value, not only for the national forecasting and warning service but, most importantly, to guide decision-making in critical services such as hydropower management and risk reduction.

1 Introduction

Regardless of the geographical setting, human society depends on water resources to satisfy basic needs and allow for social growth and development. At the same time, however, the variability of the hydrological systems, leading to extreme events such as floods or droughts, puts pressure on the viability and sustainability of many water-intensive activities. In this setting, being able to predict the future evolution of the hydrologic system may improve societal resilience by anticipating potentially hazardous events and enabling the adoption of protective and/or adaptive measures (Girons Lopez et al., 2017; Pappenberger et al., 2015b). Even if most day-to-day decisions on water-related issues are based on short- and medium-range forecasts, some activities, such as water reservoir operation and optimisation or strategic planning, benefit from long-term forecasts (Foster et al., 2018; Giuliani et al., 2020; Vigo et al., 2018). Despite their inherent uncertainties, such as hydro-meteorological model errors, future atmospheric states, or past hydro-meteorological water storages, long-term forecasts such as seasonal forecasts are a valuable tool for such applications, as they provide insights into the general trends of the hydrological system up to several months into the future, leading also to economic benefits (Bruno Soares et al., 2018; Giuliani et al., 2020).

Different techniques are available for generating seasonal forecasts, each with different strengths and weaknesses. These techniques may be based on dynamic or statistical methods, or on a weighted combination of both. Among these, the Ensemble Streamflow Prediction (ESP) methodology – originally named Extended Streamflow Prediction (Day, 1985) – has long been widely adopted for seasonal forecasting (Wang et al., 2011; Wood and Lettenmaier, 2006). Following this methodology, ensemble streamflow forecasts are generated using historical meteorological data as forcing to a hydrological model. An advantage of this method compared to methods based directly on ~~historical streamflow~~streamflow climatology, is that ESP forecasts are initialised ~~based on~~using the latest hydrological conditions ~~updated for the forecast date~~ (Crochemore et al., 2020),~~. Forecasts~~ thus benefiting from the most recent hydrological knowledge when they are initialised, which is of particular interest for unprecedented hydrological conditions. This advantage can however also lead to forecast overconfidence as this method does not consider the impact of potential uncertainties in the initial hydrologic conditions, as noted by Wood and Schaake (2008). Additionally, ~~its~~the reliance of ESP forecasts on historical meteorological forcing makes it impossible for ~~it~~them to capture hydrological responses to unprecedented meteorological events. Conversely, forecasts based on Numerical Weather Prediction (NWP) models are not constrained by the observational period as they are driven with an ensemble of dynamical meteorological forecasts, and are increasingly being used to overcome these limitations (Monhart et al., 2019). Nevertheless, ESP forecasts still offer the best study object to focus on the role of initial hydrologic conditions alone, which are best explained through catchment characteristics than by using NWP forcings.

ESP forecasts have been used by the scientific community to assess forecast skill sensitivity and uncertainties and to benchmark seasonal forecast improvements (Arnal et al., 2018; Harrigan et al., 2018), as well as for operational flood forecasting in many different settings and scales (Candogan Yossef et al., 2017). Over the years, different techniques have been developed to improve the performance of forecasting systems, such as data assimilation for improving the initial conditions of forecasts (DeChant and Moradkhani, 2011), multi-model approaches (Muhammad et al., 2018), or pre- and post-processing techniques such as using artificial neural networks for reducing the effects of model errors (Jeong and Kim, 2005; Macian-Sorribes et al., 2020), historical scenario selection and weighting (Crochemore et al., 2017; Trambauer et al., 2015), and calibration techniques (Wood and Schaake, 2008).

Evaluation efforts are typically carried out based on forecasts issued retrospectively (re-forecasts) over time periods long enough to ensure that the evaluation is statistically robust. For many operational applications it is important to understand the spatiotemporal patterns of seasonal streamflow predictability as well as the driving processes behind these patterns (Sutanto et al., 2020). Indeed, previous studies have identified different sources of forecast skill depending on hydrological characteristics; for instance, Greuell et al., (2019), Shukla et al., (2013), and Wanders et al., (2019) identified initial conditions of soil moisture and snow (during spring) as the most important sources of skill over Europe, while Singla et al.

(2012) found similar results for France. In a study over the United Kingdom, Harrigan et al. (2018) ascertained streamflow predictability was higher for slow-responding catchments, as described by the baseflow index (BFI). Some studies have even gone one step further by investigating spatiotemporal patterns in streamflow predictability in an attempt to regionalize the forecast skill. For example, Pechlivanidis et al. (2020) showed that streamflow predictability is strongly dependent on the overall hydrological regime, with limited predictability in flashy basins (low river memory) and hence, it can be regionalised based on a priori knowledge of local hydro-climatic conditions.

The Swedish Meteorological and Hydrological Institute (SMHI) has long been operationally providing streamflow forecasts (catchment outflows) and hydrological warnings for Sweden (Figure A1). to relevant actors in hydrological risk management (municipalities, county boards, Swedish Civil Contingencies Agency), as well as to the general public. Additionally, both professional actors and the general public have access to the current hydrological situation and streamflow climatology through the open access *Vattenwebb* portal (https://vattenwebb.smhi.se/). On top of that, SMHI's consultancy services provide tailored forecasts to relevant actors. These forecasts are however not included in the public service and, as of today, are limited to individual river basins. Forecasts were initially produced with the HBV model (Bergström, 1976), but in recent years operational forecasting has shifted to the Swedish implementation of the HYPE model (S-HYPE, Lindström et al., 2010), which allows for an integrated, high-resolution description of the hydrological system across the country. Where available, in-situ observations of streamflow are assimilated, which has a beneficial impact on the hydrological predictions downstream. ESP seasonal forecasts are produced operationally but have not been widely used in real-world applications due to the lack of information on their skill ~~but not generally spread to other actors due to uncertainties in their skill~~ and to the subsequent potential misinterpretation by external parties. Nevertheless, SMHI is now looking to extend the usage of long-term forecasts within its warning service, which requires a deeper understanding of forecast performance, its patterns, and controlling factors. ~~In terms of regionalisation, four main hydro-climatic regions based on hydro-climatic patterns (Lindström and Alexandersson, 2004; Pechlivanidis et al., 2018) have typically been used for water management in Sweden. However, these regions were not put forward with consideration to seasonal streamflow predictability over Sweden and might therefore be of limited use for this purpose.~~

The aim of this study is to evaluate SMHI's operational ESP seasonal forecasts by benchmarking and attributing~~e the~~ ESP forecast skill over Sweden with the operational S-HYPE model. To ~~address these questions~~achieve these objectives, we: (a) evaluate the skill of ESP seasonal forecasts generated with the operational S-HYPE model over Sweden and understand the spatiotemporal pattern of skill, (b) detect potential links between streamflow forecast skill and hydrological characteristics, and (c) attribute streamflow predictability patterns across the country to hydrological behaviour of the river systems. The paper is structured as follows: section 2 presents the data used, hydrological model setup, and methodology for the forecast evaluation; section 3 presents the results, followed by the discussion in section 4; finally, section 5 states the conclusions.

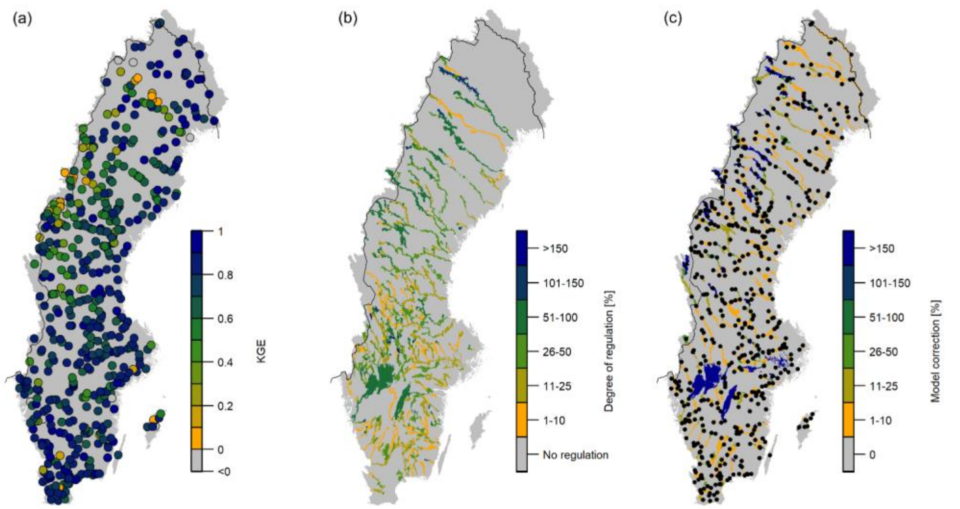
2 Methods

2.1 Data

Daily precipitation and temperature data from the PTHBV database (Johansson, 2002) were used as forcing data to the S-HYPE model. This database contains gridded data based on a weighted interpolation ~~of measured values~~ from all available station observations for any given day with a resolution of 4x4 km, and it is available from 1961 onwards. The interpolation method used for generating PTHBV considers factors such as elevation and wind frequency and direction to make interpolated values for precipitation and temperature more reliable. This dataset was processed using a weighted average method based on the area fraction of the PTHBV grid cells intersection with a given model catchment to force the semi-distributed S-HYPE model (Section 2.2). Additionally, daily stream discharge and water level data from 539 stations of SMHI’s gauge network were used to correct the model outputs for improved forecast initialisation (see Figure 1). Data availability varies greatly among stations (Figure A2a). Nevertheless, most stations have observations for the entire study period (Figure A2b).

2.2 Hydrological modelling and forecasting

The ESP re-forecasts were produced using the S-HYPE model (Strömqvist et al., 2012), which is the operational implementation of the HYPE model for Sweden (Lindström et al., 2010). ~~This allowed an analysis of model outputs for 39,493 catchments (with an average spatial resolution of 10 km^2) in the model domain.~~ The HYPE model is a process-based hydrological model for water quantity and quality which operates on a daily time step and includes both hydrological (snowpack, groundwater, surface runoff, streamflow) and anthropogenic (reservoir operation, irrigation) factors. This model framework can be used in lumped, semi-distributed, and distributed modes. More specifically, ~~T~~the S-HYPE model is semi-distributed and, in its current version, consists of 39,493 catchments (with an average spatial resolution of 10 km^2) covering the whole of Sweden as well as parts of Norway and Finland in transboundary basins. This model has a median Kling-Gupta efficiency (KGE; Gupta et al., 2009) of 0.79 for the period 1981 – 2016 at daily time step, ranging from -0.56 to 0.96 (Figure 1a). For reference, the KGE metric ranges between -∞ and 1; the closer to 1 the KGE is, the more accurate the simulations are.


(a)
(b)
(c)
KGE
1
0.8
0.6
0.4
0.2
0
<0
Degree of regulation [%]
>150
101-150
51-100
26-50
11-25
1-10
No regulation
Model correction [%]
>150
101-150
51-100
26-50
11-25
1-10
0

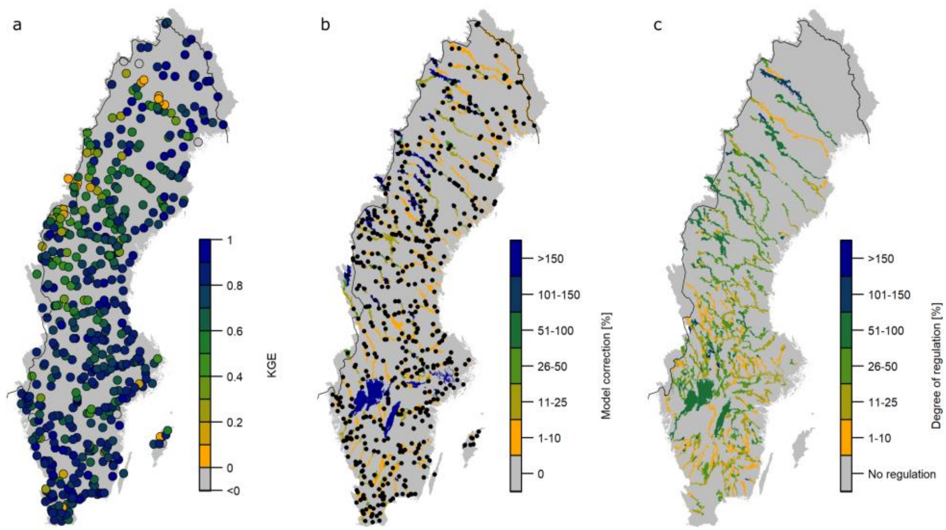

a
b
c
KGE
1
0.8
0.6
0.4
0.2
0
<0
Model correction [%]
>150
101-150
51-100
26-50
11-25
1-10
0
Degree of regulation [%]
>150
101-150
51-100
26-50
11-25
1-10
No regulation

Figure 1 Study domain: (a) S-HYPE model Kling-Gupta ~~S-HYPE model performance~~ efficiency (KGE) for the period 1981 – 2016 for each of the 539 hydrological stations, ~~(b) average model correction value for each catchment following the autoregressive correction method and,~~ (b~~c~~) degree of regulation of each S-HYPE catchment, and (c) average model correction value for each S-HYPE catchment following the autoregressive correction method.

