Authors’ response to interactive comment by Dr Louise Arnal

Black text: Reviewer comment

Blue text: Authors’ response

In this paper, the authors evaluate the performance of an ensemble streamflow prediction (ESP) hindcast dataset for seasonal streamflow forecasting in Sweden, produced with the S-HYPE hydrological model driven by resampled historical meteorological forcings. They look at the ESP hindcast skill against a benchmark, historical streamflow climatology, for 39,493 Swedish catchments. They overall found that the ESP is skilful up to 3 months ahead in Sweden, but that the skill varies in space and time, depending on: the aggregation period selected, the catchment’s hydro-climatic characteristics and regulation. They analyzed the skill against hydrological signatures, clustering basins in 7 geographical clusters in Sweden, and found that higher skill values are associated with baseflow-driven catchments. This manuscript is overall well-written and the sound methodology leads to valuable findings both for research and for operational streamflow forecasting in Sweden. Since the focus of this manuscript is on operational forecasting to guide decision-making, further context and discussion around the potential impacts of these findings on operational decision-making is crucial. Below, please find specific comments which I hope will be helpful in shaping this manuscript further for publication.

We thank Dr Louise Arnal for her valuable comments and suggestions that will undoubtedly help us improve our manuscript. Below we reply to each of the comments and explain how we will incorporate them into the manuscript.

Specific comments

Section 1:

- P1 L27: “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.” Do you have any quote or public material you could share about needs of reservoir operators in Sweden? It would help emphasize the user-oriented aspect of your paper.

Here we plan to reference the work by Foster et al. (2018), which refers to the Swedish hydropower needs. In addition, the recently accepted publication by Giuliani et al. (2020) quantifies the added economic value from incorporating seasonal forecasts in a regulated reservoir for the agriculture sector as well as for flood prevention. Finally, the public deliverable D2.2 from the S2S4E project (S2S4E, 2018) highlights the user needs from various users in the energy sector.

- P1 L29: “Despite their inherent uncertainties”. I wonder if you could very briefly here cite a few examples of the uncertainties you refer to, for readers less familiar with forecasting on longer timescales?

We will include some examples of these uncertainties, such as hydro-meteorological model errors, future atmospheric states and past hydro-meteorological water storage, in the revised version of the manuscript.

- P2 L34: I think it is important to cite Day 1985 here (Day, G. N., 1985: Extended streamflow forecasting using NWSRFS. J. Water Resour. Plann. Manage., 111, 157–170, doi: https://doi.org/10.1061/(ASCE)0733-9496(1985)111:2(157)).

We will include this reference in the revised version of the manuscript.

- P3 L63: “The Swedish Meteorological and Hydrological Institute (SMHI) has long been operationally providing streamflow forecasts and hydrological warnings to relevant actors in hydrological risk management (municipalities, county boards, Swedish Civil Contingencies Agency), as well as to the general public.” Please clarify that this is for Sweden.

We will clarify this in the revised version of the manuscript.

- P3 L69: “ESP seasonal forecasts are produced but not generally spread to other actors due to uncertainties in their skill and interpretation by external parties.” This is an interesting comment and I wonder what system actors currently use for prediction on such timescales in Sweden? Please consider mentioning this in the introduction to provide some further context.

Both actors and the general public have access to the current hydrological situation and streamflow climatology through the open access Vattenwebb portal (available at https://vattenwebb.smhi.se), which they can use to get information on the latest observed streamflow values as well as to get an estimate of the most likely discharge for any given season based on historical discharges. 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 systems. We will include this information in the revised manuscript.

- P3 L72: “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.” This appears a bit out of context here, please consider moving to the Methods section instead.

The original thought was to present this as background information to the clustering analysis, hence its placement in the introduction. However, we agree with the reviewer in that it would fit better in the methods section. We will move it there in the revised version of the manuscript.

Section 2:

- P3 L86: When you say “measured values from all available stations” do you mean station observations? Please clarify here. Same for discharge and water level data. Please clarify that these are observations.

