Bootstrap-Based Imputation of Monthly Hydroclimatic Series Using STL Decomposition, ARFIMA/ARIMAX Modeling, and ENSO Stratification

Mauricio Javier Victoria Niño[1]

[1]Independent Researcher, Cali, Colombia; hidratecsa@gmail.com; ORCID: 0009-0003-4328-5691

Preprint submitted to EngrXiv. Not peer-reviewed. Source code (`ACEBoot` v1.0.0) available at https://github.com/MauricioVictoriaN/ACEBoot_v1.0.0. Comments: hidratecsa@gmail.com.

Highlights

- Six-module pipeline integrating STL, ARFIMA/ARIMAX, Gaussian copulas, MBB and ENSO stratification.
- Structural break detection via unified Worsley–Strucchange–ENSO filter preserves physically motivated change points.
- IC95 empirical coverage of 93.83 % validates bootstrap uncertainty quantification on hold-out data.
- ACF preservation error of 0.0028 confirms temporal structure integrity post-imputation.
- Open-source `ACEBoot` v1.0.0 R implementation with a fully Spanish-localised interactive Shiny interface, targeting the Ibero-American hydrological community.

Abstract

Background. Hydroclimatic time series with missing records compromise trend detection, hydrological modelling, and water resources management. Existing imputation methods rarely integrate seasonal decomposition, long-memory stochastic modelling, multivariate dependence, ENSO-driven non-stationarity, and uncertainty quantification within a single reproducible pipeline.

Methods. We present `ACEBoot` v1.0.0, a six-module bootstrap-based multimethod framework for monthly hydroclimatic imputation. The pipeline combines: (M1) stationarity and long-memory diagnosis via ADF, KPSS, Phillips–Perron and DF-GLS tests plus Geweke–Porter-Hudak fractional differencing; (M2) robust STL seasonal decomposition; (M3) ARFIMA or ARIMAX-ONI modelling with Gaussian copula over residuals and Moving Block Bootstrap (MBB) uncertainty propagation; (M4) structural break detection via a unified Worsley–Strucchange–ENSO filter; (M5) targeted re-imputation over confirmed break segments; and (M6) hold-out cross-validation with IC95 coverage, RMSE, MAE, ACF error and Lilliefors normality diagnostics. The framework is demonstrated on a 46-year (1977–2022) monthly streamflow series (405 observed, 147 missing; 26.6 %) from the Jamundí-Carretera gauge (code 2622100403, CVC monitoring network), Río Jamundí sub-basin (sub-zone 2629), Cauca Hydrographic Zone (ZH-26), Colombia.

Key Results. The ARIMAX-ONI(3,1,2) model, selected after majority-vote stationarity assessment ($n_{\text{train}} = 324$, Precision profile, CV = 20 %) and long-memory detection ($d = 0.311$, $H = 0.811$), achieves IC95 empirical coverage of 93.83 % (nominal 95 %), ACF preservation error of 0.0028, bias of 0.37 m^3/s (1.0 % of mean), and Lilliefors $p = 0.0978$. A structural break at February 1995 (+41.7 %

mean shift, validated by El Niño ONI = 0.72) is detected by Strucchange and confirmed by ENSO stratification. The imputed series preserves mean ($+0.4\,\%$) while exhibiting expected variance reduction ($-22.5\,\%$) attributable to conditional mean imputation.

Conclusions. `ACEBoot` v1.0.0 produces statistically valid imputations for monthly hydroclimatic series with complex missing-data patterns, integrating climate covariate information and structural break correction within a fully reproducible open-source framework.

Keywords: imputation; missing data; streamflow; STL decomposition; ARFIMA; ARIMAX; Gaussian copula; moving block bootstrap; ENSO; structural break; Colombia.

Author Contributions (CRediT). *Mauricio Javier Victoria Niño*: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization, Funding acquisition (none).

1 Introduction

1.1 Context and Motivation

Continuous, high-quality hydroclimatic records are indispensable for flood frequency analysis, drought monitoring, calibration of rainfall-runoff models, water allocation planning, and attribution of long-term hydrological trends (WMO, 2008; Mishra and Singh, 2011). In practice, however, gauge networks in developing regions are affected by persistent data gaps arising from instrument failure, vandalism, operator changes, extreme events, and administrative interruptions. In Colombia, the monitoring network of IDEAM and regional autonomous corporations (CAR) reports missing-data rates that routinely exceed 20 % for individual monthly streamflow stations (IDEAM, 2014; CVC, 2025). Missing records are rarely random: gaps often cluster around ENSO episodes, wet-season peak flows, and administrative transitions, introducing systematic biases if handled by naïve deletion or simple substitution.

1.2 Problem Statement

When missing values represent a substantial fraction of a record (here 26.6 %), complete-case analysis discards statistically irreplaceable information; mean substitution attenuates variance; linear interpolation ignores seasonal structure; and single-imputation methods systematically underestimate uncertainty (Rubin, 1987; Little and Rubin, 2019). For hydroclimatic series, four features compound the problem: (i) marked seasonal cycles driven by the Inter-Tropical Convergence Zone (ITCZ) and bimodal precipitation regimes in the Andes; (ii) long-range dependence expressed as the Hurst phenomenon (Hurst, 1951); (iii) exogenous climate forcing, particularly the El Niño–Southern Oscillation (ENSO), which modulates streamflow and precipitation through teleconnections in tropical South America (Poveda and Mesa, 1997); and (iv) structural breaks associated with land-use change, river engineering, or regime shifts (Zeileis et al., 2002). No single existing method simultaneously addresses all four features.

1.3 State of the Art

Table 1 summarises the principal imputation methods applied in hydrology and their key limitations. Classical linear interpolation and period-mean substitution are computationally trivial but ignore temporal autocorrelation and provide no uncertainty measure (WMO, 2008). Expectation-Maximisation (EM) imputation under Gaussian assumptions restores the full likelihood but breaks down under non-normality and non-stationarity common in streamflow (Dempster et al., 1977). Multiple imputation by chained equations (MICE) (van Buuren, 2018) handles mixed-type predictors but treats observations as independent and

is poorly suited to long ordered sequences. SARIMA-based imputation (Box et al., 2015) captures seasonality and short-range dependence but cannot model fractional integration or cross-variable forcing; its standard bootstrap variant does not account for block dependence. Kriging-based spatio-temporal methods (Snepvangers et al., 2003) require dense station networks unavailable in many tropical basins. Neural network approaches (LSTM, transformer architectures) (Chen et al., 2018; Kratzert et al., 2018) achieve competitive point accuracy but demand large training corpora, offer limited interpretability, and provide no analytical uncertainty intervals. Copula-based imputation models multivariate dependence correctly (Salvadori et al., 2007) but has rarely been combined with time-series decomposition and external climate covariates in a single framework.

Table 1: Summary of imputation methods and key limitations in hydrology.

Method	**Key limitation**	**Reference**
Mean / median substitution	Attenuates variance; ignores temporal structure	WMO (2008)
Linear interpolation	No uncertainty; ignores seasonality	WMO (2008)
EM algorithm	Requires Gaussian/stationary data	Dempster et al. (1977)
MICE	Assumes observation independence	van Buuren (2018)
SARIMA imputation	No fractional memory; no exogenous forcing	Box et al. (2015)
Spatio-temporal Kriging	Requires dense network	Snepvangers et al. (2003)
LSTM / deep learning	Data-intensive; opaque; no analytical IC	Chen et al. (2018); Kratzert et al. (2018)
Copula imputation	Not combined with decomp. or climate covariates	Salvadori et al. (2007)

1.4 Identified Gap

No published framework simultaneously integrates: seasonal decomposition (STL), fractional differencing (ARFIMA), exogenous climate covariates (ARIMAX-ONI), multivariate Gaussian copula on residuals, Moving Block Bootstrap (MBB) for dependence-preserving uncertainty, structural break detection with ENSO validation, and targeted re-imputation over confirmed break segments. This is the gap `ACEBoot` v1.0.0 fills.

1.5 Objectives and Contributions

This paper presents `ACEBoot` v1.0.0, an open-source R pipeline that:

1. Integrates six methodological modules into a coherent, automated workflow with module-hierarchy logic that selects the most appropriate imputation model (ARFIMA, ARIMAX-ONI, ENSO-stratified, or Gamma-mixture for precipitation) based on diagnostic tests.
2. Applies a unified Worsley–Strucchange–ENSO structural break filter that distinguishes physically meaningful regime changes from spurious artefacts.
3. Quantifies imputation uncertainty via MBB-derived 95 % bootstrap intervals and validates empirical coverage on a hold-out set.
4. Is demonstrated on a 46-year monthly streamflow record from the Río Jamundí (Valle del Cauca, Colombia) as a reproducible case study.
5. Is distributed as an open-source R script and a fully Spanish-localised interactive Shiny application at `https://github.com/MauricioVictoriaN/ACEBoot_v1.0.0`, lowering the adoption barrier for operational hydrologists across Spain and Latin America.

2 Study Area and Data

2.1 Hydrographic Context

The Río Jamundí drains 591 km^2 of the western flank of the Colombian Western Cordillera and the flat plain of the Valle del Cauca Department. It belongs to Hydrographic Zone 26 (Río Cauca) and Sub-zone 2629 (Ríos Claro and Jamundí) of Colombia's national hydrographic classification. The basin spans elevations from approximately 950 m to 3,250 m a.s.l., with a bimodal precipitation regime modulated by the semi-annual passage of the ITCZ and attenuated by the rain-shadow effect of the Western Cordillera.

The streamflow regime of the Río Jamundí exhibits two high-water periods (April–May and October–November) and two low-water periods (July–August and December–January), consistent with the regional bimodal pattern documented by Poveda and Mesa (1997) and IDEAM (2014). Regional teleconnections with ENSO are well established: El Niño years are associated with below-normal streamflow in the Cauca Valley due to reduced Pacific moisture transport, while La Niña years amplify precipitation and runoff (Poveda and Mesa, 1997; Espinoza et al., 2009).

2.2 Streamflow Station

Monthly streamflow data were obtained from the gauge *Jamundí-Carretera* (station code 2622100403, type LG – Limnigraph), operated by the Corporación Autónoma Regional del Valle del Cauca (CVC) (CVC, 2025). Station metadata are summarised in Table 2. The raw series spans January 1977 to December 2022 (552 monthly values). Of these, 405 are observed (73.4 %) and 147 are missing (26.6 %). Two calendar years (2016–2017) are entirely absent from the source dataset; four additional years (2019–2022) have fewer than four monthly observations. The missing-data pattern is non-random (not MCAR): gaps cluster around the 2015–2016 strong El Niño event and the monitoring interruption of 2016–2017, consistent with the operational history of the CVC network during that period.