A large percentage of water courses in Sweden are regulated, mainly for energy production purposes; see the degree of regulation (%) in Figure 1b~~c~~; see also the definition in (Pechlivanidis et al., 2020) ~~Pechlivanidis et al., (2018)~~. This makes the simulation and prediction of water variables in the main water courses more challenging, as regulation patterns, which can largely deviate from the natural flow, need to be considered. In the operational S-HYPE model, general regulation regimes in the form of constant flow or seasonally varying sine wave shaped flow (or a combination of both) between predefined levels and, in some cases, specific dates are provided for a number of reservoirs. Nevertheless, since dam operation is continuously adapted (within certain bounds) to the changing~~present and most probable future~~ meteorological and hydrological conditions, in addition to other factors such as optimising the economic benefit and ensuring safe operation, long-range forecasts based on hydrological models with only a limited description of such complex decisions on regulation patterns will most likely be conditioned by these simplifications ~~these general regulation regimes are expected to be of little benefit for seasonal forecasting purposes~~.

We produced a series of hydrological re-forecasts up to a lead time of 190 days (approximately 6 months) at ~~the~~ daily time step for all 39,493 locations across Sweden and transboundary basins using meteorological forcing data from 25 random years for the period 1961 – 2016 so as to mimic SMHI's operational ~~forecast~~ setup (Figure 2a). When selecting the forcing data, a window of 3 years was left out around the analysis year (1 year before and ~~two~~ 2 years after) to limit the impact of interannual streamflow memory and thus avoid conditioning the forecasts. We initialised the re-forecasts on the 1^{st}, 8^{th}, 15^{th}, and 22^{nd} of each month (approximately once a week) and aggregated the daily forecast data into weekly averages.

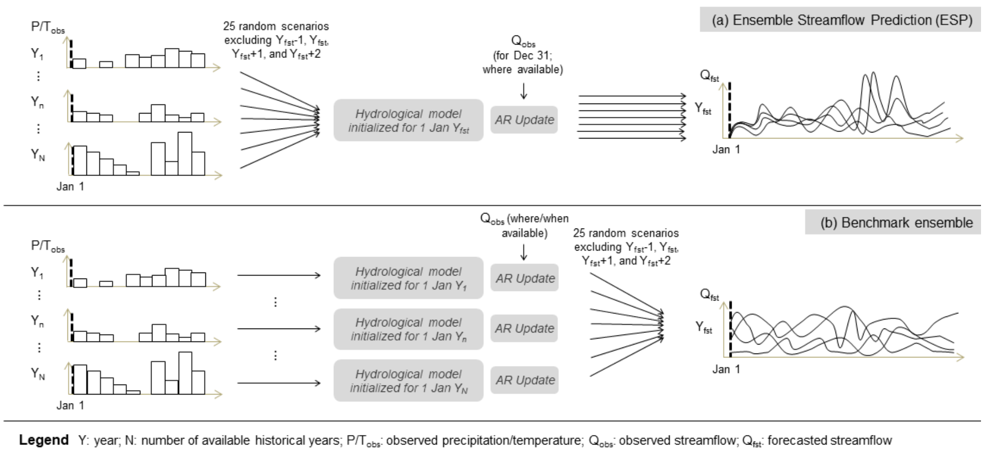


Figure 2 Schematic of the forecast generation procedure in this study: (a) ESP re-forecasts and (b) benchmark forecasts. Adapted from (Crochemore et al., 2020).

~~We used stream discharge and water level data, where available, to correct the model outputs prior to producing the re-forecasts, and thus get the best possible initialisation conditions. To this purpose, we used an autoregressive (AR) correction method~~ Following SMHI's operational setup, model outputs were corrected with stream discharge and water level observations, where and when available, to obtain the best possible initialisation conditions. When observations were no longer available, an autoregressive (AR) correction method was used (Lindström and Carlsson, 2000; Pechlivanidis et al., 2014). To illustrate this procedure, let us consider the case of a catchment which has observations throughout the analysis period. For each forecast initialisation, the model outputs were corrected up to the day before forecast initialisation and model errors were stored. Then, as observations were theoretically no longer available, the model output correction started from the latest stored model error value and exponentially decreased with time based on a calibrated factor until the model outputs converged with the simulation results. ~~Following this method, model outputs are replaced by the available observations and the model errors with respect to these observations are saved for every time step. If observations are no longer available, the output corrections converge exponentially towards the simulated values.~~ This correction only affects catchments with or downstream from ~~streamflow~~ observations and is especially relevant (at least for the first forecast lead times) for regulated water courses with low model performance where simulated streamflow can significantly deviate from actual values (Figure 1c~~b~~).

In summary, the reforecast dataset has the following size: 39,493 catchments, 43,200 forecast initialisations (4 start dates per month x 12 months x 36-year reforecast period (1981 – 2016) x 25 members) using AR-correction where and when available, and averaged weekly up to 190 days.

2.3 Forecast evaluation

We evaluated the skill of the ESP re-forecasts produced with the S-HYPE model over the period 1981 – 2016 using the Continuous Ranked Probability Skill Score (CRPSS, Appendix ~~A~~B) and a cross-validation strategy. Although studies involving large-scale models often use model simulations as reference, as this minimizes the impact of model performance on forecast skill (Arnal et al., 2018; Crochemore et al., 2020), here we followed SMHI's operational setup and therefore used a~~the~~ reference ~~was~~based on a combination of observations (for catchments with or downstream from observation points) and model simulations (also known as perfect forecasts; elsewhere) ~~as a station-corrected simulation approach was used~~ to achieve the best possible initial conditions. We assessed the skill of the ESP re-forecasts so as to highlight the added value of the ESP forecasts with respect to an ensemble forecast based on ~~historical streamflow~~streamflow climatology, which users would have access to in the absence of SMHI's forecast service (Pappenberger et al., 2015a). To this purpose, we used an ensemble whose 25 members were resampled from the ~~historical~~ station-corrected ~~model simulations~~streamflow climatology fo~~rom~~ the period 1981 – 2010 (excluding the forecast year) as a benchmark against which to derive the skill of the ESP re-forecasts (Figure 2b).

Even if hydrological models are typically run at a daily time scale, forecast results from hydroclimate prediction systems are usually post-processed and aggregated over longer periods to provide information tailored to the user needs (Bohn et al., 2010). More specifically, a temporal aggregation of one month is typically used in seasonal forecasting services (Apel et al., 2018; Bennett et al., 2017). Nevertheless, different time periods may be of interest depending on the sectorial use (e.g. water resources management, civil protection mechanisms, warning services). Therefore, in addition to using a ~~basic~~ default temporal aggregation of one week ~~(i.e. daily streamflow forecasts were aggregated to weekly averages)~~ to estimate the predictive skill of the national operational service, we were also interested in understanding how aggregating streamflow forecasts over different time periods (i.e. 2 weeks, 4 weeks, 8 weeks, 12 weeks, and 24 weeks) would impact~~s~~ forecast skill at different lead times.

2.4 Forecast skill attribution

Thereafter, we investigated which hydrological characteristics are associated with skilful forecasts. More specifically, we selected a set of 15 hydrologic signatures (statistics describing the hydrological behaviour; see Table 1) to provide diagnostics of the hydrological regime (Kuentz et al., 2017; Pechlivanidis and Arheimer, 2015). We used the non-parametric Spearman rank test to assess the correlation between forecast skill and each of the hydrologic signatures.

Table 1 Hydrologic signatures used for catchment functioning.

Signature	Abbreviation	Unit	Reference

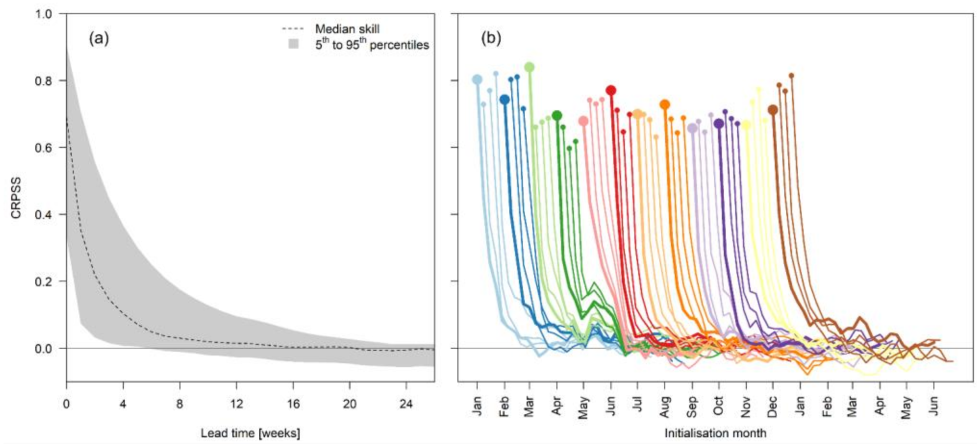
Mean annual specific runoff	Qm	mm yr^{-1}	Viglione et al. (2013)
Range of Pardé coefficients	DPar	-	Viglione et al. (2013)
Slope of streamflow duration curve	mFDC	% %$^{-1}$	Viglione et al. (2013)
Normalised low streamflow	q95	-	Viglione et al. (2013)
Normalised high streamflow	q05	-	Viglione et al. (2013)
Coefficient of variation	CV	-	Donnelly et al. (2016)
Flashiness	Flash	-	Donnelly et al. (2016)
Normalised peak distribution	PD	-	Euser et al. (2013)
Rising limb density	RLD	-	Euser et al. (2013)
Declining limb density	DLD	-	Euser et al. (2013)
Normalised relatively low streamflow	q70	-	Viglione et al. (2013)
Baseflow index	BFI	-	Kuentz et al. (2017)
Runoff coefficient	RC	-	Kuentz et al. (2017)
Streamflow elasticity	EQP	-	Sawicz et al. (2011)
High pulse count	HPC	-	Yadav et al. (2007)

Then, we applied a *k-means* clustering approach (Jin and Han, 2011) within the 15-dimension space (hydrological signatures) to group the catchments into clusters based on similarities of basin functioning and further identify the dominant streamflow generating processes for specific regions. This is not the first regionalisation effort done for Swedish catchments. Indeed, four main hydro-climatic regions based on hydro-climatic patterns (Lindström and Alexandersson, 2004; Pechlivanidis et al., 2018) have typically been used for water management in Sweden. Nevertheless, this previous regionalisation was based on different variables (e.g. marine basins) and is thus not suitable for the purposes of this study. Finally, we analysed the hydrologic predictability for each of the clusters.

3 Results

3.1 Temporal and spatial distribution of forecast skill

The skill of the ESP re-forecasts varies with the lead time and the forecast initialisation date (Figure 3~~Figure 2~~). As expected, the ESP skill for 1-week forecast averages ~~of the ESP~~ with respect to ~~historical streamflow~~streamflow climatology, as measured by the CRPSS metric, is overall very high for medium-range horizons (i.e. 1 – 2 weeks ahead), with a median skill over Sweden starting at 0.7 (Figure 3~~Figure 2~~a) and thereafter decreasing with time (CRPSS ranges between 1 (best) and -∞). After approximately three months and until the furthest horizon (190 days), the ESP provides, on average, no added value with respect to ~~historical streamflow~~streamflow climatology. Similar trends have been observed in other evaluations of forecasting systems over Sweden (Foster et al., 2018; Olsson et al., 2016). In particular, we note a rapid decrease in skill in the first forecast month (Crochemore et al., 2020; Harrigan et al., 2018). Consequently, under the common ~~monthly~~ initialisation frequency of 1 month for many climate prediction systems (Batté and Déqué, 2016; Johnson et al., 2019), streamflow predictability is expected to remain low for periods beyond a 2-week forecast horizon. By increasing the frequency of forecast initialisation (e.g. from once a month to once a week), and hence frequently updating the initial hydrological states, it is possible to maintain a high streamflow forecast skill for extended forecast horizons (Figure 3~~Figure 2~~b).