The reviewer is correct, we refer to observations here. We will revise the manuscript to ensure that the correct term is used throughout the text.

- P3 L91: Is HYPE distributed, lumped or semi-distributed? And how were the meteorological inputs prepared (e.g. interpolated) for the model to ingest?

We note here the HYPE refers to the model, while S-HYPE refers to the Swedish implementation of the HYPE model. In previous investigations, the HYPE model has been used in lumped, semi-distributed and distributed modes. That being said, the S-HYPE model setup is semi-distributed, so gridded meteorological inputs need to be averaged for each model catchment in order to be used. In this case, we follow the same methodology as in the operational service. This way, the meteorological inputs are processed using a weighted average method based on the area fraction of a given S-HYPE catchment covered by each cell of the gridded dataset (only cells which partially or totally overlap the area of the given catchment are assigned weights). We will clarify this in the revised version of the manuscript.

- P3 L93: It is unclear to me at this stage how an “analysis of model outputs” was performed for 39,493 catchments if you only have 539 observation stations? Please clarify here.

The reference used in the evaluation is based on a combination of observations and perfect forecasts, and can therefore cover all 39,493 catchments. This hybrid reference was chosen because it corresponds to SMHI’s operational setup. This is actually a common setup of operational services, which includes assimilation of available observations in order to improve the representation of local initial conditions. Therefore, the analysis is performed on all 39,493 catchments, most of them being analysed against perfect forecasts. For those catchments associated with one of the 539 observation stations, the model outputs are instead assessed against those observations (the model outputs themselves are corrected with existing observations before initialising the forecasts and AR-updating is used when observations are no longer available). Catchments located downstream observation stations partially benefit from the model corrections made upstream, and the reference thus becomes a mix of observed discharges flowing into downstream modelled catchments. We will clarify this in the revised version of the manuscript.

- P4 L96: Please provide the lowest and highest score possible for the KGE for readers not familiar with this performance metric. Out of curiosity, has a S-HYPE model evaluation been published that you could refer readers to?

We will add the KGE ranges in the revised manuscript, as suggested by the reviewer. In this investigation we used the S-HYPE 2016 version which is the latest operational model version. Since the S-HYPE model is subject to continuous efforts, the performance of the current version of the model performance has not yet been published; hence we are here firstly reporting the evaluation results.

- P4 L100: I suggest putting figures 1a-c in the same order as they are mentioned in the text. I was slightly confused and thought I had missed explanations about 1b, which in fact come after 1c.

We will modify the order of the subplots in Fig 1. following the suggestion from the reviewer.

- P4 L108: “Nevertheless, since dam operation is continuously adapted (within certain bounds) to the present and most probable future meteorological and hydrological conditions, these general regulation regimes are expected to be of little benefit for seasonal forecasting purposes.” This is a big statement which warrants further investigation (not necessarily in this paper though!).

What we tried to convey here is that, since dam operation needs to be continuously adjusted to the changing hydro-meteorological conditions, in addition to consider other factors such as optimising the economic benefit and ensuring safe operation, long-range hydrological forecasts based on models with only a limited description of such complex decisions on regulation patterns will most likely be conditioned by these simplifications. We agree with the reviewer that further investigation would be needed to justify a clear statement on this. We will clarify this in the revised version of the manuscript to avoid any misunderstandings on this matter.

- P5 L111: It may be worth explaining further how the ESP hindcasts are produced – i.e. how initial hydrological states are produced to initialize the model for each forecast start date, each meteorological forcing year corresponds to a streamflow hindcast ensemble member, etc. Perhaps a schematic would help make this clear to readers not familiar with the ESP. I also wonder what the lead time of these hindcasts is?

We will include a schematic of how ESP hindcasts and benchmark forecasts are produced in the revised manuscript. Regarding the lead time of the hindcasts, we used 190 days (~6 months). We will specify this in the revised manuscript.

- P5 L129: “as a station-corrected simulation approach was used to achieve the best possible initial conditions.” I am not sure to understand how a station-corrected simulation approach was used for catchments without station observations? Please clarify.

