https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

Hard to Beat, Not Impossible: Improving Seasonal Ensemble Streamflow Predictions (ESP) in Mountainous Catchments

Jerónimo Sota[1,2], Pablo A. Mendoza[1,2] and Miguel Lagos-Zuñiga[3]

[1]Department of Civil Engineering, Universidad de Chile, Santiago, Chile
[2]Advanced Mining Technology Center (AMTC), Universidad de Chile, Santiago, Chile
[3]Departamento de Obras Civiles, Universidad Técnica Federico Santa María, Santiago, Chile.

Correspondence to: Pablo A. Mendoza (pamendoz@uchile.cl)

Abstract. The Ensemble Streamflow Prediction (ESP) method has been extensively applied in operational forecasting and is widely considered a robust benchmark for predicting spring–summer snowmelt runoff in mountain environments. Nevertheless, it remains unclear how, when and where the implementation of data assimilation and/or post-processing techniques offer potential for improving the quality of seasonal and monthly streamflow forecasts during the snowmelt season. Here, we compare alternative ESP hindcast configurations – implemented with the conceptual rainfall-runoff model HBV.IANIGLA – across 21 Andean catchments spanning a pronounced hydroclimatic gradient (28-37° S), focusing on spring-summer (i.e., September-March) streamflow hindcasts produced for five initialization times over a 38-year period (September/1982–March/2020). Data assimilation experiments include the ensemble Kalman Filter (EnKF) and the Sequential Importance Resampling Particle Filter (SIR-PF), and the suite of post-processing techniques includes Quantile Mapping, Linear Regression, Trace Weighting and Random Forest. For completeness, the alternative ESP configurations are also compared against a simple statistical model using initial hydrologic conditions (IHCs) – computed as the sum of simulated soil and snow water storages – as the only predictand. Results show that, at the seasonal scale, the advantage of SIR-PF over EnKF becomes evident only for hindcasts initialized on August 1 or later, while post-processing can substantially degrade skill at lead times of two months or longer. However, combining data assimilation and post-processing for September 1 initializations improves skill and reliability of raw ESP monthly hindcasts, particularly at the beginning and end of the snowmelt season. The simple IHC-based statistical method also improves reliability at both seasonal and monthly scales, despite mixed results for other metrics. Finally, catchment attributes help explain where improvements are most likely: seasonal gains are favored in semi-arid basins with lower precipitation and runoff, higher aridity, and poorer hydrologic model performance, whereas monthly gains show distinct geographic and hydroclimatic patterns.

1 Introduction

Seasonal streamflow forecasts are essential for water resources planning in mountain regions, where runoff generated during the snowmelt season sustains irrigation, hydropower production, ecosystem flows, and urban water supply during subsequent dry periods. In snow-influenced basins, spring-summer runoff predictability depends strongly on the water stored in the land

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

surface – typically referred to as initial hydrologic conditions (IHCs) –, particularly snow water equivalent (e.g., Regonda et al., 2006; Mendoza et al., 2014; Arnal et al., 2024), soil moisture (e.g., Greuell & Hutjes, 2023), and antecedent streamflow (e.g., Sveinsson et al., 2008; Baker et al., 2021). This hydrological memory can provide useful forecast skill several weeks to months in advance, though its contribution varies with forecast initialization time and catchment characteristics (e.g., Araya et al., 2023). Understanding how to exploit predictability from IHCs remains especially important in mountainous environments, where streamflow generation is shaped by strong elevation gradients, seasonal snow accumulation and melt, sparse observations and, in some cases, large interannual hydroclimatic variability (e.g., Hernandez et al., 2022).

The ensemble streamflow prediction (ESP; Twedt et al., 1977; Day, 1985) method has long been a cornerstone of operational seasonal water supply forecasting. ESP initializes a process-based hydrological model with current IHCs and then forces model simulations with historically observed meteorological sequences, preserving observed climate variability while if all the streamflow forecast uncertainty is due to the spread of future seasonal meteorological forcings. This approach is attractive because it is simple, computationally efficient, and skillful when initial hydrologic conditions dominate predictability. Indeed, numerous studies have shown that ESP can be a difficult benchmark to outperform, particularly in snowmelt-dominated and highly seasonal basins where snowpack and antecedent moisture exert strong control on runoff volumes (e.g., Wood et al., 2005; Mo et al., 2012; Arnal et al., 2018; Wanders et al., 2019; Pechlivanidis et al., 2020; Peñuela et al., 2020; Petry et al., 2023). However, ESP also has well-known limitations: it does not directly use seasonal climate forecast information, it assumes perfect IHCs, and its ensemble distribution may be biased or suffer from insufficient spread.

One pathway for improving ESP forecasts is data assimilation (DA). Because snowpack and soil moisture are major sources of seasonal predictability in mountain environments (Mendoza et al., 2017), assimilating streamflow, in-situ or remotely sensed snow observations, or other land surface variables can reduce errors in hydrological model estimates before forecast initialization. DeChant & Moradkhani (2011) demonstrated the potential of assimilating in situ snow water equivalent (SWE) observations with the particle filter (PF) for improving the reliability of seasonal ESP forecasts across 15 subcatchments in the Upper Colorado River Basin. DeChant & Moradkhani (2014) later combined ESP with a PF that assimilated land surface temperature and passive microwave brightness temperature to quantify initial-condition uncertainty, together with Sequential Bayesian Combination to quantify model errors. Overall, they obtained more substantial improvements in the tails of the forecast distributions. Huang et al. (2017) tested the effects of assimilating SWE with an ensemble Kalman filter (Evensen, 1994) on the quality of seasonal (April-July) ESP streamflow forecasts in nine basins located in the western United States, concluding that most EnKF implementations improved forecast performance. More recent studies have further demonstrated the value of snow and/or streamflow assimilation for water supply forecasting in mountainous catchments (Micheletty et al., 2021; Metref et al., 2023; Uysal et al., 2026).

A second pathway for improving ESP and dynamical forecasts in general is statistical post-processing (PP), which may be viewed as a hybrid statistical-dynamical forecasting approach (Wood et al., 2018; Du & Pechlivanidis, 2025). Statistical post-processing acts on forecast outputs to correct biases, adjust ensemble spread, increase probabilistic skill, and improve forecast attributes such as reliability and discrimination. Applications in seasonal streamflow forecasting include linear regression

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

(Wood & Schaake, 2008), bias correction through scaling factors (e.g., Mendoza et al., 2017), quantile mapping (e.g., Lucatero et al., 2018) and copula-based methods (e.g., Madadgar et al., 2014). An interesting strategy is to weight ESP ensemble members – also known as "trace weighting" or post-ESP – based on the similarity of climate information between the target forecast year and the meteorological year used to generate each ESP trace (e.g., Werner et al., 2004; Najafi et al., 2012; Bradley et al., 2015; Mendoza et al., 2017; Baker et al., 2021). Although these methods have considerable potential and are relatively easy to implement, their application remains constrained by limited sample sizes for training, assumptions of stationarity, and trade-offs among forecast quality attributes when methods are optimized for different objectives (Wood et al., 2018).

Statistical (i.e., data-driven) methods offer a complementary perspective to the seasonal forecasting problem in mountainous domains. Simple linear regression using only IHCs (e.g., Mendoza et al., 2017), local regression techniques (e.g., Grantz et al., 2005; Mendoza et al., 2014), principal component regression using snow predictors (e.g., Arnal et al., 2024), and machine-learning models (e.g., Broxton et al., 2023) can produce competitive and often reliable seasonal forecasts. Additionally, statistical approaches are especially attractive where meteorological data are uncertain and/or large hydrological model errors persist. However, they may be less transferable under changing climate, land-cover, or water-management conditions, and they provide less process-level diagnostic information than dynamical ESP systems. For this reason, statistical models can be best viewed as useful complements and benchmarks for process-based forecasting rather than simple replacements.

Despite substantial progress, a remaining gap for mountainous basins is understanding when data assimilation, post-processing, or their combination improves ESP forecasts, and how such improvements compare with simple data-driven forecasting techniques. Recently, Tanguy et al. (2025) compared the effectiveness of bias correction using quantile mapping and data assimilation across a range of lead times within an operational ESP framework across 316 UK catchments. Specifically, they assessed the separate effects of updating model estimates with a particle filter (Pham, 2001), and bias-correcting ESP forecasts with quantile mapping (QM). Their results showed no universal superiority of either approach, with cases when the traditional ESP implementation outperformed the configurations implementing DA or QM. Further, they recommended the application of (i) DA at shorter lead times in summer, in catchments with high baseflow contributions, and in snow-affected catchments during winter and spring, and (ii) BC in all the remaining cases. Nevertheless, their modeling setup did not consider snow processes, and they did not assess the joint effects of DA and post-processing on ESP forecasts.

This study contributes to the existing literature by evaluating whether ESP forecasts for the snowmelt season can be improved through data assimilation and/or post-processing in mountainous basins. Specifically, we compare raw ESP with configurations using the ensemble Kalman filter, particle filtering, and four post-processing methods across 21 Andean catchments spanning a pronounced hydroclimatic gradient (28-37° S). We assess hindcast performance at seasonal and monthly scales and compare these dynamical approaches with a simple statistical model based on IHCs. By linking forecast improvements to catchment attributes, the study clarifies when (i.e., which initialization times), where (i.e., which types of basin), and how (i.e., which hindcast attribute improvements) ESP-based forecasts can be enhanced for spring-summer runoff prediction.

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

2 Study domain and data

We conduct our analyses in 21 Chilean basins located on the western slopes of the Andes Cordillera (28-37° S, 70-71° W). In this region, more than 60% of the country's population lives, underscoring the importance of accurately quantifying water resources availability. We use the same basins as in Araya et al. (2023) based on the following criteria: (i) a low (i.e., < 0.05) human intervention index, which is defined as the ratio between annual volume of water assigned for permanent consumptive uses and the observed mean annual runoff; (ii) absence of large reservoirs; (iii) snowmelt influence on runoff seasonality (i.e., they must have a snowmelt-driven, nivo-pluvial, or pluvio-nival regime, as described by Baez-Villanueva et al., 2021); (iv) at least 75% of days with streamflow observations during the period April/1980–March/2020; and (v) at least 20 water years (WYs) with seasonal (September–March) streamflow observations for hindcast verification, based on the CAMELS-CL dataset (Camila Alvarez-Garreton et al., 2018). The most restrictive conditions are (iv) and (v), which hinder the possibility of including additional mountainous catchments from CAMELS-CL.

We use daily time series of streamflow from stations maintained by the Chilean General Water Directorate (DGA, *Dirección General de Aguas*) and compiled in the CAMELS-CL dataset. Daily precipitation and mean air temperature, for the period 1980-2020, are derived from the gridded observational product CR2MET version 2.5 (Boisier, 2023), which provides time series of these variables for continental Chile at a 0.05° x 0.05° horizontal resolution.

Figures Figure **1** and Figure **2** illustrate the heterogeneity of the basins included in this study, with mean elevations (Figure 1a) ranging from 1605 to 4275 m a.s.l. and spanning a pronounced hydroclimatic gradient, as reflected in aridity indices – defined as the ratio between mean annual potential evapotranspiration (PET) and mean annual precipitation (P) – that range from 0.5 to 7. The variables in Figure 1 show a north-south transition from semi-arid, water-limited hydroclimates (with PET/P > 1) to energy-limited environments (with PET/P < 1; see Figures Figure **1**e and Figure **2**d), characterized by larger precipitation and runoff amounts.

Figure 2 shows the seasonality of precipitation, PET, and runoff, along with characteristic ratios and flow duration curves (FDCs). Precipitation over the study domain is concentrated between April and September (Figure 2a), with increasing amounts toward the south. PET peaks between November and March (late austral spring and summer; Figure 2b), while the maximum monthly runoff (Figure 2c) is reached between July and January, depending on the catchment's hydrological regime. Figure 2d further highlights the diversity of annual water and energy balances among the study basins, complementing the latitudinal gradients observed in Figure 1. Finally, the daily FDCs (Figure 2e) reveal distinct hydrological responses across basins, with variations in high and low flows, mid-segment slopes, medians, and other hydrological signatures.

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

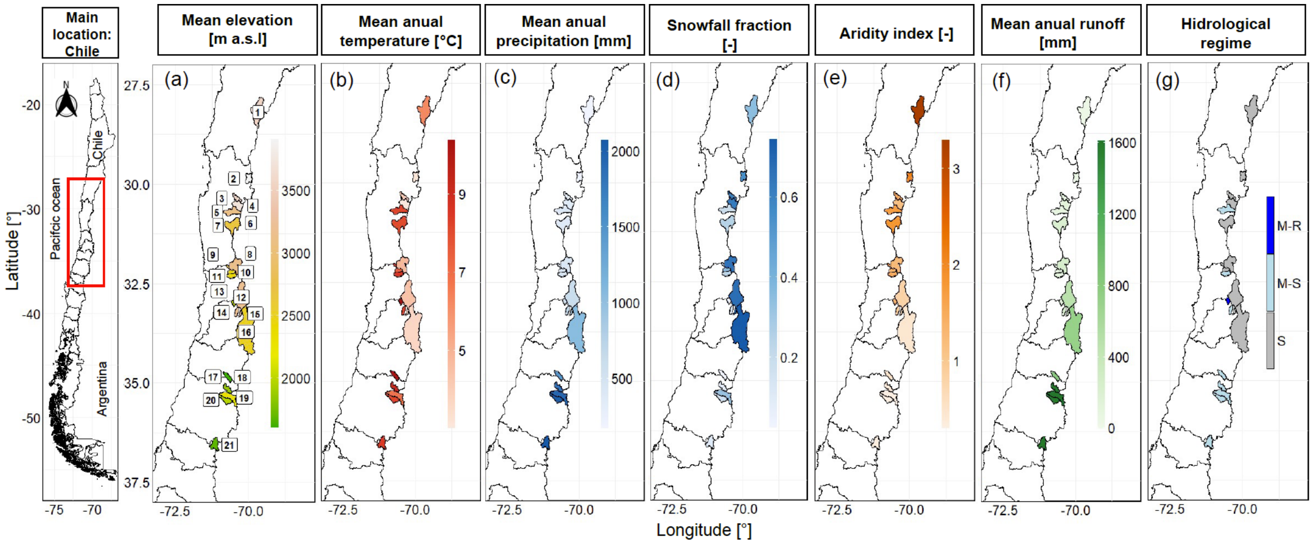


Figure 1. Basin's location and hydroclimatic gradient across the study domain: (a) mean elevation, (b) mean annual temperature, (c) mean annual precipitation, (d) fraction of precipitation falling as snow, (e) aridity index, (f) mean annual runoff, and (g) hydrological regime. Hydroclimatic attributes are computed for the period April 1980-March 2020 using data retrieved from the CAMELS-CL database (see details in Sect. 2). The numbers in panel (a) correspond to the basin IDs, and basin names are included in Figure 2.

