Seasonal sub-basin-scale runoff predictions: A regional hydrometeorological Ensemble Kalman Filter framework using global datasets

Maurus Borne [a,b,*], Christof Lorenz [b], Tanja C. Portele [b], Eduardo Sávio P. R. Martins [c], Francisco das Chagas Vasconcelos Junior [c], Harald Kunstmann [b,d]

[a] *Karlsruhe Institute of Technology, Institute of Meteorology and Climate Research (IMK-TRO), Wolfgang-Gaede-Stasse 1, Karlsruhe 76131, Baden-Württemberg, Germany*
[b] *Karlsruhe Institute of Technology, Institute of Meteorology and Climate Research (IMK-IFU), Kreuzeckbahnstraße 19, Garmisch-Partenkirchen 82467, Bavaria, Germany*
[c] *Ceará Institute for Meteorology and Water Resources (FUNCEME), Avenida Rui Barbosa 1246, Fortaleza 60001, Ceará, Brazil*
[d] *University of Augsburg, Institute of Geography, Alter Postweg 118, Augsburg 86159, Bavaria, Germany*

ABSTRACT

Study region: The São Francisco River Basin (SFRB) in Brazil
Study focus: In semi-arid regions, interannual variability of seasonal rainfall and climate change is expected to stress water availability and increase the recurrence and intensity of extreme events such as droughts or floods. Local decision makers therefore need reliable long-term hydro-meteorological forecasts to support the seasonal management of water resources, reservoir operations and agriculture. In this context, an Ensemble Kalman Filter framework is applied to predict sub-basin-scale runoff employing global freely available datasets of reanalysis precipitation (ERA5-Land) as well as bias-corrected and spatially disaggregated seasonal forecasts (SEAS5-BCSD). Runoff is estimated using least squares predictions, exploiting the covariance structures between runoff and precipitation. The performance of the assimilation framework was assessed using different ensemble skill scores.
New hydrological insights for the region: Our results show that the quality of runoff predictions are closely linked to the performance of the rainfall seasonal predictions and allows skillful predictions up to two months ahead in most sub-basins. The anthropogenic conditions such as in the Western Bahia state, however, must be taken under consideration, since non-stationary runoff time-series have poorer skill as such unnatural variations can not be captured by long-term covariances. In sub-basins which are dominated by little anthropogenic influence, the presented framework provides a promising and easily transferable approach for skillful operational seasonal runoff predictions on sub-basin scale.

* Corresponding author at: Karlsruhe Institute of Technology, Institute of Meteorology and Climate Research (IMK-TRO), Wolfgang-Gaede-Stasse 1, Karlsruhe 76131, Baden-Württemberg, Germany.
E-mail address: maurus.borne@kit.edu (M. Borne).

1. Introduction

The effect of global warming on climatic extreme events such as droughts combined with the growing water demand for cities and irrigation can result in intense water scarcity in certain regions of the world (Schewe et al., 2014; Stocker et al., 2014; Du et al., 2018). Mostly impacted are semi-arid regions due to limited water availability, accumulation and intensification of drought periods and increasing sedimentation of water reservoirs. Especially in those regions, observational data is sparse and the number of reliable and long-term operating measurement stations is decreasing (Lorenz and Kunstmann, 2012). This does not allow decision makers to rely on local statistics to anticipate those hardly predictable extreme events, and thus to plan for a proactive water allocation during droughts and to manage reservoir levels during floods. Climatic information on seasonal scales are therefore particularly important and fortunately, semi-arid climates near or at the tropics can offer reasonable predictability on seasonal timescales: In fact, the climate variability in these regions is mainly affected by large-scale forcing of thermal and dynamical atmosphere-ocean interaction from intraseasonal to multidecadal timescale (e.g., Madden Julian Oscillation, El Niõo Southern Oscillation, Pacific Decadal Oscillation, Atlantic Multi-Decadal Oscillation, Quasi-Biennal Oscillation) (Kayano and Andreoli, 2004; De Souza and Ambrizzi, 2006; Taschetto and Ambrizzi, 2012; Taschetto and Wainer, 2008; He et al., 2017; Li et al., 2020); or long-term interactions like inertial memory, soil moisture, sea surface temperatures (Koster et al., 2000), that are well represented in state-of-the-art seasonal predictions systems (Johnson et al., 2019). Hence, an adjustment of policies to sustainable water-management and the development of tools to make global information usable for individual regions is required. In this context, seasonal sub-basin-scale runoff is proposed to be estimated
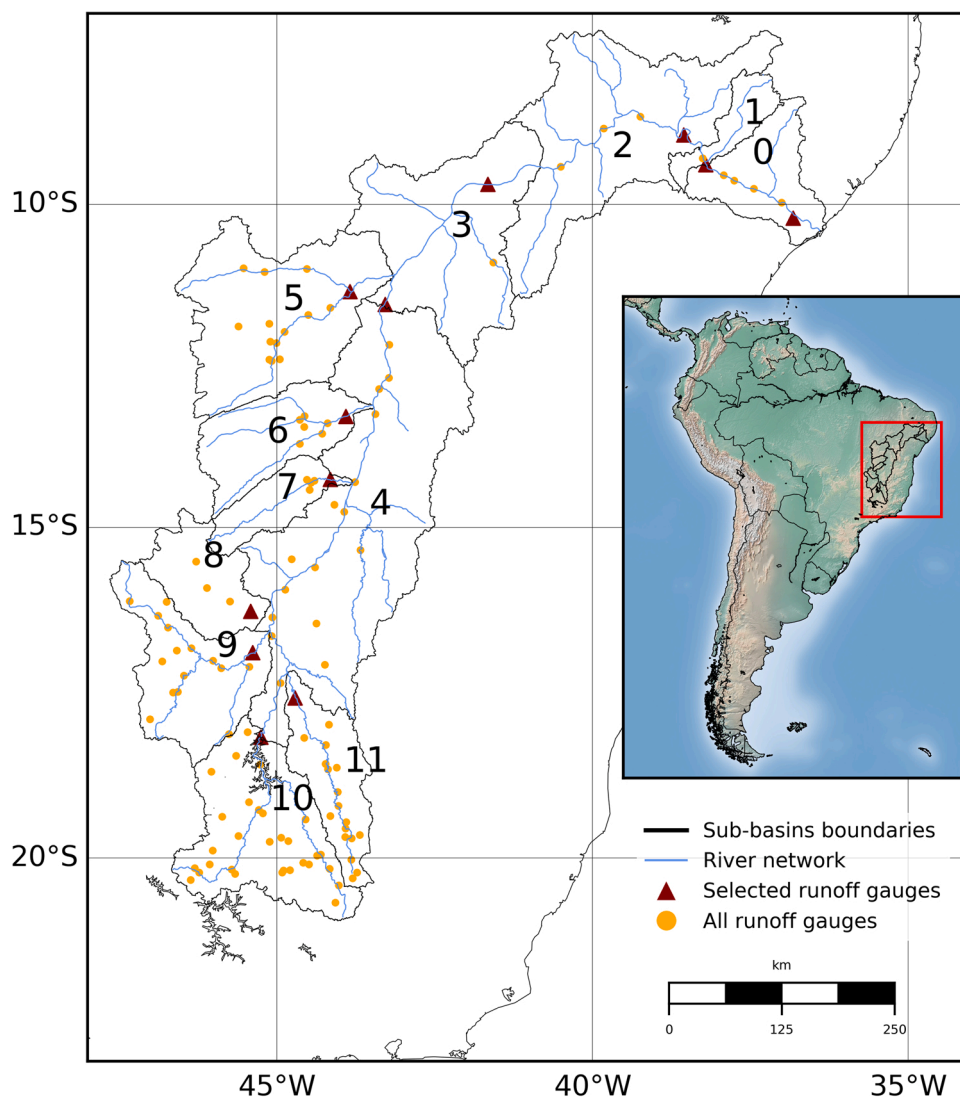

Fig. 1. The São Francisco River Basin (SFRB) in Brazil. The blue solid lines indicate the tributaries and the black solid lines show the delineation of the sub-basins. The markers represent the location of each, i.e, the selected gauges (red triangle) and all the gauges (orange circle). The numbers (0−11) are used to identify the different sub-catchments. Sub-basin N˚0 is not analysed in the EnKF framework.

applying a hydro-meteorological EnKF framework, using the statistics of global freely available datasets and global model predictions. This approach is derived from a previous study (Lorenz et al., 2015), where EnKF-based basin-scale runoff estimations proved to provide promising results on monthly time scales. More specifically, we focus on the application of publicly available global hydro-meteorological datasets and exploit the joint temporal and regional covariance structures to generate least square estimated predictions. The motivation is to develop a similar framework for seasonal runoff predictions, using seasonal precipitation predictions from the European Centre for Medium-Range Weather Forecasts (ECMWF) latest seasonal forecasting system SEAS5 as well as reanalysis-based reference precipitation from ERA5-Land, which is an offline re-run of the land-surface component from ECMWF's latest atmospheric reanalysis ERA5. We further use runoff observation data from the Brazilian National Water Agency (ANA). In fact, different studies showed the potential of seasonal forecasts to predict extreme events like droughts and to support water management in semi-arid regions (Changnon and Vonnhame, 1986; Garbrecht et al., 2006; Portele et al., 2021). In this regard, various statistical rainfall-runoff models (Misumi et al., 2001; Garbrecht et al., 2006; Ajami et al., 2016; De Paiva et al., 2020) or process-based hydrological models (Yuan et al., 2016; Crochemore et al., 2016; Meißner et al., 2017; Foster et al., 2018; Sehgal and Sridhar, 2019; Wanders et al., 2019), which convert climate or seasonal precipitation forecasts into stream flows adapted for real-time basin monitoring, already exist. In comparison, our proposed framework uses high-resolution precipitation forecasts within a statistical rainfall-runoff EnKF framework to predict computationally efficient seasonal runoff at sub-catchment level. In this study, the São Francisco River Basin (SFRB) in Brazil is investigated. Being disposed to have increasing water-related problems in the future (Cunha et al., 2018), the SFRB requires sustainable mitigation strategies and therefore provides a suitable testbed for the EnKF-based seasonal hydrological forecasting system.

