RESEARCH ARTICLE

A seamless filter for daily to seasonal forecasts, with applications to Iran and Brazil

Gerd Bürger

Institute of Earth and Environmental
Science, University of Potsdam, Potsdam,
Germany

Correspondence
Gerd Bürger, Institute of Earth and
Environmental Science, University of
Potsdam, Potsdam, Germany.
Email: gbuerger@uni-potsdam.de

Funding information
Bundesministerium für Forschung und
Technologie

Abstract

A digital filter is introduced which treats the problem of predictability versus time averaging in a continuous, seamless manner. This seamless filter (SF) is characterized by a unique smoothing rule that determines the strength of smoothing in dependence on lead time. The rule needs to be specified beforehand, either by expert knowledge or by user demand. As a result, skill curves are obtained that allow a predictability assessment across a whole range of time-scales, from daily to seasonal, in a uniform manner. The SF is applied to downscaled SEAS5 ensemble forecasts for two focus regions in or near the tropical belt, the river basins of the Karun in Iran and the São Francisco in Brazil. Both are characterized by strong seasonality and semi-aridity, so that predictability across various time-scales is in high demand. Among other things, it is found that from the start of the water year (autumn), areal precipitation is predictable with good skill for the Karun basin two and a half months ahead; for the São Francisco it is only one month, longer-term prediction skill is just above the critical level.

KEYWORDS

climate drift, ensemble prediction, seamless prediction, seasonal forecast skill

1 | INTRODUCTION

Since about a decade, the field of seamless weather prediction has experienced quite active development (Palmer *et al.*, 2008; Bauer *et al.*, 2015; Brunet *et al.*, 2015; Vitart and Robertson, 2018). This was due to both a deeper understanding of the underlying nonlinear dynamics of the system, as well as greater computational capacity to run the corresponding numerical weather prediction (NWP) models. Originally used for the interface between weather and climate (Palmer *et al.*, 2008; Shukla *et al.*, 2009), the word "seamless" now refers to efforts of unifying prediction systems (initialization, parametrization, numerics) across all time-scales. This study emphasizes predictions ranging from days to seasons, specifically from lead times of 1 to 180 days, as produced

by the latest seasonal prediction system SEAS5 of the European Centre of Medium-range Weather Forecasts (ECMWF) (Johnson *et al.*, 2019). Potential atmospheric predictability at seasonal time-scales is grounded in, mostly coupled, atmosphere–ocean processes, such as the El-Niño–Southern Oscillation or the Madden–Julian Oscillation, that are themselves predictable at those time-scales (Rowell, 1998). Land processes such as soil moisture, snow cover and vegetation also play a role (Robertson and Vitart, 2019).

Depending on variable and region, to each prediction lead time there tends to be associated a characteristic time-scale whose variations are optimally predicted. As shown in fig. 1d of Reichler and Roads (2003) or fig. 2 of Buizza *et al.* (2015) for the 30 days lead time, for example, the 60 days average is better predicted than the 30 days average, as the former

obviously better eliminates unpredictable features that propagate with periods less than those 60 days. A systematic study of the dependence of forecast skill on time-averaging and lead-time was undertaken by Zhu *et al.* (2014); they employ for a lead time of λ a time averaging window of width λ. A similar approach was taken by Ford *et al.* (2018); their averaging window is given by the Poisson distribution, with the single parameter λ as lead time. Since mean and variance of the distribution are equal to λ, their approach resembles that of Zhu *et al.* (2014). Finding the optimum aggregation scale is a non-trivial problem, since very often the underlying process time-scales are uncertain, and the problem comes down to a trial-and-error exercise. But that is not the subject here. Instead, similar to the Zhu *et al.* (2014) and Ford *et al.* (2018) approaches (cf. also Wheeler *et al.* (2017)), but more flexible, the scales are *prescribed* in terms of a *seamless filter* (SF), as detailed in section 2.2. Basically, the SF would be a standard low-pass filter that retains frequencies only below some cut-off value, but its characteristics change with growing lead time, in a way so as to filter, in the standard setting, for increasingly lower frequencies with growing lead time. But in contrast to the Zhu *et al.* (2014) and Ford *et al.* (2018) approach, the rate of increase in the SF is fully configurable and governed either by expert knowledge or by user demand.

A more objective method of specifying time-scales exists for larger-scale problems, such as hemispheric or global, by decomposing the spatial and lead-time information into dominant modes of predictability (DelSole and Tippett, 2009a; 2009b). Our focus regions are too small for this to be applied, however.

Early on, it was understood that seasonal predictability is concentrated on the Tropics (Stockdale *et al.*, 2018). But extratropical predictability, mainly through influence from stratospheric variability, has recently come into the focus as well (Kirtman and Pirani, 2008; Sigmond *et al.*, 2013; Sun and Ahn, 2015). Here we study daily to seasonal predictability in two regions in or near the tropical belt, with climates that are partly arid: the São Francisco basin (SFB, [41°W, 11°S], 617,812 km^2) in northeast Brazil and the Karun basin (KB, [49°E, 32°N], 65,230 km^2) in southwest Iran. Weather and climate of the SFB are mainly influenced by the South American monsoon system (Silva and Kousky, 2012); the semi-arid climate of the KB, on the other hand, sees influence from Mediterranean cyclones and the Indian summer monsoon (Alijani, 2002; Yadav, 2016). For both regions the rainy season, during which all the reservoirs are being refilled, lasts from October–November to March–April, for which corresponding predictions are therefore most relevant.

2 | METHODS AND DATA

The downscaling of the seasonal predictions follows a perfect-prognosis approach, which belongs to the standard procedures of climate downscaling: Using a set of observed local station data (section 2.4) and observed (i.e. analysed) atmospheric fields (section 2.5), a statistical model is calibrated (section 2.1). To obtain the local seasonal predictions, the model is applied to the same set of fields, but as predicted by an atmosphere–ocean model (section 2.5) (which is assumed to contain a *perfect* representation of the atmospheric fields). Finally, the seamless filter (section 2.2) is applied to each station record, both as observed and as predicted, and their areal mean values are compared to estimate the prediction skill.

2.1 | Expanded downscaling

It is well known that using large–scale, gridded atmospheric fields directly is not recommended when comparing with station data, but to employ some form of adjustment instead, such as downscaling. Here we use Expanded Downscaling (XDS), which is a regression-based approach that simulates local events as close to and as consistent with the prevailing atmospheric circulation as possible. In deviation from classical regression, the error minimization is done under the constraint to preserve local covariability (of variables and stations). This preservation of covariability renders XDS particularly useful for applications related to hydrological extremes, such as floods and droughts (Bürger, 2002; see also Bronstert *et al.*, 2007). XDS has been thoroughly validated, in numerous climate impact studies as well as for local weather forecasts, and it performs relatively well in comparison to other methods (Bürger *et al.*, 2012). In those studies, for local station data such as temperature and precipitation, a mix of atmospheric upper-level and surface predictor (regressor) fields has been proven most appropriate, as detailed in section 2.5. With a focus on extremes, XDS is most appropriate for the daily time-scale. In this study the daily local and atmospheric series described in sections 2.4 and 2.5 are used as predictands and predictors, respectively.

2.2 | The seamless filter

It is common folklore that daily weather becomes fundamentally unpredictable beyond about 2 weeks lead time, according to the famous result on nonlinear error growth known as the butterfly effect (Lorenz, 1972). But as mentioned in section 1, predictions beyond this limit are nevertheless possible if adequate aggregations in time (or space) are formed. For seamless predictions, therefore, one would like to have this done automatically, that is, increasingly aggregating data with growing lead time. Zhu *et al.* (2014) achieve this by categorizing lead times and aggregation levels into a few classes, such as "1d1d" and "2w2w", referring to lead- and aggregation-times of 1 day and 2 weeks, respectively. Instead of using categories, we have designed a digital filter that does the same
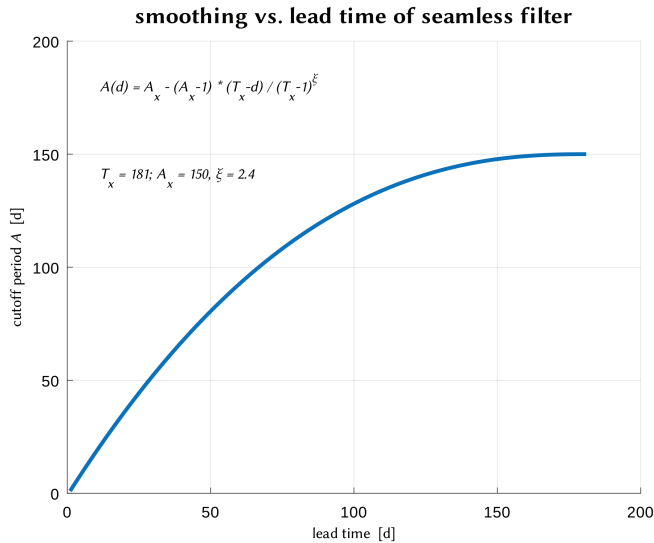
FIGURE 1 The seamless filter characteristic. It describes the dependence of the smoothing (the cut-off period A (days) of a corresponding low-pass filter) on the forecast lead time (see formula)

continuously for all lead times. This *seamless filter (SF)* is now described in more detail.