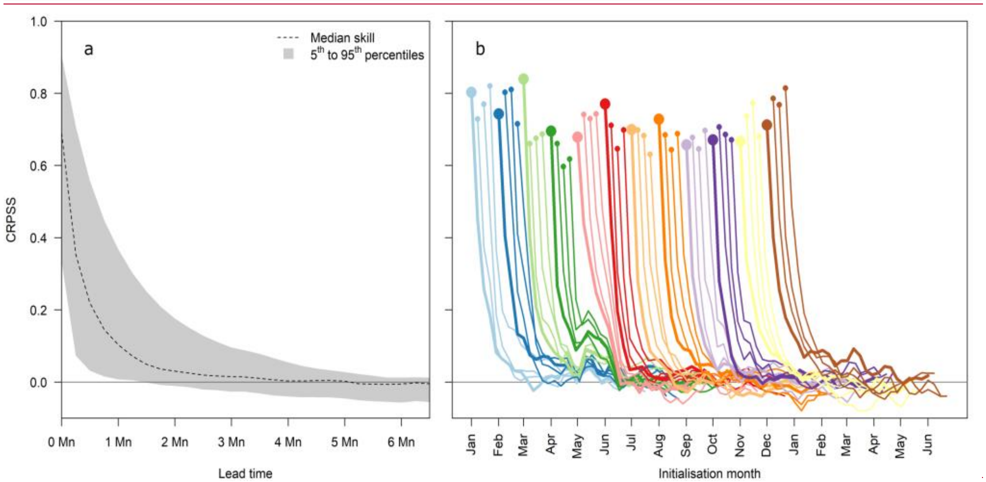
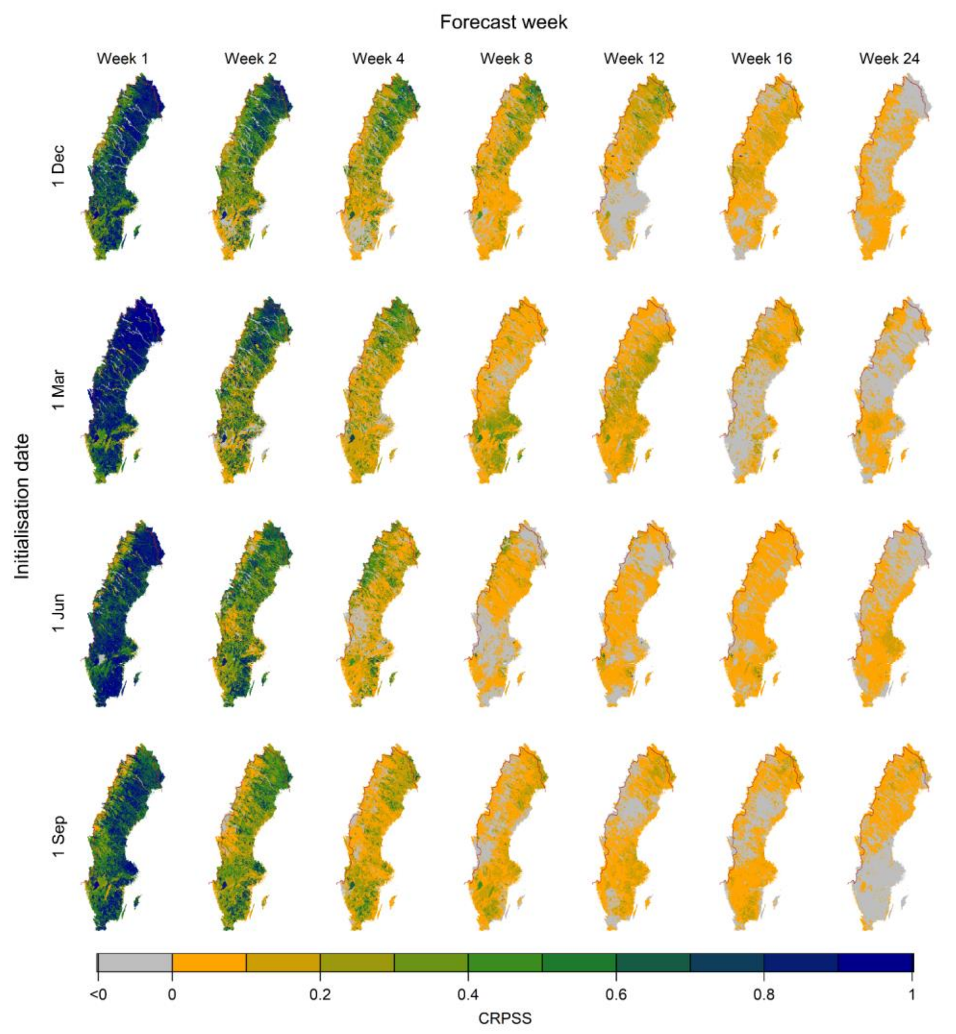


Figure 3~~2~~ Streamflow ESP forecast skill (in terms of CRPSS) for 1-week forecast averages as a function of lead time and up to 190 days: (a) median skill and 5th to 95th percentile range for the entire domain, and (b) temporal disaggregation of forecast skill per initialisation date. Initialisation dates within the same month (4 times) are represented with the same colour and the first initialisation of each month is marked with thicker markers and lines.

Even if the forecast skill follows a similar decreasing pattern for all initialisation dates, both the maximum skill value as well as the deterioration rate differ. The highest skill (greater than 0.8) is observed for forecasts initialised ~~in winter (~~between 8 December and 1 March~~February)~~, which roughly corresponds to the winter months. In the other seasons, the forecast skill starts at around 0.7, with the lowest skill value observed for initialisations in April (just under 0.6). Even if the forecast skill deteriorates quickly and reaches a predictability value close to the one of ~~historical streamflow~~streamflow climatology (CRPSS close to 0) in long forecast horizons, forecasts initialised in March and April (and to some extent also in February) show a small secondary peak in the skill in May. This may be explained by the hydrological regime at a large part of Swedish catchments, in which streamflow generally starts to increase in April-May, despite the general lack of precipitation in winter and early March~~ This is probably related to the predictable spring flood season in the northern parts of the country~~ (see also Pechlivanidis et al., 2020).

The spatial distribution of forecast skill differs significantly across initialisation dates and forecast horizons (Figure 4~~Figure 3~~). For instance, forecasts initialised in winter (e.g. December 1^{st}) maintain skill for inland forested areas of northern Sweden up to 3 months in the future. Forecasts initialised in spring (e.g. March 1^{st}) show skill up to the same forecast horizon, but most notably in the southern and eastern parts of the country. Finally, forecasts issued in summer and autumn are skilful up to 2 months except for some areas in the central-western parts of the country.

For the first forecast month, forecasts tend to have a comparatively poorer skill in the mountainous areas of north-western Sweden than in other parts of the country, except when they are initialised in the spring. Agricultural areas located around some of Sweden's largest lakes, such as Lake Mälaren and Lake Vänern (Figure A1), also have comparatively poor forecast skill. Interestingly, high predictability would have been expected in such lakes with slow hydrological response (long memory) (see Pechlivanidis et al., 2020). However, these great lakes are heavily regulated (see Figure 1), and the model correction seems to have impacted forecast skill. Streamflow forecasts in the large, highly-regulated, rivers of northern Sweden, such as River Umeälven and River Luleälven, also lack skill (Figure 1b~~e~~; Figure A1). Again, the regulation patterns that significantly differ from the natural regime of watercourses, are not adequately captured by the ESP. In these cases, the broader ensemble of ~~historical streamflow~~streamflow climatology is a better estimator of~~ the~~ future ~~trends in ~~streamflow. Conversely, streamflow forecasts show high skill in non-regulated rivers located in the same area and of similar size and hydrological regime, i.e. River Kalixälven and River Torneälven (Figure A1).


Forecast week
Week 1
Week 2
Week 4
Week 8
Week 12
Week 16
Week 24
Initialisation date
1 Dec
1 Mar
1 Jun
1 Sep
<0
0
0.2
0.4
0.6
0.8
1
CRPSS

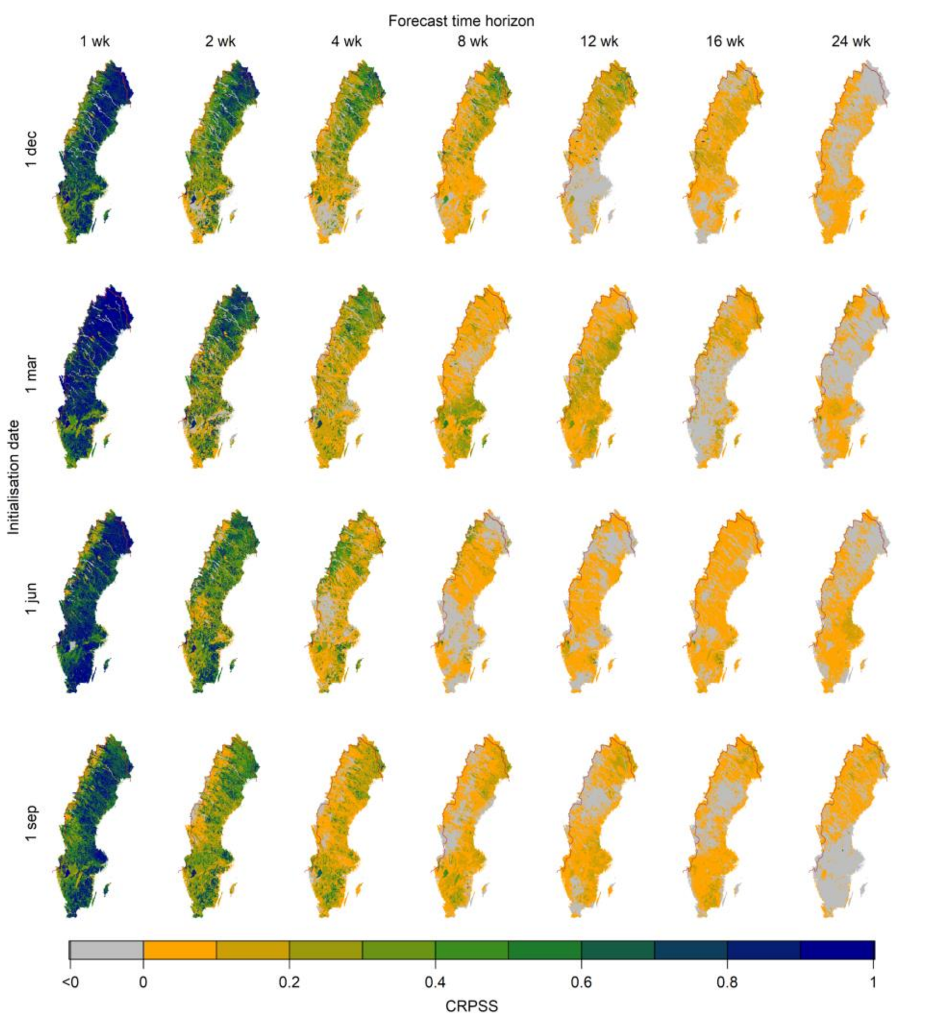
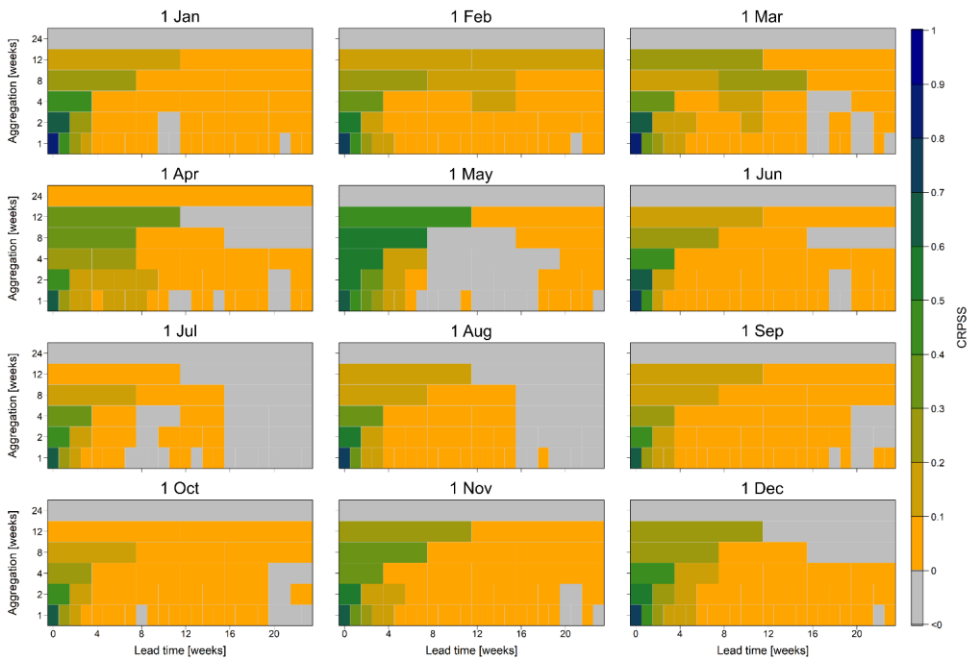


Figure 4~~3~~ Spatial distribution of streamflow forecast skill (in terms of CRPSS) for selected initialisation dates (rows) and forecast time horizons (columns).

3.2 Forecast skill as a function of temporal aggregation

We next investigate the impact of ~~the~~using different forecast aggregation periods on the forecast skill for different lead times. More specifically, in addition to the default aggregation period of 1 week, we consider the aggregation of streamflow forecasts over 2 weeks, 4 weeks, 8 weeks, 12 weeks, and 24 weeks. ~~Here,~~Since the focus here is not on the spatial patterns of skill, ~~and therefore~~ forecast skill is therefore averaged over the entire domain. Results show that, even if the average skill for the first forecast period decreases when aggregating over longer time periods, the forecasts remain skilful (CRPSS greater than 0) for aggregation periods up to 12 weeks (Figure 5~~Figure 4~~). When aggregating over 24 weeks, the ESP method generally provides no added value with respect to ~~historical streamflow~~streamflow climatology; the predictability from ESP is very similar to the one from ~~historical streamflow~~streamflow climatology. Even if, as expected, forecast skill decreases when forecasts are aggregated over long periods, a comparatively higher skill is maintained over longer time horizons than when forecasts are aggregated over short periods. In addition, forecasts initialised in February and March are skilful up to 16 weeks ahead when aggregating over long periods (e.g. 8 weeks), and the forecasts initialised in April and May show high skill values, even when aggregating over a 12-week period. This ~~seems~~ is probably ~~to be~~ due to the high predictability of the spring flood season in May, also shown by the secondary skill peak ~~in skill~~ observed for these initialisations in Figure 3~~Figure 2~~b. Many catchments and rivers, especially in the northern half of Sweden, experience the peak of the spring flood during May. Using short aggregation periods, skill is more influenced by the exact start and end times of the spring flood event while long aggregations put more emphasis on a correct total flood volume. Since the total volume linked to the accumulated snowpack is easier to model than the timing of the event, which is conditioned by meteorological variables, long aggregations tend to perform better. In the southern parts of the country, the spring flood is already over in May and low streamflow conditions start to dominate. Finally, for forecasts initialised in July and October, long aggregation periods (e.g. 12 weeks) tend to dilute the high forecast skill observed over the first weeks.