2. Material

2.1. Study area

With a total surface area of 636, 851 km^2, the SFRB (Fig. 1) is a very multifaceted region, both in terms of its climatic and physical characterization as well as its environmental diversity. With its 2860 km of length, the river originates in the Canastra mountain in the central part of the Southeast region and flows into the Atlantic Ocean in the Northeast part of the catchment. The main course of the river is supplied by the tributaries of several sub-basins with 75 % of the water flow generated in the upper reach of the state of Minas Gerais (Maneta et al., 2009; Traini et al., 2012; de Jong et al., 2018). The SFRB unites regions with two different climate regimes between the upper and lower basin. In the southern part of the basin (regions 3–11 in Fig. 1), the rainy season is during November-December-January (NDJ) and its variability is strongly influenced by the South Atlantic Convergence Zone (SACZ) (Kodama, 1992; Kodama, 1993; Carvalho et al., 2002),(Carvalho et al., 2004). The rainy season over the northern part (regions 0–2 in Fig. 1) is observed during January-Apr (JFMA) and is mainly linked to the high activity of the Intertropical Convergence Zone (ITCZ) and the emergence of easterly waves and squall lines derived by the sea breeze (ParedesTrejo et al., 2016; Sun et al., 2016). Because of its high population of approx. 17 million people, the region incorporates important economic activities, increasing the stress on the environment and the water resources (Ioris, 2001; Maneta et al., 2009). Apart from drinking water, freshwater resources are mainly used for electricity generation, urban and industrial services, intensive irrigation and mining activities. This involves a high concentration of dams and water abstractions all along the SFRB river (de Jong et al., 2018; do Vasco et al., 2019; dos Santos et al., 2020). The Western Bahia state (regions 5, 6, and 7 in Fig. 1) in particular, is a very active agricultural area which is subject to significant reduction in river discharge, rainfall and ground water level since the 1980s (Pousa et al., 2019; Santos et al., 2020). In addition to the significant environmental impact of these activities on biodiversity, soil erosion and land desertification (Cunha et al., 2015; Tomasella et al., 2018), the North East of the region was recently hit by an intense multi-year drought that caused a significant lack of water in the region and emptied most reservoirs (Marengo et al., 2018; Martins et al., 2018).

2.2. Data

The study is based on 11 sub-basins of the SFRB (Fig. 1) and monthly time series over a total period of 30 years between 1981 and 2011. The 11 sub-basins were delineated and associated with runoff gauges given the spatial distribution of the gauges and the spatial resolution of the precipitation products. Information about the different data sets used for this study are summarized in Table 1. More detailed information about the runoff gauges and missing values can be found in Table 2.

2.2.1. Runoff

The gauges located at the sub-basins outlet were used to quantify the runoff of each sub-basin (red triangles in Fig. 1). The daily discharge data of the 11 selected gauges were obtained from the Brazilian National Water Agency (ANA) database, which contains all

Table 1
Summary of the datasets used in this study.

Variable	Dataset	Grid size	Temporal resolution	Time Period
Runoff	ANA	gauges	daily	1981–2011
Precipitation	ERA5-Land	9 km	daily	1981–2011
Seasonal forecast of precipitation	SEAS5-BCSD	36 km	daily	1981–2011

Table 2
Description of the gauging stations used for each sub-basin with information on mean annual discharge, runoff and percentage of missing values in the daily time series.

Sub-basin N°	Station name	Area (km^2)	Discharge (m^3/s)	Runoff ($mm/year$)	Missing values (%)
1	P.Afonso	12,544	2567	135	0
2	Itaparcia	194,336	2564	137	0
3	Sobradinho	76,918	2457	156	0
4	Morpara	139,289	2328	213	0
5	Boqueirao	76,946	273	112	0.02
6	Porto Novo	34,082	207	192	0
7	Juvenilia	17,131	136	252	0
8	Santo Inacio	25,165	223	286	0.04
9	Porto Alegre	44,975	502	352	0.1
10	Tres Marias	56,767	709	394	0
11	Varzea Da Palma	27,860	297	337	0.01

information collected by the National Hydrometeorological Network (RHN) in Brazil. The data are averaged monthly and the few missing values (see table 2) are filled using the climatology. Because of the high dam concentration in region 0 (Fig. 1) which denies the description of the natural relationship between precipitation and runoff through simple covariances or correlations, only the N°1–11 sub-basins are considered in this study.

2.2.2. Precipitation

Reference precipitation is taken from the ERA5-Land climate reanalysis (Muñoz-Sabater et al., 2021) dataset of the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (IFS). This dataset is derived from a re-run of the land component of the ECMWF ERA5 climate reanalysis (Hersbach, 2016) and contains gridded daily global land surface precipitation at 9 km resolution from 1950 onwards. ERA5-Land has the particular advantage of including a representation of the hydrological cycle, which improves the agreement between river discharge estimates and available observations. The hydro-meteorological consistency and high spatial resolution of ERA5-Land makes it an ideal dataset for hydrological research purposes. The gridded data were monthly and spatially averaged over the different sub-basins to obtain basin scale rainfall estimates.

2.2.3. Bias-corrected and spatially disaggregated seasonal forecasts (SEAS5-BCSD)

Monthly seasonal forecasts for precipitation are obtained from ECMWFs latest seasonal forecasting system SEAS5, which are bias-corrected and spatially disaggregated towards ERA5-Land (Lorenz et al., 2021). The forecasts cover the whole (re-)forecast period from 1981 to 2019 and include bias-corrected and spatially disaggregated daily and monthly ensemble forecasts for precipitation, average, minimum and maximum temperature as well as for shortwave radiation from the issue date to the next 215 days and have a spatial resolution of 0.1°. They consist of an ensemble of 25 (until 2016) and 51 (from 2017) members, that are corrected towards ERA5-Land using a quantile-mapping approach for estimating a temporally consistent set of land-surface variables. The bias-correction of precipitation forecasts further includes a correction of precipitation intermittency to ensure the agreement of the wet- and dry-day frequencies from ERA5-Land and SEAS5-BCSD. Compared to the raw SEAS5 forecasts, SEAS5-BCSD has a higher spatial resolution, lower biases with respect to ERA5-Land and substantially reduced model drifts with forecast lead time and are, hence, more suitable for regional applications. Because this study focuses on a time period prior 2016, the SEAS5-BCSD ensemble size has only 25 members.

3. Methods

3.1. Prediction model

For this EnKF framework, the state of the system is represented as a multi-dimensional vector **X** of size $2 \times N_{reg}$ with N_{reg} being the number of sub-basins, containing the runoff **R** and the precipitation **P** state vectors to be estimated at each time step t.

$$\mathbf{X}_t = \begin{bmatrix} \mathbf{P}_t^T & \mathbf{R}_t^T \end{bmatrix}^T \tag{1}$$

The prediction equation has been derived under the assumption that the time evolution of the precipitation and runoff anomalies can be written as a first order stochastic process

$$\mathbf{r}_t = \mathbf{A}\mathbf{r}_{t-1} + \boldsymbol{\epsilon} \tag{2}$$

with the monthly residual **r** at time step t

$$\mathbf{r}_t = \mathbf{X}_t - \widetilde{\mathbf{X}}_m = \begin{bmatrix} \mathbf{P}_t - \widetilde{\mathbf{P}}_m \\ \mathbf{R}_t - \widetilde{\mathbf{R}}_m \end{bmatrix} \tag{3}$$

and the long-term mean annual cycle for the state vector $\widetilde{\mathbf{X}}$ at month m over a total period of τ months

$$\widetilde{\mathbf{X}}_m = \frac{1}{\tau/12} \sum_{i,m}^{\tau/12} \mathbf{X}_{i,m}. \tag{4}$$

The prediction is characterized by the matrix of the process dynamics $\mathbf{A}$ and is supposed to model the dynamics and statistical relationship of the temporal evolution of the two water cycle variables. The zero-mean normally distributed white prediction noise $\epsilon \sim \mathcal{N}(0, \mathbf{Q}_P)$ represents errors coming from the model data, parameters and physics with $\mathbf{Q}_P$ being the prediction error covariance matrix. Because of the complexity and the lack of physical knowledge about the dynamics of the process model, $\mathbf{A}$ is considered as unknown. Nevertheless, an estimation $\widehat{\mathbf{A}}$ can be obtained with *least square prediction* methods making use of the process stochastic information. For a complete and detailed derivation of the methodology, see (Moritz, 1980; Kurtenbach et al., 2012; Tourian, 2013; Lorenz et al., 2015). We are looking for an approximation $\widehat{\mathbf{A}}$ as a linear estimator of $\mathbf{r}_t$ given $\mathbf{r}_{t-1}$

$$\widehat{\mathbf{r}}_t = \widehat{\mathbf{A}} \mathbf{r}_{t-1} \tag{5}$$

where $\widehat{\mathbf{r}}$ is the predicted state. The estimated prediction matrix $\widehat{\mathbf{A}}$ should minimize the prediction error $\mathbf{e} = \mathbf{r}_t - \widehat{\mathbf{r}}_t$, i.e. minimize the trace of the error covariance matrix $\mathbf{Q}_P$. The total error can be developed as follow