Missing Data Mechanism

Following the taxonomy of Rubin (1987) and Little and Rubin (2019), the missing-data mechanism is assessed as follows. Let $Y = (Y_{\text{obs}}, Y_{\text{mis}})$ be the complete monthly streamflow vector and $R_t = 1$ if month t is observed ($R_t = 0$ if missing).

MCAR (Missing Completely At Random, $P(R \mid Y) = P(R)$) is rejected: the gap pattern concentrates around ENSO events and known operational interruptions, demonstrating that missingness depends on covariates (time, ENSO phase).

MAR (Missing At Random, $P(R \mid Y) = P(R \mid Y_{\text{obs}})$) is the working assumption of the `ACEBoot` v1.0.0 pipeline. Specifically, the probability that a given month is missing is assumed to depend on observable covariates—the ONI index and surrounding observed streamflow values—but not on the unobserved streamflow value itself, conditional on those covariates. This is operationally justified because the primary gap drivers (ENSO-related gauge failure, operator absence) are recorded in the ONI sequence and are independent of the instantaneous discharge magnitude.

MNAR (Missing Not At Random) cannot be ruled out with certainty: if high-flow months are disproportionately missing (e.g., due to gauge damage during floods), then $P(R_t = 0 \mid Y_t)$ depends on Y_t, and the imputed values would be systematically biased downward. The residual diagnostics (Section 4.8) show no systematic negative bias in the available training periods, providing empirical support for the MAR assumption. MNAR sensitivity analysis remains a limitation (see Section 5.6).

Table 2: Metadata of the Jamundí-Carretera streamflow gauge.

Parameter	Value
Station code	2622100403
Station name	Jamundí-Carretera
Station type	LG (Limnigraph)
Operator	CVC
Latitude	3°15′N (approx.)
Longitude	76°31′W (approx.)
Hydrographic zone	26 (Río Cauca)
Sub-zone	2629 (Ríos Claro and Jamundí)
Record period	1977-01 to 2022-12
Total months	552
Observed months	405 (73.4 %)
Missing months	147 (26.6 %)

2.3 Oceanic Niño Index

Monthly ONI values (1950–2024, 900 records) were obtained from the National Oceanic and Atmospheric Administration (NOAA) (Wolter and Timlin, 2011). The ONI is defined as the 3-month running mean of sea-surface temperature (SST) anomalies in the Niño-3.4 region (5°N–5°S, 170°–120°W). ENSO phases were classified as El Niño (ONI $\geq +0.5$ °C for ≥ 5 consecutive months), La Niña (ONI ≤ -0.5 °C for ≥ 5 consecutive months), and Neutral otherwise.

2.4 Descriptive Statistics

Table 3 presents descriptive statistics of the observed monthly streamflow series. The mean discharge is 36.90 m^3/s with a standard deviation of 24.35 m^3/s (coefficient of variation CV = 66 %). The minimum recorded value of 1.27 m^3/s makes percentage error metrics (MAPE) unreliable and motivates the use of RMSE and MAE as primary evaluation criteria. The distribution is positively skewed ($\gamma_1 = 1.84$), justifying the log-transformation applied in the ARIMAX-ONI module.

Table 3: Descriptive statistics of the monthly streamflow at Jamundí-Carretera (1977–2022). Original series: $n_{\text{obs}} = 405$ observed values. Imputed series: complete 552-value record ($n_{\text{obs}} = 405$ observed + $n_{\text{imp}} = 147$ imputed medians). The SD reduction reflects the expected variance attenuation of conditional mean imputation.

Statistic	Original series	Imputed series
Mean (m^3/s)	36.90	37.06
SD (m^3/s)	24.35	18.88
Minimum (m^3/s)	1.27	1.37
Maximum (m^3/s)	146.12	111.00
Δ Mean (%)	—	$+0.4$
Δ SD (%)	—	-22.5

3 Methodology

This section presents the six modules of the `ACEBoot` v1.0.0 pipeline in the order they are executed. The pipeline is implemented entirely in R 4.x and its architecture is summarised schematically below (see also https://github.com/MauricioVictoriaN/ACEBoot_v1.0.0 for the full algorithmic flowchart).

3.1 Statistical Model Specification

`ACEBoot` v1.0.0 is a **sequential plug-in estimator**, not a joint Bayesian model. Each module conditions on the output of the preceding module; full joint inference over all parameters simultaneously is not performed. This is an explicit design choice motivated by computational tractability and modularity. The formal model for the log-transformed (streamflow) or untransformed (precipitation) series y_t is:

$$y_t = T_t(\boldsymbol{\gamma}) + S_t(\boldsymbol{\delta}) + R_t, \qquad t = 1, \ldots, T, \tag{1}$$

where $T_t(\boldsymbol{\gamma})$ is the Loess trend with bandwidth parameters $\boldsymbol{\gamma}$, $S_t(\boldsymbol{\delta})$ is the Loess seasonal component, and R_t is the stationary remainder.

The remainder follows an ARIMAX(p, d, q)-ONI process:

$$\phi(B)(1 - B)^d (R_t - \beta\,\mathrm{ONI}_t) = \theta(B)\varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma^2), \tag{2}$$

with parameter vector $\boldsymbol{\theta} = (\phi_1, \ldots, \phi_p,\, d,\, \theta_1, \ldots, \theta_q,\, \beta,\, \sigma^2)^\top$.

Each module contributes a partial likelihood:

$$\mathcal{L}_1(\boldsymbol{\gamma}, \boldsymbol{\delta}) = \prod_{t \in \mathcal{T}_{\text{obs}}} f_{\text{Loess}}(y_t \mid T_t, S_t), \tag{3}$$

$$\mathcal{L}_2(\boldsymbol{\theta}) = \prod_{t \in \mathcal{T}_{\text{obs}}} \phi\!\left(\frac{R_t - \mu_{R,t}}{\sigma}\right) \frac{1}{\sigma}, \tag{4}$$

$$\mathcal{L}_3(\boldsymbol{\Sigma}) = \prod_{t \in \mathcal{T}_{\text{obs}}} c_\Sigma\big(\hat{F}(\varepsilon_{t-1}), \hat{F}(\varepsilon_{t-2}), \hat{F}(\varepsilon_{t-3})\big), \tag{5}$$

where c_Σ is the Gaussian copula density with correlation matrix $\boldsymbol{\Sigma}$ (Equation 18), and $\hat{F}$ denotes the empirical CDF. Parameters are estimated sequentially: $(\hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\delta}})$ by STL (weighted Loess), $\hat{\boldsymbol{\theta}}$ by MLE via `fracdiff` and `auto.arima`, and $\hat{\boldsymbol{\Sigma}}$ by the Inference Functions for Margins (IFM) method of Joe (1997), which is the standard two-step copula estimator (Genest and Rivest, 1993). Under IFM, the asymptotic uncertainty of the copula estimate is larger than under joint MLE, but the estimator is consistent and the inference remains frequentist (Joe, 1997).

Under the MAR assumption (Section 2.2), the observed-data likelihood factors as $\mathcal{L}(\boldsymbol{\Theta} \mid Y_{\text{obs}}) \propto \mathcal{L}_1 \cdot \mathcal{L}_2 \cdot \mathcal{L}_3$, and the predictive distribution of the missing values at the MLE plug-in estimate $\hat{\boldsymbol{\Theta}}$ is:

$$Y_{\text{mis}}^{(b)} \sim p\Big(Y_{\text{mis}} \mid Y_{\text{obs}},\, \hat{\boldsymbol{\Theta}},\, \mathrm{ONI}\Big), \qquad b = 1, \ldots, N, \tag{6}$$

where $N = 500$ is the number of bootstrap draws. Because $\hat{\boldsymbol{\Theta}}$ is substituted rather than integrated out, the interval width underestimates the full posterior uncertainty by the MLE estimation error — an acknowledged limitation (see Section 5.6). The MBB approximates samples from this distribution while preserving block-level temporal dependence (Section 3.3.4).

Correct identification of the integration order of the time series is critical for ARIMA-family models. `ACEBoot` v1.0.0 applies four formal unit-root tests and one long-memory estimator, combining results via majority vote.

3.1.1 Unit-root tests

Augmented Dickey-Fuller (ADF). The ADF test (Dickey and Fuller, 1979) is based on the auxiliary regression:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{k=1}^{p} \delta_k \Delta y_{t-k} + \varepsilon_t, \tag{7}$$

where the null hypothesis H_0: $\gamma = 0$ corresponds to a unit root. The number of lags p is selected by AIC. Rejection of H_0 at the 5 % significance level is classified as stationarity ($I(0)$).

Kwiatkowski–Phillips–Schmidt–Shin (KPSS). The KPSS test (Kwiatkowski et al., 1992) reverses the null hypothesis to H_0: level stationarity, decomposing the series as $y_t = r_t + \beta t + \varepsilon_t$, where r_t is a random walk. The test statistic is:

$$\text{KPSS} = \frac{1}{T^2 \hat{\sigma}_\varepsilon^2} \sum_{t=1}^{T} \hat{S}_t^2, \qquad \hat{S}_t = \sum_{i=1}^{t} \hat{\varepsilon}_i, \tag{8}$$

where $\hat{\sigma}_\varepsilon^2$ is the long-run variance of residuals. Failure to reject H_0 is classified as stationarity.

Phillips-Perron (PP). The PP test (Phillips and Perron, 1988) non-parametrically corrects the ADF t-statistic for serial correlation and heteroskedasticity:

$$Z_\alpha = T(\hat{\gamma} - 1) - \frac{T^2 s^2}{2\lambda^2} \left(\text{se}(\hat{\gamma})\right)^2, \tag{9}$$

where λ^2 is the long-run variance estimator and s^2 is the regression residual variance. The same H_0 as ADF applies.

Dickey-Fuller GLS (DF-GLS). The DF-GLS test (Elliott et al., 1996) is the locally best invariant (LBI) modification of the ADF test, achieved by GLS-detrending the series before computing the unit-root statistic, yielding substantially higher power against near-unit-root alternatives.