Noting that any aggregation level can be realized by a smoothing (low-pass) filter with characteristic cut-off period, we define a mapping, $A: [0,T_x] \rightarrow [0,A_x]$, that assigns to any forecast lead time d, $0 \leq d \leq T_x$, a suitable cut-off period $A(d)$. The mapping parameters are the two maxima of lead time T_x and cut-off period A_x, plus one shape parameter ξ that governs the growth of A with d:

$$A(d) = A_x - \frac{T_x - d}{(T_x - 1)^{\xi}}(A_x - 1). \quad (1)$$

It is designed so that $A(T_x) = A_x$ and, for $\xi = 1$, $A(1) = 1$. Similar realizations of the mapping are certainly possible; they only should roughly follow the dependence as depicted in Figure 1. In general, the SF design requires a balance between forecast skill and forecast detail (amplitude), which would translate to balancing expert knowledge and user needs. For example, the filter with ($T_x = 180$ days, $A_x = 180$ days, $\xi = 1$) is essentially the setting used by Zhu *et al.* (2014) and Ford *et al.* (2018). In this study, "standard" parameters of $T_x = 180$ days (6 months), $A_x = 150$ days, and $\xi = 2.4$ are used; what this means concretely will be shown in an example further below.

From this basic aggregation law A, the SF is constructed as follows: suppose a sequence of forecasts, $x(d)$, is given with growing lead time d, $1 \leq d \leq T_x$. For each d we form the filtered series $\tilde{x}_d(t)$, $t = 1, \ldots, T_x$, which is obtained from $x(d)$ by applying a conventional low-pass filter with cut-off period $A(d)$. We thus obtain T_x filtered copies of the original series $x(d)$ or, by taking those as rows of a quadratic matrix, a
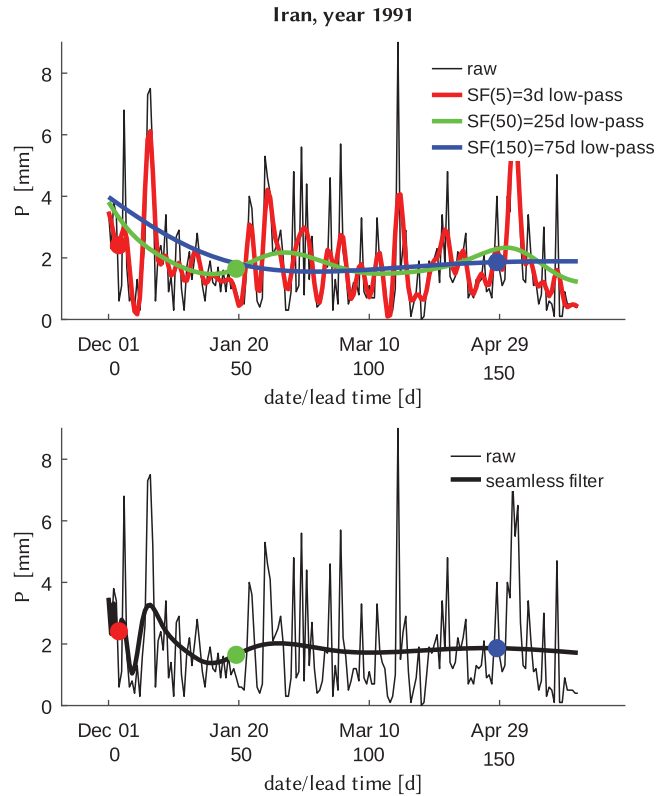

FIGURE 2 Example of the SF as applied to a 180-day forecast of areal P for Iran, issued on 1 December. Top: Raw and low-pass filtered series with three different cut-off periods, using a zero-phase (non-causal) Butterworth filter. Bottom: The SF (heavy black) of the series. The dots illustrate how SF is constructed from the filtered series above. The SF parameters are: $T_x = 180$, $A_x = 90$, $\xi = 1$

corresponding $T_x \times T_x$ matrix. The diagonal of that matrix,

$$x^{SF}(d) = \widetilde{x_d}(d) \quad (2)$$

defines the seamless filter SF. Note that by operating on the whole series, Equation 2 cannot be realized as a linear digital filter (like most smoothing filters themselves, e.g. Butterworth), it is not even finite, nor is it time invariant; its computational complexity is nevertheless feasible. The desired result of applying SF to a sequence of forecasts is that with increasing lead time, the forecasts are increasingly smoothed up to the longest forecast with maximum smoothing at cut-off period A_x. The single low-pass filters are realized by using a Butterworth filter of order 3, run in forward–backward mode (Gustafsson, 1996) to avoid any time shifts (phase lags). The content of Equation 2 is illustrated in the example of Figure 2.

As promised, I give an example of a forecast issued on 1 November and valid for 1 December, i.e. with a lead time of 30 days. For that lead, the above SF characteristic ($T_x = 180$ days, $A_x = 150$ days, $\xi = 2.4$) corresponds to a cut-off period of 52 days. This means that the user or expert is interested, for that specific lead, in the rainfall amount
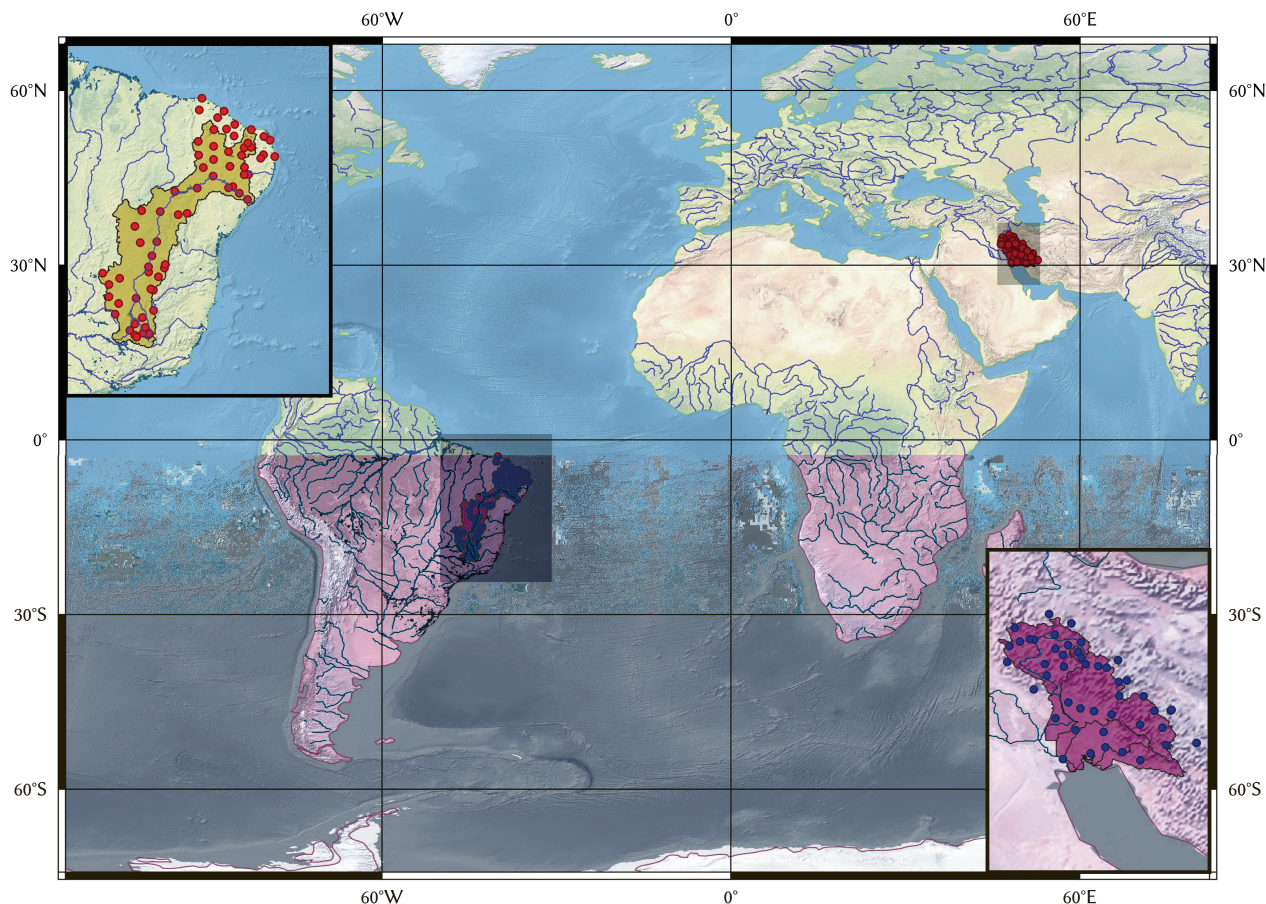
FIGURE 3 The study regions of the Karun, Iran, and the São Francisco basin, Brazil, along with their respective precipitation stations

of a 52-day period centred about 1 December (i.e. between 4 November and 26 December). This amount is obtained from the 180 days predictions of 1 November as provided by SEAS5/XDS, by applying the SF to each predicted station. The expected skill of the areal average forecast can be read from the figures below (this being the subject of the current study).