$$\begin{aligned} \mathbf{Q}_P &= E\{\mathbf{e}\mathbf{e}^T\} = E\{(\mathbf{r}_t - \widehat{\mathbf{r}}_t)(\mathbf{r}_t - \widehat{\mathbf{r}}_t)^T\} \\ &= \mathbf{\Sigma} - \mathbf{\Sigma}_\Delta \widehat{\mathbf{A}}^T - \widehat{\mathbf{A}} \mathbf{\Sigma}_\Delta^T + \widehat{\mathbf{A}} \mathbf{\Sigma} \widehat{\mathbf{A}}^T \end{aligned} \tag{6}$$

and is thus minimized for

$$\widehat{\mathbf{A}} = \mathbf{\Sigma}_\Delta \mathbf{\Sigma}^{-1} \tag{7}$$

By inserting (7) into (6) we finally obtain the error covariance of the prediction matrix $\widehat{\mathbf{A}}$

$$\mathbf{Q}_P = \mathbf{\Sigma} - \mathbf{\Sigma}_\Delta \mathbf{\Sigma}^{-1} \mathbf{\Sigma}^T. \tag{8}$$

where $\mathbf{\Sigma} = E\{\mathbf{r}_t \mathbf{r}_t^T\}$ and $\mathbf{\Sigma}_\Delta = E\{\mathbf{r}_t \mathbf{r}_{t-1}^T\}$ are the spatio-temporal auto- and cross-covariance matrices, respectively. These covariance matrices are approximated with the empirical sample covariance matrices $\widehat{\mathbf{\Sigma}}$ and $\widehat{\mathbf{\Sigma}}_\Delta$ between anomalies of precipitation, and runoff over the total time period of τ months

$$\mathbf{\Sigma} \approx \widehat{\mathbf{\Sigma}} = \frac{1}{\tau - 1} \sum_{t=1}^{\tau} \mathbf{r}_t \mathbf{r}_t^T \tag{9}$$

$$\mathbf{\Sigma}_\Delta \approx \widehat{\mathbf{\Sigma}}_\Delta = \frac{1}{\tau - 2} \sum_{t=2}^{\tau} \mathbf{r}_t \mathbf{r}_{t-1}^T. \tag{10}$$

3.2. The Ensemble Kalman Filter framework

According to the previous derived least square prediction estimate, the prediction equation can be developed by substituting (3) in (2)

$$\mathbf{X}_{t,d}^f = \widehat{\mathbf{A}} \mathbf{X}_{t-1,d}^a + \mathbf{B} \mathbf{U}_{t-1} + \boldsymbol{\epsilon}_d \tag{11}$$

where $\mathbf{B}$ and $\mathbf{U}$ represent the control input terms

$$\mathbf{B} = [-\widehat{\mathbf{A}} \quad \mathbf{I}] \quad and \quad \mathbf{U}_t = [\widetilde{\mathbf{X}}_{t-1} \quad \widetilde{\mathbf{X}}_t]^T \tag{12}$$

As a result, the predicted state vector $\mathbf{X}_t^f$ at time t can be estimated using the analysis state $\mathbf{X}_{t-1}^a$ at time t-1, the prediction matrix $\widehat{\mathbf{A}}$, describing the variables spatio-temporal relationships and the control input $\mathbf{U}$ that represents the mean annual cycle. $\mathbf{I}$ stands for the identity matrix ($\mathbf{I} = \mathbf{\Sigma}\mathbf{\Sigma}^{-1}$) and the index d refers to the ensemble member from an ensemble of size N_{ens}. The prediction and observation covariance matrices are estimated by the sample covariance matrix from the ensemble prediction of size N_{ens}

$$\widehat{\mathbf{\Sigma}}_t^f = \frac{1}{N_{ens} - 1} \sum_{d=1}^{N_{ens}} \left(\mathbf{X}_{t,d}^f - \overline{\mathbf{X}}_t^f\right)\left(\mathbf{X}_{t,d}^f - \overline{\mathbf{X}}_t^f\right)^T \tag{13}$$

with $\overline{\mathbf{X}}_t^f$, the ensemble mean. These covariance matrices describe the uncertainty in the predicted and corrected states. In the innovation step, the analysis state $\mathbf{X}_{t,d}^a$ and its associated covariance matrix $\mathbf{\Sigma}_{X,t}^a$ are obtained by updating the gain-weighted difference between the observed state $\mathbf{Y}_t$ and the predicted state $\mathbf{X}_{t,d}^f$. The correction step is formulated as follow

$$\mathbf{X}_{t,d}^a = \mathbf{X}_{t,d}^f + \mathbf{K}_t(\mathbf{Y}_t - \mathbf{H}_t \mathbf{X}_{t,d}^f + \boldsymbol{\nu}_d) \tag{14}$$

with the observation relation matrix $\mathbf{H}_t$ that maps the predicted state to the observed space and the zero-mean normally distributed

observation noise $\nu \sim \mathcal{N}\left(0, \mathbf{Q}_{t,obs}\right)$. The Kalman gain matrix $\mathbf{K}_t$ represents the optimal weighting between the error covariance of the predicted state $\boldsymbol{\Sigma}_{X,t}^-$ and the errors of the observed state $\mathbf{Q}_{t,obs}$. It is defined as

$$\mathbf{K}_t = \boldsymbol{\Sigma}_{X,t}^f \mathbf{H}_t^T \left(\mathbf{H}_t \boldsymbol{\Sigma}_{X,t}^f \mathbf{H}_t^T + \mathbf{Q}_{t,obs}\right)^{-1} \tag{15}$$

The model is then integrated recursively forward in time from the innovation until the next update cycle.

3.3. Observation update

Generally, the observation vector $\mathbf{Y}_t$ and associated observation error covariance matrix $\mathbf{Q}_{t,obs}$ are given by

$$\mathbf{Y}_t = \begin{bmatrix} \mathbf{P}_{t,obs} \\ \mathbf{R}_{t,obs} \end{bmatrix} \tag{16}$$

$$\mathbf{Q}_{t,obs} = \begin{bmatrix} \boldsymbol{\Sigma}_{P,t} & \\ 0 & \boldsymbol{\Sigma}_{R,t} \end{bmatrix} \tag{17}$$

where $\mathbf{P}_{t,obs}$ and $\mathbf{R}_{t,obs}$ are the observation vectors of precipitation and runoff at time t with $\boldsymbol{\Sigma}_{P,t}$ and $\boldsymbol{\Sigma}_{R,t}$ their corresponding error covariance matrices. For the forecast initialization, the observation vector $\mathbf{Y}_t$ of size $2 \times N_{reg}$ includes the historical precipitation (ERA5-LAND) and runoff (ANA) vectors at time *t*. Since we assume perfect observations during the initialization, the corresponding error covariance is set to zero. During the forecast, the predicted state is updated at each time step t using the SEAS5-BCSD precipitation forecasts at the corresponding forecast lead time. Since runoff observations are not assimilated, the observation vector reduces to

$$\mathbf{Y}_t = \mathbf{P}_{t,obs} \tag{18}$$

where $\mathbf{P}_{t,obs}$ represents the precipitation observation vector at time *t* including all SEAS5-BCSD ensemble members. This gives $\mathbf{Y}_t$ a total size of $N_{reg} \times N_{BCSD}$, with N_{BCSD} the SEAS5-BCSD ensemble size. Indeed, the whole ensemble is assimilated in order to benefit from all the information contained in the forecasts. Each element of the observation vector has an associated observation error covariance contained in $\mathbf{Q}_{t,obs} = \boldsymbol{\Sigma}_{P,t}$, which is estimated from the squared ensemble spread of the SEAS5-BCSD precipitation ensemble forecasts for the given region and forecast horizon.

3.4. Further assumptions

No spin-up is applied to the model states as this reduces the prediction performance. Possible reasons could be that the observation covariance matrices generated by a probabilistic climatology are not well conditioned or that the convergence to the optimal error level is too slow and would require more data.

Various ensemble sensitivity tests, not shown here, suggest that an ensemble size of $N_{ens} = 200$ members is a robust choice for the EnKF runs.

3.5. Evaluation framework

Four different skill scores, namely the Inter-Quantile Range Skill Score (IQRSS) between the 95th and 5th percentiles, the Mean Absolute Error Skill Score (MAESS), the Continuous Ranked Probability Skill Score (CRPSS) and the Brier Skill Score (BSS) for above and below normal conditions were computed. A detailed description of the used score metrics can be found in Appendix A.

The reference ensemble forecast for runoff and precipitation are leave-one-year-out climatology-based ensemble forecasts generated from the ANA runoff and the ERA5-Land precipitation data, respectively. For a given year, this established method consists of combining the observations of all other years to create an ensemble forecast, so that the validation is done with independent data.

The performance scores are then derived for both runoff and SEAS5-BCSD precipitation predictions and are computed for each region, time step and lead time over the evaluation period. In this study, we focus only on the first four lead times, with lead 0 referring to the current month's forecast and lead 3 to the fourth month's forecast.

An overview of the size of the different parameters used in the EnKF framework can be found in table 3.

Table 3
Description and value of the different EnKF parameters.