1 Jan
1 Feb
1 Mar
1 Apr
1 May
1 Jun
1 Jul
1 Aug
1 Sep
1 Oct
1 Nov
1 Dec
Aggregation [weeks]
24
12
8
4
2
1
Lead time [weeks]
0
4
8
12
16
20
CRPSS
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
<0

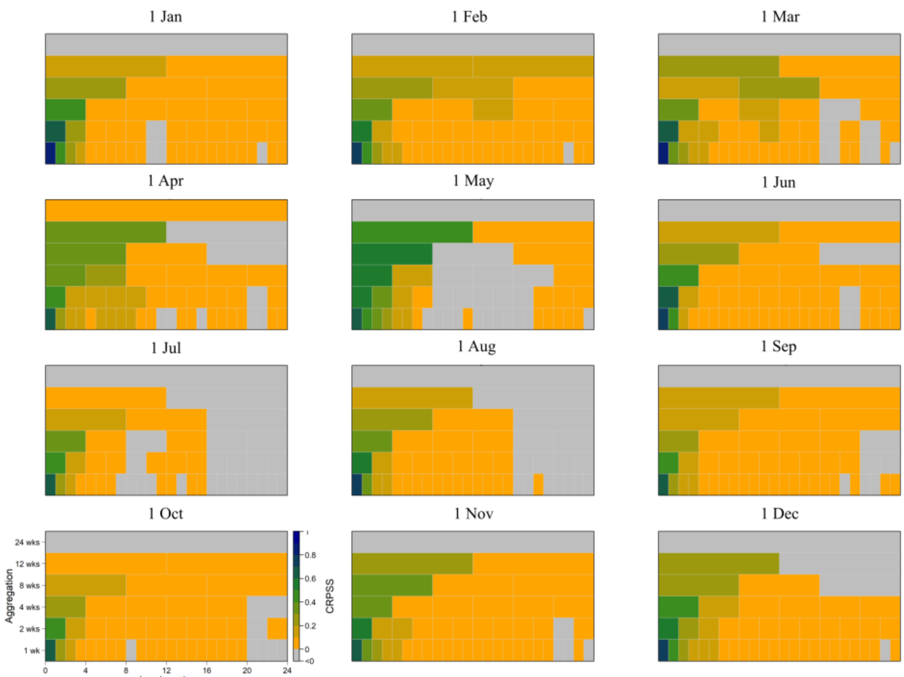
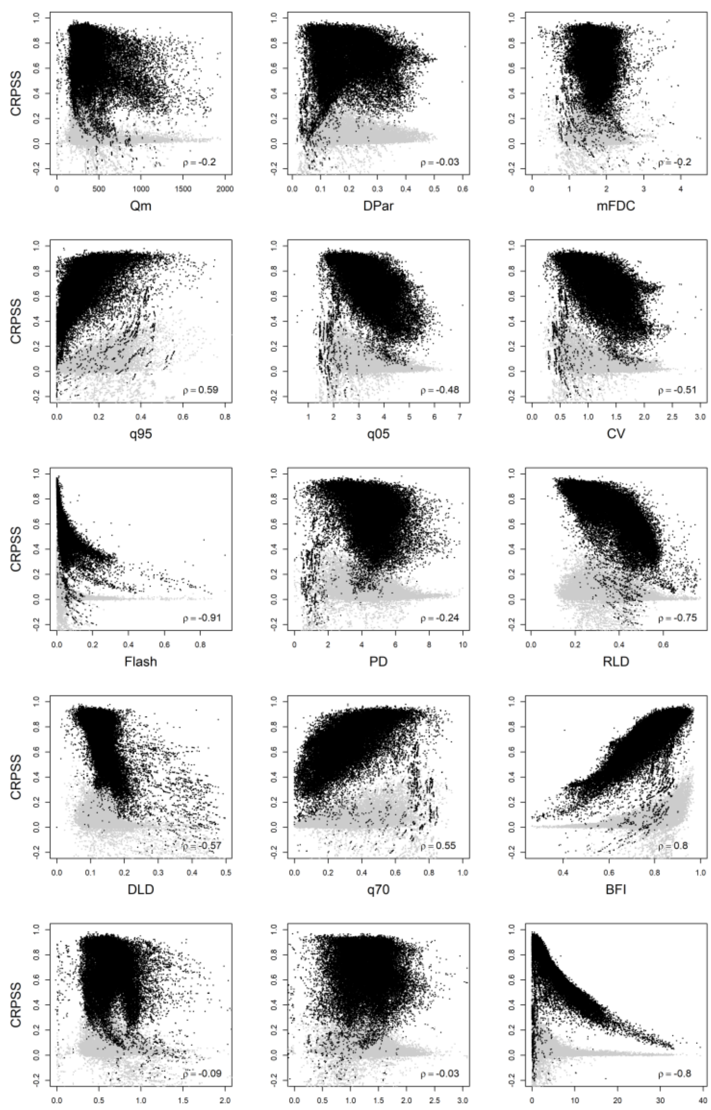


Figure 54 Skill of streamflow forecasts as a function of lead time (weeks) initialised on the 1st of each month for selected forecast aggregation periods (i.e. 1 week, 2 weeks, 4 weeks, 8 weeks, 12 weeks, and 24 weeks). The skill is averaged over Sweden.

3.3 Relating streamflow signatures and forecast skill

We next investigate potential correlations between forecast skill and the 15 streamflow signatures using the non-parametric Spearman rank test. In all cases the null hypothesis (i.e. no correlation exists between forecast skill and the streamflow signature) is rejected with a level of significance of 0.01 for lead week 0. ~~Nevertheless, d~~Different patterns emerge when comparing forecast skill ~~(first forecast week)~~ for each catchment with each of the 15 streamflow signatures (Figure 6~~Figure 5~~). More specifically, forecast skill is strongly inversely correlated (defined here as the Spearman's rank correlation coefficient (ρ) being less than -0.50) with high pulse count (HPC), flashiness (Flash), rising limb density (RLD), declining limb density (DLD), and coefficient of variation (CV). Additionally, a strong direct correlation (ρ > 0.50) is found between skill and baseflow index (BFI), normalised low streamflow (q95), and normalised relatively low streamflow (q70) indicating that slowly reacting catchments with a significant baseflow component generally experience high predictability (Harrigan et al., 2018; Pechlivanidis et al., 2020). A similar analysis has been conducted for longer forecast horizons (e.g. lead week 8~~results not shown here~~); however, since spatial patterns in forecast skill weaken and blend in with the forecast horizon, the identified correlations do not have any explanatory power ~~are not strong~~. Overall, the identified correlations highlight the existence of a generally high forecast skill in slowly reacting, baseflow-dominated catchments, while low forecast skill is predominant in flashy catchments. Although this analysis indicates the existence of dependencies between streamflow signatures and forecast skill, it can still be considered limited given that a hydrological system is generally characterized by a wider set of streamflow signatures than that considered here (Pechlivanidis and Arheimer, 2015; Sawicz et al., 2011).


CRPSS
Qm
ρ = -0.2
DPar
ρ = -0.03
mFDC
ρ = -0.2
q95
ρ = 0.59
q05
ρ = -0.48
CV
ρ = -0.51
Flash
ρ = -0.91
PD
ρ = -0.24
RLD
ρ = -0.75
DLD
ρ = -0.57
q70
ρ = 0.55
BFI
ρ = 0.8
RC
ρ = -0.09
EQP
ρ = -0.03
HPC
ρ = -0.8

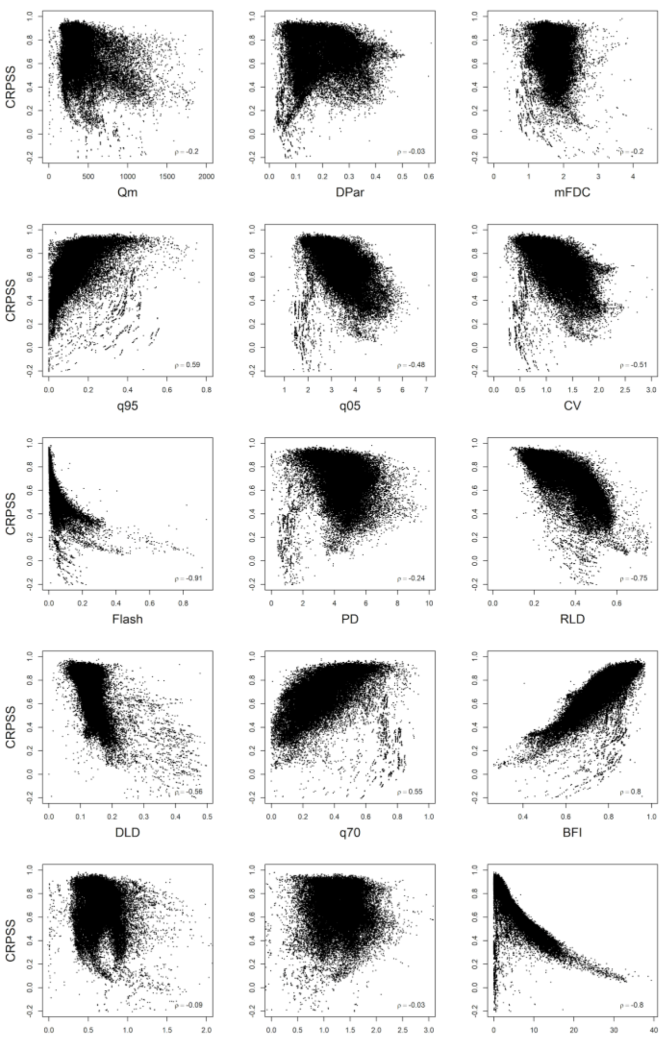
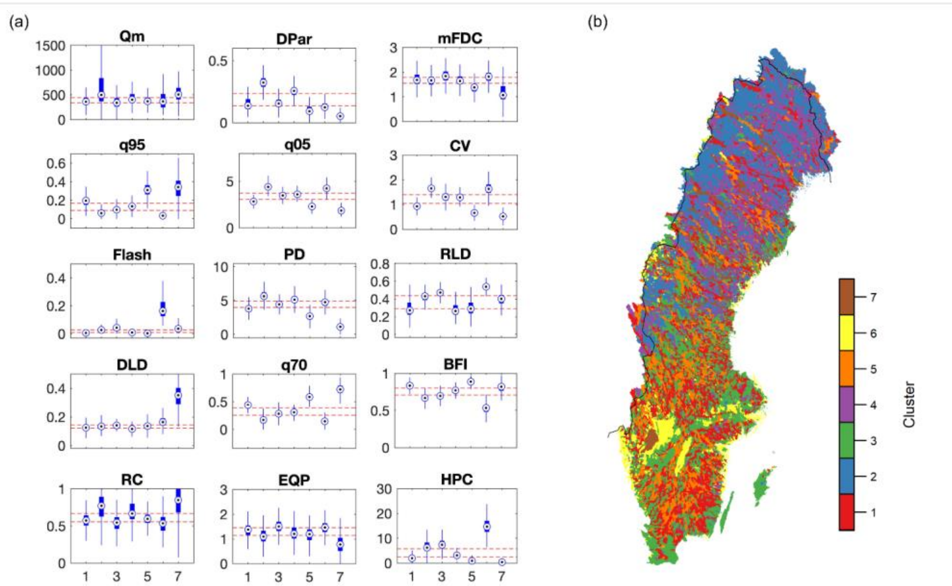


Figure 6~~5~~ Forecast skill (in terms of CRPSS) for the ~~first forecast~~ lead week 0 (black dots) and lead week 8 (light grey dots) for each of the 39,493 catchments in Sweden as a function of each of the 15 hydrological signatures. The non-parametric Spearman's rank correlation coefficient for lead week 0 (ρ) is shown for each signature.

3.4 Attributing streamflow forecast skill to hydrologic behaviour

Here, we investigate the potential attribution of streamflow predictability in the Swedish river systems to hydrological behaviour, given that such dependency has been highlighted in the previous analysis. Using the *k-means* clustering method, an optimal number of seven distinct clusters (based on a silhouette analysis using a different number of clusters; de Amorim and Hennig, 2015) have been obtained representing different hydrological regimes (Figure 7~~Figure 6~~). Table 2 provides additional information on the topographic, climatological and hydrological characteristics of each cluster while the spatial variability of each of the 15 streamflow signatures, as well as of the catchment elevation, is presented in **Fel! Hittar inte referenskälla.** ~~Appendix B~~.