The display of SF-based, or *seamless*, predictability depending on lead time, as presented here, may at first look somewhat surprising, and may violate one's expectations of prediction skill decreasing monotonically with lead time. By looking at a range of time-scales, one is faced with the complicated superposition of different physical processes that have their own predictability. For example, sea-surface temperature in the tropical Pacific has a fairly high predictability on seasonal time-scales due to the El Niño–Southern Oscillation, likely higher than corresponding variability on the daily or weekly scale. Seamless predictability, hence, does not automatically decrease with lead time; it may indeed increase, as our examples show.

One may object that smoothing will inflate the reported skill by using future information that is not available at issue time. This argument does not apply. At issue time, the (SEAS5) forecasts themselves only use information that is
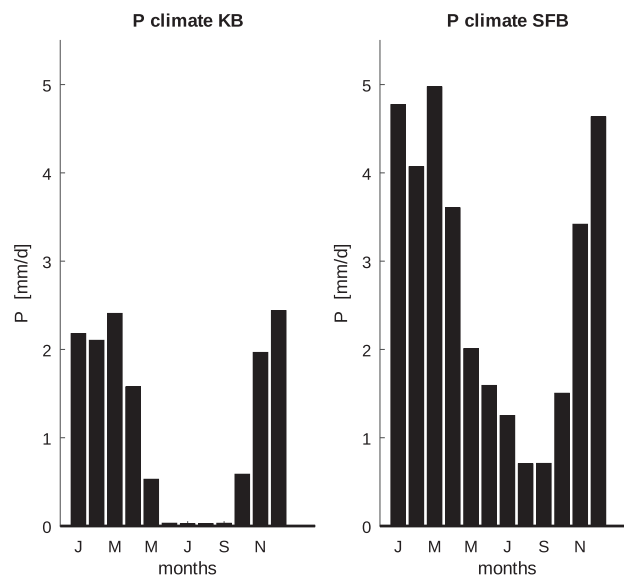

FIGURE 4 P climatology (mm·day^{-1}) for the focus regions

available then (this applies strictly only to the operational mode; in reforecast mode, future information is available but not used). The same applies to the downscaling. For the skill evaluation one can use all information from the forecasts, including data from beyond the valid time (e.g. for filtering). While such data are available for the (SEAS5/XDS) predictor
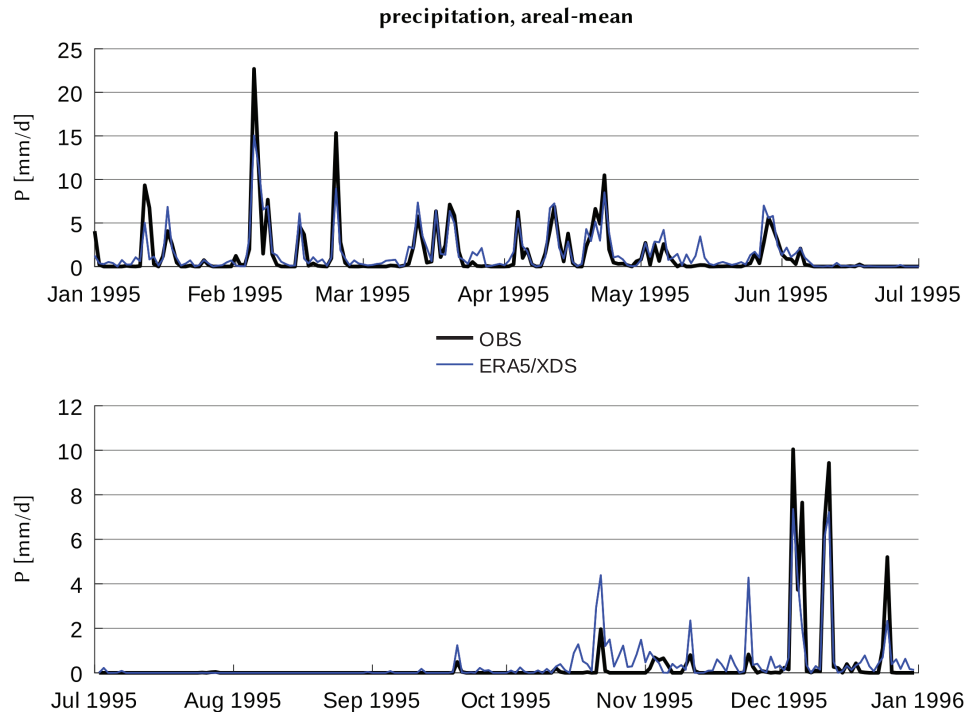
FIGURE 5 Observed and ERA5/XDS-downscaled areal P for the KB, for the year 1995

at issue time, for the predictand they become available only later, which is when the smoothing filter can be applied (i.e. in hindsight).

Note that our notion of "lead" is centred about the point of interest, extending both into the past and into the future (the reason we used a zero-phase filter). This can be circumvented by formulating a filter that is causal in a reversed sense, so that only future information (relative to valid time) is used, meaning: for the valid time 1 June and a 3-month average, use only the months June, July and August (which would be 3m3m in the terminology of Zhu *et al.* (2014) for forecasts issued on 1 March). Likewise, by using a causal filter, a "1-month lead" involves only past information (relative to valid time), like forecasting average May rainfall on 1 May, valid 1 June. As the examples show, the question of lead time is basically a semantic one.

2.3 | Potential predictability

As a validation for the downscaled ensemble spread, the approach of Buizza (1997) to estimate potential predictability (PP) may serve. It rests on the assumption that at any given lead time, the ensemble spread is a perfect representation of the prevailing uncertainty. In other words, the actual observations are indistinguishable from any of the ensemble members. This allows the assessment of PP in the ideal case of a perfect model with perfect initial uncertainty. Conversely, discrepancies between potential and actual predictive skill give a measure of the deviations from those ideal conditions. It is worth noting that while the meteorological ensemble

is designed to represent the true atmospheric uncertainty, including initial conditions and physical parametrizations, the local ensemble, as obtained from coupling a downscaling scheme to the atmospheric fields, may well diverge from the true local uncertainty (whose structure is anyway different from the grid-point errors), for example, if the downscaling scheme is mis-configured.

To estimate PP, a "perfect ensemble" system is formed from the original ensemble by selecting one member at random as the verifying analysis, cf. Buizza (1997); this provides a representative distribution of the potential skill of the prediction.

2.4 | Focus regions

We use two focus regions, the Karun basin in southwest Iran and the São Francisco basin in Brazil, as shown in Figure 3 along with the weather stations used in this study. As evident from Figure 4, the rainy season lies between October–November and March–April for both regions, with a very pronounced seasonal cycle, making their climate semi-arid. During the wet period reservoirs are re-filled, and that is the time for which seasonal predictions are most important. Both basins are near the tropical belt of best seasonal predictability. For the KB, between 1983 and 2015, 43 stations measure daily minimum (T_n) and maximum (T_x) temperature and precipitation (P). For the SFB, between 1981 and 2017, daily mean temperature (T) is also measured, altogether for 65 stations. We include temperature here because

XDS simulates *P–T* covariability, which provides an extra constraint on the local data.

2.5 | ERA5 and SEAS5 from ECMWF

As fully described by Molteni *et al.* (2011), the integrated forecasting system (IFS) of the ECMWF is a fully coupled model that predicts the global atmosphere–ocean system up to 180 days in advance. The new ERA5 reanalysis (Hersbach and Dee, 2016) (also based on IFS) serves to provide the currently best approximation of what can be called atmospheric observations. Uncertainty is modelled using a sophisticated set of perturbations, both of initial conditions and parameters, and simulating an ensemble of 25 members.