Name	Acronym	value
Number of sub-basins	N_{reg}	11
SEAS5-BCSD ensemble size	N_{BCSD}	25
EnKF ensemble size	N_{ens}	200

4. Results and discussion

4.1. Annual cycle

Fig. 2 shows the rainy season of the whole SFRB and its 12 sub-basins. The two maxima visible in December (regions 3–11) and March (regions 0–2) shows the entanglement of the two different climate regimes between the upper and lower basin. The main rainy season in the lower basin is observed during January-April (JFMA) and in the upper portion of the basin it is during November-January (NDJ). In this study, we consider the period from November to March as a good compromise between the two climate zones to describe the wet season of the entire SFRB. Folland et al. (2001), Misra (2006) and Kulikova et al. (2014) presented evidence of a relationship between Pacific and tropical Atlantic SST anomaly patterns and seasonal precipitation in northeast Brazil, contributing to the long-term predictability over the SFRB. This is confirmed by the good level of agreement between the first lead times of SEAS5-BCSD precipitation forecasts and ERA5-Land precipitation in Fig. 2. Similarities in the spatio-temporal relationship between runoff and rainfall can also be observed. However, regions 5, 6, and 7 do not seem to follow the typical rainy season cycle, it is expected therefore that predicting runoff in those regions may be more difficult.

4.2. Prediction performance

The runoff of the 11 SFRB sub-basins is reforecasted on a monthly basis over the evaluation period 1981–2011. Fig. 3 shows the reforecasts of lead time 0 (blue) and corresponding runoff gauge observations (red). In most cases, the forecasts can reproduce the observed stream flow. In particular, the upper-basin (region 8–11) and lower basin (region 1–4) show the best agreement, with a good concordance between the predicted and observed runoff peaks. However, predictions for the Western Bahia state regions 5–7 have a lower quality compared to observations, with a noisy behavior and a general underestimation of the baseflow during the first 15 years. Correspondingly, the reduction of the river streamflow described in Section 2.1 is visible in these regions over the same period. Since the control input term **U** (eq. 12) carries information on the mean annual cycle, an abnormal structure of the annual cycle of those regions can explain the observed underestimation of the runoff predictions. While most events in the catchment are well predicted, a few seem to be more difficult to predict. (e.g.. Fig. 3 region 11 in 1991 or region 8 in 1982). It is therefore important to determine the origin of the performance, whether it derives from the algorithm, the quality of the driving data or the quality of the SEAS5 BCSD rainfall predictions. Therefore, a deeper analysis was initiated where the different performance metrics described in the methods section were calculated.

. Figs. 4 and 5 show the monthly and seasonally averaged skill scores, respectively, of the EnKF runoff reforecasts and corresponding SEAS5-BCSD precipitation predictions during the main rainy season (November - March) for the first four lead months in all regions. The skill of the predictions is shown in colors from blue (better skill than the reference forecast) to orange (worse skill than the
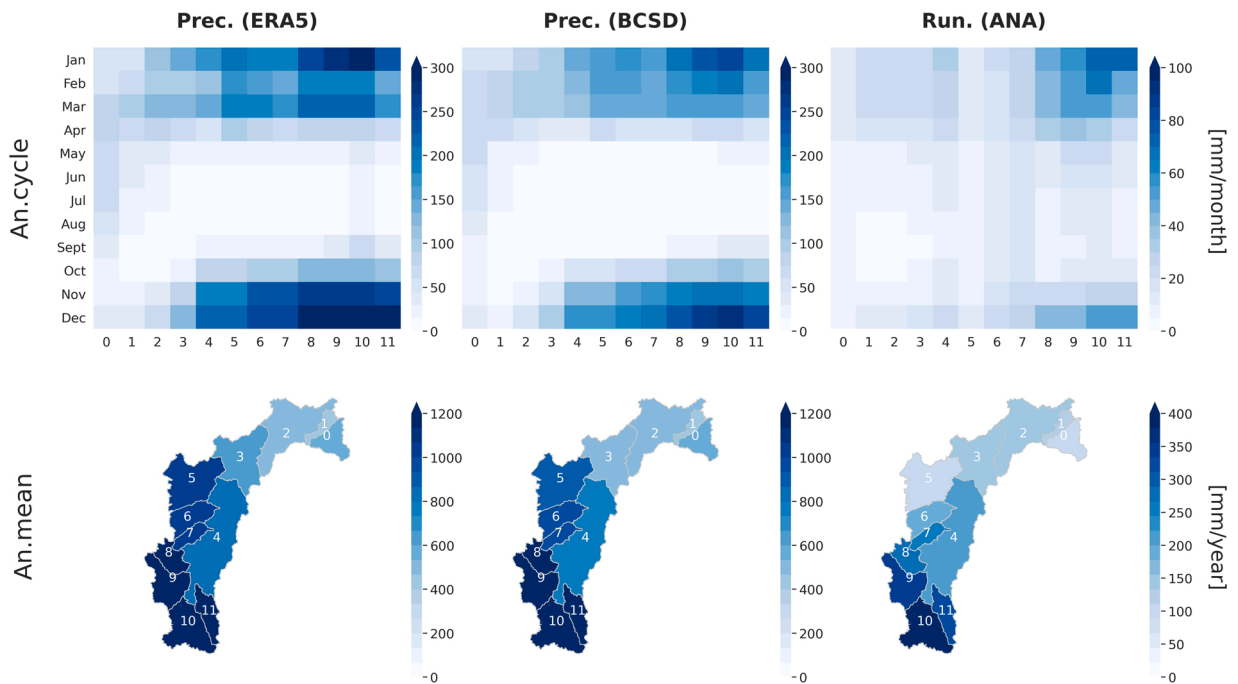

Fig. 2. Monthly (top) and annual (bottom) mean values from ERA5 land (left) and SEAS5-BCSD lead 0 (middle) precipitation and ANA (right) runoff over the different sub-basins of the SFRB.
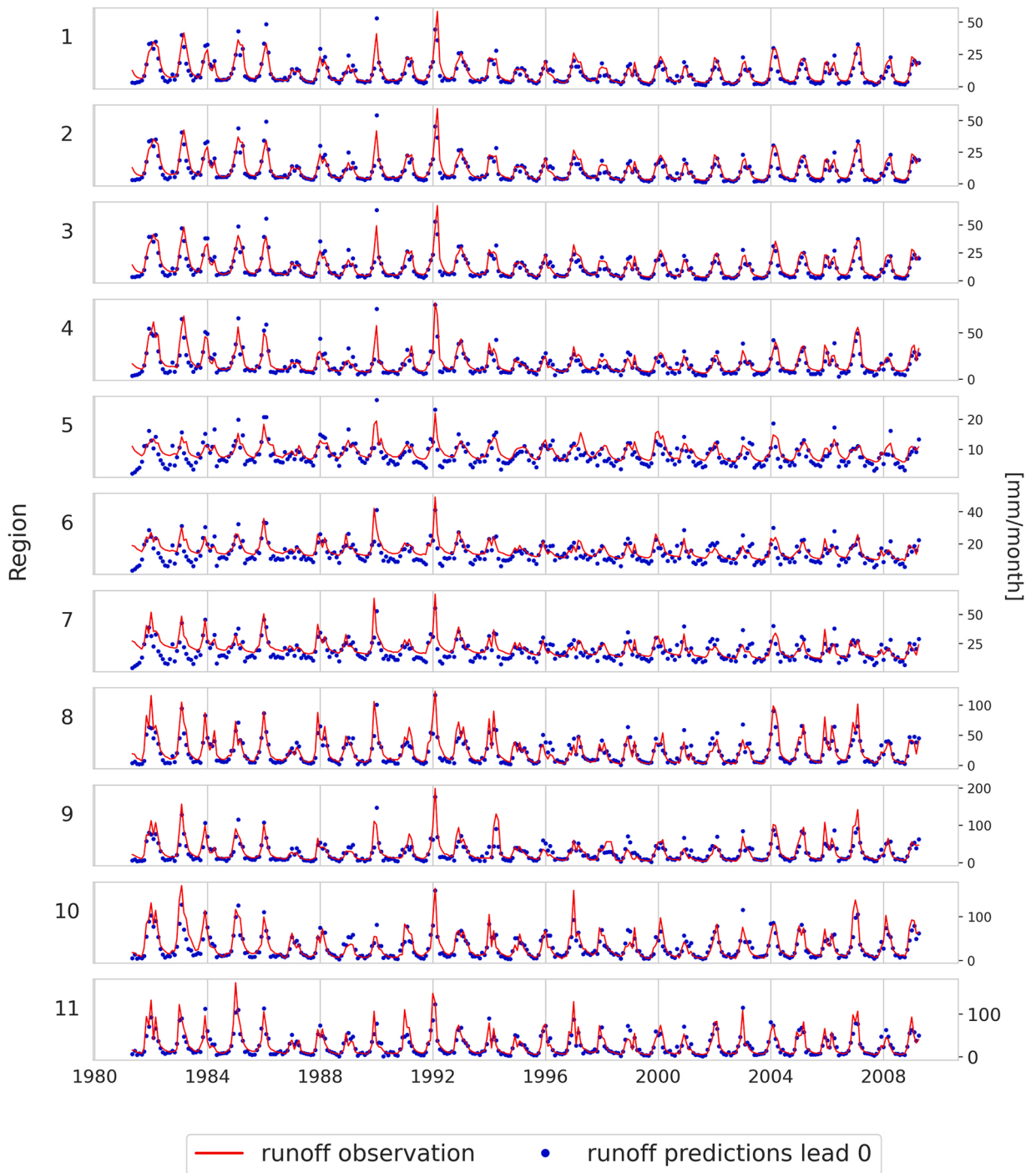
Fig. 3. Runoff reforecasts for lead 0 (blue scatter) and corresponding observations (solid red line) over the evaluation period 1981–2010 for all regions.