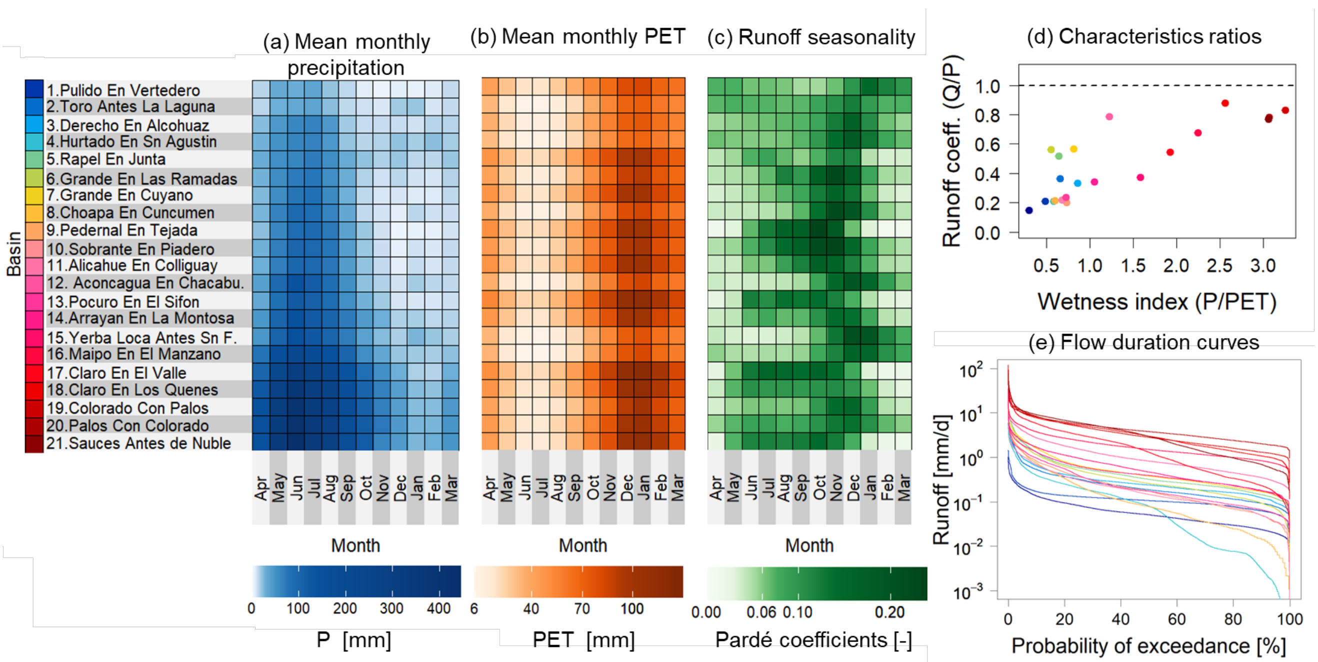


Figure 2. Catchments characteristics: (a) mean monthly precipitation, (b) mean monthly potential evapotranspiration (PET), (c) runoff seasonality, (d) runoff coefficients as a function of wetness index, and (e) daily flow duration curves (FDCs). In panels (a), (b) and (c) the catchments are listed from north to south. These graphs correspond to the period

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

April 1980-March 2020 and were produced using data retrieved from the CAMELS-CL database (see details in Sect. 2).

3 Methods

Figure 3 summarizes the methodology adopted in this study. The main steps are: (a) configuration and calibration of a hydrological model for the selected basins (Figure 3a); (b) generation of seasonal spring–summer (September–March) hindcasts for a 38-year period (September/1982–March/2020) using five initialization dates (May 1, June 1, July 1, August 1 and September 1; Figure 3b); (c) hindcast verification, using ESP as the benchmark (Figure 3c); and (d) correlation analysis between the probabilistic skill of different hindcast configurations and selected catchment descriptors (Figure 3d). In this study, we consider two main hindcasting approaches: (i) the traditional Ensemble Streamflow Prediction (ESP; Figure 3b.1), and (ii) a linear regression model relating seasonal runoff to simulated IHCs (Stat-IHC; Figure 3b.2). For ESP, we test three different initial hydrological conditions (IHCs): an open-loop (OL) configuration obtained by running the model continuously with a two-year spin-up period up to the forecast initialization date, and two additional configurations in which model estimates are updated through data assimilation (DA) using streamflow observations. Additionally, we implement and test four post-processing techniques for the ESP hindcasts.

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

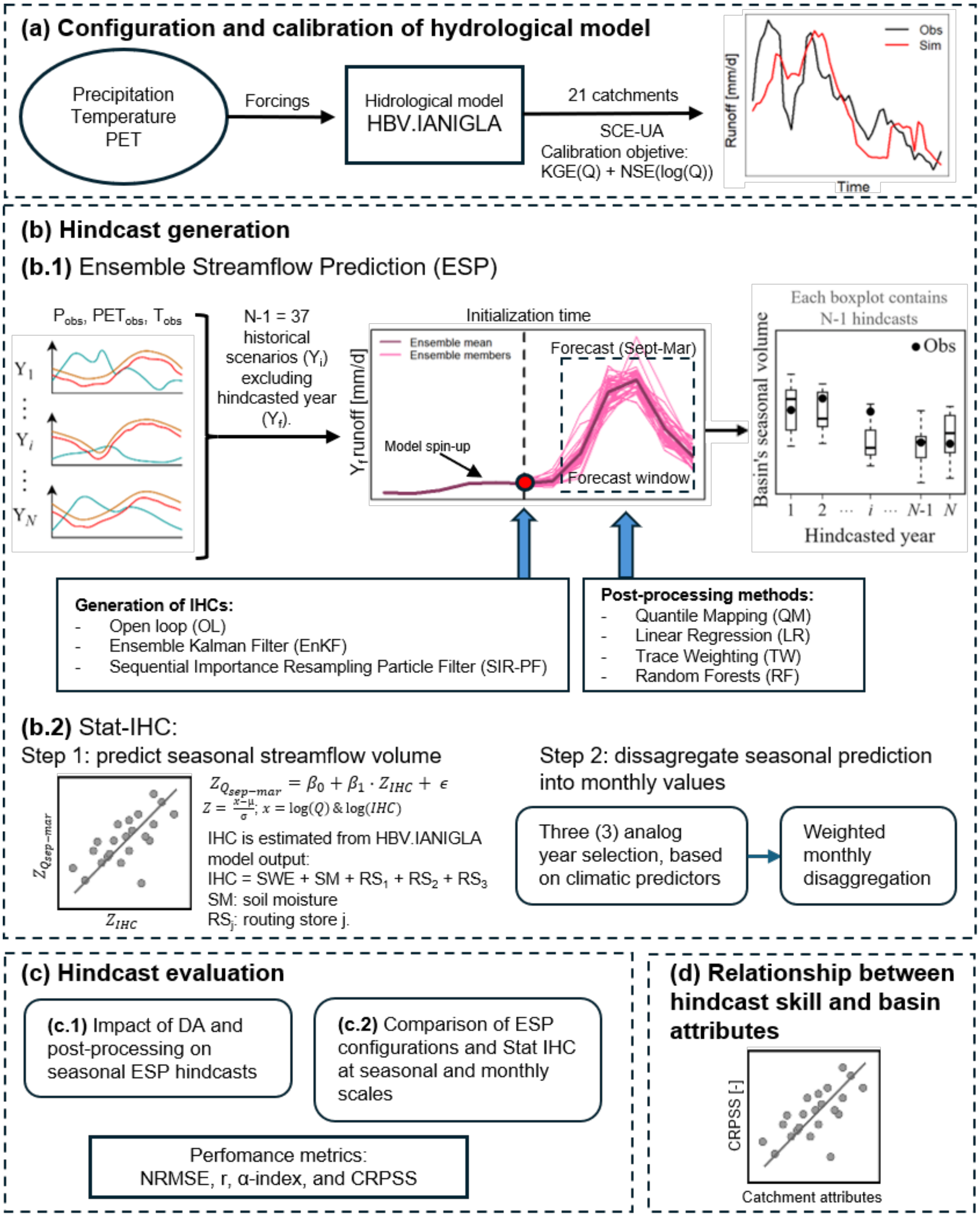


Figure 3. Flowchart describing the approach used in this study. See text for details.

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

3.1 Hydrological modeling

3.1.1 Model structure and configuration

We use the HBV.IANIGLA hydrological model (Toum et al., 2021), a modular version of HBV hydrological model (Bergström, 1976) that was developed for hydroclimatic studies in regions with steep topography and/or cryosphere processes. The modular design of the model enables the use of various spatio-temporal scales with different goals (e.g., real-time streamflow forecasting, hydrological model teaching, or glacier mass balance simulation) in the same environment (Toum et al., 2021). Importantly, previous studies have applied the HBV model and its variants to generate and assess streamflow forecasts (e.g., Pauwels & De Lannoy, 2009; Rakovec et al., 2012; Peñuela et al., 2020).

The HBV.IANIGLA model consists of the following main modules: (1) a four-parameter snow module, where precipitation is partitioned into snowfall and rainfall, and snow accumulation and melting are computed with a degree-day approach; (2) a three-parameter soil module that receives rain and melting as inputs to compute actual evapotranspiration, abstractions, antecedent conditions and effective runoff; (3) a five-parameter routine module with three interconnected buckets – i.e., an upper (STZ), an intermediate (SUZ) and a lower (SLZ) reservoir – whose components are added to compute total runoff, which subsequently feeds a one-parameter triangular transfer function that adjusts the timing of the simulated river discharge. Additionally, subunit variability can be represented through elevation bands that enable distributing air temperature and precipitation along the vertical to subsequently run the snow and soil routines. In this study, we define five equal-area elevation bands for each basin using elevation data from the ASTER Global Digital Elevation Model (DEM), version 3.0 (Abrams et al., 2020).

The model requires daily time series of precipitation, mean air temperature, and PET (Figure 3a), averaged across the elevation bands, to force hydrological simulations. PET is estimated using the formulation proposed by Oudin et al. (2005), based on air temperature and basin latitude.

3.1.2 Calibration strategy

We calibrate the model parameters (Table 1) using the Shuffled Complex Evolution (SCE-UA; Duan et al., 1992) global optimization algorithm – as implemented in the R package "rtop" (Skøien et al., 2014) – to minimize the Euclidean distance (ED) in Equation 1. The minimization of ED seeks to maximize the objective function (OF; Equation 2) recommended by Araya et al. (2023) over two hydroclimatically contrasting periods: (i) April/1990 - March/2000, which includes wetter than average years, and (ii) April/2010- March/2020, dominated by the major drought reported by Garreaud et al. (2020). The selected OF combines the Kling–Gupta efficiency (KGE; Gupta et al., 2009) and the Nash–Sutcliffe efficiency (NSE; Nash & Sutcliffe, 1970) computed on flows in log space, offering a good compromise between hydrologically consistent simulations and seasonal (spring–summer) streamflow hindcast performance (Araya et al., 2023). To compute the OF for each calibration period, the two preceding years are used for model spin-up to ensure stable IHCs. For independent model evaluation, we use

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

the period April/2000-March/2010. Note that the annual mean precipitation averaged across all basins was 945, 769, and 943 mm during the first calibration, second calibration, and evaluation periods, respectively.

$$ED = \sqrt{(OF_{period\,1} - 1)^2 + (OF_{period\,2} - 1)^2} \quad (1)$$

$$\mathrm{OF} = \mathrm{KGE(Q)} + \mathrm{NSE(log(Q))} \quad (2)$$

To maximize the data assimilation effect and ensure that the model states at time t depend only on the conditions at time t – 1, we fix the parameter Bmax – which adjusts discharge timing – to one time step (i.e., 1 d).

Table 1. Description and boundaries of the HBV.IANIGLA hydrological model parameters.

Name	Unit	Description	Range
SFCF	-	Snowfall correction factor	[1; 2]
T_r	°C	Solid and liquid precipitation threshold temperature	[-3; 3]
T_t	°C	Melt temperature	[-3; 3]
f_s	mm/°C/d	Degree-day factor	[0,2; 8]
FC	mm	Fictitious soil field capacity	[25; 1200]
LP	-	Parameter to get actual ET	[0,2; 1]
β	-	Exponential value that allows for non-linear relations between soil box water input (rainfall plus snowmelt) and the effective runoff	[1; 7]
K_0	1/d	Top bucket (STZ) storage constant	[0,08; 0,99]
K_1	1/d	Intermediate bucket (SUZ) storage constant	[0,03; 0,079]
K_2	1/d	Lower bucket (SLZ) storage constant	[0,0001; 0,029]
UZL	mm/d	Maximum flux rate between STZ and SUZ	[3, 5]
PERC	mm/d	Maximum flux rate between SUZ and SLZ	[0; 2,99]
B_{max}	d	Base of the transfer function triangle	[1]

3.2 Seasonal streamflow forecasting approaches

3.2.1 Ensemble Streamflow Prediction (ESP)

The first approach consists of producing seasonal streamflow hindcasts for spring-summer (i.e., September-March) by retrospectively applying the ESP (Day, 1985) method (Figure 3b.1). This approach relies on deterministic hydrological model simulations driven by historical meteorological inputs up to the initialization date, assuming perfect meteorological data and model representation and, therefore, perfect IHCs. The model is then forced with an ensemble of climate sequences, attributing all forecast uncertainty to the variability of future meteorological forcings. In the traditional ESP implementation, each climate

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

sequence (i.e., ensemble member) corresponds to a one-year observed meteorological record, while the traces associated with the target years are excluded from hindcast generation and verification (e.g., Mendoza et al., 2017; Araya et al., 2023).

3.2.2 Statistical forecasting using initial hydrologic conditions (Stat-IHC)

This method follows the operational approach commonly used by agencies like the U.S. Natural Resources Conservation Service (NRCS) and the Chilean Water Directorate (DGA) but replacing ground-based observations of snow water equivalent (SWE), precipitation, and streamflow with model-simulated hydrological states (e.g., Rosenberg et al., 2011; Mendoza et al., 2017). A simple linear regression model between seasonal runoff and simulated IHCs is developed and applied using a leave-one-year-out (LOYO) cross-validation framework. I.e., for a given target year, the regression model coefficients are calibrated using the remaining years, and then used to generate a deterministic seasonal runoff hindcast. IHCs are estimated as the sum of elevation-band-averaged SWE, elevation-band-averaged soil moisture (SM), and routing storages simulated by the HBV.IANIGLA hydrological model. Prior to model training, both predictors and predictands are log-transformed and standardized using z-scores ($z = \frac{x-\mu}{\sigma}$) where x is the log-transformed variable, and μ and σ are its mean and standard deviation, respectively. Then, the statistical model is calibrated using streamflow in standardized log space, and predicted values are subsequently back-transformed ($i.e., Q = \exp(z\sigma + \mu)$) to obtain streamflow hindcasts in the original units. Ensemble hindcasts are generated by adding 100 Gaussian random perturbations with zero mean and standard deviation equal to the standard error of the regression prediction (Mendoza et al., 2017).

The resulting seasonal volumetric hindcasts are subsequently disaggregated into monthly volumes using a climate analog approach that considers: (i) extraction of predictors from each WY, (ii) selection of analog WYs, and (iii) volumetric disaggregation. For each hindcast year and initialization time, a predictor vector is constructed using observed hydroclimatic variables available prior to hindcast issuance. Specifically, the predictor vector includes monthly precipitation, mean monthly air temperature, and mean monthly streamflow from April (first month of the WY in our study domain) up to the month preceding hindcast initialization, resulting in an expanding predictor set as the hindcast initialization date advances. The simulated IHCs (i.e., sum of all model-simulated states) at the hindcast date are also incorporated as an additional predictor.

For each target year, the $k = 3$ most similar historical years are identified based on the Euclidean distance between standardized predictor vectors. To avoid data leakage, the target year is excluded from the candidate pool following the LOYO cross-validation. Each analog year a_j is assigned a weight inversely proportional to the Euclidean distance (d_j) between the target year and the j-th analog:

$$w_j = \frac{1/d_j}{\sum_{l=1}^{k} 1/d_l} \tag{3}$$

The monthly distribution of seasonal streamflow volumes for each j-th analog year is represented by the fraction of the total September–March volume occurring in month m:

$$f_{m,j} = \frac{\bar{Q}_{m,j}}{\sum_{m} \bar{Q}_{m,j}} \tag{4}$$

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

where $\bar{Q}_{m,j}$ is the observed mean monthly streamflow in month m for analog year j. Monthly hindcasts are obtained by applying the weighted-average monthly distribution of the analog years to the seasonal ensemble volume hindcast ($\hat{V}$):

$$\hat{Q}_{m,i} = \left(\sum_{j=1}^{k} w_j f_{m,j}\right) \hat{V}_i \text{ with i = 1,....,100} \quad (5)$$

Note that this framework explicitly propagates two independent sources of uncertainty into the monthly ensemble hindcasts: (i) volumetric uncertainty associated with the regression model, and (ii) temporal distribution uncertainty associated with interannual variability in seasonal streamflow timing.