During a hindcast or forecast experiment by any NWP model, no observational input is allowed or available, respectively, to serve as a corrective. The model will therefore tend to evolve from its initial state at zero lead time into its own climate, which will more or less differ from observations. This *climate drift* becomes visible already after just a few days, and increasingly so with the forecast horizon approaching seasonal. This is also the case for SEAS5, although the drift is already smaller than in the predecessor version SEAS4 (Stockdale *et al.*, 2018). Obviously there is a need to correct for this drift, which was done as described in Appendix S1, cf. Figures S1–S4.

SEAS5 data are available for the years 1981 to the present (or in our case 2018). As predictors for the downscaling (cf. section 2.1) we use upper-level and surface fields from ERA5 and SEAS5: at pressure levels of 700, 850 and 1,000 hPa we use temperature, specific humidity, and the wind vectors, together with precipitation at the surface (that is, gridded ERA5 and SEAS5 precipitation). For the KB we use the area between (40°E, 20°N) and (60°E, 40°N), and for the SFB the sector between (50°W, 35°S) and (25°W, 5°S), all in 1° resolution. The fields are projected onto the dominant principal components, and from the resulting predictor time series the most important ones, 260 for SFB and 124 for KB, are selected using a least angle regression (Hastie *et al.*, 2005) for the areal mean of all local variables. These serve as input to the XDS downscaling.

3 | RESULTS AND DISCUSSION

Downscaling validation is done for the respective areal mean *P* series. The seamless seasonal predictions are evaluated using the corresponding SF versions. For all predictions (issued 1 January, 1 February, etc.) the skill of each lead time (i.e. a particular day) is accordingly estimated from those years between 1981 and 2018 when observations and predictions are available for that particular day; for the KB and the SFB that gives a sample size of 33 and 37 days, respectively, in all years. We may assume that the 33 and 37

data pairs are independent (ignoring multi-annual random variations), which results in a critical correlation value of $\rho = .40$ and $\rho = .38$ at a significance level of 1% (which we use throughout).

3.1 | Karun basin

3.1.1 | ERA5/XDS validation

Figure 5 shows for the KB and the year 1995 daily areal *P* as observed and as downscaled by ERA5/XDS. The strong annual cycle is well matched, even single events are reproduced with high accuracy; the actual scale of events is sometimes over- and sometimes underestimated, a consequence of the built-in feature of XDS to preserve local covariability, cf. section 2.1 and section S2 in Appendix S1. For the full validation period (1983–2000), this results in a correlation of $\rho = .72$ for daily de-seasonalized and normalized P using the Normal Quantile Transform (NQT) (cf. Appendix S1) *P*. The analogous result for downscaled *T* is shown in Figure S5 in Appendix S1, showing a very high accuracy with a correlation of $\rho = .96$.

For completeness, a comparison is made between the downscaled and the "raw" gridded *P* fields from ERA5. As shown in Figure S8 in Appendix S1, compared to the grid-point-based areal mean *P*, based on the convex hull of the station locations, the downscaled version performs considerably better, as revealed by correlations of $\rho = .53$ vs. $\rho = .85$, underpinning the need for downscaling.

3.1.2 | SEAS5/XDS local predictions

For the KB, the local authorities issue monthly-to-seasonal predictions (or "outlooks") in rolling releases per month, starting with the beginning of the water year in October (in the form of monthly ("O"), seasonal ("OND"), or semi-annual ("ONDJFM") forecasts. This roughly corresponds to the lead times of 15 days (SF = 27 days), 45 days (SF = 74 days), and 90 days (SF = 121 days).

We start by showing in Figure 6 the correlation skill obtained for the 180 days area and ensemble mean forecasts issued on the first of October and November. Unsurprisingly, the SF-processed forecasts are essentially all above the corresponding unfiltered daily scores (those fall below the significance level of 0.4 already after a week or so). The filtered October predictions remain skilful *(ρ > .6)* for almost a season, dropping below significance only in January and staying there. The November predictions start with a somewhat higher skill *(ρ > .8)*, then drop off to insignificant levels in December, but increase again to fairly large skill values *(ρ ∼ .7)*. Moreover, like in all other cases, after about 4 months lead time the skill curves do not change much anymore. This is a simple consequence of the SF characteristic, for which the smoothing effect becomes less distinguished after that
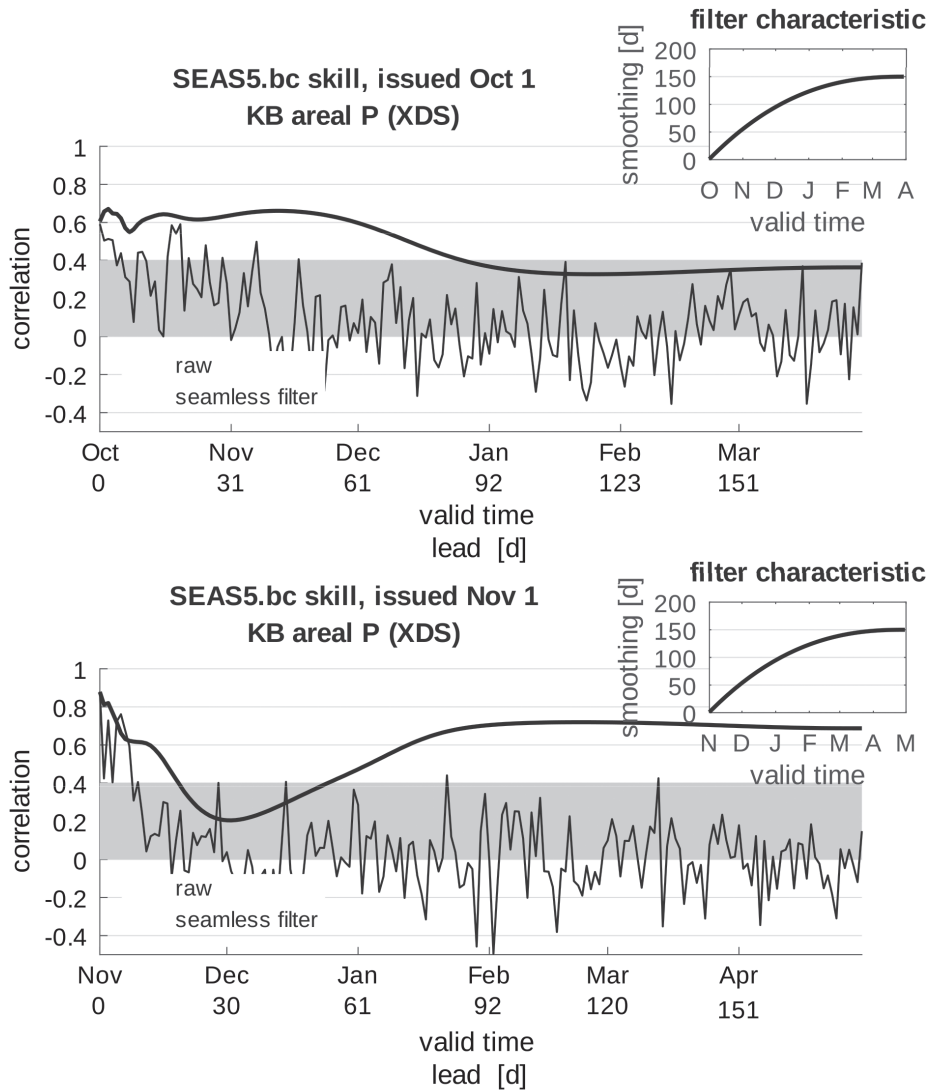
FIGURE 6 Correlation skill for the raw (thin) and seamless (heavy) areal and ensemble mean forecasts issued on 1 October (top) and 1 November (bottom). For each issue and lead time, correlations are estimated from the years 1983–2015, which results in a critical value of 0.4 (see text), displayed by the grey confidence band. The SF characteristic curve is shown as inset

time; to be able to distinguish greater lead times, a flatter SF characteristic must be used.

Why are the two predictions valid for December so different? Looking at the SF characteristic, one sees that the October predictions employ almost a 100-days smoothing for December, whereas that filtering is below 50 days for the November predictions. So apparently one can predict the *season* around 1 December more skilfully from October than (approximately) the *month* of December from November. This is supported by looking at seasonal predictions (with very strong smoothing, cf. Figure S9 in Appendix S1); they show a much higher skill especially for the November forecasts. Note, however, that due to the strong smoothing the predicted amplitude is much less in that case. The higher skill is evident from the time series shown in Figure 7; it shows normalized observations and ensemble-mean predictions valid for 1 December; due to the different smoothing (95 days vs.