reference forecast). In both figures, an overall good agreement of the three skill scores for the runoff predictions and the SEAS5-BCSD precipitation predictions can be seen, with similar patterns over months, lead times and regions, and a general trend of decreasing skill with lead time. Additionally, for both the SEAS5-BCSD precipitation forecasts and the runoff forecasts, an improved skill during the first months of the rainy season (Nov, Dec, Jan) is evident compared to the last months of the rainy season (Feb, March). This monthly trend is more pronounced for the runoff forecasts, especially at higher lead times (lead 2 and 3). The forecast sharpness, accuracy and overall performance of both the runoff and SEAS5 BCSD forecasts decrease similarly for the first two lead times. The good IQRSS and
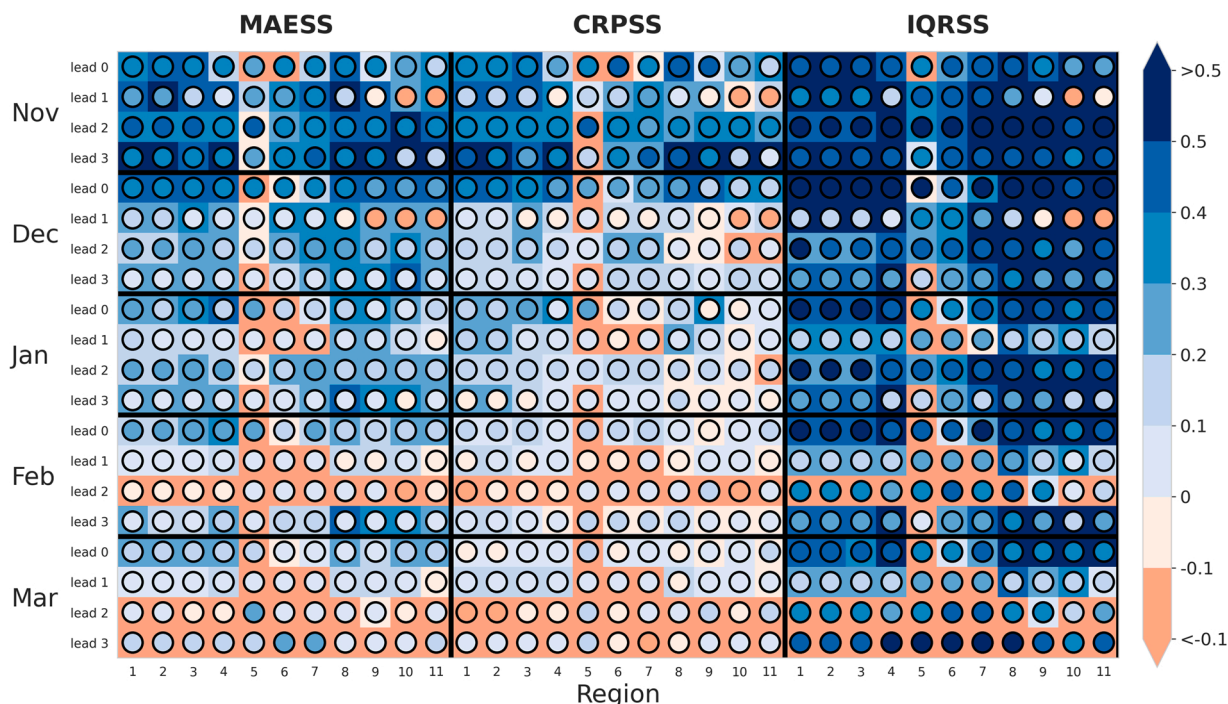
Fig. 4. Overview of the 3 skill scores (IQRSS, MAESS, CRPSS) of the runoff reforecasts (squares) and SEAS5-BCSD precipitation predictions (circles) for each region, month and lead time during the main rainy season (November - March). Blue and red colors describe negative and positive skill scores, respectively. Since scores aren't linear for negative values, all scores below the threshold of − 0.1 aren't shown.

MAESS and CRPSS values for these lead times imply that the predictions are very accurate, only slightly deviate from the reference and have a more representative distribution with respect to climatology.

However, for higher lead times, the SEAS5-BCSD predictions performance remains more uniform compared to the runoff predictions which decrease more linearly. In Fig. 4, where all three skill scores of the runoff forecasts are positive for most of the upper and lower basin, regions 5, 6 and 7 show poorer performance affecting all different lead times and months. This can also be seen in Fig. 5, where the MAESS and IQRSS of the runoff forecasts over the rainy season are negative for all lead times in region 5 and negative for regions 6–7 from lead time 2 on. This trend which isn't visible in the SEAS5-BCSD precipitation predictions skill scores is concordant with the previously mentioned erratic behavior in the runoff time series in those regions (Fig. 3). Interestingly, the lead 2 forecasts in February have a lower skill compared to the lead 3 forecast. This is also associated with a lower skill in the SEAS5-BCSD precipitation forecasts. This finding is counter-intuitive since a monotonically decreasing performance with lead time was expected. It has however been shown by several studies that central eastern Brazil can show low predictability on a seasonal scale during the austral summer as rainfall variability is related to complex land-atmosphere interactions that are not well captured by models (Marengo et al., 2003; Grimm et al., 2007).

To assess the ability of the EnKF framework to forecast events above or below the seasonal average, the BSS was calculated. To investigate which part of the skill is directly engendered by the SEAS5-BCSD precipitation forecast and which part is inherent to the EnKF framework, the BSS was also computed for SEAS5-BCSD precipitation forecast against the ERA5-Land product. Fig. 6 shows the BSS for both mentioned cases for lead 0 over the different regions and months of the rainy season. For both cases of "Above Normal" and "Below Normal", we can see positive BSS above 0.3, for most months and regions, even for the regions of the Western Bahia state. In addition, Fig. 7 shows the BSS for each month averaged over the whole rainy season and for the first four lead times. In both cases, the upper and lower tercile show better skill for lead 0 compared to higher lead times, with scores close to zero. Furthermore, lead 0 has higher skill scores in the lower basin compared to the upper basin. The features visible here show that the precipitation events in the above and below normal categories of the SEAS5-BCSD precipitation forecast match that of the predicted runoff, with higher scores in lead 0 compared to higher leads.

The skill score analysis illustrates how this framework is able to exploit the simple "rainfall-runoff" relationship, which in combination with the quality of the SEAS5-BCSD, provides skillful runoff predictions up to two months ahead. In most cases, the SEAS5-BCSD precipitation forecasts outperform or are close to the runoff forecasts, but in some cases the SEAS5-BCSD precipitation forecasts also outperform the runoff forecasts.

Greuell et al. (2019), Crochemore et al. (2016) and Yossef et al. (2013) have shown that in European catchments, most frameworks using process-based hydrological models and post-processed seasonal forecasts (ECMWF seasonal forecast Systems 3 (SEAS3) and 4 (SEAS4)) as input, hardly show skill beyond one lead month. Yossef et al. (2013) showed that in the tropical catchments of South America, the accuracy of runoff forecasts from the FEWS-World global seasonal runoff forecasting system, driven by SEAS3
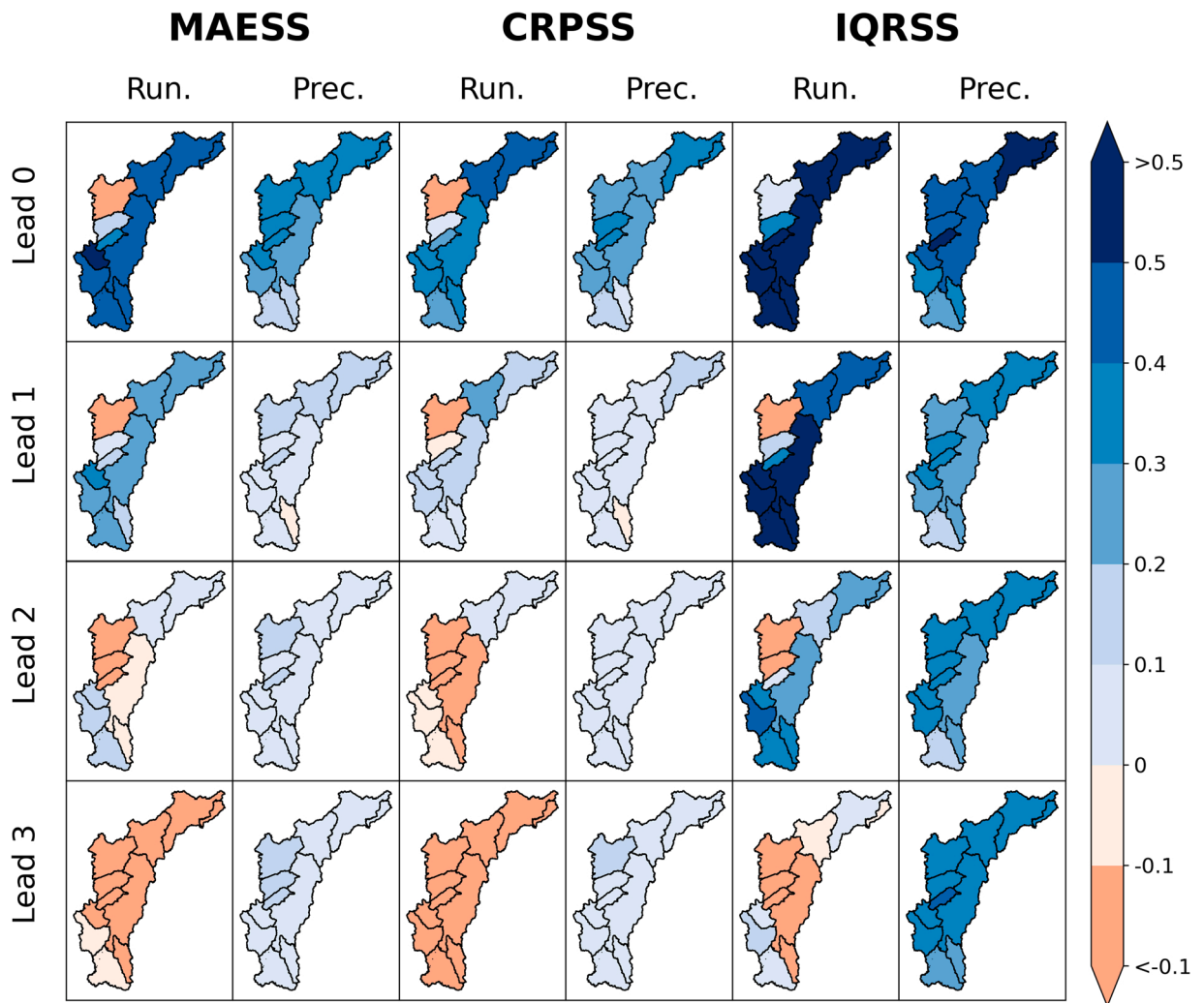
Fig. 5. Skill scores of runoff reforecasts (Run.) and SEAS5 BCSD precipitation predictions (Prec.) averaged during the main rainy season for the first four lead times over the 11 SFRB sub-catchments. Since scores aren't linear for negative values, all scores below the threshold of − 0.1 aren't shown.