This was, of course, only possible for catchments where observations were available. Nevertheless, even catchments downstream from observations were partially benefited from this station-correction approach. Elsewhere, model outputs were simply simulation results. We will clarify this in the revised manuscript. Additionally, we will also include this step (i.e. station correction) in the new schematic showing ensemble generation (see previous comment).

- P5 L130: Do you know if users in Sweden indeed use “ensemble forecast based on historical streamflow”?

As explained in an earlier comment, this information, together with the latest observations, is openly available through SMHI’s Vattenwebb portal (available at https://vattenwebb.smhi.se). The general public and other actors are encouraged to use this information to (i) get an estimation of expected discharge at any given season and, (ii) to see whether the latest observations are lower or higher than normal. That being said, it is difficult to quantify the actual use of ensemble forecasts in Sweden. From general discussions, we can say that advanced users from the hydropower sector are used to ensemble forecasts based on historical streamflow (some of them also use forecasts based on ESP and NWP techniques), while other sectors may be more familiar with deterministic information. Other tailored SMHI services using ensemble forecasts based on historical records such as the Aqua service (https://europa.eu/!bB63kr) are set up for the water supply authorities.

- P6 L151: Could you please provide some more information about the k-means clustering method, or refer the readers to publicly available material further explaining this method?

In the revised manuscript we will refer to Jin & Han (2011), which nicely summarizes the concept of k-means clustering.

Section 3.1:

- P7 L156: Please introduce Figure 2 prior to commenting on the results. What do the plots show and what is the highest/lowest score possible for the CRPSS? Same for subsequent figures.

Here we think that the figures are adequately introduced in the captions and that, therefore, including an additional introduction in the main text would lead only to redundant information in the manuscript. Nevertheless, in the revised version of the manuscript we will make an additional effort to ensure that the necessary information for understanding all figures (e.g. highest/lowest possible CRPSS) is available to the reader in an intuitive way.

- P7 L156: By lead time, do you mean the aggregation periods mentioned on P5 L142? Or are the results in Figure 2 from daily outputs, and up to what lead time? Please clarify here and in the Figure caption.

We produced daily forecasts up to 190 days into the future and then calculated weekly averages. So, on Figure 2a, “0 Mn” refers to the first forecast week, “1 Mn” to the fifth forecast week, and so on. The aggregation periods mentioned on P5 L142 refer to Section 3.2. In the revised manuscript we will clarify this both in the text and in Figure 2. Note that we are using “lead time” and “forecast time” definitions from the Copernicus Climate Change Service, i.e. for weekly aggregations, lead week 0 is the same as forecast week 1.

- P7 L162: I am not sure to understand what you mean by “the common monthly initialisation frequency of climate prediction systems”. Could you please further explain or reword?

By this we meant that, even though we now see more frequent forecast initialisations in some systems, many seasonal climate forecasts are initialised and produced once a month. We will rephrase this in the revised version of the manuscript.

- P7 L163: “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”. This is a very interesting finding and I wonder if you could comment in the Discussion on how it could be translated into operational decision-making? E.g. Would decision-makers be willing to alter their decisions regularly with each forecast initialization/update?

This is a good point which we plan to address in the Discussion section of the revised manuscript.

Here we state that the way a seasonal forecasting service is used in decision-making depends on the sector, user, and service properties. It is therefore important to evaluate a comprehensive range of possibilities in terms of seasonal information statistics (e.g. forecast aggregation, time horizons) that can technically be offered to individual decision-makers to allow flexibility in the decision process. It is also important to point out that here we can only hypothesize on the impacts of our findings on decision contexts, which are very much sector and location-dependent.

Our findings show that a frequent (i.e. weekly) initialisation can significantly improve the streamflow forecast skill, and this is expected to add value to decision-making. This is of particular high importance for periods in which decisions are subjected 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 optimize the power production for the coming months.

- P8 L184: I am not sure where these lakes are in Sweden. Perhaps it would be helpful to add a map of Sweden with a few key geographical indicators (e.g. elevation, lakes – with legends for the lakes you refer to –).