3.3 Estimation of IHCs for ESP hindcasts

To produce ESP hindcasts, we consider three types of IHCs. The first is generated by running the model in open-loop mode for the two years preceding the hindcast initialization, without assimilating any observations. The second and third are obtained by coupling the hydrological model with the Ensemble Kalman Filter (EnKF; Evensen, 1994) and the Sequential Importance Resampling Particle Filter (SIR-PF: Gordon et al., 1993; Arulampalam et al., 2002), respectively, which update the probability distribution of the model states, whenever streamflow observations are available during the five months prior to the forecast initialization. Readers are referred to Appendix A for details of the implementation of sequential DA methods (e.g., uncertainties in meteorological forcings and streamflow observations).

3.3.1 Open loop IHC

For the traditional ESP implementation, IHCs are obtained as the ensemble mean from a 50-member open loop (OL) simulation over the two years preceding the hindcast initialization (see Appendix A for details on ensemble generation for DA). I.e., we obtain a single value for each model storage as the ensemble mean from N = 50 simulated storages. That single IHC was assumed to be the perfect initial condition required in the ESP method. Note that the same approach is used to determine the IHC for hindcasting configurations that include only post-processing (without DA).

3.3.2 Ensemble Kalman Filter

The EnKF is a sequential ensemble-based data assimilation technique that estimates the analysis states by optimally combining model simulations and observations according to their respective uncertainties, under the assumption that forecast and observation errors are adequately represented by Gaussian distributions.

At each assimilation time step t, the model propagates an ensemble of N state vectors forward in time to obtain the background (forecast) ensemble:

$$\mathbf{x}_{t,i}^{b} = M\left(\mathbf{x}_{t-1,i}^{a}, \mathbf{u}_{t,i}, \boldsymbol{\theta}\right) + \boldsymbol{\varepsilon}_{t,i} \quad (6)$$

where $\mathbf{x}_{t,i}^{b}$ is the background state vector of the i-th ensemble member, $M(\cdot)$ is the hydrological model operator, $\mathbf{u}_t$ represents the meteorological forcings, $\boldsymbol{\theta}$ denotes the model parameters, and $\boldsymbol{\varepsilon}_{t,i}$ is the model error perturbation.

When observations are available, the analysis states are updated according to:

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

$$\mathbf{x}_{t,i}^{a} = \mathbf{x}_{t,i}^{b} + \mathbf{K}_t\left(\boldsymbol{y}_{t,i} - H\mathbf{x}_{t,i}^{b}\right) \tag{7}$$

where $\mathbf{x}_{t,i}^{a}$ is the updated (analysis) state vector, $\mathbf{y}_{\mathbf{t},\mathbf{i}}$ is the perturbed observation vector sampled from $\mathbf{y}_{\mathbf{t},\mathbf{i}} \sim \mathcal{N}(\mathbf{y}_t, \mathbf{R}_t)$. H is the observation operator mapping model states into the observation space, and $\mathbf{K}_t$ is the Kalman gain matrix:

$$\mathbf{K}_t = \mathbf{P}_t^b H^T (H\mathbf{P}_t^b H^T + \mathbf{R}_t)^{-1} \tag{8}$$

$$\mathbf{P}_t^b = \frac{1}{N-1}\sum_{i=1}^{N}\left(\mathbf{x}_{t,i}^{b} - \bar{\mathbf{x}}_t^b\right)\left(\mathbf{x}_{t,i}^{b} - \bar{\mathbf{x}}_t^b\right)^T \tag{9}$$

where $\mathbf{P}_t^b$ is the background error covariance matrix estimated from the ensemble, and $\mathbf{R}_t$ is the observation error covariance matrix.

In this study, the ensemble members (referred to as particles in section 3.3.3) are generated by forcing the model with an ensemble of input meteorological forcings (see details in Appendix A), and streamflow observations at the basin outlet are assimilated to update three HBV.IANIGLA model states: STZ, SUZ, and SLZ. The observation operator H maps the model states into the observation space through simulated streamflow at the basin outlet. EnKF dynamically updates the model state covariance structure through the ensemble statistics, allowing uncertainties to evolve over time. As in Piazzi et al. (2021), we perturb model states at each assimilation time step (i.e., daily) through normally distributed null-mean noise after the analysis step, preventing the spread of the ensemble hydrological simulations being reduced when perturbation of meteorological data alone is unlikely to comprehensively represent system uncertainties, especially during low flow periods. Although the EnKF has been successfully applied in hydrological forecasting (e.g., Rakovec et al., 2012; Piazzi et al., 2021), its performance may degrade in highly nonlinear systems because it relies on Gaussian assumptions and propagates only second-order moments of the probability distributions.

3.3.3 Sequential Importance Resampling Particle Filter (SIR-PF)

We also employ the SIR-PF, which – unlike the EnKF – does not require the restrictive assumption of Gaussian distributions for model errors (Weerts & El Serafy, 2006). Specifically, PFs represent non-Gaussian posterior distributions through weighted particles, making them suitable for nonlinear systems (Arulampalam et al., 2002; Salamon & Feyen, 2009). In practice, the PF represents the system using an ensemble of N possible scenarios or particles ($\boldsymbol{x}_{t,i}, i = 1, \ldots, N$), which are assigned a probability (or weight; $\boldsymbol{w}_{t,i}, i = 1, \ldots, N$) based on the proximity to the observation at a specific time step. Then, the particle weights are updated as new observations become available based on the likelihood of the simulated streamflow relative to the observed discharge. To avoid ensemble degeneracy (i.e., a few particles dominating the total weight), the Sequential Importance Resampling (SIR) procedure is applied, ensuring a well-distributed set of weights (Arulampalam et al., 2002). In this study, we use N = 50 particles, which provide a reasonable compromise between computational cost and filter stability, and equal weights as the initial condition. As in the EnKF implementation, STZ, SUZ, and SLZ are recursively estimated by updating particle weights using streamflow observations at the basin outlet.

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

3.4 Post-processing techniques

Our benchmarking framework includes four post-processing methods: Quantile Mapping (QM), linear regression (LR), Trace Weighting (TW) and Random Forest (RF). We select these techniques because they encompass simple statistical methods (QM and LR), a statistical technique incorporating climate information (TW) and a machine learning method (RF). All these methods are explained below, and details on the formulation are included in Appendix B.

3.4.1 Quantile Mapping (QM)

Quantile Mapping (QM) has been widely used for bias correction in seasonal streamflow forecasting (e.g., Chevuturi et al., 2023; Tanguy et al., 2025). The method adjusts the distribution of each forecast ensemble member according to the relationship between the Cumulative Distribution Function (CDF) of historical observations and the CDF of the reference simulation for the observed period.

In this study, QM is applied to monthly ESP streamflow hindcasts. Given an n-year hindcasting period, for each hindcast ensemble member corresponding to a given target month and year, the reference (historical) CDF is constructed from n – 1 hindcast ensemble medians (n – 1 historical monthly streamflows) for the same calendar month across the analysis period, excluding the target year. This leave-one-out approach ensures that the verification year does not influence the correction (see Section B.1 for details).

3.4.2 Linear Regression (LR)

This approach has been used to calibrate medium-range ensemble streamflow forecasts (Roulin and Vannitsem 2015). A corrected ensemble member for each lead time is obtained from a linear regression equation that requires the forecast ensemble mean and the deviation from that mean as predictors (see Section B.2 for details).

3.4.3 Trace Weighting (TW)

Trace weighting is a well-known strategy for incorporating climate information into ESP hindcasts (e.g., Smith et al., 1992; Werner et al., 2004) by exploiting the relationship between a predictor variable and seasonal streamflow to reassign probabilities across ESP ensemble members. Rather than correcting the ensemble through a parametric transformation as in quantile mapping or linear regression, this approach resamples the existing ESP traces using weights derived from the similarity between the predictor value observed at the forecast initialization time and the corresponding values recorded in each historical year of the ensemble. Here, we select three predictors for each initialization time based on their correlation with seasonal runoff volumes across the study catchments. Specifically, we use (1) the Multivariate ENSO Index (MEI; Wolter & Timlin, 2011) of April for hindcasts initialized on May 1; (2) the cumulative precipitation of May for hindcasts initialized on June 1, and (3) the preceding three-month cumulative precipitation for hindcasts initialized on July 1, August 1 and September 1 (see Section B.3 for details).

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

3.4.4 Random Forest

Random Forest (RF) is a supervised, non-parametric ensemble learning method that constructs a multitude of decision trees during training, providing the mean prediction across all trees for regression tasks (Breiman, 2001). In this study, we apply RF to correct systematic errors in the raw ESP ensemble by learning the statistical relationship between ensemble-derived predictors and historical observations (e.g., Pham et al., 2021; Du & Pechlivanidis, 2025; see Section B.4 for details).

3.5 Experimental setup

In all our experiments, streamflow hindcasts for the spring-summer (i.e., September-March) season over September/1982–March/2020 are produced for all basins in five initialization dates (May 1, June 1, July 1, August 1 and September 1). As a benchmark, we first generated ESP without DA and without PP (simply referred to as ESP). First, we assess the added value of (1) incorporating DA in seasonal ESP hindcast generation using IHCs produced with the EnKF and SIR-PF – referred to as ESP+EnKF and ESP+PF, respectively –; and (2) incorporating post-processing techniques (ESP+QM, ESP+LR, ESP+TW and ESP+RF) in seasonal ESP. The results obtained here are used to select one DA method and one PP technique for subsequent analyses.

Secondly, we assess the individual and joint effects of adding both DA and PP on the quality of seasonal and monthly ESP hindcasts, for different initialization times. In this intercomparison, we include Stat-IHC as an alternative technique given its operational relevance.

Finally, we examine where and when the methodological enhancements previously tested are likely to translate into skill improvements over the traditional ESP. To this end, we explore possible relationships between hindcast skill and catchment characteristics by computing the Spearman's rank correlation coefficient between a skill measure (using ESP as the reference; see section 3.6) and selected physiographic and hydroclimatic descriptors (Table 2).

Table 2. Selected physiographic and climatic characteristics to explore drivers of seasonal forecast quality. Hydroclimatic attributes are computed for the period April/1980 – March/2020.

Name	Description	Units	Data source	Reference
Mean annual streamflow	Mean annual streamflow for the period April/1980 – March/2020	mm/yr	DGA	-
Interannual runoff variability	Coefficient of variation for the time series of annual runoff	-	DGA	-
Baseflow index	Computed as ratio of mean daily baseflow to mean daily discharge	-	CAMELS-CL dataset	Ladson et al. (2013)
Mean annual precipitation	Mean annual precipitation for the period April/1980 – March/2020	mm/yr	CR2MET 2.5	Boisier (2023)

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

Fraction of precipitation falling as snow	Fraction of precipitation falling as snow (i.e. in days colder than 0°C)	-	CR2MET 2.5	Boisier (2023)
Aridity index (AI)	Aridity calculated as the ratio of mean annual PET to mean annual precipitation	-	CR2MET 2.5	Oudin et al. (2005) Budyko (1974)
Mean annual T	Mean annual temperature for the period April/1980 – March/2020	°C	CR2MET 2.5	Boisier (2023)
Mean elevation	Catchment mean elevation	m a.s.l.	CAMELS-CL dataset	ASTER GDEM, (Tachikawa et al., 2011)
Area	Catchment area	km^2	CAMELS-CL dataset	-
Fraction of the basin covered by forest	Fraction of the catchment covered by forest according to a land cover map. Includes native forest and forest plantation	-	CAMELS-CL dataset	Zhao et al. (2016)
Fraction of the basin covered by barren land	Fraction of the catchment covered by barren land according to a land cover map. Includes dry salt flats, sandy areas, and bare exposed rocks	-	CAMELS-CL dataset	Zhao et al. (2016)

3.6 Hindcast Verification

We evaluate ensemble seasonal (September–March) hindcasts by contrasting them against observed streamflow over a 38-year period (September/1982–March/2020), computing deterministic metrics (r, NRMSE) with the hindcast ensemble mean, and probabilistic metrics (α-index y CRPSS) using the full ensemble (Table 3). Hindcast skill refers to the relative accuracy compared to a reference hindcasting system (Wilks, 2019). In this study, the continuous ranked probability skill score (CRPSS; Hersbach, 2000) is used to assess the skill of the hindcast configurations (ESP-DA, ESP-PP, ESP-DA-PP, and Stat-IHC) compared to the benchmark ESP.

Table 3. Performance metrics used for seasonal streamflow hindcast verification.

Name	Equation	Description
Normalized root mean squared error	$NRMSE = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(q_{m,i}-o_i\right)^2}}{sd(o_i)} \cdot 100$	Deterministic metric that varies [0, ∞), with perfect score of 0. It measures the accuracy of the hindcast.
Pearson correlation coefficient	$r = \frac{\sum_{i=1}^{n}\left(q_{m,i}-\overline{q_m}\right)(o_i-\bar{o})}{\sqrt{\sum_{i=1}^{n}\left(q_{m,i}-\overline{q_m}\right)^2}\cdot\sqrt{\sum_{i=1}^{n}(o_i-\bar{o})^2}}$	Deterministic metric that varies [-1, 1] with a perfect score of 1. It measures the linear association between hindcasts and observations.

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

Name	Equation	Description
Continuous ranked probability score	$CRPSS = 1 - \frac{CRPS_{hcst}}{CRPS_{ref}}$ $CRPS = \frac{1}{N}\sum_{i=1}^{n}\int_{-\infty}^{\infty}[F(q) - F_o(q)]^2 dq$ $F_o(q) = \begin{cases} 0, q < 0 \\ 1, q > 0 \end{cases}$	Probabilistic metric that varies (-∞, 1], with perfect score of 1. It measures the skill of CRPS relative to a reference hindcast error (Hersbach, 2000). CRPS quantifies the difference between the CDF of a hindcast (F) and the corresponding CDF of the observations (Fo). The CRPSS values are interpreted as follows: CRPSS=1: the hindcast has perfect skill. CRPSS=0: the hindcast has no skill compared to the reference (the hindcast is as good as using the reference). CRPSS<0: the hindcast is less accurate than the reference (the hindcast is misleading and has no skill).
α reliability index	$\alpha = 1 - 2\left[\frac{1}{N}\sum_{i=1}^{N}\lvert P_i(o_i) - U(o_i)\rvert\right]$	Probabilistic metric that varies [0, 1]. It quantifies the closeness between the empirical CDF of sample p-values with the CDF of a uniform distribution. A value of 0 is the worst, and 1 reflects perfect reliability (Renard et al., 2010).

$q_{m,i}$: hindcast ensemble mean for year i.
$\overline{q_m}$: average over hindcast ensemble means.
o_i: observation for year i.
$\bar{o}$: average of observations.
$P_i(o_i)$: Non-exceedance probability of o_i using ensemble hindcast for year *i*
$U(o_i)$: Non-exceedance probability of o_i using the uniform distribution U [0,1]

4 Results

4.1 Hydrologic model performance

Figure 4 summarizes performance measures obtained for the two calibration periods and the evaluation period, along with other evaluation metrics of interest. Overall, the HBV.IANIGLA model provides satisfactory performance across the 21 case study basins, with a consistently higher median OF (0.79) and a smaller IQR (0.69-0.81) during the first calibration period, compared to the second calibration (median =0.69; IQR=0.53-0.79) and evaluation (median =0.66; IQR=0.58-0.77) periods (Figure 4a). Such pattern is exacerbated when analyzing the KGE (Figure 4b), with medians (IQRs) of 0.82 (0.72-0.86), 0.74 (0.6-0.81) and 0.69 (0.54-0.79) during Calib 1, Calib 2 and Eval, respectively, and also for NSE(log(Q)), with median values of (IQRs) of 0.76 (0.71-0.79), 0.71 (0.49-0.78) and 0.7 (0.51-0.79). Even more, the negative values of NSE(log(Q)) obtained for some combinations of catchment and period reflect model weaknesses in representing baseflow processes.