Catchments clustered in regions 1 and 5 are characterised by a high baseflow contribution (BFI), a slow response to precipitation (Flash) and, therefore, a generally small intra-annual variability (DPar). In terms of topography, these regions consist ~~mainly~~ of forested areas mainly located in southern Sweden. Catchments in ~~C~~cluster 2 are found in highland areas and boreal forest environments in northern Sweden, and are characterised by high seasonality (CV and DPar) due to the alternance between snow melting and accumulation. These catchments are also characterised by high runoff volumes (Qm) given that they are subject to high precipitation amounts and low evapotranspiration rates. Agricultural and coastal areas located mainly in southern and central parts of the country are found in ~~C~~cluster 3. These catchments are characterised by a highly variable streamflow regime (HPC and RLD) and a quick response to precipitation (Flash), yet exhibit a relatively long hydrograph recession (DLD). Similarly, catchments grouped in ~~C~~cluster 6, which are located in low-land coastal and lake areas, experience flashy responses (Flash), as well as high streamflow~~s~~ (q05) and seasonal variations (CV and mFDC). Boreal forest catchments in the northern part of the country are grouped in ~~C~~cluster 4, and are characterised by a generally high runoff coefficient (RC) and a slow response to precipitation events (Flash). Finally, catchments in ~~C~~cluster 7 are found along several large and highly-regulated rivers in northern Sweden. These catchments are characterised by a small variability (CV, DPar, and mFDC) but high streamflow volumes (Qm) and runoff coefficients (RC) explained by anthropogenic regulations.



(a)
Qm
DPar
mFDC
q95
q05
CV
Flash
PD
RLD
DLD
q70
BFI
RC
EQP
HPC
(b)
Cluster

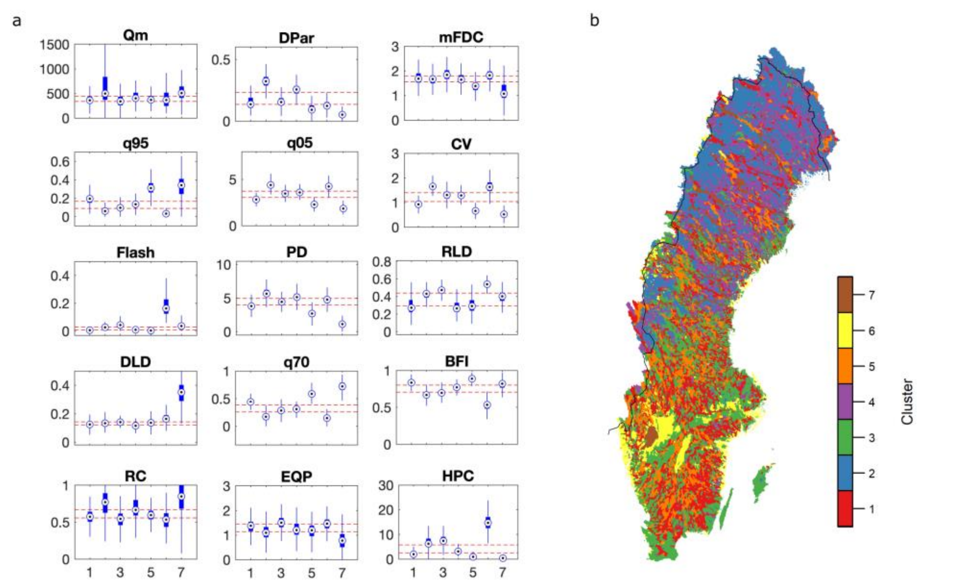


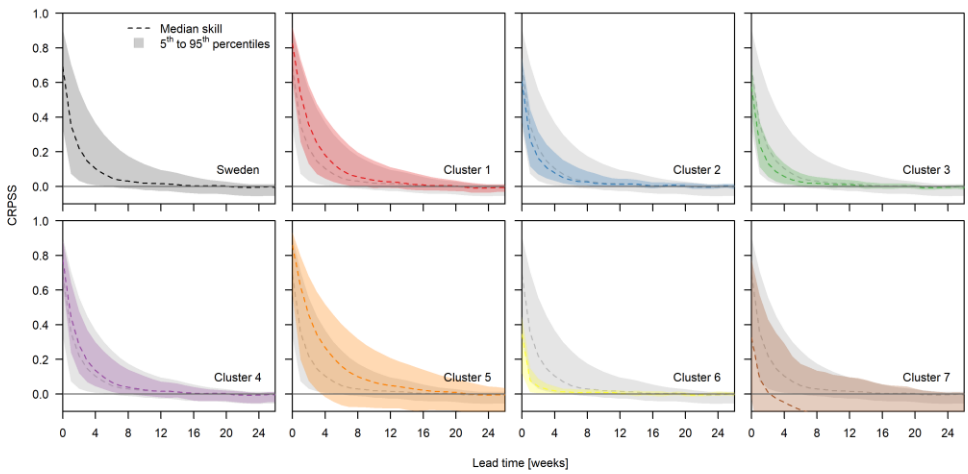
Figure 7~~6~~ Clustering analysis: (a) distribution of the 15 hydrological signatures in each clustered region. The red lines represent the 33rd and 66th percentiles for each signature; (b) geographical distribution of hydrologically similar regions over Sweden.

Table 2 Main characteristics of each of the hydrological clusters. The values provided for elevation, annual precipitation, and annual actual evapotranspiration correspond to the mean and interquartile range (25th – 75th percentiles).

Cluster region	Number of catchments	~~Mean~~ ~~e~~Elevation (m ~~-~~a.s.l.)	~~Mean~~ ~~a~~Annual precipitation (mm ~~-~~yr^{-1})	~~Mean~~ ~~a~~Annual actual evapotranspiration (mm ~~-~~yr^{-1})	Low streamflow signatures	High streamflow signatures
1	8406~~8~~	249~~7.66~~ (109 – 344)	650~~49.97~~ (557 – 694)	280~~.39~~ (232 – 334)	q05, CV, Flash, PD, RLD, HPC	q95, q70, BFI
2	6705~~4~~	478~~7.38~~ (228 – 744)	733~~2.18~~ (581 – 901)	175~~.17~~ (119 – 235)	q95, q70, BFI, EQP	Qm, DPar, q05, CV, PD, RC, HPC
3	82550~~73~~	165~~2.36~~ (6 – 278)	654~~4.22~~ (552 – 704)	294~~.87~~ (249 – 342)	BFI, RC	Flash, RLD, EQP, HPC

4	8305~~297~~	371~~0.18~~ (214 – 490)	653~~.06~~ (560 – 678)	219~~.51~~ (180 – 263)	Flash, RLD, DLD	DPar, PD
5	4329~~55~~	266~~59.55~~ (153 – 340)	642~~3.25~~ (556 – 693)	270~~.87~~ (221 – 318)	DPar, mFDC, q05, CV, Flash, PD, RLD, HPC	q95, q70, BFI
6	2405~~21~~	31~~22.47~~ (1 – 12)	732~~2.18~~ (544 – 824)	306~~5.74~~ (251 – 369)	DPar, q95, q70, BFI	mFDC, q05, CV, Flash, RLD, DLD, HPC
7	1025	224~~.10~~ (99 – 321)	619~~8.54~~ (552 – 639)	266~~.87~~ (233 – 294)	DPar, mFDC, q05, CV, PD, EQP, HPC	Qm, q95, DLD, q70, RC,

The last step is to analyse the streamflow forecast skill in each hydrological cluster (Figure 8~~Figure 7~~). Note that here we have aggregated the skill for all initialisations and hence we have not accessed the seasonal distribution of the forecast skill; however, we have focused on the detection of dependencies between skill and hydrologic regimes. Nevertheless, we note that the clusters with high (or poor) forecast skill in relation to the others are the same independently of the target month/week. According to Pechlivanidis et al. (2020), this is due to the intraannual variability of the streamflow response, which consistently varies between the catchments from the different clusters. Clusters 1, 4, and 5, which ~~are all considered to~~ have high river memory due to baseflow domination, small intraannual variability, and generally low response to precipitation (see Table 2), have a higher median skill than the country-average. Among them, cluster 5 has the highest overall median skill for all time horizons but also, interestingly, the highest spread in forecast skill as a function of lead time. This may be attributed to the large variability in rising limb density (RLD) for the catchments in this cluster (see Figure 7a). The strong but negative correlation between forecast skill and RLD means that, despite the high baseflow contribution (BFI), some of the catchments in cluster 5 experience sharp increases in their hydrographs, which translates in low forecast skill. All of these clusters correspond mainly to forested catchments across the country. Cluster 3 and, most notably cluster 6, have a lower median skill than the country-average. These catchments are characterized by short river memory with flashy responses and are strongly driven by precipitation and strong seasonal variations. Similar results are observed for cluster 2. In this case, however, the median skill is closer to the country-average skill than for clusters 3 and 6. The response from catchments in cluster 2 is highly seasonal due to snow accumulation and melting processes, and hence not as rainfall-driven as for clusters 3 and 6. Finally, cluster 7, which contains the catchments along the large regulated rivers in northern Sweden, is the only set of catchments in which the median forecast skill reaches negative values, including also a large spread in the skill values (5^{th} and 95^{th} percentiles). In these catchments, the ESP was expected to be outperformed by ~~historical streamflow~~streamflow climatology since, as previously mentioned, ~~historical streamflows~~the latter benefits from ~~the~~ AR-~~model~~ correction throughout the forecast period and thus can better reproduce regulation patterns with low intraannual variability.



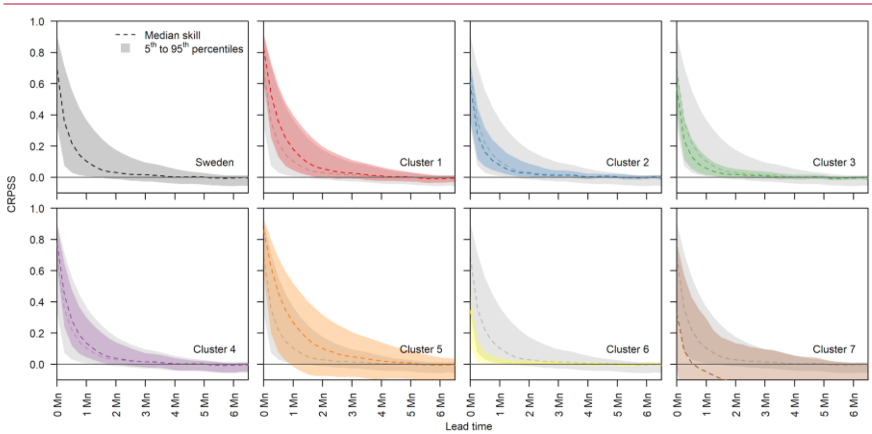


Figure 8~~7~~ Streamflow forecast skill (in terms of CRPSS) as a function of lead time for the entire country (top-left corner; also shown in Figure 3~~Figure 2~~a) and each of the 7 clusters. The median skill and range for Sweden are provided in the background of ~~C~~clusters 1-7 to provide a reference for the values of each cluster.

4 Discussion

4.1 Challenges and opportunities in an operational forecasting service

The results obtained in this study indicate that ESP seasonal forecasts produced with the operational S-HYPE hydrological model are skilful with respect to ~~historical streamflow~~streamflow climatology on average up to 3 months ahead, despite the large temporal and spatial variabilities. This positive skill would make operational seasonal forecasts, in general, suitable to guide decision-making for applications requiring long-term planning, (e.g. water resources management, agriculture). Nevertheless, issues related to the modelling setup, the forecast methodology, the hydro-climatic characteristics of the Swedish river systems (e.g. high degree of regulation in many water courses), among others, can impact the reliability of such a forecasting service.

The ESP forecasting approach is limited by its use of historical meteorological forcing data to generate the streamflow forecasts, making it unable to capture unprecedented meteorological events. Consequently, extreme events that lay outside the observed range will inevitably be misrepresented, limiting the service's predictability of extreme conditions which can be important to some decision makers. This issue may be addressed by using ~~numerical weather prediction (~~NWP~~)~~ models to predict the future climate (Monhart et al., 2019). However, although NWPs are not constrained by the observational period, they are limited by the chaotic nature of the weather system (aleatory uncertainty), which makes small errors in the initial conditions grow significant with time. In addition, NWP-based forecasts require post-processing (i.e. downscaling and bias-adjustment) to be suitable to use in impact studies. Finally, their added value for streamflow forecasting in comparison to ESP is shown to be limited in Sweden, with the possible exception of southern Sweden (Arnal et al., 2018).

As expected, ESP forecast skill decreases rapidly with time particularly in fast responding river systems. Results ~~have~~ showed~~n~~ that monthly initialization, which is the most common initialization frequency of climate prediction models, is critical to set high skill values; however, such frequency in the initialization cannot account for skill deterioration within the month. In this setting, increasing the initialisation frequency to, for instance, once a week would allow to maintain a high skill up to monthly time horizons. Nevertheless, considering that climate prediction models are not developed to represent the exact daily dynamics of the natural systems and that forecasts are therefore aggregated into long time periods, more frequent (e.g. daily) forecast initialisations are not expected to provide an added value to the forecast service in terms of useful information for decision-making at seasonal horizons since long-term decisions are in any case not taken daily. Moreover, forecast information from such frequent initialisations can easily be misinterpreted by decision-makers (Schepen et al., 2016).