precipitation forecasts, is limited to 1–2 months lead time during most of the year. The presented EnKF framework shows similar performance with skillful runoff predictions of 1–2 months ahead during the rainy season in most of the SF sub-basins. Consequently, the severe recurrent droughts in northeastern Brazil in 2011–2016 (Cunha et al., 2019; Marengo et al., 2017; Erfanian et al., 2017) could have been anticipated 1–2 months in advance, corresponding to sub-basins 1, 2 and 3. However, a direct comparison is inherently delicate as the forecast performance can depend on many factors, e.g., the hydrological model employed, the seasonal forecasts product, the post-processing techniques applied, the different geographical locations or the different regional and seasonal weather regimes affecting the study area (ENSO, MJO, North Atlantic Oscillation (NAO).

Finally, this framework does not require prior knowledge of local hydrological initial conditions generally necessary to run a model (soil moisture, snowpack, groundwater storage, regulation of lakes and dams), which makes this approach very suitable for regions where hydrological information is lacking. Furthermore, it is a transferable approach that can complement other types of statistical or more physically based models to easily monitor a single catchment at sub-basin level. The number of input variables such as evapotranspiration or temperature is also expandable, allowing for more robust correlations between all variables and thus improving prediction quality. As shown in Lorenz et al. (2015), this approach for global sub-basin forecasts can be extended to other regions, as SEAS5-BCSD forecasts for other regions already exist (Lorenz et al., 2021).

5. Summary and conclusion

In this study, a hydro-meteorological Ensemble Kalman Filter (EnKF) based data-assimilation framework is developed, with which sub-basin-scale runoff time-series are predicted. A least-squares prediction method has been applied for the prediction scheme in order to exploit the spatio-temporal relationship between precipitation and runoff. Besides reference runoff (ANA) and precipitation (ERA5-
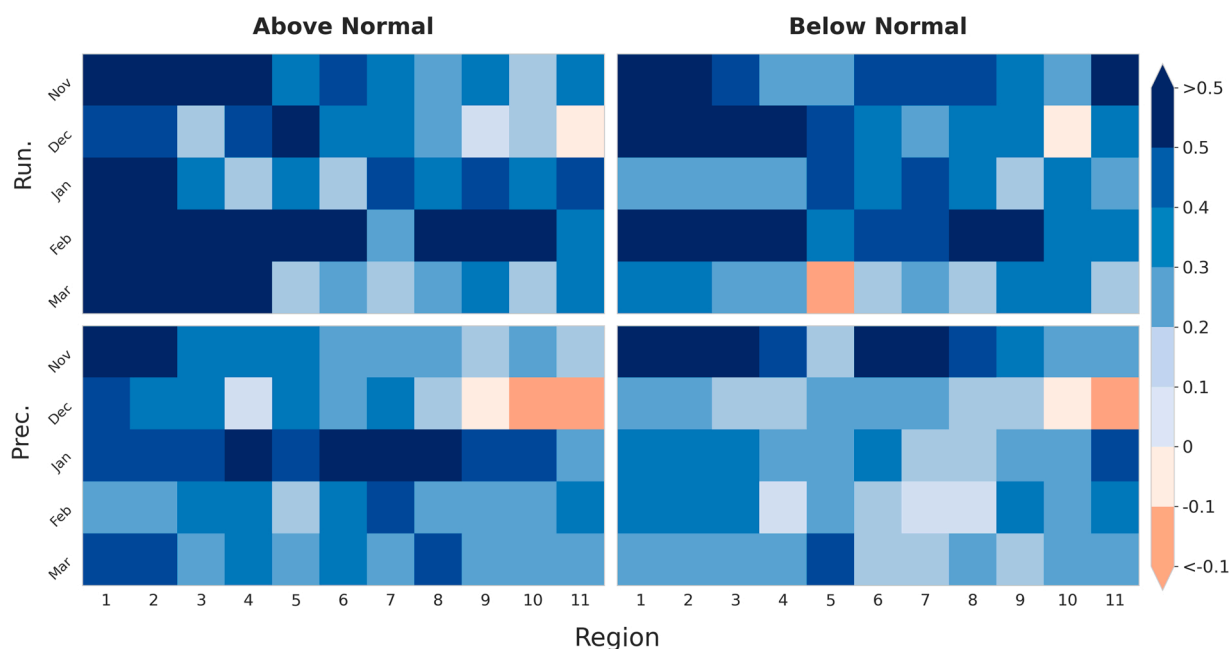
Fig. 6. Brier Skill Score (BSS) for runoff reforecasts against runoff observations (Run.) and SEAS5 BCSD precipitation predictions against ERA5-Land (Prec.). The BSS is computed for lead 0 over the different regions and months of the rainy season for Below Normal and Above Normal conditions. Since scores aren't linear for negative values, all scores below the threshold of − 0.1 aren't shown.

Land), bias-corrected and spatially disaggregated seasonal precipitation forecasts (SEAS5-BCSD) are used in order to predict runoff several months ahead. The performance of the framework was assessed for 11 sub-basins of the Rio São Francisco basin by comparing predicted against a reference leave-one-year-out climatology-based ensemble forecast. Overall, the EnKF runoff predictions have a good performance over most months with a general decrease of the skill with lead time. The positive performance in the runoff predictions for all skill scores demonstrated the ability of the assimilation framework to capture the concurrent precipitation-runoff relationship on sub-basin scale and to provide skillful estimates of runoff up to 2 months ahead.

This framework, however, also suffers from a number of limitations: (1) The skill of the SEAS5-BCSD predictions is the primary constraining factor for the performance of runoff forecasts, as these are used as input to the model. (2) The simplicity of the least-squares prediction method is limited by the insufficient knowledge of the model physics, the catchment characteristics and the local hydroclimatic conditions. For example, our current model cannot describe regions where runoff is disturbed by anthropogenic activities such as in the western state of Bahia (regions 5, 6 and 7). On the other hand, the framework is very suitable for regions where the key parameters for hydrological modeling as well as the initial and boundary conditions are missing. (3) The set of input variables, which can be extended to further variables in the SEAS5-BCSD dataset, such as evapotranspiration or temperature, could provide additional correlation statistics that have the potential to further improve the skill of the runoff predictions. Future work could include comparisons and implementation of this methodology to process-based hydrological models with more input variables and a wider range of regions.

Finally, this approach, which relies on global, freely available information for regional runoff forecasting, has great potential for regional decision support as well as seasonal water resource management and can complement other existing operational hydrological models. This includes the development of institutional systems and national strategies for drought preparedness plans, long-term flood awareness and sustainable reservoir management with a computationally cost-effective and easily transferable approach.