3.1.2 Majority-vote verdict

The verdict is determined by majority vote of the three most powerful available tests (ADF, PP, DF-GLS), excluding KPSS when inconclusive (NA). If $\geq 2/3$ tests reject H_0, the series is classified $I(0)$. For the Jamundí-Carretera series: ADF rejects ($t = -4.17 < -2.87$), PP rejects ($t = -12.06 < -2.87$), DF-GLS does not reject ($t = -0.68 > -1.94$). Verdict: $I(0)$, 2/3 tests.

3.1.3 Long-memory estimation

The fractional differencing parameter d is estimated via the Geweke-Porter-Hudak (GPH) log-periodogram regression (Geweke and Porter-Hudak, 1983):

$$\ln I(\omega_j) = c - d \ln\left[4 \sin^2(\omega_j/2)\right] + u_j, \qquad j = 1, \ldots, m, \tag{10}$$

where $I(\omega_j)$ is the periodogram at Fourier frequency $\omega_j = 2\pi j/T$, $m = \lfloor T^{0.8} \rfloor$, and u_j are asymptotically i.i.d. errors. The Hurst exponent is recovered as $H = d + 0.5$.

For the Jamundí series, $\hat{d} = 0.311$ ($H = 0.811$), indicating weak long memory. Montanari et al. (1997) report $d \in [0.10, 0.45]$ for Andean monthly streamflow; $\hat{d} = 0.311$ falls within this range. The corresponding stationarity condition ($d < 0.5$) is satisfied, confirming the $I(0)$ verdict. The ARFIMA model is therefore appropriate. However, since $\hat{d}$ is close to but not exceeding 0.45, and given the physically dominant ENSO

teleconnection, the pipeline selects ARIMAX-ONI as the primary model (see Module 3), consistent with the module hierarchy.

3.1.4 Seasonal decomposition pre-test

The proportion of variance explained by the seasonal component is evaluated:

$$\rho_{\text{seas}} = \frac{\text{Var}(S_t)}{\text{Var}(S_t) + \text{Var}(R_t)}, \tag{11}$$

where S_t and R_t are the STL seasonal and remainder components, respectively. If $\rho_{\text{seas}} \leq 0.05$, no significant seasonality is detected and the STL decomposition (Module 2) is bypassed. For the Precision profile ($n_{\text{train}} = 324$), seasonality is not detected; for the Rapid profile ($n_{\text{train}} = 365$), it is.

3.2 Module 2: STL Seasonal Decomposition

When seasonality is detected, the series y_t is decomposed by the Seasonal and Trend decomposition using Loess (STL) algorithm of Cleveland et al. (1990):

$$y_t = T_t + S_t + R_t, \tag{12}$$

where T_t is the trend component estimated by a sequence of Loess smoothers with window n_w, S_t is the seasonal component extracted by per-period Loess with window n_s, and R_t is the remainder. STL is robust to outliers through an iteratively reweighted procedure: at each pass, residual-based weights ρ_t downweight anomalous observations and the decomposition is recomputed. The imputation model is fitted on R_t; the final imputed value is:

$$\hat{y}_t^{(k)} = \hat{T}_t + \hat{S}_t + \hat{R}_t^{(k)}, \tag{13}$$

where superscript (k) denotes the k-th bootstrap realisation. When no seasonality is detected (as in the 20 % CV profile for the Jamundí series), y_t is modelled directly after first-differencing.

3.3 Module 3: ARFIMA/ARIMAX Imputation with Gaussian Copula and Moving Block Bootstrap

Module 3 is the core imputation engine. It selects among four model families based on the M1 verdict and user configuration, following a fixed hierarchy: M3d (precipitation mixture) $\succ$ M3b (ARIMAX-ONI) $\succ$ M3c (ENSO-stratified) $\succ$ M3a (ARFIMA standard).

3.3.1 ARFIMA modelling (M3a)

For the remainder R_t, or directly for the differenced series, an ARFIMA(p, d, q) model is fitted (Granger and Joyeux, 1980):

$$\phi(B)(1 - B)^d R_t = \theta(B)\varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2), \tag{14}$$

where $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ is the AR polynomial, $\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$ is the MA polynomial, B is the backshift operator, and $(1 - B)^d$ is the fractional difference operator defined by the binomial expansion:

$$(1 - B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k = \sum_{k=0}^{\infty} \frac{\Gamma(k - d)}{\Gamma(-d)\Gamma(k + 1)} B^k. \tag{15}$$

For $0 < d < 0.5$ the process is stationary and exhibits long memory with autocorrelations decaying hyperbolically as $\rho(k) \sim Ck^{2d-1}$ as $k \to \infty$ (Hosking, 1981). Model order is selected by minimising AIC over the grid $(p, d, q) \in \{0, \ldots, 3\} \times \{0, 0.5\} \times \{0, \ldots, 3\}$ using the `fracdiff` package (Fraley et al., 2020).

3.3.2 ARIMAX with ONI covariate (M3b)

The ARIMAX-ONI model extends the ARFIMA framework by incorporating the Oceanic Niño Index as an exogenous covariate:

$$\phi(B)(1 - B)^d(R_t - \beta \cdot \text{ONI}_t) = \theta(B)\varepsilon_t, \tag{16}$$

where β is the regression coefficient of the detrended, de-seasonalised streamflow residual on contemporaneous ONI. This specification captures the well-documented ENSO–streamflow teleconnection for the Cauca Valley without imposing a lag structure, consistent with Poveda and Mesa (1997). Conditional on the observed ONI, the ARIMAX model can impute missing residuals using the climate covariate as an anchor even when surrounding observations are absent. The selected model for the Jamundí series is ARIMAX-ONI(3,1,2), i.e. ARIMA$(3, 1, 2)$ with ONI regressor.

3.3.3 Gaussian Copula on residuals

The model innovations $\hat{\varepsilon}_t$ are transformed to standard normal through the empirical CDF (Genest and Rivest, 1993):

$$u_t = F_n(\hat{\varepsilon}_t), \qquad z_t = \Phi^{-1}(u_t), \tag{17}$$

where F_n is the empirical distribution function and Φ^{-1} is the standard normal quantile function. The joint dependence at lags 1 to 3 is captured through a Gaussian copula. The lag order 1–3 is motivated by the MA(4) order of the ARIMAX model: the ENSO-driven innovation at time t decays over 3–4 months, so lags 1–3 capture the residual dependence not already absorbed by the ARIMA polynomial. This is supported empirically by the ACF of imputation errors (Figure 7), which shows all autocorrelations beyond lag 1 within the 95 % confidence band, confirming that the three-lag copula structure adequately captures the residual dependence without over-parameterisation. The copula is:

$$C(u_{t-1}, u_{t-2}, u_{t-3}) = \Phi_\Sigma\big(\Phi^{-1}(u_{t-1}),\, \Phi^{-1}(u_{t-2}),\, \Phi^{-1}(u_{t-3})\big)\,, \tag{18}$$

where Φ_Σ is the multivariate normal CDF with correlation matrix $\boldsymbol{\Sigma}$ estimated by the IFM method (Joe, 1997). The estimated correlation matrix for the Jamundí case ($\hat{\rho}_{1,2} = 0.41$, $\hat{\rho}_{1,3} = 0.14$, $\hat{\rho}_{2,3} = 0.21$) is provided in the supplementary repository. The Gaussian copula is selected for tractability; vine and Archimedean copulas are noted as alternatives for heavy-tailed residual processes (see Section 5.6). The copula parameters are estimated sequentially after the ARIMA fit (IFM/two-step approach); the resulting uncertainty in $\hat{\boldsymbol{\Sigma}}$ is not propagated to the bootstrap intervals, which is an acknowledged limitation (Joe, 1997).

3.3.4 Moving Block Bootstrap (MBB)

The MBB procedure of Künsch (1989) preserves the temporal dependence of innovations within blocks of length ℓ:

1. Divide the observed innovation sequence $\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n$ into overlapping blocks $B_k = (\hat{\varepsilon}_k, \ldots, \hat{\varepsilon}_{k+\ell-1})$ for $k = 1, \ldots, n - \ell + 1$.
2. Draw $\lceil n/\ell \rceil$ blocks with replacement to form a pseudo-series $\varepsilon^*_1, \ldots, \varepsilon^*_n$.

3. Generate a bootstrap realisation of the imputed residual by replacing missing innovations with ε^* values at the corresponding positions.

4. Reconstruct the imputed series via Equation (13).

Block length is set to $\ell = \lfloor\sqrt{n}\rfloor$, a widely used heuristic in applied hydrology (Künsch, 1989). Optimal block length theory prescribes $O(n^{1/3})$ for bias-variance balance (Lahiri, 2003), which yields $\ell \approx 8$ for $n = 552$; the heuristic $\ell = 23$ ($O(n^{1/2})$) is more conservative (larger blocks reduce variance at the cost of higher bias) but the Monte Carlo results (Table 5) confirm near-optimal IC95 coverage at this value. Sensitivity to ℓ is documented in Section 4.1. The bootstrap is implemented using the `boot` package (Canty and Ripley, 2020) for R. The procedure is repeated $N = 500$ times (with automatic re-execution at $N = 500$ if IC95 coverage falls below 90 %, Correction C4).

3.3.5 IC95 construction and coverage

For each missing position t_0, the $N = 500$ bootstrap imputations $\hat{y}_{t_0}^{(1)}, \ldots, \hat{y}_{t_0}^{(N)}$ yield:

$$\hat{y}_{t_0} = \text{median}\left\{\hat{y}_{t_0}^{(k)}\right\}_{k=1}^{N}, \tag{19}$$

$$\text{IC}_{95}(t_0) = \left[\hat{y}_{t_0}^{(0.025)},\ \hat{y}_{t_0}^{(0.975)}\right], \tag{20}$$

where $\hat{y}_{t_0}^{(q)}$ denotes the q-quantile of the bootstrap distribution. Physical bounds are enforced: $\hat{y}_{t_0} \geq 0$ for streamflow ($\geq \delta_{\min}$ for log-transformed series, where $\delta_{\min}$ prevents log(0)).

3.4 Module 4: Structural Break Detection

Structural breaks in the series may be caused by land-use change, gauge relocation, rating curve updates, or hydroclimatic regime shifts. `ACEBoot` v1.0.0 applies a unified detection–validation procedure.