In the revised version of the manuscript we will include an additional figure in the appending in which we will show the elevation, and hydrography of Sweden. Additionally, we will locate the main Swedish rivers and lakes that are named throughout the manuscript in this new figure as well.

- P8 L188: While I can see lower skill for the regulated rivers, it is hard to identify which rivers you refer to on L191-192. Another plot, such as a zoomed in plot, might be necessary to show these results more clearly.

As mentioned in the previous comment, we plan to include a figure with the main Swedish rivers and lakes. This would allow a clearer identification of the river systems we refer to.

- P8 L191: "future trends in streamflow". This sounds like you are looking at events (e.g. high/low flows). It is perhaps better to rephrase to "future streamflow".

We will reformulate this following the reviewer's suggestion, as it may indeed be clearer for readers.

- It is clever to aggregate forecasts for different periods (Figure 3). This enables to retain some skill for longer lead times than otherwise possible when looking at Figure 2. I wonder if users are interested in such time aggregations, or if they would prefer weekly/monthly aggregations instead? Could you perhaps comment on that in the Discussion, as this is important for the user-oriented analysis you are trying to achieve.

As mentioned earlier, the temporal aggregations depend on the sector and user. For instance, for the energy sector, the 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 (Figure 4) for different aggregations, initializations, and lead times allows communication of skill to various users depending on their needs. We will include these considerations in the revised version of the manuscript.

Section 3.2:

- Figure 4:

- Before looking at this figure, it wasn't clear to me that the analysis was performed for different aggregation periods as well as lead times. Could you please clarify this in the Methods section?

We explained briefly the analysis using different aggregation periods in P5 L139-143. In the revised manuscript we will reformulate this so it is clearer for the reader that we also perform this type of analysis.

- Could you please add ticks (and perhaps tick labels where possible) to all subplots of this figure as it is difficult to follow the results clearly without.

As suggested by the reviewer, we will add ticks for all subplots in this figure in the revised manuscript. Additionally, we will add labels to the y-axes of the subplots for January, April, July, and October, and to the x-axes of the subplots for October, November, and December.

- Do you have an explanation for the sudden increase in skill for hindcasts initialized on 1 March, with a 8- vs 12-week aggregation period? Is it because you are predicting streamflow for the summer with the 12-week aggregation period, which is "easier" to predict as levels are generally low during this season? Please consider reflecting on this briefly in the paper.

This increase in skill, which is particularly obvious in March, can in fact be observed for hindcasts initialised between 1 February to 1 May when looking at the 4-week aggregation period, and corresponds roughly to the month of May. Many catchments and rivers, especially in the northern half of the country, see the peak of the spring flood during this month. With shorter aggregation periods, the focus is more influenced on the start/end of the event, while longer aggregations put more emphasis on having a correct total volume, regardless of the exact start/end dates. Since this total volume linked to the accumulated snowpack is easier to model than the timing of the event, which is conditioned by meteorological variables, longer aggregations perform better. In the southern parts of the country, in the month of May the spring flood has already passed and low flow conditions start to dominate. We will include these considerations in the revised version of the manuscript.

- P10 L198: Could you please remind us here which aggregation periods were used for this analysis?

We will follow the reviewer's suggestion and add the aggregation periods we used in the analysis here.

- P10 L203: "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." It would be interesting if you could add an indication of the lead time at which the skill is 0 for shorter aggregation periods (results from Figure 2) on this figure.

The bottom row of each subplot in Figure 4 contains already the same information as Figure 2b, as the aggregation period (i.e. 1 week) is exactly the same. So, the first grey box in the bottom line of each subplot already shows this information. So, after discussion with the co-authors, we decided to avoid making this figure heavier than it is, as the objective here is to depict how the skill changes as a function of the aggregation window, and not only when this drops below 0.

Section 3.3:

- I would argue that results for longer forecast horizons would be good to show as well as the focus of this paper is on seasonal forecasting. Perhaps correlations could be stronger when calculated against another performance metric which might not weaken so much over time (e.g. CRPS instead of its skill score)?