54 days), observations slightly differ between the panels. It is obvious that the longer-lead October predictions are much better than those of November. Especially around the years 1985, 2000 and 2010, differences are noticeable, with large discrepancies in the November predictions.

It turns out that the failure to predict December rainfall from November is actually caused by the downscaling. As shown in Figure S10 in Appendix S1, there is no drop in skill for the (raw) gridded SEAS5 rainfall data, even without drift correction. Although this failure seems to be an exception (see Appendix S1), further analysis is needed to understand that discrepancy.

It is of interest whether and how strong the skill depends on the fact that the ensemble mean is considered. To that effect, we re-plot the correlation skill for the October ensemble mean predictions, and compare it to the skill of the single members. For comparison, we do the same for the February
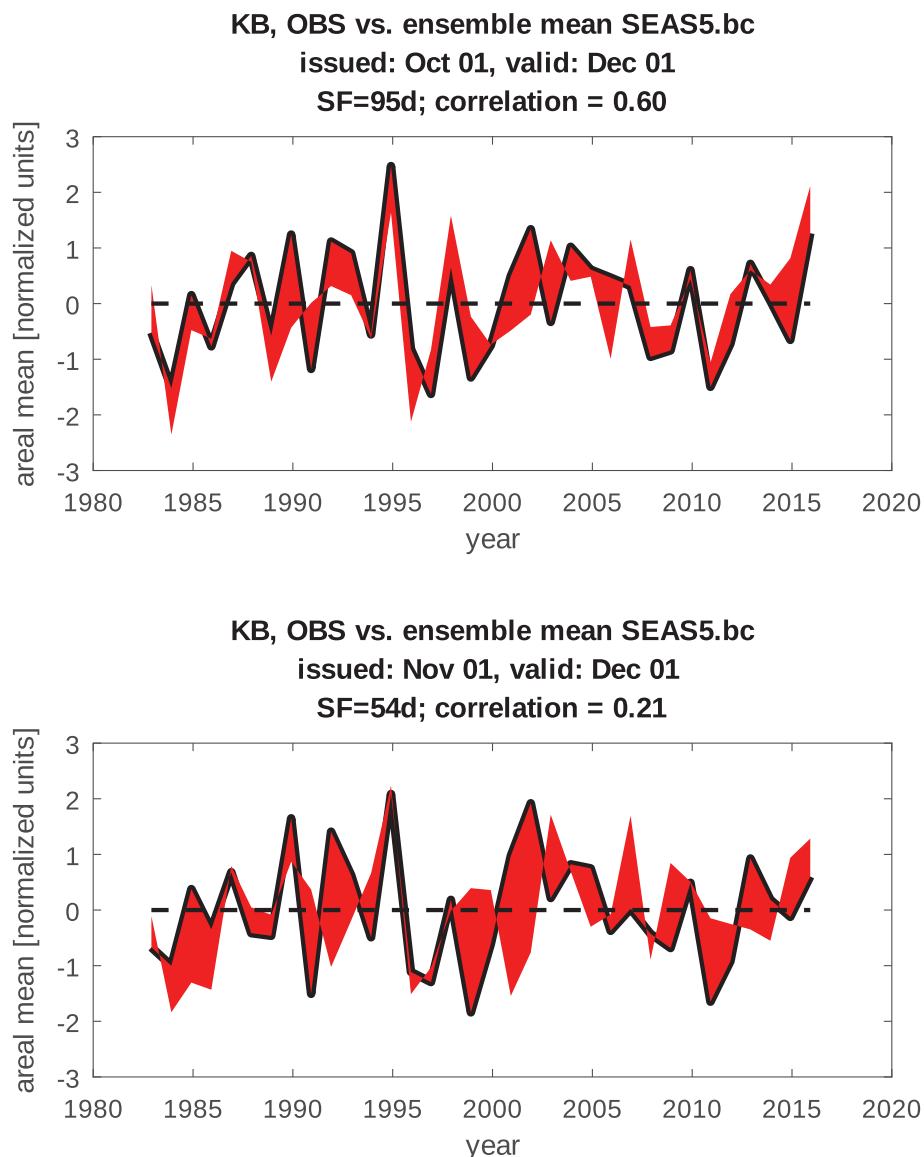
KB, OBS vs. ensemble mean SEAS5.bc
issued: Oct 01, valid: Dec 01
SF=95d; correlation = 0.60

KB, OBS vs. ensemble mean SEAS5.bc
issued: Nov 01, valid: Dec 01
SF=54d; correlation = 0.21

FIGURE 7 Example observations (black line) and deviation to the ensemble mean forecasts (filled area) for the fixed valid time 1 December, for forecasts issued 1 October (top) and 1 November (bottom), with correlations of .60 and .21, respectively. The data are normalized, so that the deviation (area) can be used to compare correlations (inversely)

predictions. In Figure 8 the October predictions show that for short leads of under about a week, the ensemble mean performs like an average member, as is to be expected. After that, however, up until about 2 months, the ensemble mean is better than *any single member* (except for one short period for one member around late October). The single and ensemble mean skills settle to their final value after about 4 months, which for the mean is not significant ($\rho < .4$) and, probably randomly, within the spread of the single skills. With respect to PP, the initial skill is expectedly much larger, this gap indicating the error in initial conditions. The November–December high skill of the mean is strikingly different from the fairly low PP skill, pointing again to a mismatch between real and predicted uncertainty. In a way, the February ensemble mean prediction shows the reverse behaviour: starting with very good initial

skill ($\rho > .8$), the skill drops early below significance (also for single members) but regains significant values after mid-April and ultimately converges to high values of about 0.6. Unlike the October case, the PP skill is highest at all lead times, which indicates an overall higher consistency between real and predicted uncertainty (i.e. prediction error and ensemble spread). Most interestingly, for long leads the ensemble mean again outperforms all the members, whereas in the shorter range (mid-February to March) the mean has average skill, similar to the long-lead October predictions. It appears, thus, that skilful prediction of the ensemble mean is directly linked to that mean being more skilful than all members.

This may be verified when looking at all months, as in Figure 9. One notices that when the ensemble mean skill
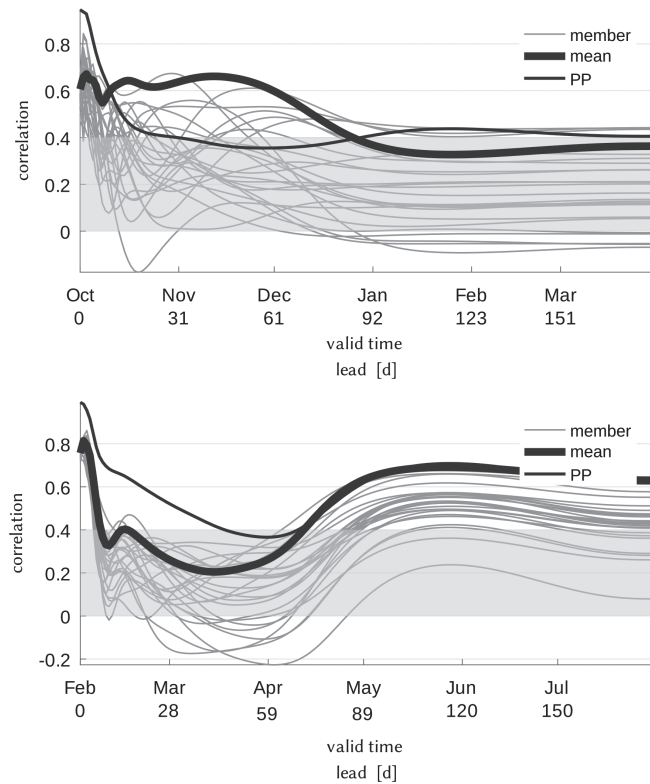
FIGURE 8 Comparison of seamless correlation skill of single members (grey) vs. the ensemble mean (thick black), for the 1 October (top) and 1 February (bottom) predictions. For comparison, the potential predictability skill is also shown (thin black)

is significantly high for a range of leads, as in the February, March, October and November predictions, it is indeed also higher than all members. Vice versa, if it is low, then it tends to be lower than several members, as is evident from the long-lead predictions of June, August and December. For almost all months and lead times, PP skill is at least as high as the ensemble mean, with two major exceptions: the shorter-lead October and the longer-lead November predictions (see above), which again points to inconsistent prediction ensembles for those months.