The KGE components provide additional insights into hydrologic model performance. Regarding correlation (Figure 4d), the highest median (0.84) and the smallest IQR (0.79-0.88) were obtained during the first calibration period, followed by Eval (median =0.77; IQR=0.67-0.85) and Calib 2 (median =0.77; IQR=0.67-0.85). β values (Figure 4e) remain centered close to 1

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

during both calibration periods – indicating that runoff volumes are generally well replicated –, with shifts in median β to 0.97 (<1, indicating overall underestimation of runoff volumes) and larger IQR during the evaluation period. Similar results are obtained for the variability ratio (α), though with larger IQRs. In general, the relatively larger inter-basin differences and worse median results obtained for all metrics during the evaluation period reflect parameter transferability issues arising from hydroclimate differences between periods, as well as differences in catchment characteristics.

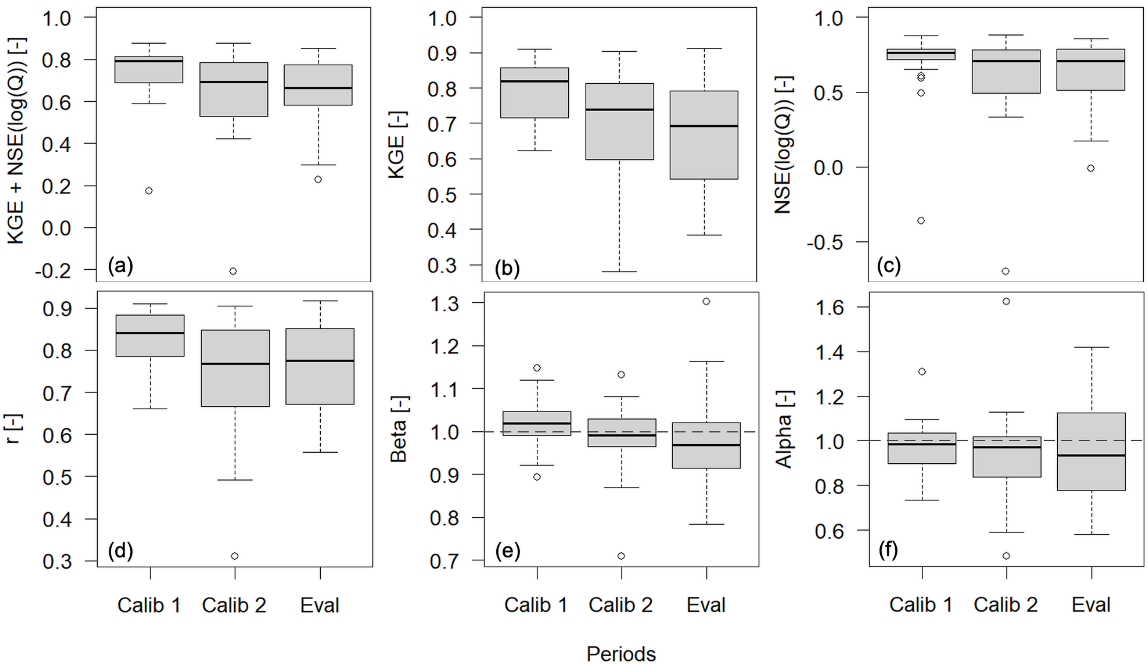


Figure 4. Streamflow performance metrics for the two calibration periods – Apr/1990-Mar/2000 (Calib 1) and Apr/2010-Mar/2020 (Calib 2) – and the evaluation period April/2000-March/2010 (Eval). The metrics displayed are: (a) combined KGE + NSE(log(Q)), which was used as the objective function for calibration; (b) KGE; (c) NSE(log(Q)); (d) correlation (*r*); (e) bias ratio (*β*), and (f) variability ratio (*α*). Each boxplot summarizes results across the 21 case study basins. The boxes represent the interquartile range (IQR; 25th to 75th percentiles), the horizontal line indicates the median, whiskers extend to 1.5 × IQR, and circles denote outliers falling outside the whiskers.

4.2 Impact of DA and post-processing on ESP Spring-Summer hindcast volumes

Figure 5 shows the effects of the DA strategies implemented here on the quality of seasonal streamflow hindcasts across the study domain. The results for NRMSE (Figure 5a) reflect marginal differences (IQR bounds within ±1%) with raw ESP for early initializations (i.e., May 1 and June 1 hindcasts), and negative median ΔNRMSE – along with a larger number of basins showing improvements – for August 1 (-0.9%) and September 1 (-1.8%) hindcasts generated with SIR-PF. The results for CRPSS (Figure 5b) reinforce the benefits of implementing DA – especially SIR-PF – for improving probabilistic skill toward

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

the beginning of the snowmelt season, with basins reaching values as high as 0.3 and 0.75 on September 1 using EnKF and SIR-PF, respectively. We also examined the effects of DA on correlation and the α-index, finding very small differences with respect to raw ESP (not shown).

Figure 6 illustrates the impacts of post-processing techniques on different attributes that describe seasonal hindcast quality. Overall, the results reveal tradeoffs among techniques and verification metrics across initialization times, without an overall "best" method. For example, although QM and LR degrade considerably hindcasts initialized on May 1, June 1 or July 1 in terms of NRMSE (Figure 6a) and correlation (Figure 6b), both methods improve the α-index (Figure 6c) for the same initialization times. On the other hand, the application of RF and TW to May 1 ESP hindcasts helps to reduce NRMSE and improve r in most basins; further, the same methods degrade the α-index at all lead times. Finally, none of the post-processing techniques can improve probabilistic skill in most basins at any lead time, being TW (QM) the best post-processing method for May 1 and June 1 (Aug 1 and Sep 1) hindcasts.

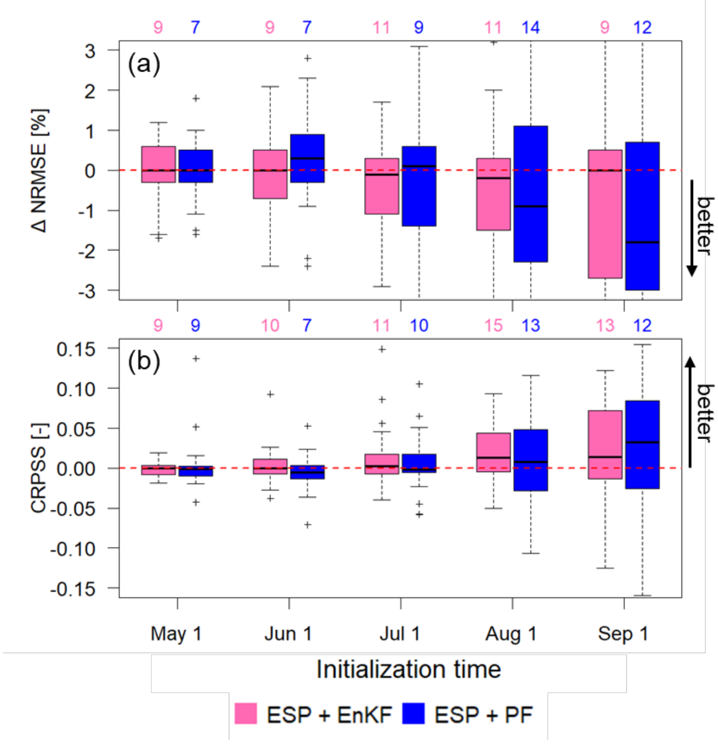


Figure 5. Benefits of implementing DA on the quality of seasonal (September-March) streamflow hindcasts over the period 1982-2019: (a) difference between the NRMSE obtained with ESP+DA (EnKF and PF) and raw ESP; and (b) CRPSS obtained with ESP+DA, using the traditional ESP as the reference. Each boxplot comprises results for our 21 case study basins. The boxes correspond to the interquartile range (IQR, i.e., 25th and 75th percentiles), the horizontal line in each box is the median, and whiskers extend to the $\pm 1.5 \cdot IQR$ of the ensemble. The numbers located over each

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

boxplot indicate the number of catchments where the implementation of a specific DA technique improves the metric displayed in comparison with raw ESP.

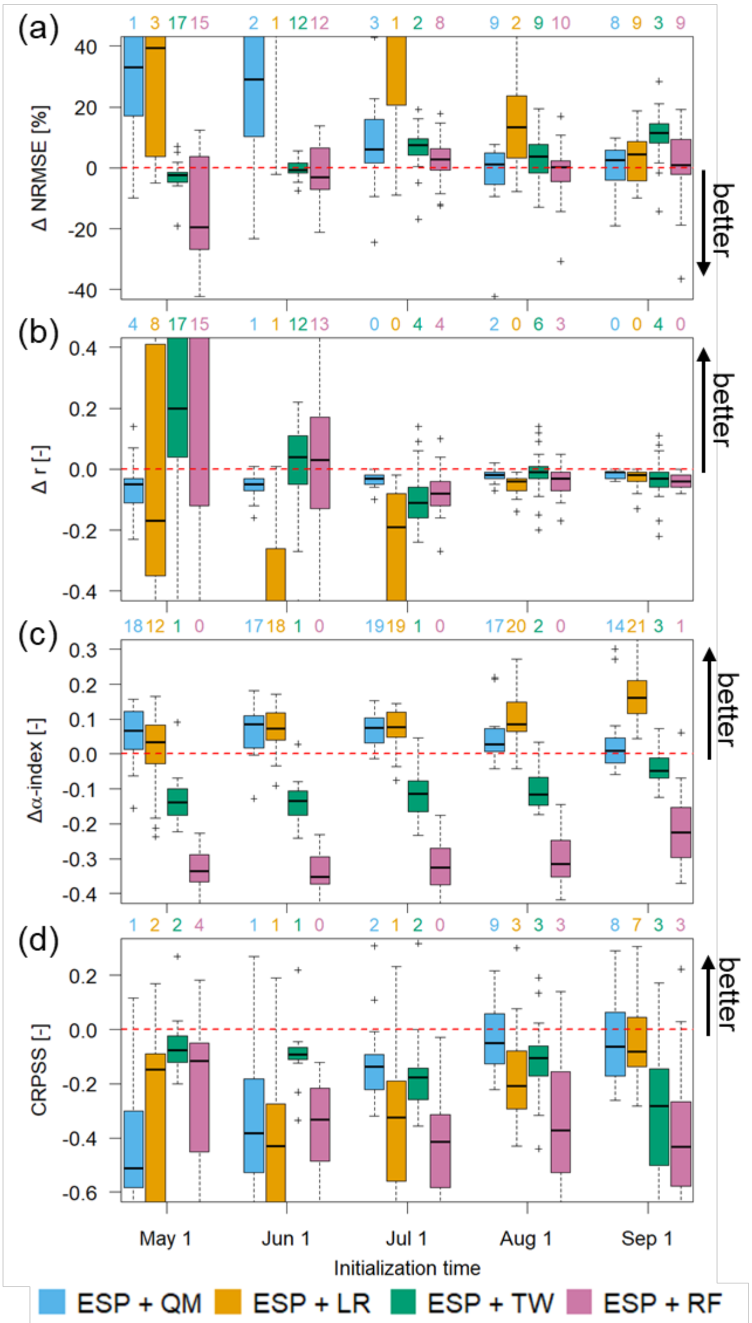


Figure 6. Benefits of implementing different post-processing techniques on the quality of seasonal (September-March) streamflow hindcasts over the period September/1982–March/2020. The panels display differences between the (a) NRMSE, (b) Pearson correlation coefficient, and (c) α reliability index obtained with post-processed ESP and raw ESP

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

output. Panel (d) displays the CRPSS obtained with post-processed ESP output, using traditional ESP as the reference. Each boxplot comprises results for our 21 case study basins. The boxes correspond to the interquartile range (IQR, i.e., 25th and 75th percentiles), the horizontal line in each box is the median, and whiskers extend to the $\pm 1.5 \cdot IQR$ of the ensemble. The numbers located over each boxplot indicate the number of catchments where the implementation of a specific post-processing technique improves the metric displayed in comparison with raw ESP.

4.3 Evaluating ESP enhancements at seasonal and monthly time scales

Given the superiority of SIR-PF over EnKF (Figure 5), and the capability of QM to preserve the correlation Aug 1 and Sep 1 ESP hindcasts and improve the α-index without deteriorating probabilistic skill, hereafter we explore the individual and combined potential of SIR-PF and QM to enhance seasonal hindcasts across our study domain. Figure 7 illustrates this assessment for the Maipo at El Manzano River basin (4839 km^2), which provides nearly 70 % of municipal water supply for Santiago (Chile's capital city) and is also the primary source of water for agriculture, hydropower, and industry in the area (Ayala et al., 2020). The results show that none of the alternative configurations can beat the CRPS and NRMSE achieved by the traditional ESP on July 1, and that the SIR+PF is the only addition that can preserve r in both initialization times. Stat-IHC is the configuration with the highest (best) α-index achieved on July 1 and Sep 1, and on Sep 1 SIR+PF is the only configuration that can beat ESP in terms of CRPS.

Figure 8 expands the previous intercomparison to all the case study basins. The results show that most configurations do not reduce NRMSE during the earliest initialization dates (Figure 8a). The two configurations including QM provide median ΔNRMSE > 25% for May 1 and June 1 hindcasts, indicating larger errors than traditional ESP in most basins. Improvements become more frequent toward the beginning of the snowmelt season, especially for ESP+PF, which reduces the NRMSE of Aug 1 and Sep 1 hindcasts in 14 and 12 basins, respectively. ESP+QM and ESP+PF+QM also improve NRMSE in more basins after July 1 (9-11 in Aug 1 and 8-10 in Sep 1), whereas Stat-IHC yields limited improvements before Sep 1. Figure 8b shows that improving correlation across the sample of catchments is a challenging task, with SIR-PF being the most effective strategy, providing r increments in 9, 7, 6, 6, and 10 basins for the five initializations from May 1 to September 1, respectively. Figure 8c shows that Stat-IHC improves hindcast reliability in all basins for every initialization date, with median increments in α-index around 0.46-0.55, whereas the ESP-based configurations yield smaller reliability gains. ESP+QM (ESP+PF+QM) improves the α-index in 18, 17, 19, 17, and 14 (17, 19, 20, 17, and 13) basins from May 1 to September 1, respectively, whereas ESP+PF improves fewer basins, ranging from 8 to 12 depending on initialization date. CRPSS results are mostly negative for configurations involving quantile mapping at early initialization dates (Figure 8d). Indeed, ESP+QM and ESP+PF+QM show median CRPSS values below -0.3 for May 1 and June 1 (i.e., lower skill than traditional ESP in most basins). Since July 1, CRPSS values move closer to zero, and ESP+PF improves CRPSS in 10-13 basins between Jul 1 and Sep 1. Overall, improvements over traditional ESP are metric-dependent: Stat-IHC consistently improves reliability, ESP+PF provides the most frequent late-season reductions in NRMSE and CRPSS gains, and QM improves reliability but often increases NRMSE and reduces CRPSS at longer lead times.