The results presented in Section 3.3 correspond to an exploratory investigation connecting the first part of the analysis (i.e. temporal and spatial variability of ESP forecast skill) with the second part (i.e. attribution of skill to hydrological behaviour). By focusing on the CRPSS, we look at the "added value" of the ESP with respect to streamflow climatology, which is in line with the idea of evaluating/understanding the use of ESP for decision making (against an alternative system). Looking at the CRPS or any score without a benchmark would be a different analysis completely which would undoubtedly be very interesting but which is outside the scope of this study. That being said, in the revised version of the manuscript we will address this comment by adding the results for a further lead time in light grey in the same figure.

- To what extent do you think these results are dependent on your hydrological model? Please consider commenting on this in the Discussion section.

Different aspects of the S-HYPE modelling and forecasting chain in this study, such as the model setup and data, the model structure, and its parameters may convey uncertainty to the forecast results (see also the discussion in Pechlivanidis et al., 2020). However, the impact of model errors for our particular setup is especially complex as we used a combination of observations and perfect forecasts as reference. While we can expect model errors to be minimal for those catchments in which forecasts are purely evaluated against perfect forecasts, they become relevant for catchments at 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 have all observations at dams or other river regulation structures. The correction of model outputs with observations and, when these are no longer available (e.g. at forecast initialization), with an exponentially decreasing factor based on the last known model error (i.e. AR correction) effectively minimises model uncertainties, especially at forecast initialisation and during the first time steps of the forecast. Nevertheless, any model errors will tend to become more significant for further lead times. The downstream distance of a given catchment with respect to an observation is also relevant in this case, as the model correction would only affect a fraction of the simulated/forecasted streamflow at that location.

The most important model errors, though, can be expected for heavy regulated catchments with or downstream of observations. Complex river regulation routines which depend on factors external to hydrological models cannot be adequately reproduced by these models. In these cases, even if the correction of model outputs with observations may minimise model errors at forecast initialisation, these errors will rapidly spread due to the inability of the model to reproduce the modified hydrological regime.

We will include these considerations in the revised version of the manuscript.

- Could you please increase the font size of the correlation coefficient on each subplot of Figure 5? It took me a bit of time to notice them.

Following the reviewer's advice, we will increase the font size of the text of Figure 5 to make it more readable.

Section 3.4:

- Table 2: It would be good to show the range of elevation, annual precipitation, etc. instead of just the mean values, to show the catchments variability within each cluster region. This might become a bit messy and could be clearer in a figure rather than a table.

In the revised version of the manuscript we will include the interquartile ranges (Q25 - Q75) in addition to the mean values for each of the variables. Following a comment by another reviewer, we will remove potential evapotranspiration from the table, which will give more space for the additional information.

- P14 L241-254: It may be easier to follow by having these observations as bullet points in Table 2. It might also make it easier to link the results presented in Figure 7 with the cluster characteristics.

The text in L241-254 refers to the dominant hydrological processes and topographic characteristics, while Table 2 summarizes the streamflow signatures which define the clusters. We propose not to add similar information in Table 2 and hence introduce redundancy in the manuscript. We will nevertheless make an effort to make this paragraph easier to follow by the reader in the revised version of the manuscript.

- Could the large/small spread in forecast skill shown in Figure 7 be caused by large/small basin differences within these clusters? E.g. spread in the topographic, climatological or hydrological characteristics (from Table 2) within each cluster. It would be interesting to hear your thoughts on this here on in the Discussion. For example, cluster 5 catchments appear more spread out throughout Sweden (Figure 6b) compared to cluster 6 catchments.

The hydrological characteristics are the end-product of climatological and physiographic properties and can therefore not be assessed together. Some combinations of climatological and physiographic properties can be found in very specific areas of the country, while others are more widespread. For instance, from a physiographic perspective, cluster 6 consists mainly of agricultural and coastal catchments, in addition to big lakes, which are quite limited geographically in Sweden. Conversely, cluster 5 contains mostly slowly-responding forested catchments, which can be found throughout the country.