3.2 | São Francisco basin

3.2.1 | ERA5/XDS validation

Figure 10 shows observed and ERA5/XDS simulated areal *P* for the SFB. It is obvious that the seasonal cycle is somewhat weaker as compared to the KB (with its extended dry period), but still considerable, and it is well reproduced by the downscaled ERA5. This includes the main events, which are sometimes over- and sometimes underestimated, like for the KB. The respective de-seasonalized, NQT-normalized correlations for the validation period 1981–1999 are: $\rho = .70$ (P), $\rho = .78$ (T_n), $\rho = .78$ (T), $\rho = .79$ (T_x). Time series are shown in Appendix S1 (Figures S6 and S7).

3.2.2 | SEAS5/XDS local predictions

Compared to the important October/November skill of the KB, corresponding skill for the SFB (cf. Figure 11) is less satisfactory. As skill is already low for zero lead time, this may indicate a rather bad initialization for those months; note that predictions issued in autumn are by far the worst for all variants (gridded with and without drift-correction, XDS downscaled, cf. Figure S10 in Appendix S1). The medium-term November forecasts are somewhat better, while the long-term forecasts are just above the significance level. For the other months, skill values are often better, as demonstrated in Figure 12. Predictions are skilful when issued in the months from February to June, almost equally for all leads (except for the February case), in accordance with previous studies (Hastenrath *et al.*, 2009 and references therein). The mentioned behaviour, that the ensemble mean prediction skill is significantly high if and only if it is higher than all members, is violated only for the May and July predictions, which are skilful but slightly smaller than some members. PP skill is higher in practically all significant cases. As an example, we show in Figure 13 a comparison of the predictions valid for 1 November, issued in June (SF = 148 days) and October (SF = 55 days). The higher skill is visible for the longer-lead June forecasts. The medium-term October forecasts are especially bad in the years before 2000.

A figure comparing the ensemble mean skill to that of the single members and to the PP, as in Figure 8, is shown in Appendix S1 (Figure S12).

3.3 | Discussion

We need to understand the high prediction skill for long leads that is seen for both basins. A good example where both predictions are similarly high is for February. Both start quite skilful $(\rho > .8)$, then level off in March, and regain high skill $(\rho > .6)$ up until the end of the forecast period. Forecasting for June involves a lead-time of 4 months or ~120 days, which translates into an SF smoothing of about 140 days, representing a valid time from spring to the summer months. Inspecting climatology, rainfall for that period is to a larger portion determined by spring rainfall. If that is well predicted, the corresponding skill would be extended into the summer months and "smoothed" with their own skill. We have checked prediction skill confined to April alone, by employing a seamless filter with a much weaker, roughly monthly, smoothing, using $(A_x = 30, \xi = 100.0)$, and compare it to our standard filter $(A_x = 150, \xi = 2.4)$ in Figure 14. For the KB, there is indeed monthly skill for March and, marginally, for July. For the SFB, on the other hand, monthly skill is quite strong throughout from February to July. So at least for SFB, the high skill values that we have seen for the original SF can be explained from the monthly skill already.
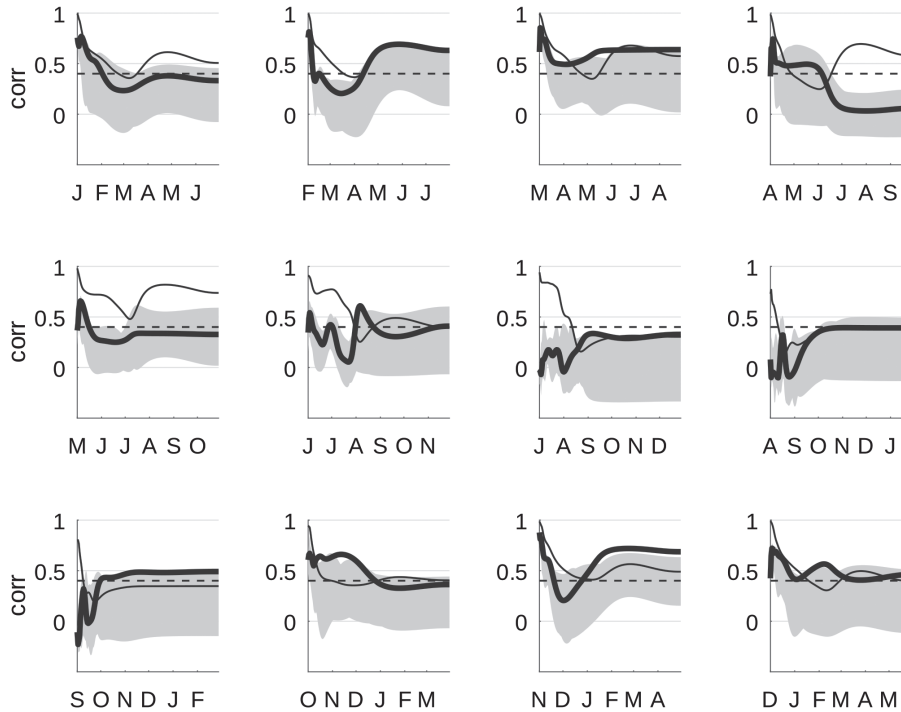
FIGURE 9 For KB, all 12 monthly seamless prediction skills for the ensemble mean (thick black) and potential predictability (thin black), along with the range of the single members (grey)
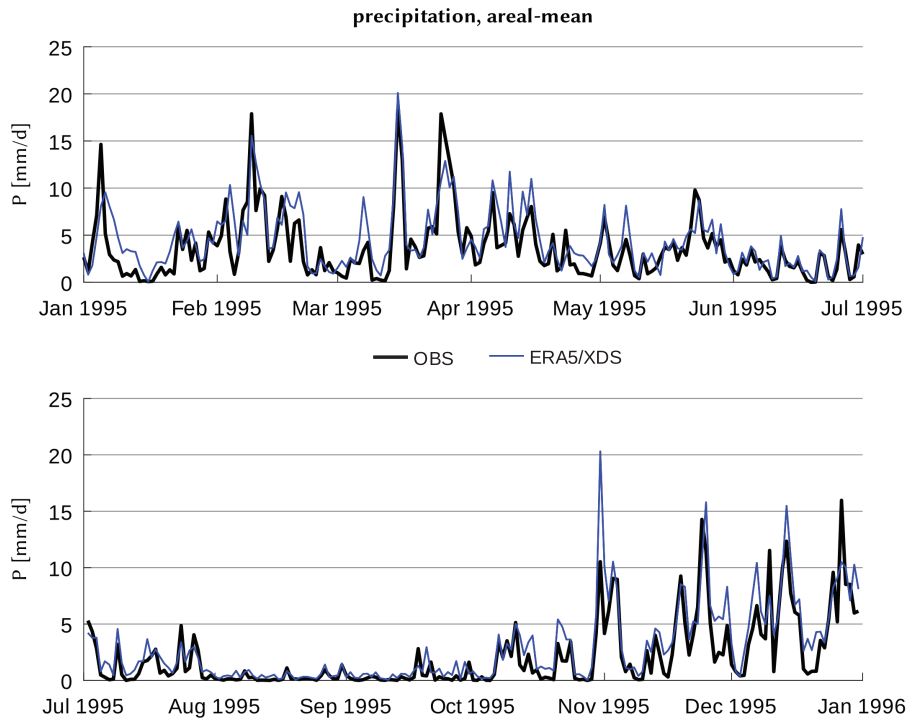

FIGURE 10 Similar to Figure 5, for the SFB

4 | CONCLUSIONS

The seamless filter (SF) is introduced here as a tool to present daily to seasonal forecasts in a concise and continuous way, with lead time and time-scale varying consistently. The rule