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

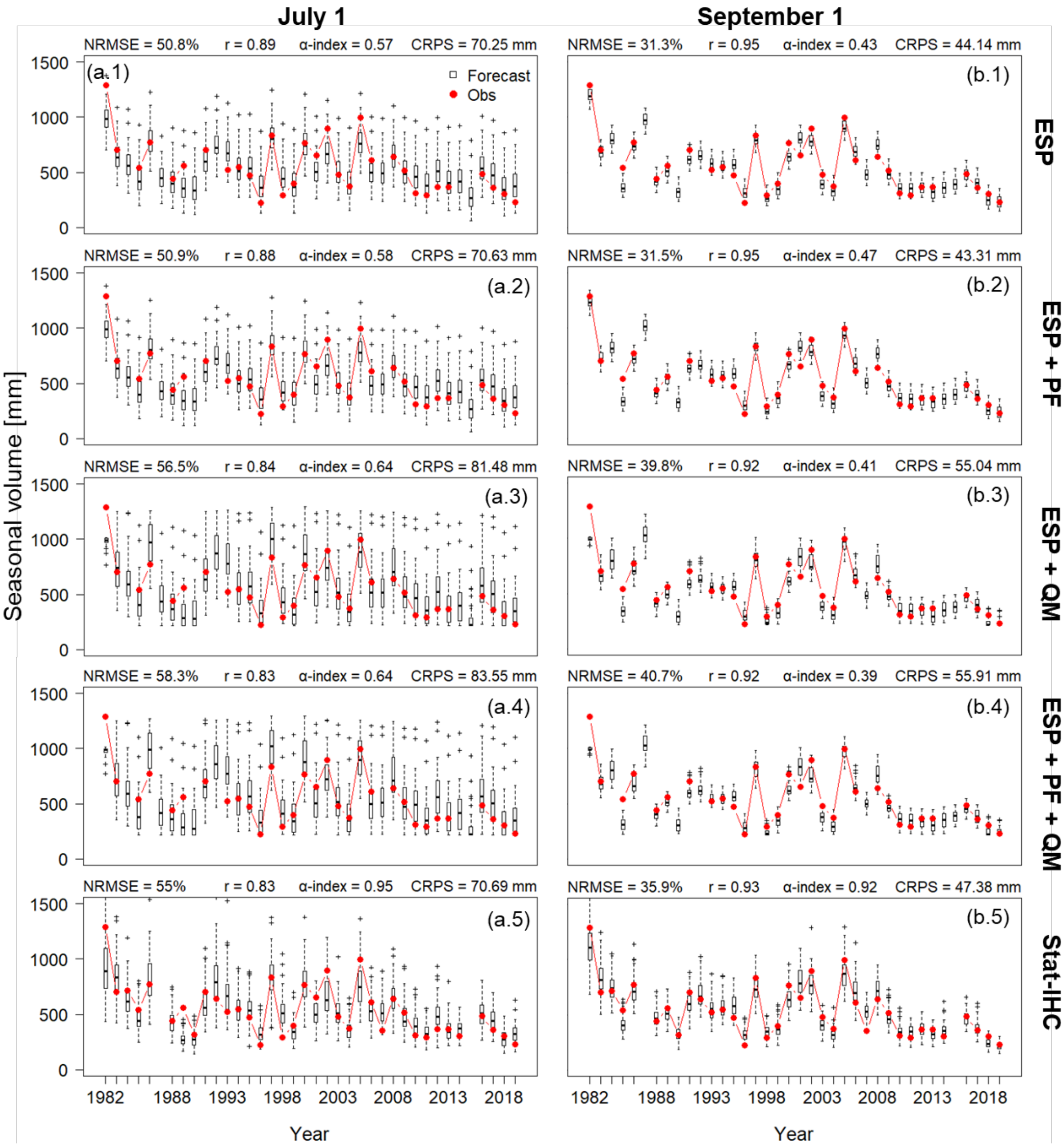


Figure 7. Time series with ESP seasonal hindcasts (i.e., September-March runoff) over the period over the period September/1982–March/2020, initialized on July 1 (left panels), and September 1 (right panels) for the Maipo at El Manzano River basin. The boxes correspond to the interquartile range (IQR, i.e., 25th and 75th percentiles); the horizontal line in each box is the median, whiskers extend to the ±1.5·IQR of the ensemble, and the red dots represent observations. The results were produced with the ESP technique and a few variants that include (from top to bottom) IHCs estimated with particle filtering (ESP+PF), ESP hindcasts post-processed with Quantile Mapping (ESP+QM),

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

the combination of the two (ESP+PF+QM) and the linear regression model (Stat-IHC). Each panel displays the NRMSE and r (computed using the ensemble hindcast mean), the α-index, and CRPS.

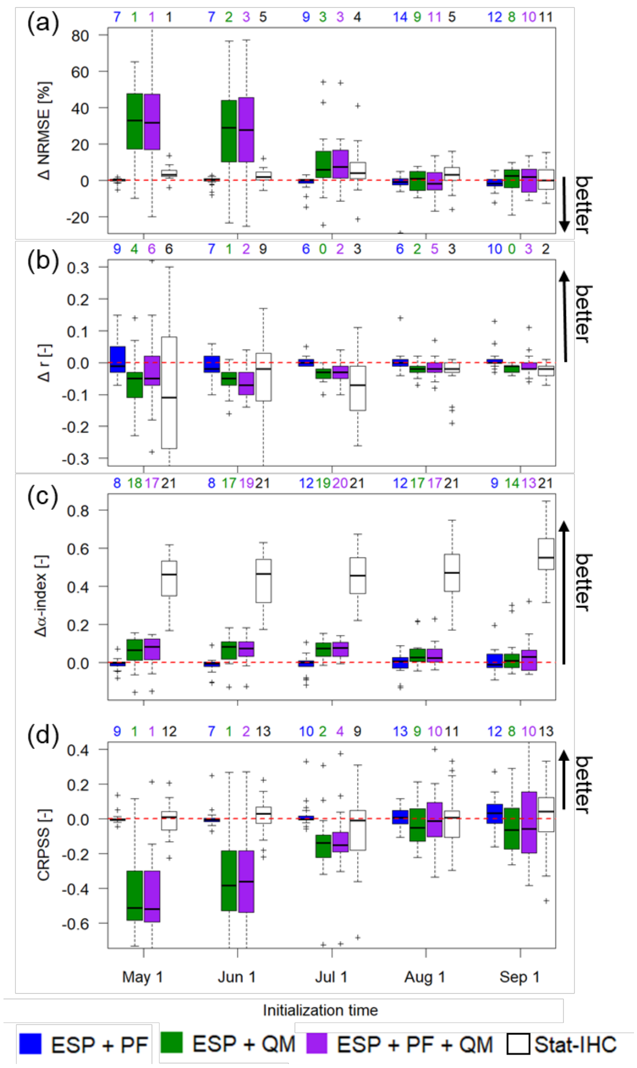


Figure 8. Same as in Figure 6, but testing the benefits of alternative ESP implementations and Stat-IHC relative to raw ESP. Each boxplot comprises results for our 21 case study basins. The numbers located over each boxplot indicate the number of catchments where the implementation of a specific hindcast configuration improved the metric displayed in comparison with raw ESP.

Figure 9 shows the effects of alternative ESP configurations on the quality of monthly seasonal hindcasts. For July 1 initializations (Figure 9a), ESP+PF yields the most frequent NRMSE reductions (e.g., decrease in 12 basins for Sep and Oct flows, and in 11 basins for Feb and Mar flows) among the ESP-based configurations, with median ΔNRMSE values are close

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

or below zero in several months (Figure 9a.1). In contrast, ESP+QM and ESP+PF+QM show positive median ΔNRMSE values in many months (ranging 3-11%), indicating larger errors than ESP in most basins. Stat-IHC shows a month-dependent behavior, with median ΔNRMSE values near zero in Sep and Oct, and positive values Jan, Feb and Mar. For hindcasts initialized on September 1 (Figure 9b.1), NRMSE improvements are more frequent, with ESP+PF, ESP+QM, and ESP+PF+QM providing negative ΔNRMSE medians in several months, especially at the beginning and the end of the snowmelt season. The results in Figures 9a.2 and 9b.2 show that the alternative configurations have limited effects on linear correlation, with most median Δr values between -0.05 and 0.05. ESP+PF preserves the original ESP correlation in most cases, whereas Stat-IHC tends to reduce *r* more often than the ESP-based configurations – particularly for July 1 hindcasts –, with negative medians in all months.

Regarding the α-index (Figure 9c), Stat-IHC yields the largest reliability gains in both initialization dates, with median Δα-index > 0 and improvements in nearly all basins. The ESP-based configurations also improve reliability (especially those using QM), though their medians range between 0 and 0.18. CRPSS results for July 1 hindcasts (Figure 9a.4) indicate mostly negative median values for the ESP-based configurations, often between about -0.2 and 0, meaning that monthly probabilistic skill is lower than traditional ESP in many months. For September 1 hindcasts panels (Figure 9b.4), CRPSS distributions shift upward, with median values ≥ 0 in several months, especially in Sep, Feb and Mar flows. Overall, Figure 9 reaffirms that monthly improvements over ESP depend on initialization date, forecast month, and metric: September 1 initializations provide more frequent improvements in NRMSE and CRPS, whereas Stat-IHC provides the most consistent reliability gains, with the cost of reduced correlation.

In summary, the results presented in this section show that no method always improves all hindcast attributes. ESP+PF is the most promising ESP-based enhancement, mainly for later seasonal initializations and monthly hindcasts initialized on September 1. QM enhances reliability at the cost of degrading NRMSE and probabilistic skill, especially at longer lead times. Stat-IHC provides metric-specific improvements, especially in terms of reliability. From a practical perspective, forecast enhancement should be conditional: ESP alternatives can be used more confidently for late-winter to early-spring initializations – especially August 1 and September 1 at the seasonal scale and September 1 at the monthly scale –, whereas early-season enhancements should be used with caution.

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

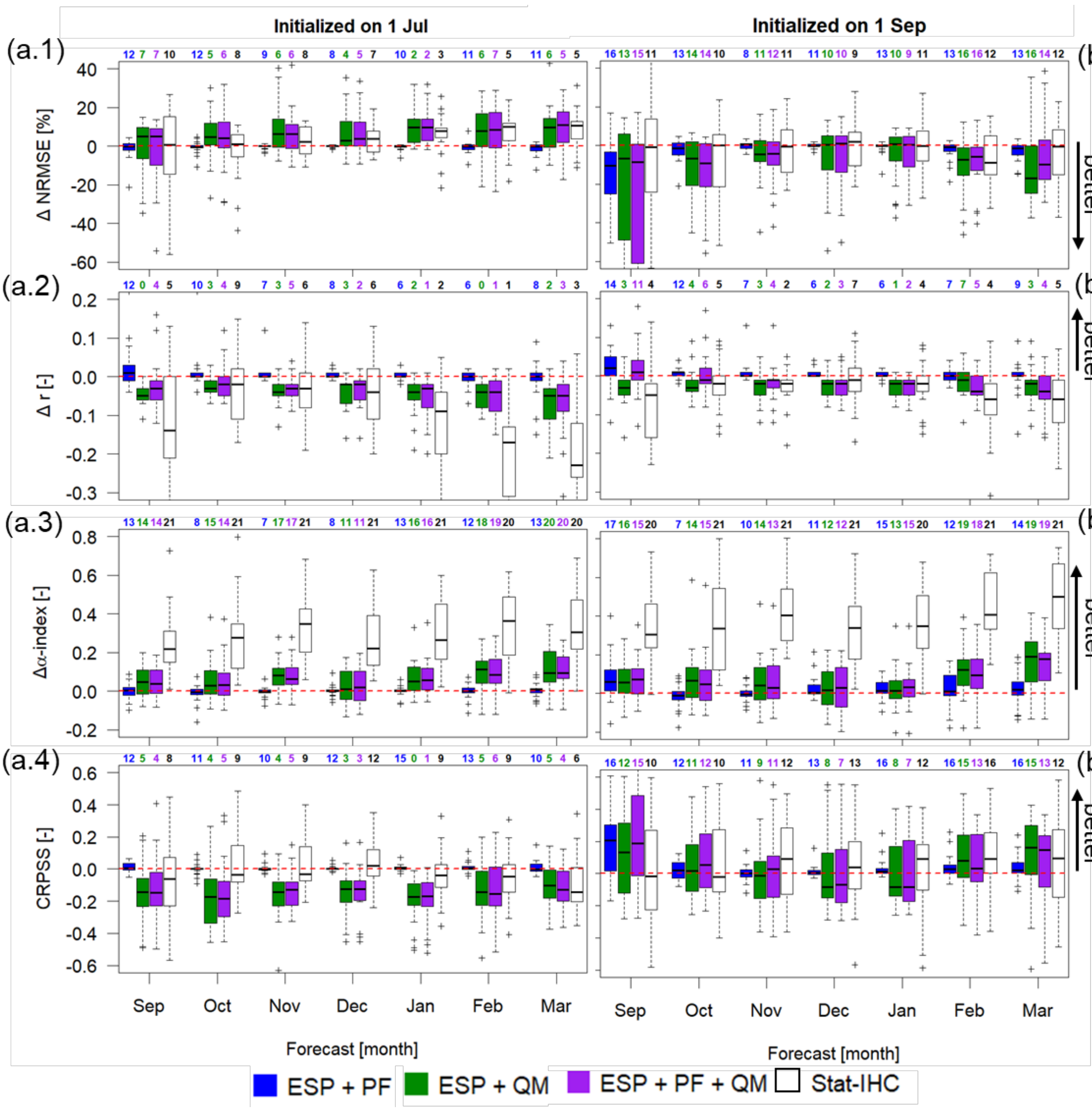


Figure 9. Same as in Figure 8, but for monthly streamflow forecasts. Each boxplot comprises results for our 21 case study basins. The numbers located over each boxplot indicate the number of catchments where the implementation of a specific hindcast configuration improved the metric displayed in comparison with raw ESP.

4.4 Hindcast skill vs. catchment characteristics

Given the potential for improving ESP skill at both seasonal (Figure 8d) and monthly (Figure 9b.4) time scales, we now examine where and when the methodological enhancements tested here are likely to translate into skill improvements. Based on the previous results, we focus on hindcasts initialized on Aug 1 and Sep 1, which provided positive CRPSS values in several

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

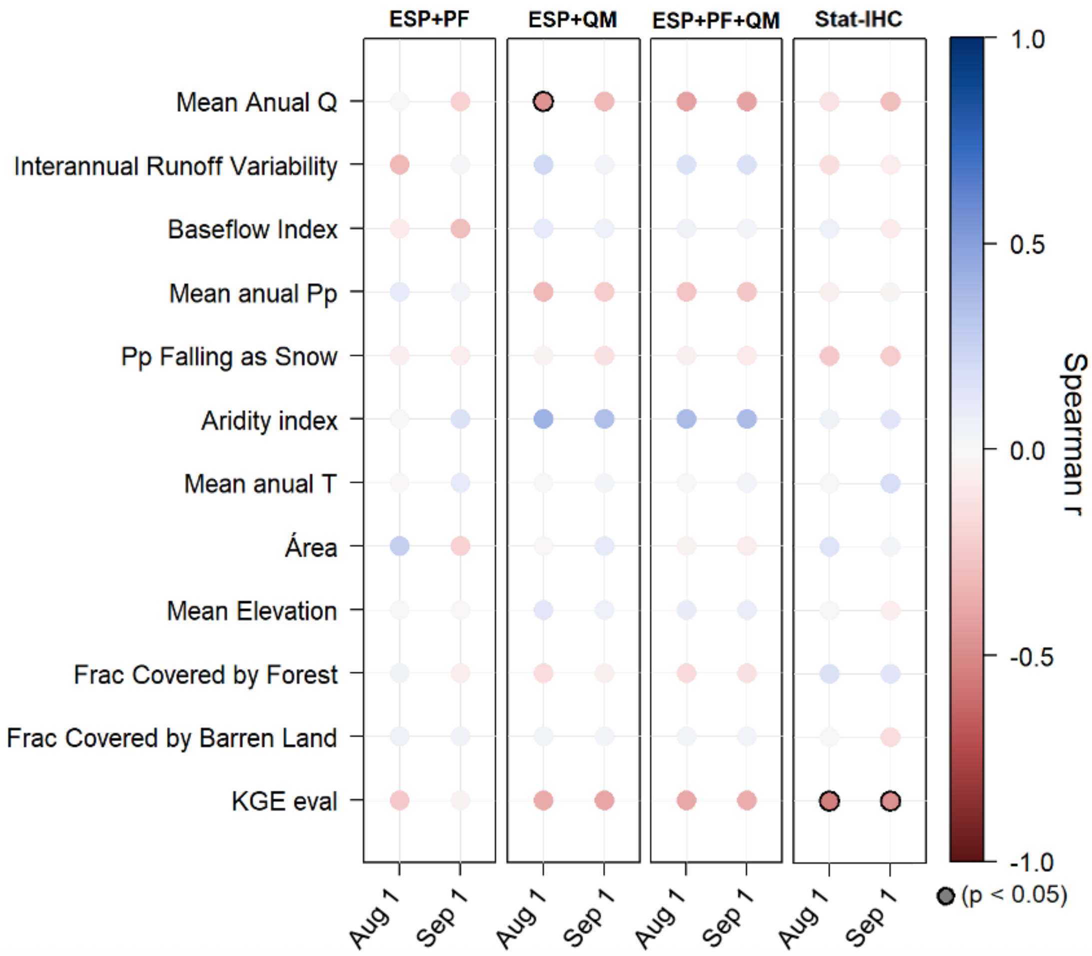
catchments at both seasonal and monthly time scales. Figure 10 shows that relationships between catchment attributes and seasonal (i.e., Sep-Mar volumes) hindcast skill are generally weak to moderate, and only a few correlations are statistically significant. The clearest signal is associated with model performance, represented by KGE in evaluation. For Stat-IHC, CRPSS is negatively and significantly correlated with KGE eval for both August 1 (ρ = -0.54) and September 1 (ρ = -0.48) initializations, indicating that Stat-IHC tends to provide higher probabilistic skill in catchments where the hydrological model has lower evaluation performance. For the ESP-based configurations, significant correlations are scarce, with the largest one obtained with mean annual runoff (ρ = -0.46) for ESP+QM initialized on August 1. More broadly, adding QM to the hindcast generation process yields increments in skill in more arid basins (ρ = 0.42) and in basins with lower precipitation (ρ = -0.31), lower runoff (ρ = -0.46), and lower KGE eval (ρ = - 0.39).