Regarding the aggregation of forecast outputs, most studies have focused on a 1-month aggregation period as it is reported to provide an *"appropriate forecast at the seasonal scale and a proxy of the underlying distribution"* (Emerton et al., 2018; Meißner et al., 2017; Yossef et al., 2013). Nevertheless, since the way a seasonal forecasting service is used in decision-making depends on the sector, user, and service properties, there may be value in considering aggregation periods different from the standard monthly aggregation (or even adaptive aggregation periods) for providing guidance on the usability of the forecasts for decision-making. For instance, for the energy sector, Swedish hydropower companies tend to be interested in a fixed 3-month aggregation over the period May-July. Alternatively, crop water needs can be assessed over the entire summer season to get estimates of required water volumes for irrigation. The produced matrix for different aggregations, initialisations, and lead times (Figure 5) allows communication of skill to various users depending on their needs. ~~This choice should be driven by user needs and o~~Our findings suggest~~is~~ that aggregations over periods longer than the default 1-month do not necessarily mean a loss in skill. On the contrary, ~~H~~here~~,~~ we ~~have~~ observed that, in Sweden, long aggregations~~ed~~ of streamflow forecasts covering the spring flood season tend to gain in skill. Overall, however, from time horizons of, on average, 4 months into the future, forecasts have very low or no skill regardless of the aggregation period of choice.

Another important factor driving hydrological predictability at the seasonal scale is the adequate knowledge of the initial hydrological conditions (Shukla et al., 2013). In many cases, ESP forecasts are initialized based on the latest available model state (modelled reality), which may significantly deviate from the actual hydrological state (observed reality). Incorporating the latest available observations into forecast initialisation can thus be especially important to bridge the gap between modelled and observed reality. Here, ~~an AR-updating method is used to correct~~ the model outputs were corrected whenever observations ~~are~~ were available (and using AR-correction thereafter), with the objective to generate forecasts which are as close as possible to observed reality (see Section 2.2). This method is straightforward and easy to implement, and takes advantage of streamflow memory to not only correct the initial forecast state but also the following forecast horizons when observations are no longer available. More advanced data assimilation methods could be considered in further developments of the presented operational forecast system, such as Kalman Filters (Sun et al., 2016), allowing not only for a correction of model outputs, but also an adjustment of model states and thus of process representation (Musuuza et al., 2020). Additionally, observations other than streamflow, such as soil moisture or snow water equivalent, could also be assimilated into the model (Huang et al., 2017; Musuuza et al., 2020). Regarding snow water equivalent, snow is a key component of the hydrological cycle in many Swedish catchments and therefore, the impact of snow accumulation and melting on ESP forecast skill would deserve further investigation. Snow processes play an important role in river memory together with other processes such as groundwater/baseflow contribution, or hydrograph dampening from lakes. Contrary to the other two, however, snow processes tend to define the catchment dynamics only seasonally (e.g. precipitation in the form of snow in early December may be accumulated and further released as meltwater during the spring flood period), and hence the role of snow on ESP forecast skill is expected to have a seasonal pattern too.

Another approach to obtaining updated knowledge on the initial hydrological conditions is through frequent forecast initialisation. Our findings suggest that using weekly forecast initialisation instead of the more common monthly initialisation may significantly improve the streamflow ESP forecast skill, which is expected to add value to decision-making in different contexts. This may be of particular importance for periods in which decisions are subject to hydrological responses that alter in a short time window. For instance, in Sweden it is important to be able to predict the onset of the spring flood due to a combination of snow melting and precipitation, and adjust the reservoir regulation accordingly to optimise the power production for the coming months.

Different components of the S-HYPE modelling and forecasting chains, such as the model setup, forcing data, model structure and model parameters, lead to uncertainties in the forecasts results. Moreover, the setup used in this study, which uses a combination of observations and perfect forecasts as reference, makes the assessment of these uncertainties particularly complex. The contribution of model error to the total uncertainties in the results obtained here is removed from those catchments in which forecasts are evaluated against perfect forecasts. This non-represented contribution of model errors can nonetheless be considered minimal due to the high KGE performances of S-HYPE (see Figure 1a), which ensure a fair representation of temporal dynamics in non-regulated Swedish rivers. However, these errors may become significant for catchments with – or downstream of – observations, especially due to the interplay between correction of model outputs with observations and streamflow regulation. While model outputs are corrected with all available observations, not all watercourses with observations are regulated, and even those that are regulated do not necessarily have observations downstream from dams or other river regulation structures. The correction of model outputs with observations and, when these are no longer available (e.g. beyond forecast initialisation), with an exponentially decreasing factor based on the last known model error (i.e. AR-correction) may effectively reduce corresponding uncertainties, especially in the first time steps of the forecast. The downstream distance of a given catchment with respect to an observation is also relevant in this case, as model correction will only affect a fraction of the streamflow forecast at that location. The largest uncertainties, though, can be expected for heavily-regulated catchments with or downstream of observations. In these locations, complex river regulation routines, which depend on factors external to hydrological models, make it almost impossible for these models to adequately reproduce streamflow dynamics. Consequently, even if the correction of model outputs with observations may minimise uncertainties at forecast initialisation, these errors will rapidly spread due to the inability of the model to reproduce the modified hydrological regime.

4.2 Impact of regulation on forecasting skill

One of the main applications of long-term forecasting in Sweden is for planning reservoir operation during the spring flood season (May-July) (Foster et al., 2018). However, forecast skill is low for the main hydropower-producing heavily-regulated rivers in the northern parts of the country, where the highest spring flood peak volumes occur. Nevertheless, high ESP forecast skill, and subsequently valuable forecast information, can still be expected in the upstream reaches of these rivers that are not affected by other upstream regulation. Following this assumption, ~~I~~in these locations, even if ESP forecasts ~~may be used to adequately predict the water inflows to the reservoirs from the headwaters but they would~~ have no value for predicting reservoir outflows with respect to using the ensemble of ~~historical streamflow~~streamflow climatology, they may adequately predict the water inflows from the headwaters to the reservoirs. In order to further understand the impact of streamflow regulation on the results, we evaluated the ESP forecasts using model simulations (without AR-~~–~~correction) as reference. Forecast skill was in this case very high for the highly regulated rivers where low forecast skill was obtained in the main analysis. This exercise shows that the regulation routines in some river stations in the S-HYPE model still need improvement in order to correctly represent the management rules dominating regulated streamflow patterns. This issue is not as obvious in less heavily regulated rivers elsewhere in the country, where ESP forecasts are generally skilful.

With the exception of River Luleälven and other comparatively smaller rivers in the Swedish mountains, the S-HYPE model performance is generally high for most locations, including the large rivers in the northern parts of the country. Similarly, ESP seasonal forecasts are skilful for non-regulated rivers in that area that also benefit from long-term planning. More specifically, River Torneälven and, to a lesser extent, River Kalixälven are susceptible to severe ice break-up events in connection to the spring melt season and subsequent spring flooding (Zachrisson, 1989). An important factor in predicting the timing of the ice break-up is the onset of spring flood due to snowmelt. Skilful ESP seasonal forecasts for these rivers should allow for early planning and allocation of resources that could greatly contribute to mitigate potentially severe ice break-ups.

4.3 Regionalisation of skill in other domains

Besides streamflow regulation patterns, certain characteristics of the hydrological regime have a high impact on hydrological predictability. Here we have shown that forecast skill is high in baseflow-dominated catchments where past hydrologic conditions drive the catchment response, while it is low in flashy catchments where rainfall drives the streamflow dynamics and hence accurate rainfall forecasts are crucial. This corresponds well with findings from similar studies over different geographical domains (Harrigan et al., 2018; Pechlivanidis et al., 2020). However, contrary to the findings by (Harrigan et al., 2018), who identified a specific streamflow signature (i.e. the baseflow index) as the main driver of hydrological predictability, we have found that, for Sweden, it is instead the result of the overall hydrological behaviour, even if some specific streamflow signatures may have a greater impact than others. Additionally, the seven clusters not only differ in terms of hydrological response, but also in terms of climatological patterns and physiographic characteristics.

The results obtained here may contribute to guiding in which areas, seasons, and how long into the future ESP hydrological forecasts provide an added value not only for SMHI's forecasting and warning service, but most importantly for guiding decision-making in critical services such as hydropower management and risk reduction. Here, we note that, even if the hydro-climatic gradient of Sweden does not fully represent the equivalent gradients over the continent or the globe, our results are however transferable to other locations with similar climatological and hydrological conditions as it has also been highlighted in Pechlivanidis et al. (2020).

5 Conclusions

Herein, we analysed the skill of ESP re-forecasts using the operational S-HYPE hydrological model over Sweden in an effort to evaluate the suitability of this methodology for producing ~~reliable~~ skilful forecasts at the seasonal scale within SMHI's hydrological forecasting and warning service as well as for other activities requiring long-term planning. In addition, we aimed at understanding the underlying patterns and drivers behind skilful forecasts and attributed the seasonal predictability to hydrological characteristics. ~~About~~ Approximately 39,400 catchments~~, which~~ lying~~ie~~ along Sweden's strong hydroclimatic gradient~~,~~ were investigated. The main conclusions of this study are:

- The skill of the ESP forecasts varies both geographically and seasonally, and depends on the initialization month and aggregation period. Moreover, the skill decreases rapidly with time particularly in fast responding river systems; however, the ESP forecasts are generally skilful up to 3 months into the future. Forecasts are most skilful during the winter months for the northern parts of the country, except for the highly-regulated hydropower-producing rivers.
- Initialization frequency is a key driver affecting streamflow forecasting skill. Monthly initialisations are critical to preserve high forecast skill values without, however, addressing the skill deterioration over the first forecast month. Increasing the initialisation frequency to once a week allows maintaining the high skill up to monthly time horizons.
- The river systems in Sweden can be categorised into 7 clusters based on similarities in streamflow signatures. This results in an improved understanding of the dominating hydrological processes, which are shown to vary spatially and seasonally. Particularly, dominant streamflow generation processes over the mountainous regions, including baseflow and snow accumulation/melting, dampening from lakes, and reservoir alterations could explain the hydrological clustering across the country.
- A link between forecast skill and streamflow signatures has been detected. Over the 15 streamflow signatures investigated here, baseflow index, flashiness, rising limb density, coefficient of variation and high pulse count show strong correlations with forecast skill. Streamflow forecasts are most skilful for slowly-reacting catchments due to snow-related processes and/or dampening from lakes and baseflow-dominated catchments (river systems with long memory). Conversely, forecasts are least skilful for catchments with a flashy response to rainfall (river systems with short memory).

510 **Appendix A: Study domain and data availability**

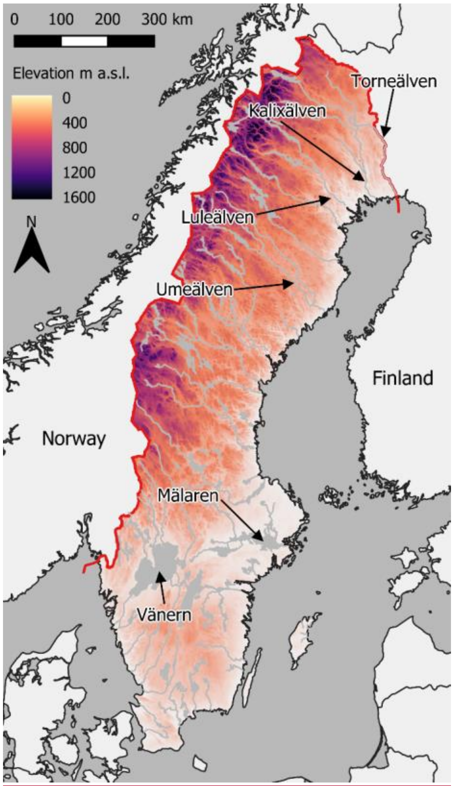


Figure A1 Map of Sweden showing its topography and main hydrographic network. The rivers and lakes referred to in the main text are indicated here for convenience.

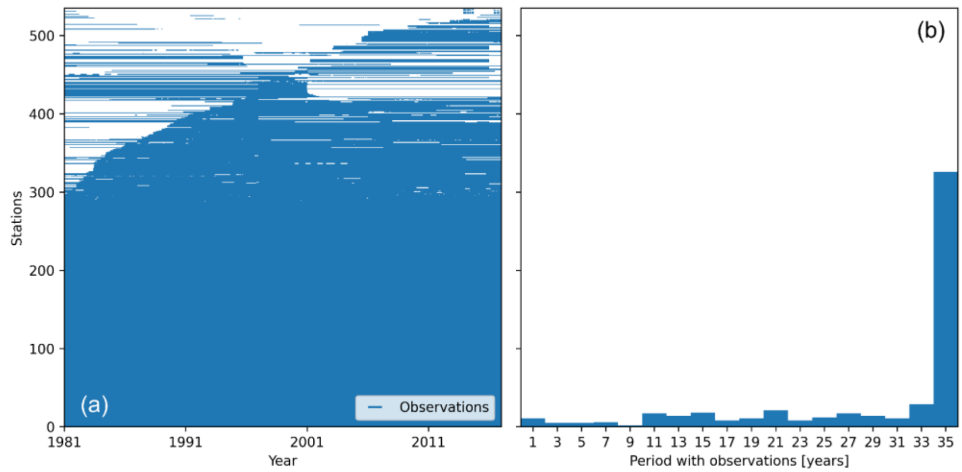


Figure A2 Streamflow observations used in this study: (a) Temporal availability of observations for each of the 539 stations; sorted from longer (bottom) to shorter availability (top); (b) Histogram showing the total number of years of observations for all the stations.