after which they vary – how time-scale (smoothing) increases with lead time – needs to be specified beforehand from expert knowledge or user demand. Applicability of the SF, and corresponding skill, was tested using ECMWF's SEAS5 system and 180-day forecasts for two river basins in Iran (KB) and Brazil (SFB). Being station-based, the forecasts, after
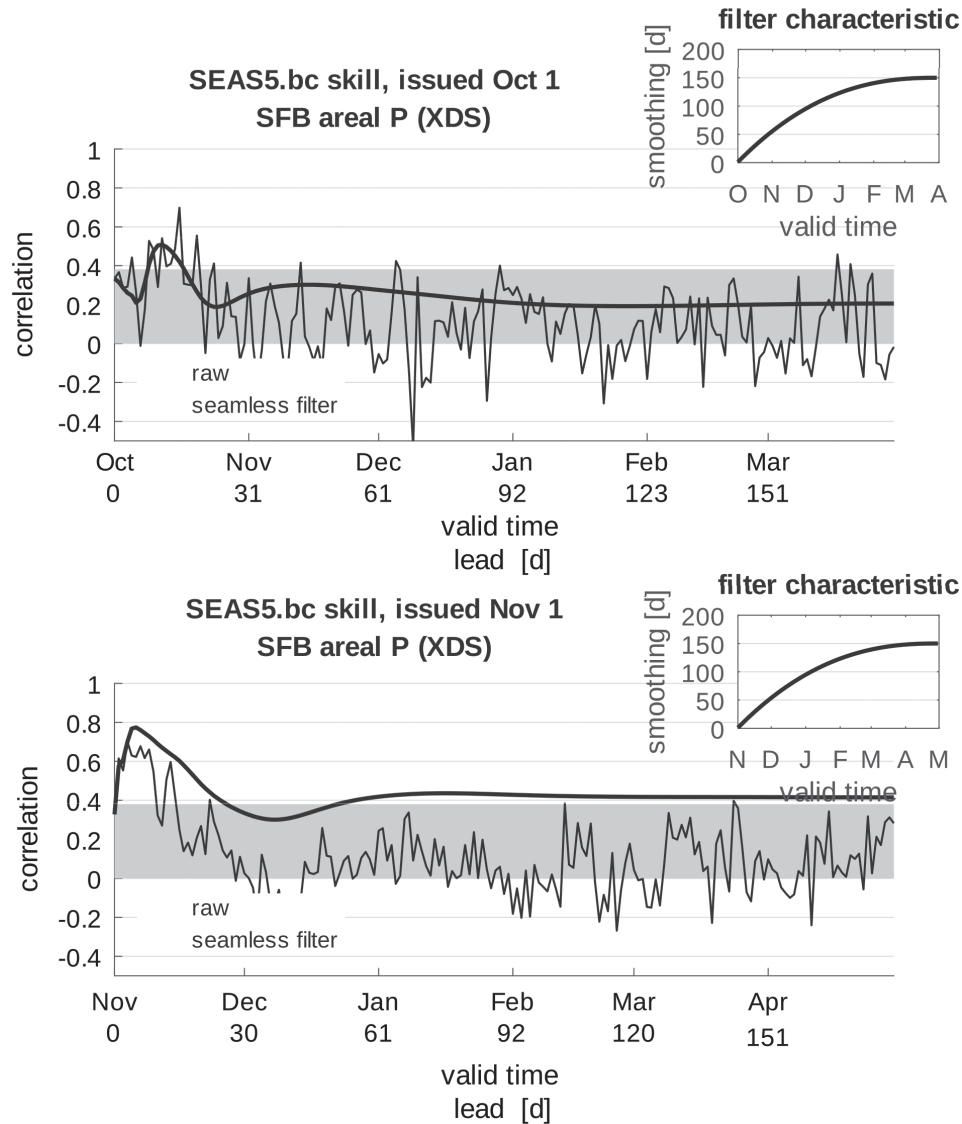
FIGURE 11   Like Figure 6, for the SFB

the required downscaling (here done by XDS), fit easily to existing water management strategies.

The water year for both basins starts in October–November and ends in March–April the following year. Ideally, one would like to have skilful long-term predictions starting some time in autumn. According to Johnson *et al.* (2019), fig. 20, both regions lie at the boundary of skilful tropical predictions (DJF, 1-month lead). This is consistent with Figures 6 and 12 here. For the KB, predictability extends further in to the water year with increased smoothing up until April. For the SFB, the same is not true since predictability is not clearly above the significance level. Greater predictability is seen starting with the January forecasts and higher medium-term skill, and extending to the full 180-day range (and strong smoothing) for the February to June predictions.

The character of this study is mostly methodological, yet a word of caution is in place with respect to so many forecast skill outcomes, across multiple locations, lead times, and time-scales, that entail enough statistical degrees of freedom from which "skilful" predictions are easily detected based on pure chance. One must shy away from overemphasizing them, or if not, scrutinizing them carefully with respect to physical plausibility and internal consistency. This was done only briefly here (e.g. Figures 7 and 13) but should be done elsewhere with greater diligence.

The study can be extended in three obvious directions. (a) By focusing solely on areal mean precipitation, spatial patterns varying within the season and, correspondingly, within the forecast period are ignored. For example, in the SFB (with more than 20° latitude) the rainy season travels from Tres Marias, then to Sobradinho, and finally to Itaparica (personal communication, Francisco Vasconcelos Júnior, FUNCEME); this should be accounted for in a refined assessment. (b) The current study evaluates ensemble mean forecasts, which helped to somewhat reduce the complexity of the problem. The next step is to extend the evaluation to
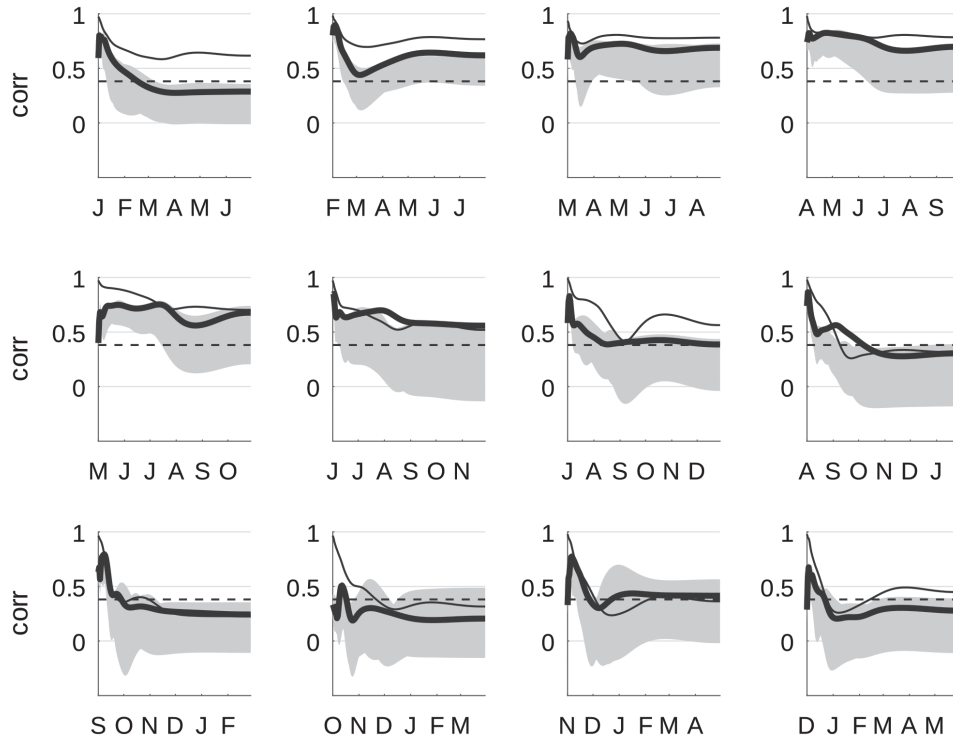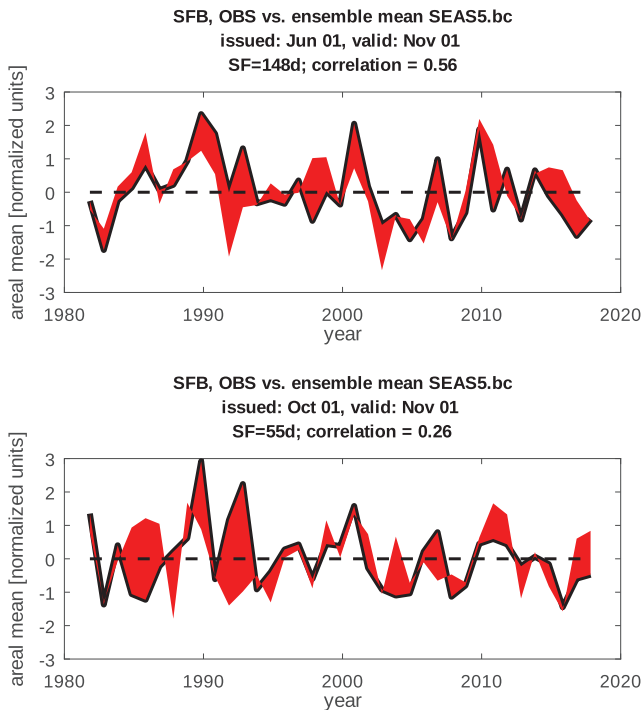
FIGURE 12 Like Figure 9, for the SFB