CRediT authorship contribution statement

Maurus Borne: Conceptualization, Methodology, Software, Investigation, Visualization, Formal analysis, Writing - Original Draft. **Christof Lorenz**: Conceptualization, Methodology, Software, Data Curation, Writing - Original Draft, Supervision, Project administration, Funding acquisition. **Tanja C. Portele**: Methodology, Formal analysis, Investigation, Writing - Original Draft. **Eduardo Sávio P. R. Martins**: Conceptualization, Methodology, Writing - Review & Editing. **Francisco das Chagas Vasconcelos Junior**: Conceptualization, Methodology, Writing - Review & Editing. **Harald Kunstmann**: Conceptualization, Methodology, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to
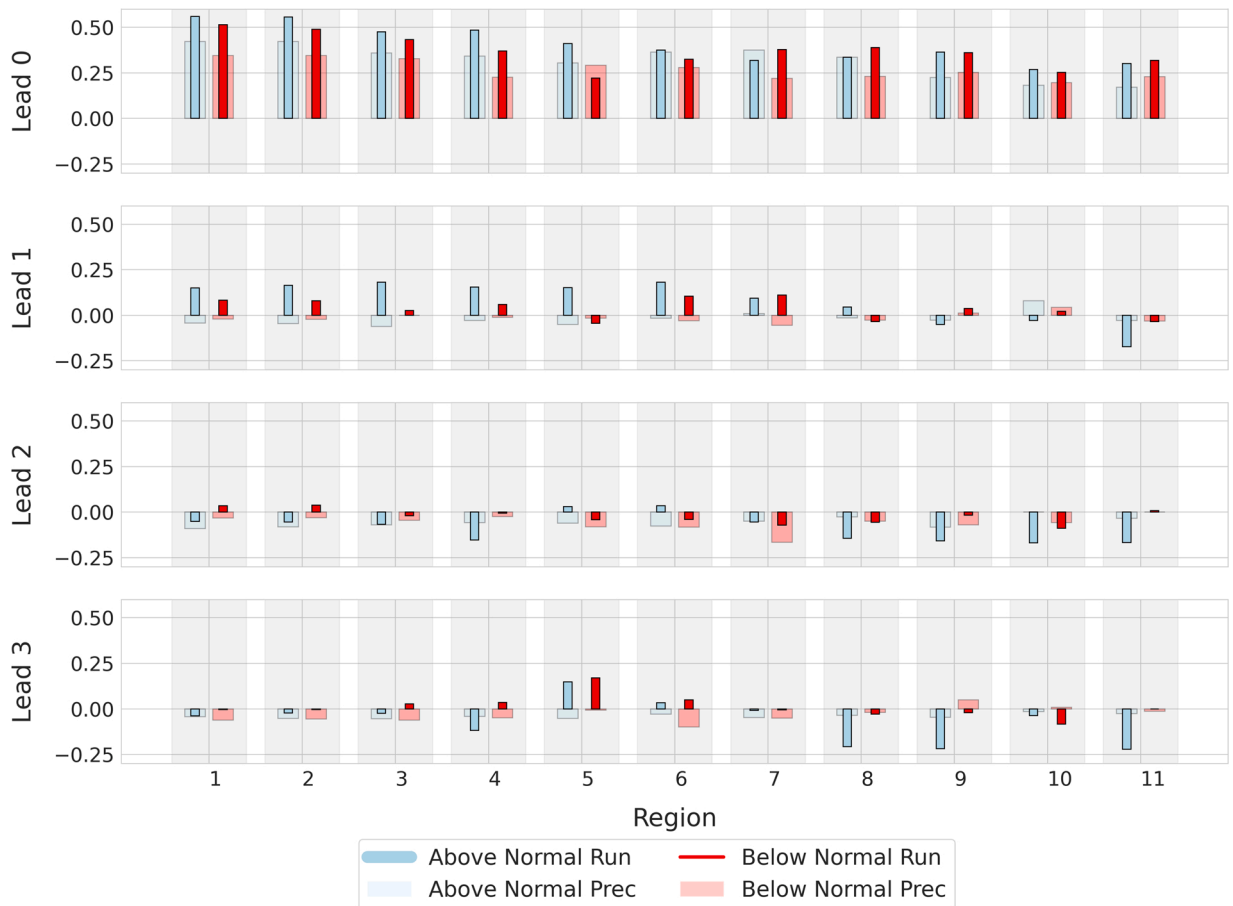
Fig. 7. Brier Skill Score (BSS) for runoff reforecasts against runoff observations (Run.) and SEAS5 BCSD precipitation predictions against ERA5-Land (Prec.). The BSS was computed for the upper and lower terciles for the first four lead times, for each month and averaged over the full rainy season.

influence the work reported in this paper.

Acknowledgments

Appendix A. Score metrics

The forecasts are evaluated by comparing the performance of a given forecast against a reference forecast. In general, the skill score of a given region, month and lead time is expressed as

$$Skillscore = 1 - \frac{Skill^f}{Skill^r} \tag{A.1}$$

Skill scores are obtained by the median of the individual scores and range from $-\infty$ to $+1$ with scores larger (smaller) than zero indicating that the forecast system is more (less) skillful than the reference forecast.

A.1. Inter-quantile range (IQR)

Sharpness is an intrinsic property of the forecast and indicates the spread of the ensemble forecast. When the ensemble spread is too large, it indicates uncertain predictions. When the ensemble spread is small, the ensemble members are consistent and the forecast is more confident. In this study, sharpness is evaluated as the difference between the 95th and 5th percentiles of the forecast distribution defined as the Inter-Quantile Range (IQR).

A.2. Mean absolute error (MAE)

The accuracy of the forecasts is assessed with the mean absolute error (MAE). The MAE computes the absolute difference between the forecast ensemble mean and the observed value and is expressed as

$$MAE = \frac{1}{D} \sum_{d=1}^{D} \left| x_{f,d} - x_o \right| \tag{A.2}$$

where x_f is the ensemble forecast of size D and x_o denotes the associated observation. Smaller MAE values correspond to more accurate forecasts.

A.3. Brier score (BS)

The Brier Score (BS) is generally used for scoring binary forecasts and is defined as the mean squared error of the forecast probability and a observed category:

$$BS = \frac{1}{T} \sum_{t=1}^{T} (x_{f,t} - x_{o,t})^2 \tag{A.3}$$

where $x_{o,t}$ is the actual observed category at instance t, and $x_{f,t}$ is the categorical forecast probability of that same event. The BS indicates how a forecast can estimate the occurrence of an observed category. In our study, we evaluated the skill of tercile forecasts, i.e., of forecasts for above and below normal conditions that were defined from the ensemble mean being above the 0.66 and below the 0.33 reference percentile, respectively.

A.4. Continuous ranked probability score (CRPS)

The overall performance of the forecast is evaluated using the continuous ranked probability score (CRPS) (Hersbach, 2000). The CRPS measures the difference between the predicted and the observed cumulative distributions functions (CDF) of a given variable of interest x. The CDF of the observation $H(x)$ is represented by a delta function at the observed value x_o and is defined as the Heaviside function.

$$CRPS = \int_{-\infty}^{+\infty} [F(x) - H(x - x_o)]^2 dx \tag{A.4}$$

The CRPS is a probabilistic generalization of the MAE, since it incorporates the uncertainty of the forecasts.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ejrh.2022.101146.

References

Ajami, H., Khan, U., Tuteja, N.K., Sharma, A., 2016. Development of a computationally efficient semi-distributed hydrologic modeling application for soil moisture, lateral flow and runoff simulation. Environ. Model. Softw. 85, 319–331.

Carvalho, L.M., Jones, C., Liebmann, B., 2002. Extreme precipitation events in southeastern south america and large-scale convective patterns in the south atlantic convergence zone. J. Clim. 15, 2377–2394.

Carvalho, L.M., Jones, C., Liebmann, B., 2004. The south atlantic convergence zone: Intensity, form, persistence, and relationships with intraseasonal to interannual activity and extreme rainfall. J. Clim. 17, 88–108.

Changnon Jr., S.A., Vonnhame, D.R., 1986. Use of climate predictions to decide a water management problem 1. J. Am. Water Resour. Assoc. 22, 649–652.

Crochemore, L., Ramos, M.H., Pappenberger, F., 2016. Bias correcting precipitation forecasts to improve the skill of seasonal streamflow forecasts. Hydrol. Earth Syst. Sci. 20, 3601–3618.

Cunha, A., Alvalá, R.C., Nobre, C.A., Carvalho, M.A., 2015. Monitoring vegetative drought dynamics in the brazilian semiarid region. Agric. For. Meteorol. 214, 494–505.

Cunha, A.P., Zeri, M., DeusdaráLeal, K., Costa, L., Cuartas, L.A., Marengo, J.A., Tomasella, J., Vieira, R.M., Barbosa, A.A., Cunningham, C., et al., 2019. Extreme drought events over brazil from 2011 to 2019. Atmosphere 10, 642.

Cunha, A.P.M., Tomasella, J., Ribeiro-Neto, G.G., Brown, M., Garcia, S.R., Brito, S.B., Carvalho, M.A., 2018. Changes in the spatial-temporal patterns of droughts in the Brazilian Northeast. Atmos. Sci. Lett. 19, e855.

De Paiva, L.F.G., Montenegro, S.M., Cataldi, M., 2020. Prediction of monthly flows for três marias reservoir (são francisco river basin) using the cfs climate forecast model. RBRH 25.

De Souza, E.B., Ambrizzi, T., 2006. Modulation of the intraseasonal rainfall over tropical brazil by the madden-julian oscillation. Int. J. Climatol.: A J. R. Meteorol. Soc. 26, 1759–1776.

dos Santos, A., Lopes, P.M.O., da Silva, M.V., Jardim, A.M.d.R.F., de Albuquerque Moura, G.B., Fernandes, G.S.T., de Oliveira Silva, D.A., da Silva, J.L.B., de Moraes Rodrigues, J.A., Silva, E.A., 2020. Causes and consequences of seasonal changes in the water flow of the são francisco river in the semiarid of brazil. Environ. Sustain. Indic. 8, 100084.

Du, L., Mikle, N., Zou, Z., Huang, Y., Shi, Z., Jiang, L., McCarthy, H.R., Liang, J., Luo, Y., 2018. Global patterns of extreme drought-induced loss in land primary production: identifying ecological extremes from rain-use efficiency. Sci. Total Environ. 628, 611–620.

Erfanian, A., Wang, G., Fomenko, L., 2017. Unprecedented drought over tropical south america in 2016: significantly under-predicted by tropical sst. Sci. Rep. 7, 1–11.

Folland, C.K., Colman, A.W., Rowell, D.P., Davey, M.K., 2001. Predictability of northeast brazil rainfall and real-time forecast skill, 1987-98. J. Clim. 14, 1937–1958.

Foster, K., Bertacchi Uvo, C., Olsson, J., 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, 2953–2970.

Garbrecht, J.D., Schneider, J.M., Van Liew, M.W., 2006. Monthly runoff predictions based on rainfall forecasts in a small oklahoma watershed. J. Am. Water Resour. Assoc. 42, 1285–1295.

Greuell, W., Franssen, W.H., Hutjes, R.W., 2019. Seasonal hydro-meteorological forecasts for europe: sources of skill. Hydrol. Earth Syst. Sci. 23, 1–21.

Grimm, A.M., Pal, J.S., Giorgi, F., 2007. Connection between spring conditions and peak summer monsoon rainfall in south america: role of soil moisture, surface temperature, and topography in eastern brazil. J. Clim. 20, 5929–5945.

He, J., Deser, C., Soden, B.J., 2017. Atmospheric and oceanic origins of tropical precipitation variability. J. Clim. 30, 3197–3217.