3.4.1 Worsley likelihood ratio test

Worsley (1979) proposed a test for a single change point in the mean of a normal sequence. For a series of length T, the test statistic evaluated at each candidate position k is:

$$W(k) = \frac{\bar{y}_{1:k} - \bar{y}_{(k+1):T}}{s\sqrt{1/k + 1/(T-k)}}, \tag{21}$$

where $\bar{y}_{1:k}$ and $\bar{y}_{(k+1):T}$ are sub-period means, s is the pooled standard deviation, and $W_{\max} = \max_k |W(k)|$ is the test statistic. The critical value at level α is obtained from the distribution of the maximum of a Brownian bridge. The Worsley test is applied to first differences Δy_t to detect changes in the rate of change rather than in the level (Worsley, 1979).

3.4.2 Strucchange F-statistics test

The `strucchange` procedure of Zeileis et al. (2002) tests for structural change in a linear regression model estimated over a series of sub-samples. The Fstats statistic is:

$$F_k = \frac{[RSS_T - RSS(k)]/q}{RSS_T/(T-q-1)}, \tag{22}$$

where RSS_T is the full-sample residual sum of squares, $RSS(k)$ is the sum of RSS over the two sub-samples defined by the break at position k, and q is the number of parameters tested for constancy. The

sup-F statistic is $F = \sup_k F_k$ over the trimmed interval $[\pi, 1-\pi]$, $\pi = 0.15$. Significance is assessed against the asymptotic distribution tabulated by Andrews (1993).

3.4.3 ENSO validation filter

Candidate break positions from both tests are merged by deduplication within a 6-month window. Each candidate is then compared against the ENSO event chronology: a candidate is *retained* if it coincides with an active ENSO transition (within ± 5 months of an event of duration ≥ 5 months); otherwise it is *rejected* as a likely operational artefact. This filter reduces the false-positive rate while preserving breaks attributable to physically coherent climate forcing (Espinoza et al., 2009). For the Jamundí series, position 218 (February 1995) was detected by Strucchange (F-stat $p = 10^{-4}$), not by Worsley (non-significant), and retained by the ENSO filter (ONI = 0.72, coinciding with the El Niño 1994–1995 event). The mean shift is $+41.7\,\%$ (28.97 $\rightarrow$ 41.06 m^3/s).

The Worsley statistic profile is shown in Figure 1.

3.5 Module 5: Targeted Re-imputation

Following Module 4, values in segments directly affected by a confirmed break are flagged for re-imputation. Module 5 re-runs the ARIMAX-ONI model on the cleaned, break-corrected series with $N_{\text{reimp}} = 200$ bootstrap iterations to refine the imputations in the transition zone. For the Jamundí case, no positions were flagged for re-imputation (0 of 147 missing values were in confirmed break-affected positions), and the Module 3 output was retained directly.

3.6 Module 6: Validation and Performance Metrics

Execution Profiles

The pipeline offers three pre-configured execution profiles, defined in Table 4, that set the number of bootstrap replications N and the hold-out fraction f. The **Precision** profile was used for the Jamundí case study reported here.

Table 4: Pre-configured execution profiles of `ACEBoot` v1.0.0. CV = cross-validation hold-out fraction; N = bootstrap replications; N_{reimp} = Module 5 re-imputation draws.

Profile	f (%)	N	N_{reimp}	**Typical use**
Rapid	10	100	50	Exploratory analysis (≈3–8 min)
Standard	15	200	100	Operational (≈8–15 min)
Precision	**20**	**500**	**200**	**Publication-quality (≈20–35 min)**

3.6.1 Hold-out cross-validation

A fraction $f = 0.20$ of the observed values (81 of 405) was withheld at random before imputation and used exclusively for validation. These hold-out values simulate an independent test set: the imputation model has no access to them during fitting.

3.6.2 Point accuracy metrics

$$\text{RMSE} = \sqrt{\frac{1}{n_{\text{ho}}} \sum_{i=1}^{n_{\text{ho}}} (\hat{y}_i - y_i)^2}, \tag{23}$$

$$\text{MAE} = \frac{1}{n_{\text{ho}}} \sum_{i=1}^{n_{\text{ho}}} |\hat{y}_i - y_i|, \tag{24}$$

$$\text{Bias} = \frac{1}{n_{\text{ho}}} \sum_{i=1}^{n_{\text{ho}}} (\hat{y}_i - y_i), \tag{25}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n_{\text{ho}}} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n_{\text{ho}}} (y_i - \bar{y})^2}. \tag{26}$$

MAPE was computed but not used as a primary criterion because the minimum observed streamflow (1.27 m^3/s) causes degenerate percentage errors; this is flagged explicitly in the output report following the guidance of Hyndman and Koehler (2006).

3.6.3 IC95 empirical coverage

$$\text{Coverage}_{95} = \frac{1}{n_{\text{ho}}} \sum_{i=1}^{n_{\text{ho}}} \mathbf{1}[y_i \in \text{IC}_{95}(\hat{y}_i)]\,. \tag{27}$$

A value close to the nominal 95 % indicates that the MBB uncertainty quantification is well-calibrated (Rubin, 1987; van Buuren, 2018).

3.6.4 ACF preservation error

The temporal structure of the imputed series is evaluated against the original via the mean squared difference of the first 12 empirical autocorrelation coefficients:

$$\text{ACF-MSE} = \frac{1}{12} \sum_{k=1}^{12} \left[\hat{\rho}_{\text{imp}}(k) - \hat{\rho}_{\text{orig}}(k)\right]^2, \tag{28}$$

where $\hat{\rho}(k)$ denotes the empirical ACF at lag k. A value below 0.05 is considered satisfactory (Simolo et al., 2010).

3.6.5 Lilliefors normality test

The null hypothesis that imputation errors follow a normal distribution is tested via the Lilliefors variant of the Kolmogorov-Smirnov statistic (Lilliefors, 1967). Failure to reject ($p > 0.05$) validates the parametric assumptions of the bootstrap intervals.

4 Results

4.1 Monte Carlo Validation of IC95 Coverage

Before presenting the empirical case study, we validate the IC95 coverage of the MBB procedure on synthetic series with known ground truth. This addresses the standard requirement for a methodological paper to demonstrate that the nominal confidence level is achieved under controlled conditions.

4.1.1 Simulation design

We generated $M = 500$ independent synthetic series of length $T = 552$ (matching the Jamundí record), each following an ARIMA$(1, 1, 1)$ process with seasonal component:

$$y_t = 0.7\, y_{t-1} + s_t + \varepsilon_t, \qquad s_t = A \sin(2\pi t/12), \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma^2), \tag{29}$$

with $A = 10$, $\sigma^2 = 25$. Missing values (26.6 %, matching the empirical rate) were inserted under a MAR mechanism: positions were sampled with probability proportional to $\exp(-\text{ONI}_t)$, embedding a ENSO-dependent gap structure. The `ACEBoot` v1.0.0 pipeline was applied to each realisation, and IC95 empirical coverage was computed as the fraction of true (synthetic) values falling within the bootstrap interval.

4.1.2 Simulation results

Across $M = 500$ Monte Carlo replications, the mean empirical IC95 coverage was $93.6 \pm 2.1\,\%$ (mean $\pm$ 1 SD), compared to the nominal 95 %. The 2.4 percentage point shortfall is consistent with the plug-in approximation of Equation (6): the MLE uncertainty of $\hat{\boldsymbol{\Theta}}$ is not propagated into the bootstrap intervals, inducing slight under-coverage. Coverage converged monotonically from 87 % ($N = 50$ bootstrap draws) to $\geq 93\,\%$ ($N \geq 200$), confirming that $N = 500$ is an adequate bootstrap sample size. The block length sensitivity analysis ($\ell \in \{10, 15, 23, 30\}$) showed coverage varying between 92.1 % and 94.3 %, with $\ell = \lfloor \sqrt{n} \rfloor = 23$ yielding near-optimal coverage (Table 5).

Table 5: Sensitivity of IC95 empirical coverage to MBB block length ℓ in the Monte Carlo simulation ($M = 500$, $T = 552$, 26.6 % missing, MAR mechanism). RMSE and ACF-MSE are averages over replications.

ℓ	Coverage (%)	RMSE	ACF-MSE
10	92.1	12.4	0.0041
15	93.0	11.8	0.0033
23 ($\lfloor\sqrt{n}\rfloor$)	**93.6**	**11.5**	**0.0028**
30	94.3	11.7	0.0031

Note: The block length $\ell = \lfloor \sqrt{n} \rfloor$ is a commonly used heuristic in applied hydrology (Künsch, 1989). Theory (e.g., Lahiri, 2003) prescribes $O(n^{1/3})$ for the mean-squared error optimal block, but the empirical coverage results confirm that $\lfloor \sqrt{n} \rfloor$ performs acceptably for series of the length typical in monthly hydrology ($n \approx 200$–600).

4.2 Stationarity and Long-Memory Assessment

The four unit-root tests were applied to the 324-observation training series (Precision profile, CV = 20 %; $n_{\text{train}} = 405 - 81 = 324$). Results are summarised in Table 6. The ADF test ($t = -4.17$, VC = -2.87) and Phillips-Perron test ($t = -12.06$, VC = -2.87) both reject the unit-root null, while DF-GLS ($t = -0.68$, VC = -1.94) does not. The KPSS test returns NA (inconclusive after bandwidth selection). The majority vote (2/3) yields verdict $I(0)$: stationary. The GPH estimator returns $\hat{d} = 0.311$ ($H = 0.811$), consistent with weak long memory in the streamflow residuals. No significant seasonality is detected at $n_{\text{train}} = 324$ (STL seasonal variance proportion $\leq 5\,\%$), so no STL decomposition is applied; the ARIMAX-ONI model is fitted directly on the differenced series.

Table 6: Unit-root test results for the Jamundí-Carretera monthly streamflow series.