FIGURE 13 Similar to Figure 7, the June (top) and October (bottom) seamless predictions valid 1 November, for the SFB
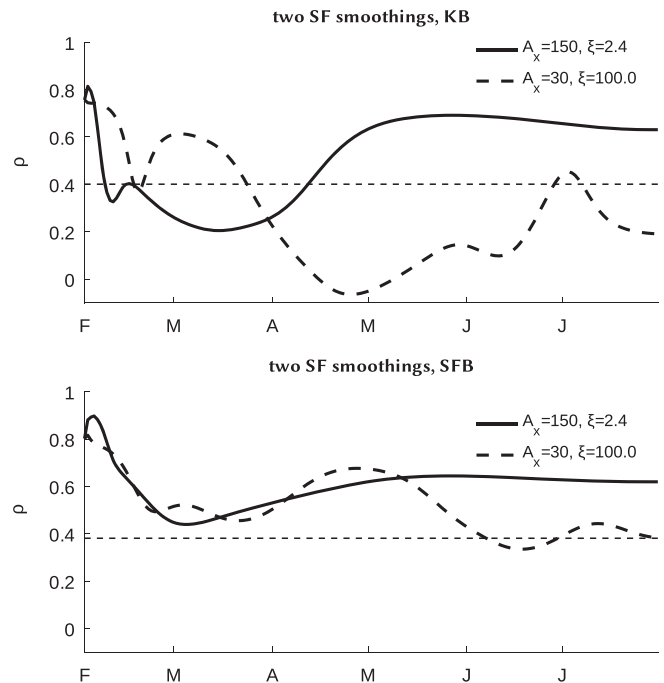

FIGURE 14 For both basins, seamless February forecast skills with two different SF characteristics, one with strong (solid) and one with weak smoothing (dashed)

a fully probabilistic setting that assesses local forecasts along with their uncertainty. (c) The SF aggregation is currently formulated as a simple form of averaging (or Butterworth smoothing). When a greater focus on real events is intended, that may not be enough. In that case, one may employ a more

refined statistic, such as the likelihood of events per time range, or their expected intensity, or any statistical measure that may fit to that time range. With the current set-up this can be implemented easily.

ACKNOWLEDGEMENTS, SAMPLES, AND DATA

ORCID

Gerd Bürger 🄳 https://orcid.org/0000-0003-3539-2975

REFERENCES

Alijani, B. (2002) Variations of 500 hPa flow patterns over Iran and surrounding areas and their relationship with the climate of Iran. *Theoretical and Applied Climatology*, 72, 41–54. https://doi.org/10.1007/s007040200011.

Bauer, P., Thorpe, A. and Brunet, G. (2015) The quiet revolution of numerical weather prediction. *Nature*, 525, 47–55.

Bronstert, A., Kolokotronis, V., Schwandt, D. and Straub, H. (2007) Comparison and evaluation of regional climate scenarios for hydrological impact analysis: general scheme and application example. *International Journal of Climatology*, 27, 1579–1594.

Brunet, G., Jones, S. and Ruti, P.M. (2015) *Seamless Prediction of the Earth System: From minutes to months*. Geneva, Switzerland: World Meteorological Organization.

Buizza, R. (1997) Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF ensemble prediction system. *Monthly Weather Review*, 125, 99–119. https://doi.org/10.1175/1520-0493(1997)125<0099:PFSOEP>2.0.CO;2.

Buizza, R., Leutbecher, M. and Thorpe, A. (2015) *Living with the butterfly effect: a seamless view of predictability*. ECMWF Newsletter 145, 18–23.

Bürger, G. (2002) Selected precipitation scenarios across Europe. *Journal of Hydrology*, 262, 99–110.

Bürger, G., Murdock, T., Werner, A., Sobie, S. and Cannon, A. (2012) Downscaling extremes – An intercomparison of multiple statistical methods for present climate. *Journal of Climate*, 25, 4366–4388.

DelSole, T. and Tippett, M.K. (2009a) Average predictability time. Part I: Theory. *Journal of the Atmospheric Sciences*, 66, 1172–1187.

DelSole, T. and Tippett, M.K. (2009b) Average predictability time. Part II: Seamless diagnoses of predictability on multiple time scales. *Journal of the Atmospheric Sciences*, 66, 1188–1204. https://doi.org/10.1175/2008JAS2869.1.

Ford, T.W., Dirmeyer, P.A. and Benson, D.O. (2018) Evaluation of heat wave forecasts seamlessly across subseasonal timescales. *npj Climate and Atmospheric Science*, 1, 1–9. https://doi.org/10.1038/s41612-018-0027-7.

Gustafsson, F. (1996) Determining the initial states in forward-backward filtering. *IEEE Transactions on Signal Processing*, 44, 988–992.

Hastenrath, S., Sun, L. and Moura, A.D. (2009) Climate prediction for Brazil's Nordeste by empirical and numerical modeling methods. *International Journal of Climatology*, 29, 921–926. https://doi.org/10.1002/joc.1770.

Hastie, T., Tibshirani, R., Friedman, J. and Franklin, J. (2005) *The elements of statistical learning: data mining, inference and prediction*, 2nd ed. New York, NY: Springer-Verlag.

Hersbach, H. and Dee, D. (2016) ERA5 reanalysis is in production. *ECMWF Newsletter*, 147(7), 5–6.

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. (2019) SEAS5: the new ECMWF seasonal forecast system. *Geoscientific Model Development*, 12, 1087–1117. https://doi.org/10.5194/gmd-12-1087-2019.

Kirtman, B., and Pirani, A. (2008) *WCRP position paper on seasonal prediction*. WCRP Informal Report No. 3/2008, ICPO Publication No. 127.

Silva, V.B.S. and Kousky, V.E. (2012) The South American monsoon system: climatology and variability. In: Wang, S.-W. and Gillies, R.R. (Eds.) *Chapter 5 in Modern Climatology*. Norderstedt, Germany: IntechOpen.

Lorenz, E.N. (1972) *Predictability: Does the flap of a butterfly's wings in Brazil set off a tornado in Texas?* . Washington, DC: American Association for the Advancement of Science.

Molteni, F., Stockdale, T., Balmaseda, M., Balsamo, G., Buizza, R., Ferranti, L., Magnusson, L., Mogensen, K., Palmer, T.N. and Vitart, F. (2011) *The New ECMWF Seasonal Forecast System (System 4)*. Technical Memorandum 656, Reading: European Centre for Medium-Range Weather Forecasts.

Palmer, T.N., Doblas-Reyes, F., Weisheimer, A. and Rodwell, M. (2008) Toward seamless prediction: calibration of climate change projections using seasonal forecasts. *Bulletin of the American Meteorological Society*, 89, 459–470.

Reichler, T. and Roads, J. (2003) The role of boundary and initial conditions for dynamical seasonal predictability. *Nonlinear Processes in Geophysics*, 10, 211–232.

Robertson, A.W. and Vitart, F. (2019) *Sub-seasonal to Seasonal Prediction: The gap between weather and climate forecasting*. Amsterdam, Netherlands: Elsevier.

Rowell, D.P. (1998) Assessing potential seasonal predictability with an ensemble of multidecadal GCM simulations. *Journal of Climate*, 11, 109–120.

Shukla, J., Hagedorn, R., Miller, M., Palmer, T.N., Hoskins, B., Kinter, J., Marotzke, J. and Slingo, J. (2009) Strategies: revolution in climate prediction is both necessary and possible: a declaration at the world modelling summit for climate prediction. *Bulletin of the American Meteorological Society*, 90, 175–178.

Sigmond, M., Scinocca, J.F., Kharin, V.V. and Shepherd, T.G. (2013) Enhanced seasonal forecast skill following stratospheric sudden warmings. *Nature Geoscience*, 6, 98–102. https://doi.org/10.1038/ngeo1698.

Stockdale, T., Johnson, S., Ferranti, L., Balmaseda, M. and Briceag, S. (2018) *ECMWF's new long-range forecasting system SEAS5*. Reading: ECMWF Newsletter, 154, 15–20.

Sun, J. and Ahn, J. (2015) Dynamical seasonal predictability of the Arctic Oscillation using a CGCM. *International Journal of Climatology*, 35, 1342–1353.

Vitart, F. and Robertson, A.W. (2018) The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. *npj Climate and Atmospheric Science*, 1, 3.

Wheeler, M.C., Zhu, H., Sobel, A.H., Hudson, D. and Vitart, F. (2017) Seamless precipitation prediction skill comparison between two global models. *Quarterly Journal of the Royal Meteorological Society*, 143(702), 374–383. https://doi.org/10.1002/qj.2928.

Yadav, R.K. (2016) On the relationship between Iran surface temperature and northwest India summer monsoon rainfall. *International Journal of Climatology*, 36, 4425–4438.

Zhu, H., Wheeler, M.C., Sobel, A.H. and Hudson, D. (2014) Seamless precipitation prediction skill in the tropics and extratropics from a global model. *Monthly Weather Review*, 142, 1556–1569.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.