Appendix B~~A~~: Continuous Ranked Probability Skill Score

The Continuous Ranked Probability Score (CRPS; Hersbach, 2000) is a common measure of ensemble forecast performance. It is formulated as the integral squared distance between the forecast ensemble and the observation step function. The CRPS is then averaged over all forecasts of the evaluation period. Its dimension is that of the forecast variable being assessed, here $m^3 s^{-1}$, and its value is equivalent to the mean absolute error when applied to deterministic forecasts.

The Continuous Ranked Probability Skill Score (CRPSS) is then assessed by comparing the CRPS value of the investigated forecast system (here, ESP) to that of a selected benchmark (here, an ensemble of ~~historical streamflow~~streamflow climatology selected from the period 1981 – 2010). Given $CRPS_{sys}$, the CRPS of the forecasting system, $CRPS_{bench}$ the CRPS of the benchmark and $CRPS_{pft}$ the optimal CRPS value (0), the CRPSS is formulated according to Eq. A1.

$$CRPSS = \frac{CRPS_{bench}\cancel{CRPS^{bench}} - CRPS_{sys}\cancel{CRPS^{sys}}}{CRPS_{bench}\cancel{CRPS^{bench}} - CRPS_{pft}\cancel{CRPS^{pft}}} \qquad \text{(A1)}$$

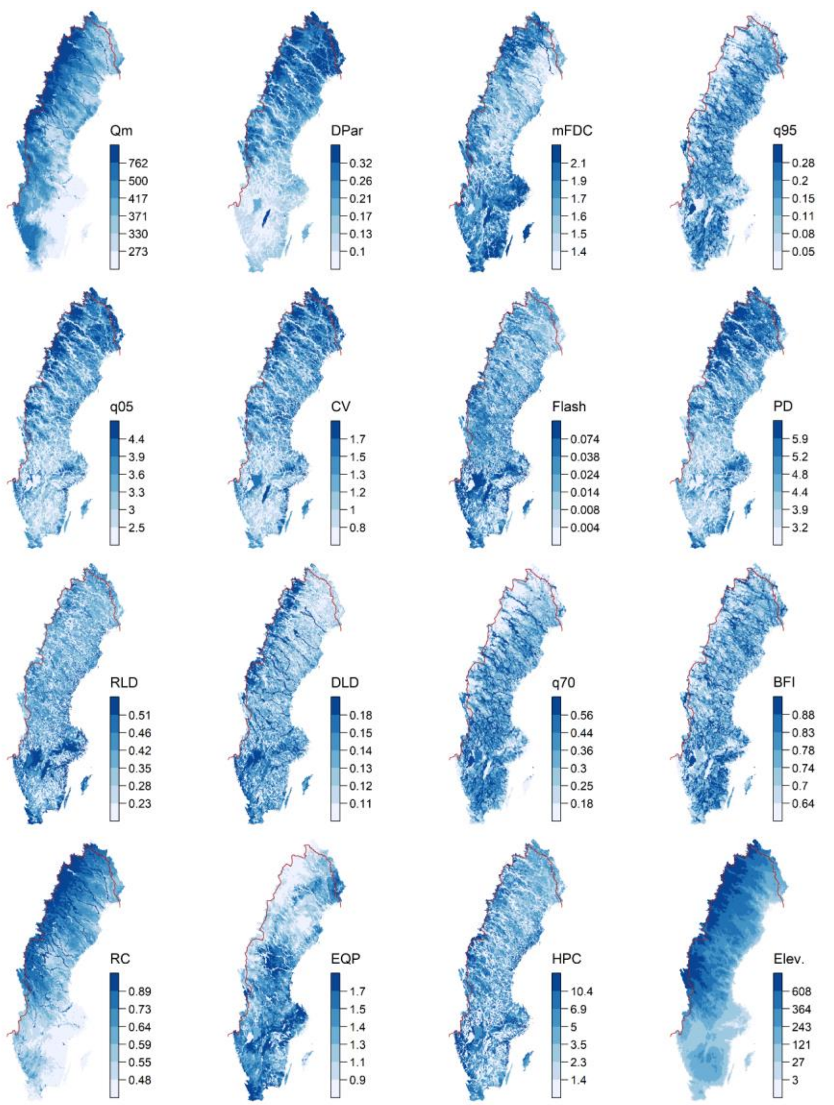
This metric is non-dimensional and takes values between 1 (optimum) and ~~low negative values~~ $-\infty$. Positive (negative) skill scores indicate that the forecast system performs better (worse) than the benchmark in terms of CRPS. Skill scores close to 0 indicate that the evaluated forecast system has equivalent performance to that of the benchmark.

Formaterat: Nedsänkt

Formaterat: Nedsänkt

Formaterat: Nedsänkt

Appendix C~~B~~: Spatial variability of hydrological signatures


Qm
762
500
417
371
330
273
DPar
0.32
0.26
0.21
0.17
0.13
0.1
mFDC
2.1
1.9
1.7
1.6
1.5
1.4
q95
0.28
0.2
0.15
0.11
0.08
0.05
q05
4.4
3.9
3.6
3.3
3
2.5
CV
1.7
1.5
1.3
1.2
1
0.8
Flash
0.074
0.038
0.024
0.014
0.008
0.004
PD
5.9
5.2
4.8
4.4
3.9
3.2
RLD
0.51
0.46
0.42
0.35
0.28
0.23
DLD
0.18
0.15
0.14
0.13
0.12
0.11
q70
0.56
0.44
0.36
0.3
0.25
0.18
BFI
0.88
0.83
0.78
0.74
0.7
0.64
RC
0.89
0.73
0.64
0.59
0.55
0.48
EQP
1.7
1.5
1.4
1.3
1.1
0.9
HPC
10.4
6.9
5
3.5
2.3
1.4
Elev.
608
364
243
121
27
3

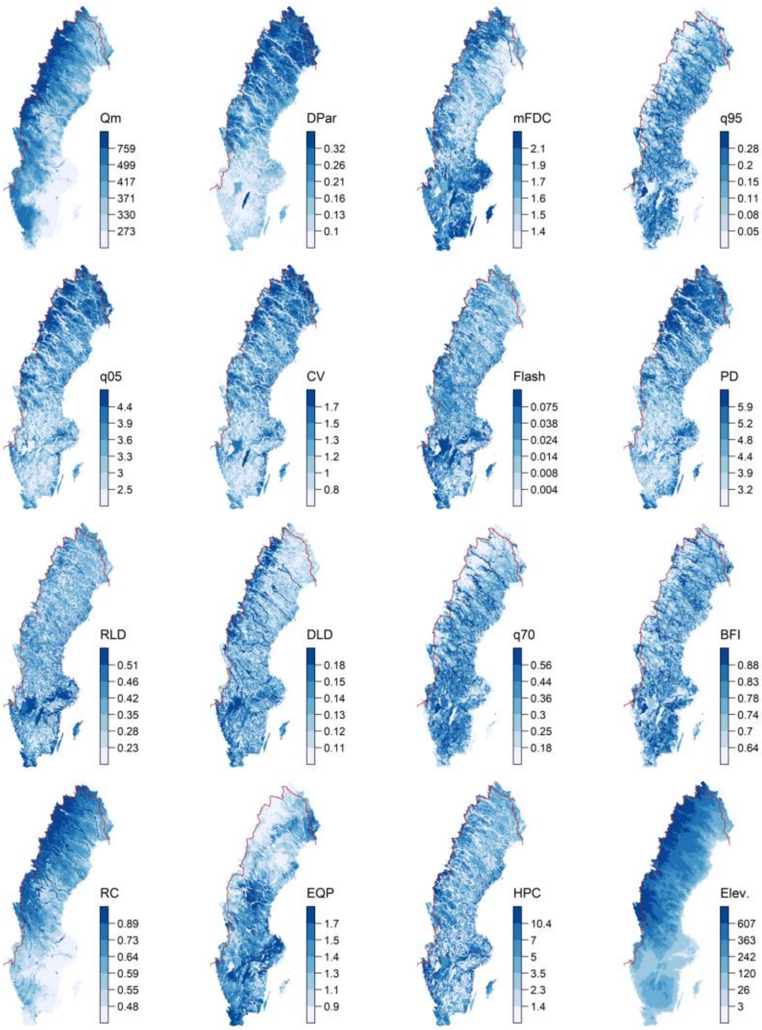


Figure C1 ~~Figure B1~~ Spatial variability of the 15 modelled hydrological signatures including the catchment mean elevation. The colour intervals are based on the quantiles (15% intervals) of each signature (and elevation) distribution. A clarification of the abbreviations used here can be found in Table 1 in the main text.

Data availability

The HYPE model code is available from the *HYPEweb* portal (https://hypeweb.smhi.se/model-water/). The meteorological data used for driving the ESP re-forecasts (PTHBV) is available from the ~~l~~*Luft*~~w~~*Webb* portal (https://luftweb.smhi.se/), and the hydrological data used for model correction is available from the ~~v~~*Vattenwebb* portal (https://vattenwebb.smhi.se/).

Author contribution

M.G.L. contributed with the study design, model runs, result analysis and figures, interpretation of the results, and writing the manuscript; L.C. contributed with the study design, code development for post-processing of results, the interpretation of results, and writing the manuscript; I.G.P. was responsible for the project management and funding acquisition, and contributed with the basic idea, the study design, clustering analysis and figures, interpretation of the results, and writing the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Financial support

This work was funded by the project “Long-term forecasts of wind and hydropower supply in a fluctuating climate – Importance for production planning and investments in energy storage and power transmission” granted by the Swedish Energy Agency under grant agreement No. 46412-1. Funding was also received from the EU Horizon 2020 project S2S4E (Subseasonal to seasonal forecasting for the energy sector) under Grant Agreement 776787. This study was also partially funded by the EU Horizon 2020 project PrimeWater (Delivering advanced predictive tools from medium to seasonal range for water dependent industries exploiting the cross-cutting potential of EO and hydroecological modeling) under the Grant Agreement 870497.

References

de Amorim, R. C. and Hennig, C.: Recovering the number of clusters in data sets with noise features using feature rescaling factors, Inf. Sci., 324, 126–145, https://doi.org/10.1016/j.ins.2015.06.039, 2015.

Apel, H., Abdykerimova, Z., Agalhanova, M., Baimaganbetov, A., Gavrilenko, N., Gerlitz, L., Kalashnikova, O., Unger-Shayesteh, K., Vorogushyn, S. and Gafurov, A.: Statistical forecast of seasonal discharge in Central Asia using observational records: development of a generic linear modelling tool for operational water resource management, Hydrol. Earth Syst. Sci., 22(4), 2225–2254, https://doi.org/10.5194/hess-22-2225-2018, 2018.

Arnal, L., Cloke, H. L., Stephens, E., Wetterhall, F., Prudhomme, C., Neumann, J., Krzeminski, B. and Pappenberger, F.: Skilful seasonal forecasts of streamflow over Europe?, Hydrol. Earth Syst. Sci., 22(4), 2057–2072, 2018.

Batté, L. and Déqué, M.: Randomly correcting model errors in the ARPEGE-Climate v6.1 component of CNRM-CM: applications for seasonal forecasts, Geosci. Model Dev., 9(6), 2055–2076, https://doi.org/10.5194/gmd-9-2055-2016, 2016.

Bennett, J. C., Wang, Q. J., Robertson, D. E., Schepen, A., Li, M. and Michael, K.: Assessment of an ensemble seasonal streamflow forecasting system for Australia, Hydrol. Earth Syst. Sci., 21(12), 6007–6030, https://doi.org/10.5194/hess-21-6007-2017, 2017.

Bergström, S.: Development and Application of a Conceptual Runoff Model for Scandinavian Catchments, Swedish Meteorological and Hydrological Institute (SMHI), Norrköping, Sweden., 1976.

Bohn, T. J., Sonessa, M. Y. and Lettenmaier, D. P.: Seasonal Hydrologic Forecasting: Do Multimodel Ensemble Averages Always Yield Improvements in Forecast Skill?, J. Hydrometeorol., 11(6), 1358–1372, https://doi.org/10.1175/2010JHM1267.1, 2010.

Bruno Soares, M., Alexander, M. and Dessai, S.: Sectoral use of climate information in Europe: A synoptic overview, Clim. Serv., 9, 5–20, https://doi.org/10.1016/j.cliser.2017.06.001, 2018.

Candogan Yossef, N., Van Beek, R., Weerts, A., Winsemius, H. and Bierkens, M. F. P.: Skill of a global forecasting system in seasonal ensemble streamflow prediction, Hydrol. Earth Syst. Sci., 21(8), 4103–4114, https://doi.org/10.5194/hess-21-4103-2017, 2017.

Crochemore, L., Ramos, M. H., Pappenberger, F. and Perrin, C.: Seasonal streamflow forecasting by conditioning climatology with precipitation indices, Hydrol. Earth Syst. Sci., 21(3), 1573–1591, https://doi.org/10.5194/hess-21-1573-2017, 2017.

Crochemore, L., Ramos, M. -H. and Pechlivanidis, I. G.: Can Continental Models Convey Useful Seasonal Hydrologic Information at the Catchment Scale?, Water Resour. Res., 56(2), https://doi.org/10.1029/2019WR025700, 2020.

Day, G. N.: Extended Streamflow Forecasting Using NWSRFS, J. Water Resour. Plan. Manag., 111(2), 157–170, https://doi.org/10.1061/(ASCE)0733-9496(1985)111:2(157), 1985.