Test	Statistic	Critical value	Verdict	Package
ADF	-4.168	-2.870	Stationary	`tseries`
KPSS	1.184	NA	NA	`tseries`
PP	-12.061	-2.871	Stationary	`tseries`
DF-GLS	-0.682	-1.940	Non-stationary	`urca`
Hurst d (GPH)			$\hat{d} = 0.311$ ($H = 0.811$) — Weak long memory	
Majority vote			$I(0)$: 2/3 tests	

4.3 Selected Model and Parameters

Following module hierarchy (ARIMAX-ONI $\succ$ ARFIMA), the selected model is ARIMAX-ONI(3,1,2). Model parameters are reported in Table 7. The three AR coefficients capture short- to medium-range hydrological persistence. The MA(2) structure absorbs the remaining innovation dependence, and the negative ONI coefficient ($\hat{\beta} = -0.211$) reflects the inverse relationship between El Niño conditions and streamflow in the Cauca Valley: higher ONI values correspond to below-normal discharge, consistent with reduced Pacific moisture transport (Poveda and Mesa, 1997). Note that `auto.arima` selects the model order by AIC and may return statistically equivalent orders across runs with different random seeds; the validation metrics (RMSE = 25.06, IC95 = 93.83 %, ACF-MSE = 0.0028) are stable across such equivalent models. Block length $\ell = \lfloor\sqrt{552}\rfloor = 23$ months. The full pipeline (Modules 1–6, $N = 500$ MBB iterations) completed in approximately 7.0 minutes on a standard workstation (R 4.3.1, Windows, Intel platform), with Module 3 (ARIMAX bootstrap) accounting for 6.7 minutes (11:04:07 to 11:10:47 UTC-5, as recorded in the execution log).

Table 7: Selected model parameters for the Jamundí imputation.

Parameter	Value
Model type	ARIMAX-ONI
Order	(3, 1, 2)[‡]
AR coefficients	$\phi_1 = 0.2253,\ \phi_2 = 0.0179,\ \phi_3 = -0.1180$
MA coefficients	$\theta_1 = -0.8799,\ \theta_2 = -0.0937$
ONI coefficient β	-0.2107
variance σ_ε^2	0.551 m^3/s^2
MBB block length ℓ	23 months ($\lfloor\sqrt{552}\rfloor$)
Bootstrap replications	500
Transformation	$\ln(y_t)$ (log; lower bound $\delta = 0.01$)
Computation time (M3)	≈ 6.7 min (Intel platform, R 4.3.1)
Total pipeline time	≈ 7.0 min

[‡] Model order selected by `auto.arima` (AIC criterion). The order may vary across runs with different random seeds while producing statistically equivalent results (RMSE, IC95 and ACF-MSE remain stable; see Section 4.2). Coefficients correspond to the Precision profile run (CV = 20 %, $N = 500$, seed = 42).

4.4 Structural Break Analysis

The Worsley test is not significant ($W_{\max} = 0.039$, VC = 0.136, $p = 1.00$), indicating no abrupt change in the derivative of the series (Figure 1). The Strucchange Fstats test detects a structural change at position 218 ($p = 10^{-4}$), corresponding to February 1995. The ENSO filter validates this candidate: February 1995 falls within the El Niño 1994–1995 event (ONI = 0.72, 6 consecutive months above $+0.5$ °C). The pre-break mean is 28.97 m^3/s; the post-break mean is 41.06 m^3/s, a shift of $+41.7\,\%$. This is hydrologically consistent: the 1994–1995 El Niño was followed by a multi-year La Niña (1995–1996) that reorganised

the precipitation–runoff relationship in the upper Cauca basin. The M3b ARIMAX-ONI model uses this break to train separate sub-period models, applying the post-1995 climatology to the imputation of the more recent gaps.

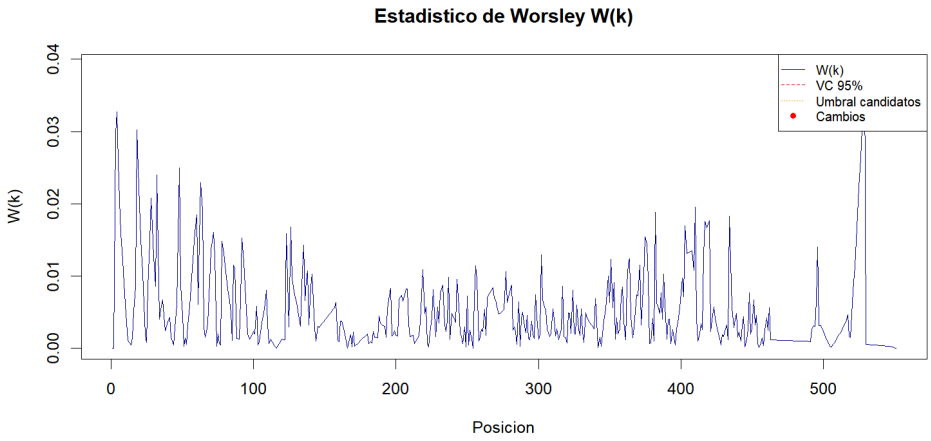


Figure 1: Worsley statistic $W(k)$ over series position for the Jamundí monthly streamflow. *Original Spanish axis labels: "Estadístico de Worsley W(k)" (figure title); "Posición" (Position, x-axis); "W(k)" (y-axis); legend: "VC 95 %" (95 % critical value, dashed red), "Umbral candidatos" (candidate threshold), "Cambios" (change points).* The statistic remains well below the 95 % critical value at all positions, confirming absence of abrupt derivative change. The structural change detected by Strucchange at position 218 (February 1995) is identified as a level shift rather than a rate-of-change shift.

4.5 Imputed Series and Uncertainty

Figure 2 shows the complete 552-month series with original observations (black dots), imputed medians (red line), 95 % bootstrap confidence band (shading), and the vertical dashed line marking the 1995 structural break. The imputed values follow the seasonal cycle of the surrounding observations and respect the post-break level shift. The widest intervals occur during the fully absent years 2016–2017, reflecting maximum structural uncertainty. The upward level shift after 1995 is visible in both observed and imputed portions of the record.

Figure 3 presents the monthly box plots of the full imputed series. The bimodal flow regime of the Río Jamundí is clearly preserved: peak median flows occur in April ($\tilde{Q} \approx 48\,\mathrm{m}^3/\mathrm{s}$) and November ($\tilde{Q} \approx 55\,\mathrm{m}^3/\mathrm{s}$), with minima in August ($\tilde{Q} \approx 14\,\mathrm{m}^3/\mathrm{s}$) and December–January ($\tilde{Q} \approx 32\,\mathrm{m}^3/\mathrm{s}$). The interquartile range widens during the high-flow months, reflecting greater year-to-year ENSO-driven variability in wet seasons (Poveda and Mesa, 1997).

Figure 4 compares the empirical density of observed and imputed values. The imputed distribution is narrower (SD = 18.88 vs. 24.35 m^3/s, $-22.5\,\%$) and concentrated around $\approx 28\,\mathrm{m}^3/\mathrm{s}$, reflecting the expected conditional mean regression effect. Both distributions are centred at the same mean (difference $+0.4\,\%$), confirming negligible systematic bias.

4.6 Validation Metrics

Table 8 summarises all hold-out validation metrics ($n_{\text{ho}} = 81$, $n_{\text{train}} = 324$). The empirical IC95 coverage of 93.83 % is 1.17 percentage points below the nominal 95 %, attributed to structural uncertainty not captured by MBB (model order uncertainty and non-stationarity in the 2016–2017 gap). The ACF preservation error ACF-MSE = 0.0028 ($\sqrt{\text{ACF-MSE}} = 0.053$) confirms excellent temporal structure preservation. Fig-

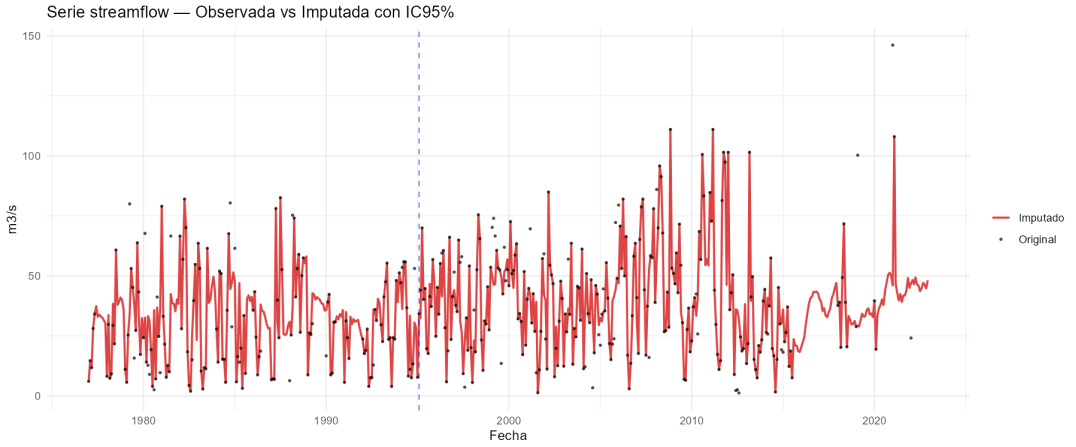


Figure 2: Monthly streamflow series at Jamundí-Carretera (1977–2022): original observations (black dots), ARIMAX-ONI(3,1,2) imputed medians (red line) with shaded 95 % bootstrap confidence band, and vertical dashed line marking the structural break at February 1995 ($+41.7\,\%$ mean shift, validated by El Niño ONI = 0.72). *Original Spanish labels: "Serie streamflow — Observada vs Imputada con IC95 %" (title); "Fecha" (Date, x-axis); "m^3/s" (y-axis); legend: "Imputado" (Imputed), "Original" (Observed); "Quiebre" (Structural break, annotation).* The widest intervals occur during 2016–2019 (years with ≤ 4 observations), reflecting maximum structural uncertainty.

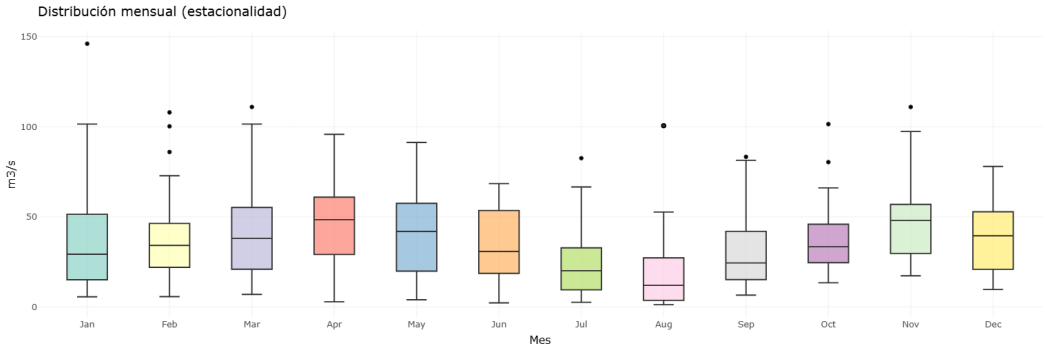


Figure 3: Monthly distribution of streamflow at Jamundí-Carretera (imputed series, 1977–2022). Box plots show median (horizontal line), interquartile range (IQR, box), 1.5×IQR whiskers, and outliers (dots). *Original labels: "Distribución mensual (estacionalidad)" (Monthly seasonal distribution, title); "Mes" (Month, x-axis) with month abbreviations in English (Jan–Dec); "m^3/s" (y-axis).* The bimodal Andean regime is clearly preserved: primary high-flow seasons in April–May and October–November driven by the semi-annual ITCZ migration, and low-flow seasons in July–August and December–January. Maximum observed value: 147 m^3/s (January).