Figure 11 shows stronger relationships between CRPSS for monthly flows and catchment descriptors compared to the seasonal results, especially for configurations that include QM. Specifically, the implementations using QM tend to increase the skill of September monthly flows in catchments with lower precipitation and runoff, and these relationships are most evident for hindcasts initialized on August 1. Additionally, the CRPSS obtained with ESP+QM and ESP+PF+QM for September and/or October is positively and significantly correlated with precipitation falling as snow (ρ = 0.57), mean elevation (ρ = 0.68), and fraction covered by barren land (ρ =0.64), suggesting that post-processing is more beneficial in high-elevation, snow-influenced, and sparsely vegetated catchments. Conversely, correlations with mean annual temperature (ρ =-0.64) and forest cover (ρ = -0.46) are often negative for the same configurations, consistent with the idea that the gains are concentrated in colder, less forested, more snow-dominated basins.

The results in Figure 11 also show less significant relationships for ESP+PF. With this setup, the CRPSS of October hindcasts show some significant positive correlations with mean annual runoff (ρ =0.49), precipitation (ρ =0.52), and forest cover (ρ =0.46), and negative correlations with baseflow index (ρ = - 0.63) and elevation (ρ = -0.45). This suggests that particle filtering alone may provide more skill in relatively wetter or lower-elevation catchments for October flows, contrasting with the snow-dominated signal obtained with setups including QM. Stat-IHC shows weaker and less systematic correlations than the ESP-based post-processed configurations, with only a few significant relationships.

In summary, catchment controls on seasonal CRPSS are method-dependent, with improved relative skill in drier or lower-performing basins, particularly for Stat-IHC and, to a lesser extent, ESP+QM initialized on August 1. On the other hand, improvements in monthly skill are method- and month-dependent: QM performs better for September flows in arid, high-elevation, snow-dominated catchments, whereas PF-only gains appear more limited and associated with wetter hydroclimatic conditions.

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026



535 **Figure 10. Spearman's rank correlation coefficients between catchment characteristics (shown in different rows) and the CRPSS for seasonal (September-March) streamflow hindcasts (period September/1982–March/2020) obtained with different setups (columns) and two different initialization times (August 1 and September 1). CRPSS values were computed using raw ESP output as the reference. Circles with bold outlines indicate statistically significant (p < 0.05) correlations.**

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

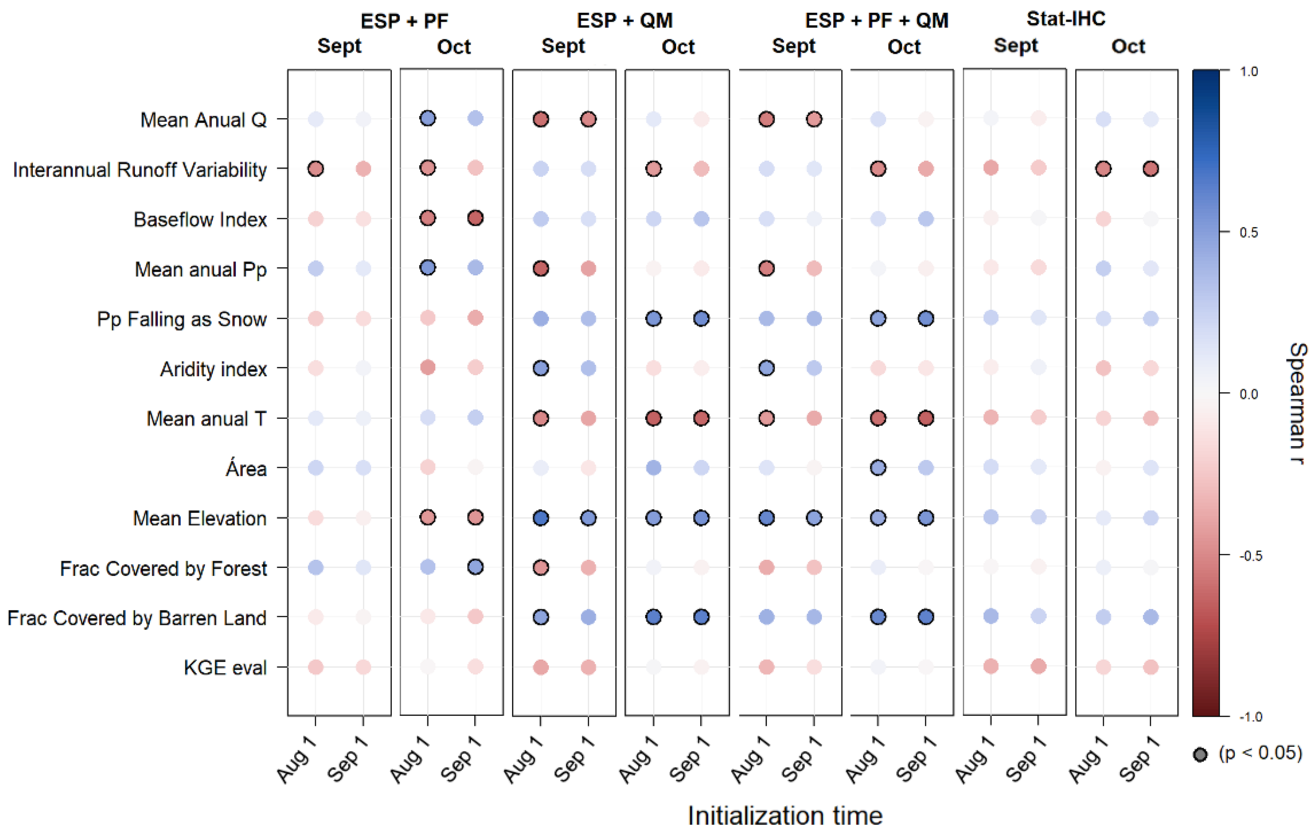
540 

Figure 11. Same as in Figure 10, but for September and October streamflow hindcasts (period September/1982–March/2020) obtained with different setups (columns) and two different initialization times (August 1 and September 1).

5 Discussion

545 ### 5.1 Impact of DA and post-processing

The results presented here confirm that traditional ESP is a difficult benchmark to outperform in snow-influenced mountainous catchments, in agreement with previous studies showing that ESP often remains highly competitive when seasonal runoff is strongly controlled by IHCs (e.g., Arnal et al., 2018; Wanders et al., 2019; Peñuela et al., 2020; Petry et al., 2023). Nevertheless, the hindcasting experiments also revealed opportunities to improve some quality attributes.

550 Despite our results reaffirm that better IHC estimates can improve seasonal ESP reliability and skill (e.g., DeChant & Moradkhani, 2011; Caleb M. DeChant & Moradkhani, 2014; Huang et al., 2017; Micheletty et al., 2021; Metref et al., 2023; Uysal et al., 2026), they also underscore that DA benefits are timing dependent. In our experiments, streamflow assimilation was most useful when the updated states were close enough to the hindcast period to influence September-March runoff

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

volumes. Further, the SIR-PF produced more consistent gains in hindcast quality than the EnKF for August 1 and September 1 initializations, suggesting that particle filtering may better handle nonlinear and non-Gaussian snowmelt-runoff dynamics. Previous studies have also highlighted the benefits of particle filters over Kalman filters for short-range (Jin Noh et al., 2013) and medium-range (Piazzi et al., 2021) streamflow forecasting, though not – to the best of our knowledge – for long-range forecasting applications.

Post-processing showed more pronounced tradeoffs among hindcast attributes. QM generally improved reliability, but often degraded NRMSE, correlation, and CRPSS, especially at longer lead times. TW and RF reduced NRMSE and improved correlation for some early (i.e., May 1 and June 1) hindcasts, but tended to reduce reliability compared to raw ESP. These results are consistent with previous work showing that post-processing can improve selected forecast attributes while degrading others (Gobena & Gan, 2010; Wood et al., 2018). The joint application of PF and QM was useful mainly for August 1 and September 1 hindcasts, though their effects were not additive since DA and post-processing target different error sources. Our results also align with Tanguy et al. (2025), who found no universal superiority of bias correction or DA within an operational ESP framework across the UK.

Finally, Stat-IHC produced the most consistent reliability gains, reinforcing the value of simple statistical benchmarks in seasonal water-supply forecasting, and aligning with previous studies showing that statistical or hybrid methods based on IHCs, climate indices, and/or antecedent streamflow can be competitive with ESP (Mendoza et al., 2017; Broxton et al., 2023). However, Stat-IHC did not consistently improve NRMSE, correlation, or CRPSS, indicating that high reliability alone does not guarantee superior overall hindcast quality.

5.2 Skill improvements vs. catchment characteristics

Relationships between skill improvements at the seasonal scale (i.e., Sep-Mar runoff) and catchment attributes were generally weak to moderate. The clearest signal was the negative relationship between the CRPSS from Stat-IHC and hydrological model performance (KGE eval), suggesting that the statistical hindcasting approach added more value in catchments where HBV.IANIGLA performed less well. This may indicate that simple statistical models can partly compensate for structural or parameter-transferability limitations of process-based models. QM also tended to provide larger skill improvements in more arid basins with lower precipitation, lower runoff, and worse model performance.

At the monthly scale, catchment controls on ESP skill gains were stronger and more physically interpretable. QM-based configurations tended to improve September and October hindcasts in colder, high-elevation, snow-influenced, sparsely vegetated catchments. In contrast, PF-only gains for October flows were more associated with wetter and lower-elevation catchments. It should be noted, however, that these findings are not comparable with previous studies exploring drivers of seasonal forecast skill (e.g., Pechlivanidis et al., 2020; Araya et al., 2023), since their ESP implementations, hindcasting benchmarks and/or observational references are different to ours.

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

5.3 Implications for operational forecasting

The robust performance of the traditional ESP implementation – especially for early initializations compared to alternative configurations – supports its utility as a central benchmark for operational forecasting systems in mountainous domains. Hence, new DA or post-processing components should be adopted after hindcast-based evaluation against raw ESP. Because the benefits of SIR-PF were evident only during the first month of the snowmelt season (Figure 9), the potential of pre-processed seasonal climate forecasts (e.g., Lucatero et al., 2018; Wanders et al., 2019; Petry et al., 2023) should be explored in Andean catchments, as meteorological uncertainty becomes increasingly important for longer lead times (Shukla & Lettenmaier, 2011). Since no single configuration optimized all hindcast attributes, operational systems may benefit from a small multi-method forecasting suite that considers raw ESP as the benchmark, ESP+PF for late-winter initializations, QM-adjusted forecasts when reliability is prioritized, and Stat-IHC as a simple and complementary statistical reference. Such a conditional strategy is consistent with Tanguy et al. (2025), who recommended selecting DA or bias correction according to catchment characteristics, season, and lead time.

5.4 Limitations and future work

Our results are based on a single hydrological model structure and calibration strategy. Although HBV.IANIGLA overall performed satisfactorily, the evaluation metrics varied across basins, with some of them having difficulties in low flow simulations. Multi-model experiments (e.g., Saavedra et al., 2022; Knoben et al., 2025) and the explicit treatment of parameter uncertainty (e.g., Thébault et al., 2025) would help assess the extent to which the relative benefits of DA and post-processing depend on hydrological modeling decisions. Additionally, we did not assimilate any snow observations. Because spring-summer runoff across the study domain depends strongly on snow accumulation and melt, future work should test the assimilation of in situ SWE (Huang et al., 2017), remotely sensed snow-covered area (Micheletty et al., 2021), SWE reanalysis products (e.g., Muñoz-Castro, 2021), or combined snow-streamflow observations (e.g., Metref et al., 2023). Importantly, the ESP method cannot predict extreme events with magnitudes outside the historical record (Sabzipour et al., 2021), and its skill may be reduced under non-stationary climate conditions (Peñuela et al., 2020).

Despite the post-processing techniques and Stat-IHC are easy to implement, they depend on relatively small training samples and stationarity assumptions that may be fragile under prolonged drought and climate change. Additionally, this study focused on ESP-based enhancements and did not incorporate dynamical seasonal climate forecasts, which can outperform ESP in some regions and seasons after appropriate bias correction or downscaling (e.g., Wanders et al., 2019; Petry et al., 2023). Future research should test hybrid forecasting approaches that combine improved IHC estimation with seasonal climate forecasts and a post-processing step, particularly for early initializations. Climate forecast information could be used within a Historic Weather Analog framework (e.g., Chan et al., 2026), in which plausible scenarios of precipitation and temperature at the local scale can be inferred by selecting independent "analogue" months from the historical observations that resemble the

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

atmospheric circulation patterns forecasted by a dynamical climate model. Finally, ESP enhancements could be tested in other mountainous regions across the globe to verify the transferability of our findings.

6 Conclusions

In this study, we evaluated how data assimilation and post-processing affect the quality of seasonal streamflow forecasts generated with the traditional ensemble streamflow prediction (ESP) method in 21 snow-influenced Andean catchments. The alternative ESP hindcast configurations, together with a simple linear regression model using initial hydrologic conditions (IHCs) as the only predictor, were benchmarked against raw ESP forecasts at seasonal and monthly time scales for the spring-summer runoff season. Our results confirm that traditional ESP is a difficult benchmark to outperform in snowmelt-dominated catchments. Nevertheless, they also reveal important opportunities, limitations, and tradeoffs among forecast enhancement strategies.

At the seasonal scale, the benefits of data assimilation depended strongly on initialization date and filtering method. The superiority of SIR-PF over EnKF was not evident until August 1, while post-processing tended to degrade hindcast skill for lead times of two months or longer. In contrast, the joint use of data assimilation and post-processing in forecasts initialized on September 1 produced notable improvements in NMRSE, CRPSS, and the α-index, particularly near the beginning of the snowmelt season, in September and October, and near its end, in February and March. An interesting result is that Stat-IHC, the simplest method tested here, produced considerable improvements over the traditional ESP in terms of hindcast reliability (α-index) at both seasonal and monthly time scales, although its performance was mixed for other metrics and runoff aggregations.

Our catchment-scale analyses also show that improvements over traditional ESP are not spatially uniform but depend on basin hydroclimatic characteristics. For seasonal flow volumes initialized on August 1 and September 1, probabilistic skill was more likely to improve through data assimilation, post-processing, or Stat-IHC in semi-arid basins with higher aridity indices, lower precipitation and runoff amounts, and poorer hydrologic model performance. At the monthly scale, improvements at the beginning of the snowmelt season were geographically clustered. Data assimilation alone tended to be more beneficial in relatively humid catchments with lower baseflow contributions, whereas post-processing and, to a lesser extent, Stat-IHC showed greater potential in high-altitude semi-arid basins with larger snowfall fractions and more barren land cover.

Overall, these results provide new insights into when, where, and how IHC-based seasonal streamflow forecasts can improve upon traditional ESP in mountainous regions. By showing how data assimilation and post-processing effects propagate through monthly forecasts during the snowmelt season, and by identifying the basin types where ESP skill is most likely to be improved, this study offers guidance for future forecasting research in mountain environments and for operational water management across the Andes.