Focusing on the hydrological characteristics, results from cluster 5 are indeed interesting. The forecasts in the catchments clustered here generally show the highest skill (for all lead times) among all cluster groups, yet results are widely spread. In this paper we conclude that forecast skill is strongly linked to the various hydrological regimes (see also a more detailed investigation in Pechlivanidis et al. 2020), and hence we argue that the reason for this spread lies in a deeper understanding of the hydrological signatures in cluster 5. As we state in P14 L241-242, the catchments in cluster 5 are characterized by a high baseflow contribution (BFI), a slow response to precipitation (Flash) and a generally small intra-annual variability (DPar). In Figure 6a we observe that, although the mean values for RLD (rising limb density) are below the 33rd percentile of this signature (which represent 'below normal' signature values), the variability among the 4355 catchments composing this cluster is high (as indicated by the boxplot), with some catchments experiencing 'normal' RLD values and yet some others with values even higher than the 66th percentile of this signature. Consequently, this indicates that, despite their high baseflow contribution, some catchments in cluster 5 experience sharp increases in their hydrographs, which is an indication of low skill as seen in Figure 5 (CRPSS and RLD are strongly, but negatively, correlated). We will explain the above argument for the large spread in cluster 5 in the revised manuscript.

Section 4:

- P18 L306: "forecast initialisations are not expected to provide an added value to the forecast service." I would argue the opposite. You have shown in your paper that more frequent forecast initializations could substantially increase the forecast skill. The added value is potentially immense for decision-makers. The challenge remains to translate this into actionable outputs for the users, as you mention it briefly. Please consider rephrasing and elaborating on this.

Frequent initialization as seen in this manuscript (i.e. weekly with respect to monthly), does provide added skill. However, we argue that daily initialization (when compared to weekly initialization) is unlikely to convey any further useful information for decision making at seasonal horizons, since long-term decisions are also not taken daily. In such services, due to high uncertainty, results are aggregated into weekly values, which further smooth the potentially high streamflow dynamics. We will clarify this in the revised version of the manuscript.

- P19 L332: Would you be able to add a figure to the paper to support these very interesting findings?

Here we want to clarify that the sentence in P19 L332 does not correspond to actual findings presented in this manuscript, which build on an analysis of the operational forecasting setup from the perspective of public service, thus focusing on catchment outflows. Instead, this statement is based on the assumption that, since forecast skill is shown to be consistently lower in highly regulated catchments than elsewhere, the fraction of the inflows to a given reservoir that are not affected by other regulation upstream may be more predictable and therefore convey higher forecast skill when compared to the outflows, which would be very relevant for the hydropower sector. This is indeed a very interesting analysis that we plan to investigate further in the future. We will clarify this in the revised version of the manuscript.

- P19 L344: “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.” To evaluate this, a different performance metric, such as the brier or ROC score for high flow events, might be better adapted than the CRPS. Do you plan to look at this in the future?

The severity of ice break-ups is determined by the interplay of different factors and processes over a long period, usually starting in late autumn. The main drivers are meteorological (defining the ice build-up during the winter months and meltdown during spring) and hydrological (regarding the timing of the streamflow increase marking the start of the spring flood). So, here we argue that scores which evaluate the overall performance, including biases in volume, such as the CRPS are also suitable for decision-making on the allocation of resources. That being said, we do plan to explore this further into the future, including looking at the metrics suggested by the reviewer.

References

Foster, K., C. B. Uvo, and J. Olsson (2018), 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, doi:10.5194/hess-22-2953-2018.

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., https://doi.org/10.5194/hess-2019-659, accepted, 2020.

Jin X., Han J. (2011) K-Means Clustering. In: Sammut C., Webb G.I. (eds) Encyclopedia of Machine Learning. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-30164-8_425

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

S2S4E (2018) Deliverable 2.2 User needs and decision-making processes that can benefit from S2S forecasts.