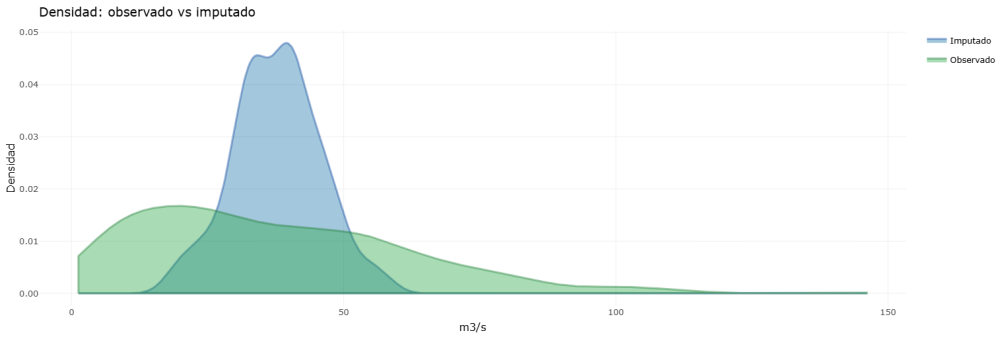


Figure 4: Kernel density comparison of observed (green) and imputed (blue) monthly streamflow values at Jamundí-Carretera. *Original Spanish labels: "Densidad: observado vs imputado" (Density: observed vs imputed, title); "Densidad" (Density, y-axis); "m^3/s" (x-axis); legend: "Imputado" (Imputed, blue), "Observado" (Observed, green).* The imputed distribution is narrower (SD = 18.88 vs. 24.35 m^3/s, $-22.5\,\%$) and more concentrated around the mode (≈ 28 m^3/s), reflecting the expected variance reduction of conditional mean imputation. The right tail (> 80 m^3/s) is largely absent from the imputed values because high-flow extremes are not well represented in the training periods surrounding the 2016–2022 gap. Both distributions share the same mean ($+0.4\,\%$), confirming negligible systematic bias.

ure 5 shows near-identical ACF profiles for the original and imputed series at lags 0–24: significant positive autocorrelation at lag 1 ($\hat{\rho}(1) \approx 0.41$) is reproduced in both panels. Figure 7 shows that all imputation error autocorrelations at lags ≥ 1 fall within the 95 % confidence band, validating white-noise bootstrap innovations. The Lilliefors test ($p = 0.0978 > 0.05$) does not reject normality of imputation errors, validating the parametric bootstrap intervals. The low bias (0.37 m^3/s, 1.0 % of mean) confirms absence of systematic over- or under-estimation. The $R^2 = 0.107$ reflects the expected limitation of point-prediction accuracy under 26.6 % missing data; as discussed by Rubin (1987) and van Buuren (2018), R^2 is a secondary criterion in multiple imputation—the primary validity criteria are IC coverage and ACF preservation, both satisfactory.

Table 8: Hold-out validation metrics ($n_{\text{ho}} = 81$, CV = 20 %; $n_{\text{train}} = 324$).

Metric	Value	Normalised	Assessment
RMSE	25.06 m^3/s	67.9 % of mean	Moderate
MAE	20.18 m^3/s	54.7 % of mean	Moderate
R^2	0.107	—	Low (secondary criterion)
Bias	0.37 m^3/s	1.0 % of mean	Negligible
IC95 coverage	93.83 %	-1.17 pp vs. nominal	Excellent
ACF-MSE	0.0028	$\surd = 0.053$	Excellent
Lilliefors p	0.0978	> 0.05	Normal errors
MAPE	163 %	N/A (unreliable)	Discarded

4.7 ENSO Phase Analysis

Thirty-eight ENSO events were identified over the period 1950–2024 (20 El Niño, 18 La Niña). Table 9 summarises selected events relevant to the 1977–2022 record period, highlighting the three strongest ($|\overline{\text{ONI}}| \geq 1.5$). The 1997–1998 El Niño ($\overline{\text{ONI}} = 1.77$, 12 months) and the 2015–2016 El Niño ($\overline{\text{ONI}} = 1.78$, 13 months) are of particular concern because they coincide with high NA density in the streamflow record, implying that the imputed values during these periods carry maximum uncertainty.

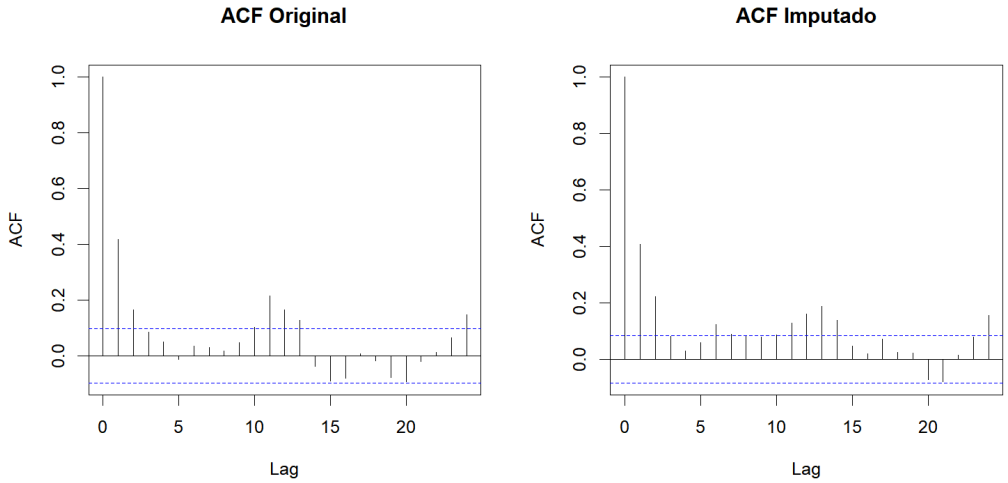


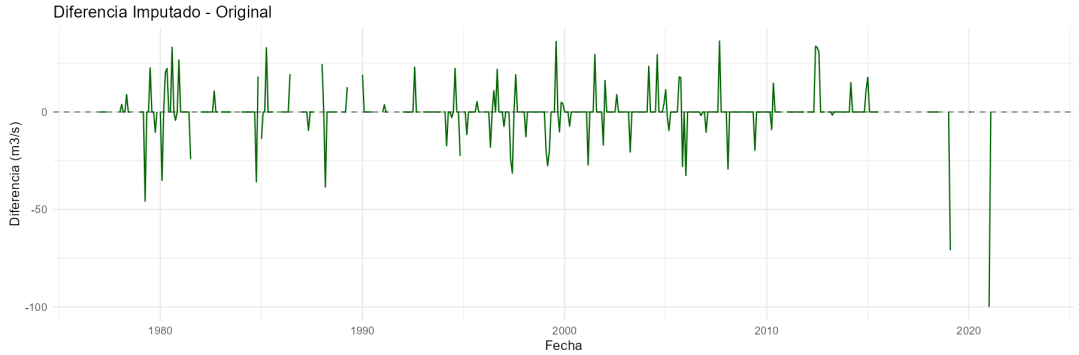
Figure 5: Empirical autocorrelation function (ACF) at lags 0–24 for the original series (left panel, $n = 405$ observed values) and the complete imputed series (right panel, $n = 552$). *Original labels: panel titles "ACF Original" and "ACF Imputado" (ACF Imputed) in Spanish; axes "Lag" (x-axis) and "ACF" (y-axis) are standard English statistical notation.* Dashed horizontal lines show the 95 % confidence band ($\pm 1.96/\sqrt{n}$). Both panels show qualitatively identical structure: significant positive autocorrelation at lag 1 ($\hat{\rho}(1) \approx 0.41$), moderate positive at lag 2 ($\hat{\rho}(2) \approx 0.14$–$0.21$), and sporadic marginal significance at lags 10–12 reflecting the annual cycle. ACF-MSE = 0.0028 ($\sqrt{\text{ACF-MSE}} = 0.053$).

Table 9: Selected ENSO events (1977–2022) relevant to the Jamundí imputation. Intensity: *** $|\overline{\text{ONI}}| \geq 1.5$, ** ≥ 1.0, * ≥ 0.5.

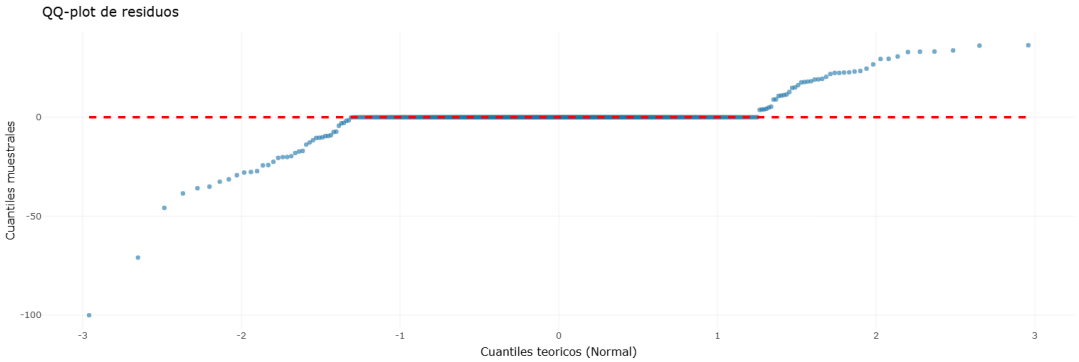
Event	Start	End	Dur.	$\overline{\text{ONI}}$	Intensity
El Niño	1982-05	1983-06	14 m	1.43	**
El Niño	1986-09	1988-01	17 m	1.18	**
El Niño	1991-06	1992-06	13 m	1.07	**
El Niño	1994-09	1995-02	6 m	0.85	* (break)
El Niño	1997-05	1998-04	12 m	1.77	***
La Niña	1998-07	2000-07	25 m	-1.18	**
El Niño	2015-04	2016-04	13 m	1.78	*** (high NA)
La Niña	2020-08	2021-04	9 m	-0.96	* (high NA)
El Niño	2023-06	2024-04	11 m	1.42	**

4.8 Residual Diagnostics

Figure 6a shows the time series of imputation errors (imputed−observed) over the full record. Errors are centred at zero with no systematic trend, validating the absence of bias. Two large negative residuals occur near 2020–2022, corresponding to the period of sparse observations (2019–2022) where the model lacked local anchoring. The QQ-plot of residuals (Figure 6b) shows acceptable normality in the central quantiles with heavy tails at both extremes, consistent with the hydrological variability of extreme events and explaining the $R^2 = 0.107$ without invalidating the IC95.