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

Appendix A: Implementation of sequential DA methods

A.1. Uncertainty in Meteorological Forcings

An effective ensemble based DA implementation requires maintaining an adequate ensemble spread (Piazzi et al., 2021). If the ensemble is too narrow, the resampling will be inefficient because particles exhibit similar likelihood values, limiting the filter's ability to prioritize the most likely members. Conversely, an overly dispersed ensemble may result in an excessive reliance on observations, despite their associated errors. In this study, we generate the particles by creating an ensemble of plausible weather scenarios. Specifically, we apply a multiplicative stochastic noise to precipitation and additive noise to temperature, following Clark et al. (2008):

$$P' = P\varphi_p\ [N_{Mbr}, N_{Time}] \qquad \text{(A.1)}$$

$$\varphi_p = \left(1 - \mathcal{E}_p\right) + 2u_p\mathcal{E}_p \qquad \text{(A.2)}$$

where u_p is a uniform random number, such that φ_p is a realization from a uniform distribution ranging from 1- $\mathcal{E}_p$to 1+ $\mathcal{E}_p$.

$$T' = T + \varphi_t\ [N_{Mbr}, N_{Time}] \qquad \text{(A.3)}$$

$$\varphi_t = -\mathcal{E}_t + 2t\mathcal{E}_t \qquad \text{(A.4)}$$

where u_t is a uniform random number, such that φ_t is a realization from a uniform distribution ranging from - $\mathcal{E}_t$to $\mathcal{E}_t$.

Random perturbations were generated using a first-order autoregressive model to preserve temporal correlation and physical consistency. This approach represents time-varying hydrological uncertainty more realistically than temporally constant perturbations. The fractional error parameters $(\mathcal{E}_p, \mathcal{E}_t)$ were optimized for each basin to satisfy two ensemble reliability conditions over the period April/1990-March/2020: (i) if the ensemble spread is large enough, the temporal average of the ensemble skill (sk_t) should be similar to the temporal average of the ensemble spread sp_t, i.e., $\overline{sk}\ /\ \overline{sp} = 1$ (e.g., De Lannoy et al., 2006; Alvarez-Garreton et al., 2014), and (ii) if the observation is indistinguishable from a member of the ensemble, the ratio between $\overline{sk}$ and the ensemble mean squared error ($\overline{mse}$), normalized by $\sqrt{(n+1)/2n}$ should be equal to one (e.g., Mendoza et al., 2012; Alvarez-Garreton et al., 2014). We use a temporal decorrelation length of one day for precipitation and two days for temperature, as in Piazzi et al. (2021).

A.2. Observational uncertainty

The characterization of observational errors is critical to determine the level of confidence assigned to observations and, therefore, the extent to which they are assimilated into hydrological models. In the case of streamflow, observational uncertainty primarily arises from instrumental errors and inaccuracies in the rating curve Clark et al. (2008). In this study, measurement noise is represented by a normal distribution with zero mean and a variance (σ_{obs}^2) that depends on the observed discharge Q_{obs} (Weerts & El Serafy, 2006; Clark et al., 2008; Piazzi et al., 2021):

$$\sigma_{obs}^2 = (\mathcal{E}_Q \cdot Q_{obs})^2 \qquad \text{(A.5)}$$

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

The parameter $\mathcal{E}_Q$ is derived from an uncertainty analysis of the rating curves used to estimate discharge at the streamflow gauging stations of the DGA network (DGA, 2024). In general, most discharge measurements fall within the mid-range of flows, where uncertainty is typically lower. In contrast, extreme flow events often lie in the extrapolated portion of the rating curves or are supported by a limited number of measurements, which inevitably increases their uncertainty.

Appendix B: Post-processing techniques

B.1. Quantile Mapping (QM)

The forecasted quantile at a given time step (daily, monthly, or seasonal) is first identified from the hindcast CDF, and the corresponding quantile in the observed CDF is then used to adjust the hindcast (Wood & Lettenmaier, 2006; Hashino et al., 2007; Madadgar et al., 2014):

$$Z_j = {F_{obs_j}}^{-1}(F_{fcst_j}(Y_j)) \tag{B.1}$$

Where Z_j is the corrected hindcast at time j; F_{obs_j} is the CDF of observations at time j; F_{fcs_j} is the CDF of hindcast at time j; and Y_j is the raw hindcast at time j.

B.2. Linear Regression (LR)

The hindcast ensemble member *i* and lead time *n* ($Q_{i,n}$), with $n = 1,\ldots,N_h$, can be expressed as:

$$Q_{i,n} = \overline{Q_n} + e_{i,n} \tag{B.2}$$

Where $\overline{Q_n}$ is the hindcast ensemble mean. If μ_f and μ_o are the respective means of the hindcast ensemble means and the observations over the training period, the corrected hindcast ensemble member for lead time *n* is given by:

$$Q_{i,n}^{LR} = \mu_o + \alpha(\overline{Q_n} - \mu_f) + \beta e_{i,n} \tag{B.3}$$

Where the parameters α and β are obtained following:

$$\alpha = \rho_{o\bar{f}} \frac{\sigma_o}{\sigma_{\bar{f}}} \tag{B.4}$$

$$\beta = \frac{\sigma_o}{\sigma_e} \sqrt{1 - \rho_{o\bar{f}}^2} \tag{B.5}$$

where $\rho_{o\bar{f}}$ is the correlation between the time series of observations and hindcast ensemble means, σ_o and $\sigma_{\bar{f}}$ are the standard deviations of observations and hindcast ensemble means, and σ_e is the square root of the average ensemble variance, all of them associated to lead time *n* and computed over the training period. All calculations are performed in standard normal space following the application of the Normal Quantile Transform (NQT).

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

B.3. Trace Weighting (TW)

For a given hindcast issued at initialization time t_0, let I^* denote the predictor value observed at t_0 and $\{I_i\}_{i=1}^{n}$ the predictor values corresponding to the n historical years that constitute the ESP ensemble. The absolute distance between the current predictor and each historical analog is computed as:

$$x_i = \mid I^* - I_i \mid, i = 1, \dots, n \tag{B.6}$$

and the resulting distances are sorted in ascending order to form the vector $\mathcal{X} = [x_{(1)}, x_{(2)}, \dots, x_{(n)}]$. A weight is then assigned to each historical year according to:

$$w_i = \left(1 - \frac{x_i}{x_{(k)}}\right)^{\lambda}, x_i \leq x_{(k)} \tag{B.7}$$

$$w_i = 0, x_i > x_{(k)} \tag{B.8}$$

where $k = \mathrm{round}(n/\alpha)$ defines the number of nearest neighbors retained and $x_{(k)}$ is the corresponding distance threshold. The parameter $\lambda > 0$ controls the sensitivity of the weights to distance — larger values concentrate probability on the closest analogs — whereas $\alpha \geq 1$ determines the fraction of ensemble members assigned nonzero weight. Weights are normalized to sum to unity, $p_i = w_i / \sum_j w_j$, and the post-processed ensemble is obtained by resampling the historical ESP traces with probabilities $\{p_i\}$.

The parameters λ and α are optimized independently for each catchment and initialization month by maximizing the median CRPSS evaluated over the full hindcast period in a leave-one-out cross-validation framework, using the unweighted ESP as the reference forecast. In this study, we test as predictor I the widely used Multivariate ENSO Index (Wolter & Timlin, 2011), as well as the observed streamflow and precipitation of 1, 2 and 3 months preceding the hindcast initialization, selected based on its correlation with seasonal runoff volumes across the study catchments.

B.4. Random Forest (RF)

For each hindcast initialization date (May 1 through September 1) and each lead time within the September– hindcast forecast horizon, a separate RF model is trained in a leave-one-year-out (LOYO) cross-validation framework, ensuring that no information from the verification year leaks into the training set. At each training step, the raw ESP ensemble hindcast for a given month and lead time is summarized into a set of statistical descriptors — the ensemble mean, median, standard deviation, and the 10th, 25th, 75th, and 90th percentiles — which serve as the predictor variables. The corresponding monthly observed streamflow values are the response variables. The RF model then learns a non-linear mapping from ensemble statistics to observations, capturing both conditional bias and spread errors that simpler linear methods may fail to represent.

During verification, the trained RF model receives the same set of ensemble-derived predictors computed from the current-year forecast and produces a bias-corrected deterministic estimate, corresponding to the average prediction across all trees in the forest. To ensure comparability across the 21 study catchments and to prevent overfitting given the limited sample size of the historical record (approximately 37 years per initialization month), the model configuration is held constant throughout the

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

study domain: the forest consists of 500 trees, the maximum number of terminal nodes per tree is fixed at 10, and the minimum node size is set to 5 observations. For the computational implementation we use the *randomForest* R package.

Data availability

The CR2METv2.5 dataset is available at https://www.cr2.cl/datos-productos-grillados (Boisier, 2023). Daily streamflow records from the Chilean Water Directorate (DGA) are available at https://www.cr2.cl/datos-de-caudales/ (Center for Climate and Resilience Research (CR2), 2026). Time series of the Multivariate ENSO Index (Wolter & Timlin, 2011) are available at https://psl.noaa.gov/enso/mei/.

Author contributions

JS, PM and MLZ conceptualized the study and designed the overall approach. JS conducted all the model simulations, generated the hindcasts, analyzed the results and created all the figures. JS and PM wrote the first draft of the paper. All the authors contributed to refine the analysis framework, discussed the results and contributed reviewing and editing the manuscript.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank James McPhee for his feedback during earlier stages of this work.

Financial support

JS and PM received financial support from ANID-PIA Project AFB230001 (AMTC). MLZ acknowledges support of the Center for Interdisciplinary Research on Disaster Risk, Resilience and Recovery (CIGIDEN R+), ANID/CIN250023.

References

Abrams, M., Crippen, R., & Fujisada, H. (2020, April 1). ASTER Global Digital Elevation Model (GDEM) and ASTER Global Water Body Dataset (ASTWBD). *Remote Sensing*. MDPI AG. https://doi.org/10.3390/rs12071156

Alvarez-Garreton, C., Ryu, D., Western, A. W., Crow, W. T., & Robertson, D. E. (2014). The impacts of assimilating satellite soil moisture into a rainfall-runoff model in a semi-arid catchment. *Journal of Hydrology*, *519*(PD), 2763–2774. https://doi.org/10.1016/j.jhydrol.2014.07.041

Alvarez-Garreton, Camila, Mendoza, P. A., Pablo Boisier, J., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., et al. (2018). The CAMELS-CL dataset: Catchment attributes and meteorology for large sample studies-Chile dataset. *Hydrology and Earth System Sciences*, *22*(11), 5817–5846. https://doi.org/10.5194/hess-22-5817-2018

Araya, D., Mendoza, P. A., Muñoz-castro, E., & McPhee, J. (2023). Towards robust seasonal streamflow forecasts in mountainous catchments: impact of calibration metric selection in hydrological modeling. *Hydrology and Earth System Sciences*, *27*(24), 4385–4408. https://doi.org/10.5194/hess-27-4385-2023

Arnal, L., Cloke, H. L., Stephens, E., Wetterhall, F., Prudhomme, C., Neumann, J., et al. (2018). Skilful seasonal forecasts of streamflow over Europe? *Hydrology and Earth System Sciences*, *22*(4), 2057–2072. https://doi.org/10.5194/hess-22-2057-2018

Arnal, L., Clark, M. P., Pietroniro, A., Vionnet, V., Casson, D. R., Whitfield, P. H., et al. (2024). FROSTBYTE: a reproducible data-driven workflow for probabilistic seasonal streamflow forecasting in snow-fed river basins across North America. *Hydrology and Earth System Sciences*, *28*(17), 4127–4155. https://doi.org/10.5194/hess-28-4127-2024

Arulampalam, M. S., Maskell, S., Gordon, N., & Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, *50*(2), 174–188. https://doi.org/10.1109/78.978374

Ayala, Á., Farías-Barahona, D., Huss, M., Pellicciotti, F., McPhee, J., & Farinotti, D. (2020). Glacier runoff variations since 1955 in the Maipo River Basin, semiarid Andes of central Chile. *The Cryosphere*, 1–39. https://doi.org/10.5194/tc-2019-233

Baez-Villanueva, O. M., Zambrano-Bigiarini, M., Mendoza, P. A., McNamara, I., Beck, H. E., Thurner, J., et al. (2021). On the selection of precipitation products for the regionalisation of hydrological model parameters. *Hydrology and Earth System Sciences*, *25*(11), 5805–5837. https://doi.org/10.5194/hess-25-5805-2021

Baker, S. A., Rajagopalan, B., & Wood, A. W. (2021). Enhancing Ensemble Seasonal Streamflow Forecasts in the Upper Colorado River Basin Using Multi-Model Climate Forecasts. *Journal of the American Water Resources Association*, *57*(6), 906–922. https://doi.org/10.1111/1752-1688.12960

Bergström, S. (1976). *Development and application of a conceptual runoff model for Scandinavian catchments*. Report RHO 7, SMHI, Norrköping, Sweden.

Boisier, J. P. (2023). CR2MET: A high-resolution precipitation and temperature dataset for the period 1960-2021 in continental Chile. *Zenodo*, *(v2.5) [Da*.

Bradley, A. A., Habib, M., & Schwartz, S. S. (2015). Climate index weighting of ensemble streamflow forecasts using a simple Bayesian approach. *Water Resources Research*, *51*(9), 7382–7400. https://doi.org/10.1002/2014WR016811

Breiman, L. (2001). Random Forests. *Machine Learning*, (45), 5–32. https://doi.org/10.1023/A:1010933404324

Broxton, P. D., van Leeuwen, W. J. D., Svoma, B. M., Walter, J., & Biederman, J. A. (2023). Subseasonal to seasonal streamflow forecasting in a semiarid watershed. *Journal of the American Water Resources Association*, *59*(6), 1493–1510. https://doi.org/10.1111/1752-1688.13147

Budyko, M. I. M. I. (1974). *Climate and Life*. *Journal of Range Management* (Vol. 28). London: Academic Press. https://doi.org/10.2307/3897455

Center for Climate and Resilience Research (CR2). (2026). Datos de caudales. https://www.cr2.cl/datos-de-caudales/.

Chan, W., Facer-Childs, K. A., Tanguy, M., Magee, E., Bulut, B., Stringer, N., et al. (2026). UK Hydrological Outlook using Historic Weather Analogues. *Hydrology and Earth System Sciences*, *30*(4), 905–927. https://doi.org/10.5194/hess-30-905-2026

Chevuturi, A., Tanguy, M., Facer-Childs, K., Martínez-de la Torre, A., Sarkar, S., Thober, S., et al. (2023). Improving global hydrological simulations through bias-correction and multi-model blending. *Journal of Hydrology*, *621*. https://doi.org/10.1016/j.jhydrol.2023.129607

Clark, M. P., Rupp, D. E., Woods, R. a., Zheng, X., Ibbitt, R. P., Slater, A. G., et al. (2008). Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model. *Advances in Water Resources*, *31*(10), 1309–1324. https://doi.org/10.1016/j.advwatres.2008.06.005

Day, G. N. (1985). Extended Streamflow Forecasting Using NWSRFS. *Journal of Water Resources Planning and Management*, *111*(2), 157–170. https://doi.org/10.1061/(ASCE)0733-9496(1985)111:2(157)

DeChant, C. M., & Moradkhani, H. (2011). Improving the characterization of initial condition for ensemble streamflow prediction using data assimilation. *Hydrology and Earth System Sciences*, *15*(11), 3399–3410. https://doi.org/10.5194/hess-15-3399-2011

DeChant, Caleb M., & Moradkhani, H. (2014). Toward a reliable prediction of seasonal forecast uncertainty: Addressing model and initial condition uncertainty with ensemble data assimilation and Sequential Bayesian Combination. *Journal of Hydrology*, *519*, 2967–2977. https://doi.org/10.1016/j.jhydrol.2014.05.045

DGA. (2024). *Actualización de modelos para pronóstico de caudales de deshielo integrando red nival DGA y su disposición en aplicación informática*. Santiago de Chile.