DeChant, C. and Moradkhani, H.: Improving the Characterization of Initial Condition for Ensemble Streamflow Prediction Using Data Assimilation, Hydrol. Earth Syst. Sci., https://doi.org/10.5194/hess-15-3399-2011, 2011.

Donnelly, C., Andersson, J. C. M. and Arheimer, B.: Using flow signatures and catchment similarities to evaluate the E-HYPE multi-basin model across Europe, Hydrol. Sci. J., 61(2), 255–273, https://doi.org/10.1080/02626667.2015.1027710, 2016.

Emerton, R., Zsoter, E., Arnal, L., Cloke, H. L., Muraro, D., Prudhomme, C., Stephens, E. M., Salamon, P. and Pappenberger, F.: Developing a global operational seasonal hydro-meteorological forecasting system: GloFAS-Seasonal v1.0, Geosci. Model Dev., 11(8), 3327–3346, https://doi.org/10.5194/gmd-11-3327-2018, 2018.

Euser, T., Winsemius, H. C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S. and Savenije, H. H. G.: A framework to assess the realism of model structures using hydrological signatures, Hydrol. Earth Syst. Sci., 17(5), 1893–1912, https://doi.org/10.5194/hess-17-1893-2013, 2013.

Foster, K., Bertacchi Uvo, C. and Olsson, J.: The development and evaluation of a hydrological seasonal forecast system prototype for predicting spring flood volumes in Swedish rivers, Hydrol. Earth Syst. Sci., 22(5), 2953–2970, https://doi.org/10.5194/hess-22-2953-2018, 2018.

Girons Lopez, M., Di Baldassarre, G. and Seibert, J.: Impact of social preparedness on flood early warning systems, Water Resour. Res., 53(1), 522–534, https://doi.org/10.1002/2016WR019387, 2017.

Giuliani, M., Crochemore, L., Pechlivanidis, I. and Castelletti, A.: From skill to value: isolating the influence of end-user behaviour on seasonal forecast assessment, Hydrol. Earth Syst. Sci. Discuss., 1–20, https://doi.org/10.5194/hess-2019-659, 2020.

Greuell, W., Franssen, W. H. P. and Hutjes, R. W. A.: Seasonal streamflow forecasts for Europe – Part 2: Sources of skill, Hydrol. Earth Syst. Sci., 23(1), 371–391, https://doi.org/10.5194/hess-23-371-2019, 2019.

Gupta, H. V., Kling, H., Yilmaz, K. K. and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J. Hydrol., 377(1), 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

Harrigan, S., Prudhomme, C., Parry, S., Smith, K. and Tanguy, M.: Benchmarking ensemble streamflow prediction skill in the UK, Hydrol. Earth Syst. Sci., 22(3), https://doi.org/10.5194/hess-22-2023-2018, 2018.

Hersbach, H.: Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems, Weather Forecast., 15(5), 559–570, https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2, 2000.

Huang, C., Newman, A. J., Clark, M. P., Wood, A. W. and Zheng, X.: Evaluation of snow data assimilation using the ensemble Kalman filter for seasonal streamflow prediction in the western United States, Hydrol. Earth Syst. Sci., 21(1), 635–650, https://doi.org/10.5194/hess-21-635-2017, 2017.

Jeong, D.-I. and Kim, Y.-O.: Rainfall-runoff models using artificial neural networks for ensemble streamflow prediction, Hydrol. Process., 19(19), 3819–3835, https://doi.org/10.1002/hyp.5983, 2005.

Jin, X. and Han, J.: K-Means Clustering, in Encyclopedia of Machine Learning, edited by C. Sammut and G. I. Webb, pp. 563–564, Springer US, Boston, MA, https://doi.org/10.1007/978-0-387-30164-8_425, , 2011.

Johansson, B.: Estimation of areal precipitation for hydrological medelling in Sweden, https://gupea.ub.gu.se/handle/2077/15575, last access: 24 June 2020, 2002.

Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., Tietsche, S., Decremer, D., Weisheimer, A., Balsamo, G., Keeley, S. P. E., Mogensen, K., Zuo, H. and Monge-Sanz, B. M.: SEAS5: the new ECMWF seasonal forecast system, Geosci. Model Dev., 12(3), 1087–1117, https://doi.org/10.5194/gmd-12-1087-2019, 2019.

Kuentz, A., Arheimer, B., Hundecha, Y. and Wagener, T.: Understanding hydrologic variability across Europe through catchment classification, Hydrol. Earth Syst. Sci., 21(6), 2863–2879, https://doi.org/10.5194/hess-21-2863-2017, 2017.

Lindström, G. and Alexandersson, H.: Recent Mild and Wet Years in Relation to Long Observation Records and Future Climate Change in Sweden, AMBIO J. Hum. Environ., 33(4), 183–186, https://doi.org/10.1579/0044-7447-33.4.183, 2004.

Lindström, G. and Carlsson, B.: Evaluation of updating methods for short range forecasting by the HBV model, in Proceedings XXI Nordic Hydrological Conference, vol. 46, pp. 380–386, Nordic Hydrological Program, Uppsala, Sweden, , 2000.

Lindström, G., Pers, C., Rosberg, J., Strömqvist, J. and Arheimer, B.: Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales, Hydrol. Res., 41(3–4), 295–319, https://doi.org/10.2166/nh.2010.007, 2010.

Macian-Sorribes, H., Pechlivanidis, I., Crochemore, L. and Pulido-Velazquez, M.: Fuzzy Postprocessing to Advance the Quality of Continental Seasonal Hydrological Forecasts for River Basin Management, J. Hydrometeorol., 21(10), 2375–2389, https://doi.org/10.1175/JHM-D-19-0266.1, 2020.

Meißner, D., Klein, B. and Ionita, M.: Development of a monthly to seasonal forecast framework tailored to inland waterway transport in central Europe, Hydrol. Earth Syst. Sci., 21(12), 6401–6423, https://doi.org/10.5194/hess-21-6401-2017, 2017.

Monhart, S., Zappa, M., Spirig, C., Schär, C. and Bogner, K.: Subseasonal hydrometeorological ensemble predictions in small- and medium-sized mountainous catchments: benefits of the NWP approach, Hydrol. Earth Syst. Sci., 23(1), 493–513, https://doi.org/10.5194/hess-23-493-2019, 2019.

Muhammad, A., Stadnyk, T. A., Unduche, F. and Coulibaly, P.: Multi-Model Approaches for Improving Seasonal Ensemble Streamflow Prediction Scheme with Various Statistical Post-Processing Techniques in the Canadian Prairie Region, Water, 10(11), 1604, https://doi.org/10.3390/w10111604, 2018.

Musuuza, J. L., Gustafsson, D., Pimentel, R., Crochemore, L. and Pechlivanidis, I.: Impact of Satellite and In Situ Data Assimilation on Hydrological Predictions, Remote Sens., 12(5), 811, https://doi.org/10.3390/rs12050811, 2020.

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(2), 659–667, https://doi.org/10.5194/hess-20-659-2016, 2016.

Pappenberger, F., Ramos, M. H., Cloke, H. L., Wetterhall, F., Alfieri, L., Bogner, K., Mueller, A. and Salamon, P.: How do I know if my forecasts are better? Using benchmarks in hydrological ensemble prediction, J. Hydrol., 522, 697–713, https://doi.org/10.1016/j.jhydrol.2015.01.024, 2015a.

Pappenberger, F., Cloke, H. L., Parker, D. J., Wetterhall, F., Richardson, D. S. and Thielen, J.: The monetary benefit of early flood warnings in Europe, Environ. Sci. Policy, 51, 278–291, https://doi.org/10.1016/j.envsci.2015.04.016, 2015b.

Pechlivanidis, I. and Arheimer, B.: Large-scale hydrological modelling by using modified PUB recommendations : the India-HYPE case, Hydrol. Earth Syst. Sci., 19(11), 4559–4579, 2015.

Pechlivanidis, I., Bosshard, T., Spångmyr, H., Lindström, G., Gustafsson, D. and Arheimer, B.: Uncertainty in the Swedish Operational Hydrological Forecasting Systems, in Vulnerability, Uncertainty, and Risk, pp. 253–262, Liverpool, UK, https://doi.org/10.1061/9780784413609.026, , 2014.

Pechlivanidis, I. G., Gupta, H. and Bosshard, T.: An Information Theory Approach to Identifying a Representative Subset of Hydro-Climatic Simulations for Impact Modeling Studies, Water Resour. Res., 54(8), 5422–5435, https://doi.org/10.1029/2017WR022035, 2018.

Pechlivanidis, I. G., Crochemore, L., Rosberg, J. and Bosshard, T.: What Are the Key Drivers Controlling the Quality of Seasonal Streamflow Forecasts?, Water Resour. Res., 56(6), e2019WR026987, https://doi.org/10.1029/2019WR026987, 2020.

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A. and Carrillo, G.: Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA, Hydrol. Earth Syst. Sci., 15(9), 2895–2911, https://doi.org/10.5194/hess-15-2895-2011, 2011.

Schepen, A., Zhao, T., Wang, Q. J., Zhou, S. and Feikema, P.: Optimising seasonal streamflow forecast lead time for operational decision making in Australia, Hydrol. Earth Syst. Sci., 20(10), 4117–4128, https://doi.org/10.5194/hess-20-4117-2016, 2016.

Shukla, S., Sheffield, J., Wood, E. F. and Lettenmaier, D. P.: On the sources of global land surface hydrologic predictability, Hydrol. Earth Syst. Sci., 17(7), 2781–2796, https://doi.org/10.5194/hess-17-2781-2013, 2013.

Singla, S., Céron, J.-P., Martin, E., Regimbeau, F., Déqué, M., Habets, F. and Vidal, J.-P.: Predictability of soil moisture and river flows over France for the spring season, Hydrol. Earth Syst. Sci., 16(1), 201–216, https://doi.org/10.5194/hess-16-201-2012, 2012.

Strömqvist, J., Arheimer, B., Dahné, J., Donnelly, C. and Lindström, G.: Water and nutrient predictions in ungauged basins: set-up and evaluation of a model at the national scale, Hydrol. Sci. J., 57(2), 229–247, https://doi.org/10.1080/02626667.2011.637497, 2012.

Sun, L., Seidou, O., Nistor, I. and Liu, K.: Review of the Kalman-type hydrological data assimilation, Hydrol. Sci. J., 61(13), 2348–2366, https://doi.org/10.1080/02626667.2015.1127376, 2016.

Sutanto, S. J., Van Lanen, H. A. J., Wetterhall, F. and Llort, X.: Potential of Pan-European Seasonal Hydrometeorological Drought Forecasts Obtained from a Multihazard Early Warning System, Bull. Am. Meteorol. Soc., 101(4), E368–E393, https://doi.org/10.1175/BAMS-D-18-0196.1, 2020.

Trambauer, P., Werner, M., Winsemius, H. C., Maskey, S., Dutra, E. and Uhlenbrook, S.: Hydrological drought forecasting and skill assessment for the Limpopo River basin, southern Africa, Hydrol. Earth Syst. Sci., 19(4), 1695–1711, https://doi.org/10.5194/hess-19-1695-2015, 2015.

Viglione, A., Parajka, J., Rogger, M., Salinas, J. L., Laaha, G., Sivapalan, M. and Blöschl, G.: Comparative assessment of predictions in ungauged basins – Part 3: Runoff signatures in Austria, Hydrol. Earth Syst. Sci., 17(6), 2263–2279, https://doi.org/10.5194/hess-17-2263-2013, 2013.

Vigo, I., Badal, M., Jiménez, I., Soret, A., Kielmanowicz, D., Salel, M., Aaheim, H. A., Hermansen, E. A., Sillmann, J., Photiadou, C., Pechlivanidis, I. G. and Hernandez, K.: User needs and decision-making processes that can benefit from S2S forecasts, Deliverable D2.1, S2S4E Climate Services for Clean Energy., 2018.

Wanders, N., Thober, S., Kumar, R., Pan, M., Sheffield, J., Samaniego, L. and Wood, E. F.: Development and Evaluation of a Pan-European Multimodel Seasonal Hydrological Forecasting System, J. Hydrometeorol., 20(1), 99–115, https://doi.org/10.1175/JHM-D-18-0040.1, 2019.

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(5), https://doi.org/10.1029/2010WR009922, 2011.

Wood, A. W. and Lettenmaier, D. P.: A Test Bed for New Seasonal Hydrologic Forecasting Approaches in the Western United States, Bull. Am. Meteorol. Soc., 87(12), 1699–1712, https://doi.org/10.1175/BAMS-87-12-1699, 2006.

Wood, A. W. and Schaake, J. C.: Correcting Errors in Streamflow Forecast Ensemble Mean and Spread, J. Hydrometeorol., 9(1), 132–148, https://doi.org/10.1175/2007JHM862.1, 2008.

Yadav, M., Wagener, T. and Gupta, H.: Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, Adv. Water Resour., 30(8), 1756–1774, https://doi.org/10.1016/j.advwatres.2007.01.005, 2007.

Yossef, N. C., Winsemius, H., Weerts, A., Beek, R. van 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(8), 4687–4699, https://doi.org/10.1002/wrcr.20350, 2013.

Zachrisson, G.: Svåra islossningar i Torneälven - Förslag till skadeförebyggande åtgärder, SMHI, Norrköping, Sweden., 1989.

Formaterat: Svenska (Sverige)