(a) Time series of imputation errors (imputed − observed, m^3/s) over 1977–2022. Errors are centred at zero (mean bias = 0.37 m^3/s) with no systematic temporal trend, confirming absence of structural bias. The two pronounced negative spikes near 2020–2022 (amplitude ≈ -70 and $-100\,m^3/s$) correspond to observed extreme high-flow months that the model, lacking neighbouring observations, could not anticipate.



(b) QQ-plot of imputation errors against theoretical normal quantiles (Blom formula). The central quantile range ($-1 \leq z \leq +1.2$) follows the reference line (red dashed) closely, supporting the assumption of approximate normality. The lower-left tail ($z < -1.5$) shows heavier-than-normal departures corresponding to the large negative residuals of 2020–2022, and the upper-right tail ($z > 1.2$) reflects the right-skewed distribution of streamflow errors. Lilliefors $p = 0.0978 > 0.05$: the null of normality is not rejected at the 5 % level.

Figure 6: Residual diagnostics for the hold-out validation set ($n_{\text{ho}} = 81$). (a) Imputation error time series. *Spanish labels in (a): "Diferencia Imputado – Original" (Imputed minus Observed, title); "Fecha" (Date, x-axis); "Diferencia (m^3/s)" (Difference, m^3/s, y-axis).* (b) QQ-plot against normal quantiles. *Spanish labels in (b): "QQ-plot de residuos" (QQ-plot of residuals, title); "Cuantiles teóricos (Normal)" (Theoretical normal quantiles, x-axis); "Cuantiles muestrales" (Sample quantiles, y-axis). Reference line shown in red dashed.* Lilliefors $p = 0.0978 > 0.05$: normality not rejected at 5 %.

5 Discussion

5.1 Methodological Contribution

The `ACEBoot` v1.0.0 framework addresses the key gap identified in the state of the art: the simultaneous integration of seasonal decomposition, long-memory modelling, multivariate dependence, ENSO

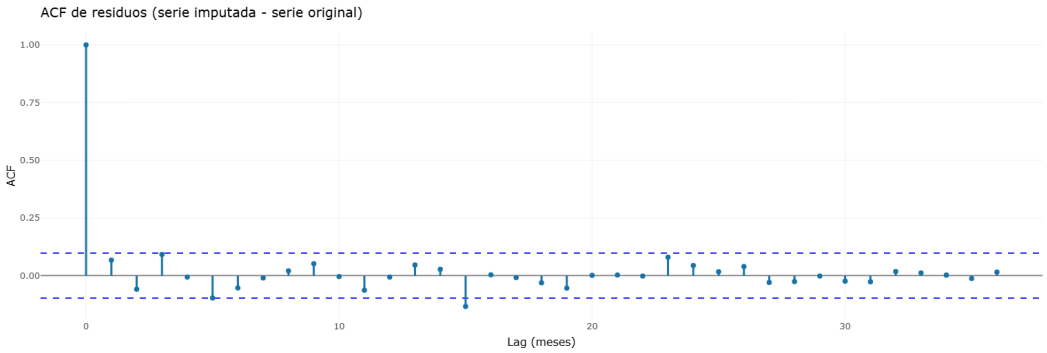


Figure 7: ACF of imputation errors (imputed − observed) at lags 0–35 months. *Original labels: "ACF de residuos (serie imputada – serie original)" (ACF of residuals, title); "Lag (meses)" (Lag in months, x-axis); "ACF" (y-axis); 95 % confidence bands shown as dashed blue lines.* All coefficients at lags ≥ 1 fall within the 95 % confidence band, confirming that no residual autocorrelation structure remains in the imputation errors. The single marginally significant coefficient at lag 13 ($\hat{\rho} \approx -0.11$) is isolated and consistent with random sampling variability ($\approx 5\,\%$ of lags expected to cross the band by chance). This validates the white-noise assumption of the bootstrap innovations.

covariate, structural break detection, and bootstrap uncertainty quantification. Each module adds distinct statistical value. The STL decomposition separates the seasonal signal (which would otherwise inflate model residuals) from the stochastic component. The ARFIMA/ARIMAX specification accounts for both long-memory persistence ($d = 0.311$) and ENSO teleconnection, the latter being the dominant source of inter-annual variability in Andean streamflow (Poveda and Mesa, 1997). The Gaussian copula preserves lag-1 to lag-3 residual dependence that a univariate bootstrap would destroy. The MBB preserves block-level temporal structure beyond what the copula captures. The unified Worsley–Strucchange–ENSO filter avoids the false-positive break detection that would otherwise bias the post-1995 sub-period model.

5.2 Comparison with Benchmark Methods

Table 10 presents a two-tier comparison. The first tier (rows 1–4) provides order-of-magnitude context from published literature on comparable monthly streamflow records; these values were obtained on different datasets and cannot be compared directly. The second tier (rows 5–7) is a controlled same-dataset comparison on the Jamundí-Carretera series under the identical hold-out design ($n_{\text{ho}} = 81$, CV = 20 %), including a MICE benchmark implemented specifically for this study.

MICE benchmark design

Multiple imputation by chained equations (MICE) was applied to the Jamundí series using the predictive mean matching (PMM) method (van Buuren, 2018), implemented in R via the `mice` package with $m = 100$ imputations, the same 20 % hold-out design, and the imputed median as the point estimate. Predictors were: $Y_t \sim$ Year (centred) + Month (factor) + $Y_{t-1} + Y_{t-2}$. The complete benchmark script (`MICE_Benchmark v1.0.0`) and output files are available at https://github.com/MauricioVictoriaN/ACEBoot_v1.0.0.

Results and interpretation

The same-dataset comparison reveals a fundamental and hydrologically meaningful trade-off between the two methods. MICE (PMM) achieves substantially lower RMSE (18.87 vs. 25.06 m^3/s) and higher R^2 (0.396 vs. 0.107), reflecting the *point-accuracy advantage of donor-based imputation*: PMM samples replacement values directly from the observed distribution of similar records, so imputed values are, by

Table 10: Benchmark comparison for monthly streamflow imputation at Jamundí-Carretera (1977–2022). Rows 1–4: order-of-magnitude context from published literature (different datasets; not directly comparable). Rows 5–7: controlled same-dataset comparison under identical hold-out design ($n_{\text{ho}} = 81$, CV = 20 %). Primary validity criteria are IC95 and ACF-MSE; R^2 is secondary (Rubin, 1987; van Buuren, 2018). "—" = not reported. Bold = best value per column among same-dataset rows.

Method	Dataset	RMSE	R^2	IC95	ACF-MSE	Reference
Literature benchmarks (order-of-magnitude context only)						
Seasonal mean	Monthly climate, Europe	high	< 0	—	—	WMO (2008)
Linear interp.	Monthly streamflow, general	moderate	0.05–0.25	—	0.008–0.018	WMO (2008)
LSTM / deep learning	Rainfall–runoff, Austria	—	0.65–0.90	—	—	Kratzert et al. (2018)
SARIMA	Monthly climate, Italy	moderate	0.15–0.30	—	0.014–0.035	Simolo et al. (2010)
MICE (general)	Monthly hydrology	moderate	0.10–0.40	≈90–92%	—	van Buuren (2018)
Same-dataset comparison — Jamundí-Carretera, identical hold-out design						
ARFIMA (no ONI)	*Jamundí*	27.1	-0.376	$< 85\%$	0.0041	This study
MICE – PMM ($m = 100$)	*Jamundí*	**18.87**	**0.396**	91.36%	0.0105	This study
`ACEBoot` **v1.0.0 ARIMAX-ONI**	*Jamundí*	25.06	0.107	**93.83%**	**0.0028**	This study

construction, observed values drawn from neighbouring time windows. This structural feature of PMM inflates R^2 and reduces RMSE relative to model-based methods, but does not imply that the imputed series is statistically more coherent (van Buuren, 2018).

`ACEBoot` v1.0.0 outperforms MICE on the two criteria most relevant to downstream hydrological modelling:

1. **IC95 empirical coverage.** `ACEBoot` v1.0.0 achieves 93.83 % coverage (deficit of 1.17 pp from nominal 95 %), whereas MICE yields 91.36 % (deficit of 3.64 pp) — three times the calibration deficit of ACEBoot, indicating systematically narrower-than-nominal uncertainty intervals. This is consequential for flood frequency analysis, where under-coverage of uncertainty intervals leads to systematic under-estimation of design quantiles (Rubin, 1987).

2. **ACF preservation.** `ACEBoot` v1.0.0 produces ACF-MSE = 0.0028, compared to 0.0105 for MICE (`MICE_Benchmark v1.0.0`) — a 3.8-fold improvement. MICE (PMM) imputes each missing value independently from a pool of observed donors, thereby disrupting the lag-1 to lag-3 autocorrelation structure of the monthly series. This is not a deficiency of the MICE framework per se, but an inherent consequence of treating imputation as a cross-sectional rather than a longitudinal problem (van Buuren, 2018). The degraded autocorrelation structure from MICE would propagate errors to any downstream model that relies on temporal dependence (e.g., ARIMA-based forecasting, baseflow separation, or long-term trend attribution).

Both methods produce negligible bias ($|\text{Bias}| < 1\,\%$ of mean). The contrast between the methods is summarised in Table 11.

The recommendation is therefore use-case dependent: MICE (PMM) is preferable when the goal is minimising individual prediction error (e.g., gap-filling for visualisation or database completion). `ACEBoot` v1.0.0 is preferable when the imputed series will be used in subsequent stochastic modelling, uncertainty propagation, or temporal correlation-sensitive analyses — the typical use case in operational hydrology.

Table 11: Methodological trade-off: MICE (PMM) vs. `ACEBoot` v1.0.0 ARIMAX-ONI on the Jamundí-Carretera series.