Du, Y., & Pechlivanidis, I. G. (2025). Hybrid approaches enhance hydrological model usability for local streamflow prediction. *Communications Earth and Environment*, *6*(1). https://doi.org/10.1038/s43247-025-02324-y

Duan, Q., Sorooshian, S., & Gupta, V. (1992). Effective and Efficient Global Optimization for Conceptual Rainfal-Runoff Models. *Water Resources Research*, *28*(4), 1015–1031.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*. https://doi.org/10.1029/94jc00572

Gobena, A. K., & Gan, T. Y. (2010). Incorporation of seasonal climate forecasts in the ensemble streamflow prediction system. *Journal of Hydrology*, *385*(1–4), 336–352. https://doi.org/10.1016/j.jhydrol.2010.03.002

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

Gordon, N. J., Salmond, D. J., & Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F Radar and Signal Processing*, *140*(2), 107. https://doi.org/10.1049/ip-f-2.1993.0015

Grantz, K., Rajagopalan, B., Clark, M., & Zagona, E. (2005). A technique for incorporating large-scale climate information in basin-scale ensemble streamflow forecasts. *Water Resources Research*, *41*, W10410. https://doi.org/10.1029/2004WR003467

Greuell, W., & Hutjes, R. W. A. (2023). Skill and sources of skill in seasonal streamflow hindcasts for South America made with ECMWF's SEAS5 and VIC. *Journal of Hydrology*, *617*. https://doi.org/10.1016/j.jhydrol.2022.128806

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Hashino, T., Bradley, A. A., & Schwartz, S. S. (2007). Evaluation of bias-correction methods for ensemble streamflow volume forecasts. *Hydrology and Earth System Sciences*, *11*(2), 939–950. https://doi.org/10.5194/hess-11-939-2007

Hernandez, D., Mendoza, P. A., Boisier, J. P., & Ricchetti, F. (2022). Hydrologic Sensitivities and ENSO Variability Across Hydrological Regimes in Central Chile (28°–41°S). *Water Resources Research*, *58*(9). https://doi.org/10.1029/2021WR031860

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

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

Jin Noh, S., Tachikawa, Y., Shiiba, M., & Kim, S. (2013). Sequential data assimilation for streamflow forecasting using a distributed hydrologic model: particle filtering and ensemble Kalman filtering. *Floods: From Risk to Opportunity*, *357*, 341–349.

Knoben, W. J. M., Raman, A., Gründemann, G. J., Kumar, M., Pietroniro, A., Shen, C., et al. (2025). Technical note: How many models do we need to simulate hydrologic processes across large geographical domains? *Hydrology and Earth System Sciences*, *29*(11), 2361–2375. https://doi.org/10.5194/hess-29-2361-2025

Ladson, A., Brown, R., Neal, B., & Nathan, R. (2013). A standard approach to baseflow separation using the Lyne and Hollick filter. *Australian Journal of Water Resources*, *17*(1). https://doi.org/10.7158/W12-028.2013.17.1

De Lannoy, G. J. M., Houser, P. R., Pauwels, V. R. N., & Verhoest, N. E. C. (2006). Assessment of model uncertainty for soil moisture through ensemble verification. *Journal of Geophysical Research*, *111*(D10), D10101. https://doi.org/10.1029/2005JD006367

Lucatero, D., Madsen, H., Refsgaard, J. C., Kidmose, J., & Jensen, K. H. (2018). Seasonal streamflow forecasts in the Ahlergaarde catchment, Denmark: The effect of preprocessing and post-processing on skill and statistical consistency. *Hydrology and Earth System Sciences*, *22*(7), 3601–3617. https://doi.org/10.5194/hess-22-3601-2018

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

Madadgar, S., Moradkhani, H., & Garen, D. (2014). Towards improved post-processing of hydrologic forecast ensembles. *Hydrological Processes*, *28*(1), 104–122. https://doi.org/10.1002/hyp.9562

Mendoza, P. A., McPhee, J., & Vargas, X. (2012). Uncertainty in flood forecasting: A distributed modeling approach in a sparse data catchment. *Water Resources Research*, *48*(9), W09532. https://doi.org/10.1029/2011WR011089

Mendoza, P. A., Rajagopalan, B., Clark, M. P., Cortés, G., & McPhee, J. (2014). A robust multimodel framework for ensemble seasonal hydroclimatic forecasts. *Water Resources Research*, *50*(7), 6030–6052. https://doi.org/10.1002/2014WR015426

Mendoza, P. A., Wood, A. W., Clark, E., Rothwell, E., Clark, M. P., Nijssen, B., et al. (2017). An intercomparison of approaches for improving operational seasonal streamflow forecasts. *Hydrology and Earth System Sciences*, *21*(7), 3915–3935. https://doi.org/10.5194/hess-21-3915-2017

Metref, S., Cosme, E., Le Lay, M., & Gailhard, J. (2023). Snow data assimilation for seasonal streamflow supply prediction in mountainous basins. *Hydrology and Earth System Sciences*, *27*(12), 2283–2299. https://doi.org/10.5194/hess-27-2283-2023

Micheletty, P., Perrot, D., Day, G., & Rittger, K. (2021). Assimilation of Ground and Satellite Snow Observations in a Distributed Hydrologic Model for Water Supply Forecasting. *Journal of the American Water Resources Association*. https://doi.org/10.1111/1752-1688.12975

Mo, K. C., Shukla, S., Lettenmaier, D. P., & Chen, L.-C. (2012). Do Climate Forecast System (CFSv2) forecasts improve seasonal soil moisture prediction? *Geophysical Research Letters*, *39*(23), n/a-n/a. https://doi.org/10.1029/2012GL053598

Muñoz-Castro, E. (2021). *Asimilación de observaciones in situ e información satelital para estimaciones de balance hídrico en Chile continental* (Thesis to obtain a MSc degree in Water Resources and Environment). Universidad de Chile.

Najafi, M. R., Moradkhani, H., & Piechota, T. C. (2012). Ensemble Streamflow Prediction: Climate signal weighting methods vs. Climate Forecast System Reanalysis. *Journal of Hydrology*, *442–443*, 105–116. https://doi.org/10.1016/j.jhydrol.2012.04.003

Nash, J., & Sutcliffe, J. (1970). River flow forecasting through conceptual models part I - A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005). Which potential evapotranspiration input for a lumped rainfall–runoff model? *Journal of Hydrology*, *303*(1–4), 290–306. https://doi.org/10.1016/j.jhydrol.2004.08.026

Pauwels, V. R. N., & De Lannoy, G. J. M. (2009). Ensemble-based assimilation of discharge into rainfall-runoff models: A comparison of approaches to mapping observational information to state space. *Water Resources Research*, *45*(8), W08428. https://doi.org/10.1029/2008WR007590

Pechlivanidis, I. G., Crochemore, L., Rosberg, J., & Bosshard, T. (2020). What Are the Key Drivers Controlling the Quality of Seasonal Streamflow Forecasts? *Water Resources Research*, *56*(6), 1–19. https://doi.org/10.1029/2019WR026987

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

Peñuela, A., Hutton, C., & Pianosi, F. (2020). Assessing the value of seasonal hydrological forecasts for improving water resource management: insights from a pilot application in the UK. *Hydrology and Earth System Sciences*, *24*(12), 6059–6073. https://doi.org/10.5194/hess-24-6059-2020

Petry, I., Fan, F. M., Siqueira, V. A., Collishonn, W., de Paiva, R. C. D., Quedi, E., et al. (2023). Seasonal streamflow forecasting in South America's largest rivers. *Journal of Hydrology: Regional Studies*, *49*. https://doi.org/10.1016/j.ejrh.2023.101487

Pham, D. T. (2001). Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Monthly Weather Review*, *129*(5), 1194–1207. https://doi.org/10.1175/1520-0493(2001)129<1194:SMFSDA>2.0.CO;2

Pham, L. T., Luo, L., & Finley, A. (2021). Evaluation of random forests for short-term daily streamflow forecasting in rainfall-And snowmelt-driven watersheds. *Hydrology and Earth System Sciences*, *25*(6), 2997–3015. https://doi.org/10.5194/hess-25-2997-2021

Piazzi, G., Thirel, G., Perrin, C., & Delaigue, O. (2021). Sequential Data Assimilation for Streamflow Forecasting: Assessing the Sensitivity to Uncertainties and Updated Variables of a Conceptual Hydrological Model at Basin Scale. *Water Resources Researche*, *57*(4), 1–28. https://doi.org/10.1029/2020WR028390

Rakovec, O., Weerts, a. H., Hazenberg, P., Torfs, P. J. J. F., & Uijlenhoet, R. (2012). State updating of a distributed hydrological model with Ensemble Kalman Filtering: effects of updating frequency and observation network density on forecast accuracy. *Hydrology and Earth System Sciences*, *16*(9), 3435–3449. https://doi.org/10.5194/hess-16-3435-2012

Regonda, S. K., Rajagopalan, B., Clark, M., & Zagona, E. (2006). A multimodel ensemble forecast framework: Application to spring seasonal flows in the Gunnison River Basin. *Water Resources Research*, *42*, W09404. https://doi.org/10.1029/2005WR004653

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., & Franks, S. W. (2010). Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research*, *46*(5), W05521. https://doi.org/10.1029/2009WR008328

Rosenberg, E. A., Wood, A. W., & Steinemann, A. C. (2011). Statistical applications of physically based hydrologic models to seasonal streamflow forecasts. *Water Resources Research*, *47*(3), W00H14. https://doi.org/10.1029/2010WR010101

Roulin, E., & Vannitsem, S. (2015). Post-processing of medium-range probabilistic hydrological forecasting: Impact of forcing, initial conditions and model errors. *Hydrological Processes*, *29*(6), 1434–1449. https://doi.org/10.1002/hyp.10259

Saavedra, D., Mendoza, P. A., Addor, N., Llauca, H., & Vargas, X. (2022). A multi-objective approach to select hydrological models and constrain structural uncertainties for climate impact assessments. *Hydrological Processes*, *36*(1). https://doi.org/10.1002/hyp.14446

Sabzipour, B., Arsenault, R., & Brissette, F. (2021). Evaluation of the potential of using subsets of historical climatological data for ensemble streamflow prediction (ESP) forecasting. *Journal of Hydrology*, *595*(October 2020), 125656. https://doi.org/10.1016/j.jhydrol.2020.125656

Salamon, P., & Feyen, L. (2009). Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter. *Journal of Hydrology*, *376*(3–4), 428–442. https://doi.org/10.1016/j.jhydrol.2009.07.051

Shukla, S., & Lettenmaier, D. P. (2011). Seasonal hydrologic prediction in the United States: Understanding the role of initial hydrologic conditions and seasonal climate forecast skill. *Hydrology and Earth System Sciences*, *15*(11), 3529–3538. https://doi.org/10.5194/hess-15-3529-2011

Skøien, J. O., Blöschl, G., Laaha, G., Pebesma, E., Parajka, J., & Viglione, A. (2014). rtop: An R package for interpolation of data with a variable spatial support, with an example from river networks. *Computers & Geosciences*, *67*, 180–190. https://doi.org/10.1016/j.cageo.2014.02.009

Smith, J. A., Day, G. N., & Kane, M. D. (1992). Nonparametric Framework for Long-range Streamflow Forecasting. *Journal of Water Resources Planning and Management*, *118*(1), 82–92. https://doi.org/10.1061/(ASCE)0733-9496(1992)118:1(82)

Sveinsson, O. G. B., Lall, U., Gaudet, J., Kushnir, Y., Zebiak, S., & Fortin, V. (2008). Analysis of climatic states and atmospheric circulation patterns that influence québec spring streamflows. *Journal of Hydrologic Engineering*, *13*(6), 411–425. https://doi.org/10.1061/(ASCE)1084-0699(2008)13:6(411)

Tachikawa, T., Hato, M., Kaku, M., & Iwasaki, A. (2011). Characteristics of ASTER GDEM version 2. *International Geoscience and Remote Sensing Symposium (IGARSS)*, (January), 3657–3660. https://doi.org/10.1109/IGARSS.2011.6050017

Tanguy, M., Eastman, M., Chevuturi, A., Magee, E., Cooper, E., Johnson, R. H. B., et al. (2025). Optimising ensemble streamflow predictions with bias correction and data assimilation techniques. *Hydrology and Earth System Sciences*, *29*(6), 1587–1614. https://doi.org/10.5194/hess-29-1587-2025

Thébault, C., Perrin, C., Legrand, S., Andréassian, V., Thirel, G., & Delaigue, O. (2025). What can be expected from a semi-distributed multi-model approach for streamflow forecasting? Tailoring the structure and size of a super-ensemble on the Rhône basin. *Journal of Hydrology*, *661*. https://doi.org/10.1016/j.jhydrol.2025.133589

Toum, E., Masiokas, Mariano, H., Villalba, R., Pitte, P., & Ruiz, L. (2021). The HBV.IANIGLA Hydrological Model. *The R Journal*, *13*(1), 378. https://doi.org/10.32614/RJ-2021-059

Twedt, T. M., Schaake, J. C. Jr., & Peck, E. L. (1977). National Weather Service extended streamflow prediction. In *Western Snow Conference* (pp. 52–57).

Uysal, G., Boucher, M. A., Mathieu, C., Alvarado-Montero, R., & Şorman, A. A. (2026). A data-assimilated SEAS5 forecasting framework for seasonal hydropower inflows in a snow-dominated basin. *Journal of Hydrology*, *671*. https://doi.org/10.1016/j.jhydrol.2026.135279

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. *Journal of Hydrometeorology*, *20*(1), 99–115. https://doi.org/10.1175/JHM-D-18-0040.1

https://doi.org/10.5194/egusphere-2026-4067
Preprint. Discussion started: 28 August 2026

Weerts, A. H., & El Serafy, G. Y. H. (2006). Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models. *Water Resources Research*, *42*(9), 1–17. https://doi.org/10.1029/2005WR004093

Werner, K., Brandon, D., Clark, M., & Gangopadhyay, S. (2004a). Climate Index Weighting Schemes for NWS ESP-Based Seasonal Volume Forecasts. *Journal of Hydrometeorology*, *5*(6), 1076–1090. https://doi.org/10.1175/JHM-381.1

Werner, K., Brandon, D., Clark, M., & Gangopadhyay, S. (2004b). Climate Index Weighting Schemes for NWS ESP-Based Seasonal Volume Forecasts. *Journal of Hydrometeorology*, *5*(6), 1076–1090. https://doi.org/10.1175/JHM-381.1

Wilks, D. S. (2019). *Statistical Methods in the Atmospheric Sciences* (Fourth Edi). Elsevier.

Wolter, K., & Timlin, M. S. (2011). El Niño/Southern Oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index (MEI.ext). *International Journal of Climatology*, *31*(7), 1074–1087. https://doi.org/10.1002/joc.2336

Wood, A. W., & Lettenmaier, D. P. (2006). A Test Bed for New Seasonal Hydrologic Forecasting Approaches in the Western United States. *Bulletin of the American Meteorological Society*, *87*(12), 1699–1712. https://doi.org/10.1175/BAMS-87-12-1699

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

Wood, A. W., Kumar, A., & Lettenmaier, D. P. (2005). A retrospective assessment of National Centers for Environmental prediction climate model-based ensemble hydrologic forecasting in the western United States. *Journal of Geophysical Research D: Atmospheres*, *110*(4), 1–16. https://doi.org/10.1029/2004JD004508

Wood, A. W., Sankarasubramanian, A., & Mendoza, P. (2018). Seasonal Ensemble Forecast Post-processing. In *Handbook of Hydrometeorological Ensemble Forecasting* (pp. 1–27). Berlin, Heidelberg, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-40457-3_37-1

Zhao, Y., Feng, D., Yu, L., Wang, X., Chen, Y., Bai, Y., et al. (2016). Detailed dynamic land cover mapping of Chile: Accuracy improvement by integrating multi-temporal data. *Remote Sensing of Environment*, *183*, 170–185. https://doi.org/10.1016/j.rse.2016.05.016