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

Hersbach, H., 2016. The era5 atmospheric reanalysis. In: Proceedings of the AGU fall meeting abstracts.

Ioris, A.A., 2001. Water resources development in the sáto francisco river basin (brazil): conflicts and management perspectives. Water Int. 26, 24–39.

Johnson, S.J., Stockdale, T.N., Ferranti, L., Balmaseda, M.A., Molteni, F., Magnusson, L., Tietsche, S., Decremer, D., Weisheimer, A., Balsamo, G., et al., 2019. Seas5: the new ecmwf seasonal forecast system. Geosci. Model Dev. 12.

de Jong, P., Tanajura, C.A.S., Sánchez, A.S., Dargaville, R., Kiperstok, A., Torres, E.A., 2018. Hydroelectric production from brazil's são francisco river could cease due to climate change and inter-annual variability. Sci. Total Environ. 634, 1540–1553.

Kayano, M.T., Andreoli, R.V., 2004. Decadal variability of northern northeast brazil rainfall and its relation to tropical sea surface temperature and global sea level pressure anomalies. J. Geophys. Res.: Oceans 109.

Kodama, Y., 1992. Large-scale common features of subtropical precipitation zones (the baiu frontal zone, the spcz, and the sacz) part i: Characteristics of subtropical frontal zones. J. Meteorol. Soc. Jpn. Ser. II 70, 813–836.

Kodama, Y.M., 1993. Large-scale common features of sub-tropical convergence zones (the baiu frontal zone, the spcz, and the sacz) part ii: conditions of the circulations for generating the stczs. J. Meteorol. Soc. Jpn. Ser. II 71, 581–610.

Koster, R.D., Suarez, M.J., Heiser, M., 2000. Variance and predictability of precipitation at seasonal-to-interannual timescales. J. Hydrometeorol. 1, 26–46.

Kulikova, I., Fedorova, N., Levit, V., Cordeiro, E., 2014. Sea surface temperature anomaly and precipitation distribution in the Alagoas State of the Brazilian Northeast. Nat. Sci. 6, 1159.

Kurtenbach, E., Eicker, A., Mayer-Gürr, T., Holschneider, M., Hayn, M., Fuhrmann, M., Kusche, J., 2012. Improved daily GRACE gravity field solutions using a Kalman smoother. J. Geodyn. 59, 39–48.

Li, Y., Gupta, A.S., Taschetto, A.S., Jourdain, N.C., DiLuca, A., Done, J.M., Luo, J.J., 2020. Assessing the role of the ocean-atmosphere coupling frequency in the western maritime continent rainfall. Clim. Dyn. 54, 4935–4952.

Lorenz, C., Kunstmann, H., 2012. The hydrological cycle in three state-of-the-art reanalyses: Intercomparison and performance analysis. J. Hydrometeorol. 13, 1397–1420.

Lorenz, C., Tourian, M.J., Devaraju, B., Sneeuw, N., Kunstmann, H., 2015. Basin-scale runoff prediction: an ensemble Kalman filter framework based on global hydrometeorological data sets. Water Resour. Res. 51, 8450–8475.

Lorenz, C., Portele, T.C., Laux, P., Kunstmann, H., 2021. Bias-corrected and spatially disaggregated seasonal forecasts: a long-term reference forecast product for the water sector in semi-arid regions. Earth Syst. Sci. Data 13, 2701–2722.

Maneta, M., Torres, M., Wallender, W.W., Vosti, S., Kirby, M., Bassoi, L.H., Rodrigues, L., 2009. Water demand and flows in the são francisco river basin (brazil) with increased irrigation. Agric. Water Manag. 96, 1191–1200.

Marengo, J.A., Cavalcanti, I., Satyamurty, P., Trosnikov, I., Nobre, C., Bonatti, J., Camargo, H., Sampaio, G., Sanches, M., Manzi, A., 2003. Assessment of regional seasonal rainfall predictability using the cptec/cola atmospheric gcm. Clim. Dyn. 21, 459–475.

Marengo, J.A., Alves, L.M., Alvala, R., Cunha, A.P., Brito, S., Moraes, O.L., 2017. Climatic characteristics of the 2010-2016 drought in the semiarid northeast brazil region. An. da Acad. Bras. De. Ciências 90, 1973–1985.

Marengo, J.A., Alves, L.M., Alvala, R., Cunha, A.P., Brito, S., Moraes, O.L., 2018. Climatic characteristics of the 2010-2016 drought in the semiarid northeast brazil region. An. da Acad. Bras. De. Ciências 90, 1973–1985.

Martins, E.S., Coelho, C.A., Haarsma, R., Otto, F.E., King, A.D., Jan van Oldenborgh, G., Kew, S., Philip, S., Vasconcelos Júnior, F.C., Cullen, H., 2018. A multimethod attribution analysis of the prolonged northeast brazil hydrometeorological drought (2012-16). Bull. Am. Meteorol. Soc. 99, S65–S69.

Meißner, D., Klein, B., Ionita, M., 2017. Development of a monthly to seasonal forecast framework tailored to inland waterway transport in central europe. Hydrol. Earth Syst. Sci. 21, 6401–6423.

Misra, V., 2006. Understanding the predictability of seasonal precipitation over northeast brazil. Tellus A: Dyn. Meteorol. Oceanogr. 58, 307–319.

Misumi, R., Bell, V., Moore, R., 2001. River flow forecasting using a rainfall disaggregation model incorporating small-scale topographic effects. Meteorol. App. A J. Forecast., Pract. Appl., Train. Tech. Model. 8, 297–305.

Moritz, H., 1980. Advanced physical geodesy. Adv. Planet. Geol.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., et al., 2021. Era5-land: a state-of-the-art global reanalysis dataset for land applications. Earth Syst. Sci. Data Discuss. 1–50.

ParedesTrejo, F., Brito-Castillo, L., BarbosaAlves, H., Guevara, E., 2016. Main features of large-scale oceanic-atmospheric circulation related to strongest droughts during rainy season in Brazilian São Francisco River Basin. Int. J. Climatol. 36, 4102–4117.

Portele, T.C., Lorenz, C., Dibrani, B., Laux, P., Bliefernicht, J., Kunstmann, H., 2021. Seasonal forecasts offer economic benefit for hydrological decision making in semi-arid regions. Sci. Rep. 11, 1–15.

Pousa, R., Costa, M.H., Pimenta, F.M., Fontes, V.C., Brito, V.F.A.d., Castro, M., 2019. Climate change and intense irrigation growth in western bahia, brazil: the urgent need for hydroclimatic monitoring. Water 11, 933.

Santos, A.B., HeilCosta, M., Chartuni Mantovani, E., Boninsenha, I., Castro, M., 2020. A remote sensing diagnosis of water use and water stress in a region with intense irrigation growth in brazil. Remote Sens. 12, 3725.

Schewe, J., Heinke, J., Gerten, D., Haddeland, I., Arnell, N.W., Clark, D.B., Dankers, R., Eisner, S., Fekete, B.M., Colón-González, F.J., et al., 2014. Multimodel assessment of water scarcity under climate change. Proc. Natl. Acad. Sci. USA 111, 3245–3250.

Sehgal, V., Sridhar, V., 2019. Watershed-scale retrospective drought analysis and seasonal forecasting using multi-layer, high-resolution simulated soil moisture for southeastern us. Weather Clim. Extrem. 23, 100191.

Stocker, T.F., Qin, D., Plattner, G.K., Tignor, M.M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M., 2014.Climate change 2013: The physical science basis. contribution of working group i to the fifth assessment report of ipcc the intergovernmental panel on Climate Change.

Sun, T., Ferreira, V., He, X., Andam-Akorful, S., 2016. Water availability of São Francisco river basin based on a space-borne geodetic sensor. Water 8, 213.

Taschetto, A.S., Ambrizzi, T., 2012. Can indian ocean sst anomalies influence south american rainfall? Clim. Dyn. 38, 1615–1628.

Taschetto, A.S., Wainer, I., 2008. The impact of the subtropical south atlantic sst on south american precipitation. Ann. Geophys., Copernic. GmbH 3457–3476.

Tomasella, J., Vieira, R.M.S.P., Barbosa, A.A., Rodriguez, D.A., de Oliveira Santana, M., Sestini, M.F., 2018. Desertification trends in the northeast of brazil over the period 2000-2016. Int. J. Appl. Earth Obs. Geoinf. 73, 197–206.

Tourian, M.J., 2013. Application of spaceborne geodetic sensors for hydrology, PhD thesis, Institute of Geodesy, Univ. Stuttgart, Stuttgart, Germany.

Traini, C., Schrottke, K., Stattegger, K., Dominguez, J.L., Guimaraes, J.K., Vital, H., Beserra, D., daSilva, A.A., 2012. Morphology of subaqueous dunes at the mouth of the dammed river são francisco (brazil). J. Coast. Res. 28, 1580–1590.

do Vasco, A.N., de Oliveira Aguiar Netto, A., da Silva, M.G., 2019. The influence of dams on ecohydrological conditions in the são francisco river basin, brazil. Ecohydrol. Hydrobiol. 19, 556–565.

Wanders, N., Thober, S., Kumar, R., Pan, M., Sheffield, J., Samaniego, L., Wood, E.F., 2019. Development and evaluation of a pan-european multimodel seasonal hydrological forecasting system. J. Hydrometeorol. 20, 99–115.

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

Yuan, X., Ma, F., Wang, L., Zheng, Z., Ma, Z., Ye, A., Peng, S., 2016. An experimental seasonal hydrological forecasting system over the yellow river basin-part 1: understanding the role of initial hydrological conditions. Hydrol. Earth Syst. Sci. 20, 2437–2451.