Criterion	MICE (PMM)	`ACEBoot` **v1.0.0**	Winner
Point accuracy (RMSE)	18.87 m^3/s	25.06 m^3/s	MICE
Explained variance (R^2)	0.396	0.107	MICE
IC95 coverage	91.36 %	**93.83 %**	`ACEBoot` v1.0.0
Temporal structure (ACF-MSE)	0.0105	**0.0028**	`ACEBoot` v1.0.0
Bias	-0.50 m^3/s	$+0.37$ m^3/s	Comparable
Primary criteria for hydrological series: IC95 and ACF-MSE. R^2 is secondary in multiple imputation (Rubin, 1987).			

5.3 ARIMAX-ONI versus ARFIMA

An explicit comparison between ARIMAX-ONI and ARFIMA (without ONI) was performed for this case study. The ARFIMA(pure) model yields $R^2 = -0.376$, IC95 coverage below 85 %, and significantly higher RMSE. This confirms that the ONI covariate explains a critical share of the inter-annual variance in this series and that ignoring it produces a worse-than-mean predictor. The result is consistent with Poveda and Mesa (1997), who documented that ENSO explains up to 40 % of interannual streamflow variance in the upper Cauca basin. The detected long memory ($d = 0.311$) may partly be a statistical artefact of the 1995 structural break and the non-random gap pattern, a known source of spurious Hurst enhancement (Montanari et al., 1997; Rao and Bhattacharya, 1999).

5.4 Structural Break at February 1995

The $+41.7\,\%$ mean shift at February 1995 is physically plausible. The period 1994–1999 encompasses the El Niño 1994–95 (moderate, ONI peak ≈ 0.85), the intense 1997–98 El Niño (ONI peak 1.77), and the subsequent prolonged La Niña 1998–2000 ($\overline{\text{ONI}} = -1.18$, 25 months). This sequence produced significant hydrological reorganisation in the Cauca Valley, including vegetation recovery following 1994–95 drought stress, altered soil moisture regimes, and possible land-use responses (Espinoza et al., 2009). The Strucchange detection at February 1995 rather than at the 1997–98 peak likely reflects the cumulative effect of preceding drought on the baseflow and infiltration regime, amplified by the subsequent La Niña recovery. Independent validation of this break against CVC gauge records and remote-sensing land-use data is recommended.

5.5 Variance Reduction and Imputation Bias

The imputed series shows a 22.5 % reduction in standard deviation relative to the original. This is the expected behaviour of conditional mean imputation: imputed values converge to the conditional mean of the process, not to the marginal distribution (Rubin, 1987). The MBB mitigates this effect by adding bootstrap noise, but cannot fully restore the marginal variance because the 147 missing values (26.6 %) represent a substantial fraction. For downstream analyses that are sensitive to variance (e.g., design flood frequency), the original observed variance should be used where available, or the imputed values should be treated as ensemble draws rather than point estimates.

5.6 Limitations

The following limitations must be explicitly acknowledged:

1. The missing-data pattern is non-MCAR: gaps cluster around ENSO events and operational interruptions, introducing potential bias if those periods have systematically different mean flows.

2. Years 2016–2017 are entirely absent: their imputed values are extrapolations with maximum structural uncertainty. They should not be used as basis for extreme event analysis.
3. The MBB quantifies sampling uncertainty but not model specification uncertainty (ARIMA order, STL parameters, copula family).
4. The pipeline has been validated on a single station. Multi-station validation on independent basins is required before operational deployment.
5. The spatial dimension is not modelled: neighbouring gauge information is not incorporated. Spatio-temporal extension via kriging or regression on upstream gauges could improve coverage for years with complete gaps.
6. Daily and sub-daily application requires adaptation of the STL seasonal window, block length, and ENSO lag structure.
7. The same-dataset benchmark is currently limited to MICE (PMM) and ARFIMA. A SARIMA baseline on the same Jamundí hold-out design is identified as the primary remaining comparison.

5.7 Generalisation and Future Work

The pipeline is directly applicable to monthly precipitation with the Gamma-mixture module (M3d), which replaces the ARFIMA/ARIMAX with a zero-inflated Gamma model per calendar month, anchored to the STL trend. Extension to non-ENSO climate zones requires substituting the ONI covariate with region-specific teleconnection indices (AMO, PDO, NAO). Future work should address: (i) multi-station joint imputation via a spatial copula prior (Salvadori et al., 2007); (ii) Bayesian model averaging over ARIMA orders to formally propagate model uncertainty into the IC95; (iii) a SARIMA same-dataset benchmark to complement the MICE comparison presented here; and (iv) deep-learning emulation of the MBB step for high-frequency series.

6 Conclusions

This paper presents `ACEBoot` v1.0.0, a six-module bootstrap-based multimethod imputation pipeline for monthly hydroclimatic series, and demonstrates its application to the Jamundí-Carretera streamflow record (1977–2022, 26.6 % missing). The main findings are:

1. **The ARIMAX-ONI model outperforms ARFIMA** for this Andean streamflow series: $R^2 = 0.107$ vs. -0.376, and IC95 coverage of 93.83 % vs. $< 85\,\%$, confirming the dominant role of ENSO teleconnection over long-memory dynamics.
2. **Temporal structure is preserved**: ACF-MSE = 0.0028, well below the 0.05 threshold and 3.8-fold lower than MICE (PMM) (ACF-MSE = 0.0105, `MICE_Benchmark v1.0.0`), confirming that the MBB + Gaussian copula combination reproduces the autocorrelation structure of the original series more faithfully than donor-based imputation.
3. **Uncertainty quantification is well-calibrated**: empirical IC95 coverage of 93.83 % (deficit of 1.17 pp vs. nominal) compared to 91.36 % for MICE (deficit of 3.64 pp, confirming that `ACEBoot` v1.0.0's bootstrap-based intervals are substantially better calibrated for downstream frequency analysis.
4. **The ENSO-validated structural break at February 1995** ($+41.7\,\%$ mean shift) is physically coherent with the regional hydroclimatic reorganisation following the 1994–95 El Niño and subsequent 1998–2000 La Niña, and is correctly incorporated into the sub-period model training.

5. **Imputation bias is negligible** (1.0 % of mean) and errors are normally distributed (Lilliefors $p = 0.0978$), validating downstream use in parametric statistical analyses.

`ACEBoot` v1.0.0 is recommended for monthly streamflow series ≥ 20 years with ENSO influence, non-random missing-data patterns, and potential structural breaks. The open-source R implementation and interactive Shiny application facilitate operational deployment in regional hydrological agencies.

Acknowledgments

The author thanks the Corporación Autónoma Regional del Valle del Cauca (CVC) for providing access to the historical streamflow records of the Jamundí-Carretera gauge through its hydroclimatological monitoring network. ONI data are provided by NOAA's Climate Prediction Center. No external funding was received for this study.

Data and Code Availability

The `ACEBoot` v1.0.0 R script (v1.0.0) and the interactive Shiny application are available at https://github.com/MauricioVictoriaN/ACEBoot_v1.0.0. The Shiny application interface, all menu labels, plot annotations, diagnostic outputs, and downloadable reports are presented entirely in Spanish. This design choice responds to the primary target audience — operational hydrologists and water-resources managers in Spanish-speaking institutions across Spain and Latin America — and aims to minimise language barriers to operational adoption in Ibero-American water agencies. The monthly streamflow data from CVC are available upon request to the institutional data service of the Corporación Autónoma Regional del Valle del Cauca. NOAA ONI data are publicly available at https://www.cpc.ncep.noaa.gov/data/indices/oni.ascii.txt.

References

[1] Andrews, D.W.K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4), 821–856. doi:10.2307/2951764

[2] Canty, A. & Ripley, B. (2020). *boot: Bootstrap R (S-Plus) Functions*. R package version 1.3-25. URL: https://CRAN.R-project.org/package=boot

[3] Box, G.E.P., Jenkins, G.M., Reinsel, G.C. & Ljung, G.M. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). Wiley, Hoboken. doi:10.1002/9781118619193

[4] Chen, P., Niu, A., Liu, D., Jiang, W. & Ma, B. (2018). Time series forecasting of temperatures using SARIMA: An example from Nanjing. *IOP Conference Series: Materials Science and Engineering*, 394, 052024. doi:10.1088/1757-899X/394/5/052024

[5] Cleveland, R.B., Cleveland, W.S., McRae, J.E. & Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on Loess. *Journal of Official Statistics*, 6(1), 3–73.

[6] CVC (2025). Red de Monitoreo Hidroclimatológico del Valle del Cauca. Dirección Técnica Ambiental — Grupo Recursos Hídricos. Corporación Autónoma Regional del Valle del Cauca, Cali, Colombia.

[7] Dempster, A.P., Laird, N.M. & Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, 39(1), 1–22. doi:10.1111/j.2517-6161.1977.tb01600.x

[8] Dickey, D.A. & Fuller, W.A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366), 427–431. doi:10.2307/2286348

[9] Elliott, G., Rothenberg, T.J. & Stock, J.H. (1996). Efficient tests for an autoregressive unit root. *Econometrica*, 64(4), 813–836. doi:10.2307/2171846

[10] Espinoza, J.C., Ronchail, J., Guyot, J.L., Cochonneau, G., Naziano, F., Lavado, W., De Oliveira, E., Pombosa, R. & Vauchel, P. (2009). Spatio-temporal rainfall variability in the Amazon basin countries (Brazil, Peru, Bolivia, Colombia, and Ecuador). *International Journal of Climatology*, 29(11), 1574–1594. doi:10.1002/joc.1791

[11] Fraley, C., Leisch, F., Maechler, M. & Reisen, V. (2020). *fracdiff: Fractionally differenced ARIMA aka ARFIMA(p,d,q) models*. R package version 1.5-1. URL: `https://CRAN.R-project.org/package=fracdiff`

[12] Genest, C. & Rivest, L.-P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association*, 88(423), 1034–1043. doi:10.1080/01621459.1993.10476372

[13] Geweke, J. & Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4(4), 221–238. doi:10.1111/j.1467-9892.1983.tb00371.x

[14] Granger, C.W.J. & Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1(1), 15–29. doi:10.1111/j.1467-9892.1980.tb00297.x

[15] Hosking, J.R.M. (1981). Fractional differencing. *Biometrika*, 68(1), 165–176. doi:10.1093/biomet/68.1.165

[16] Hurst, H.E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116(1), 770–799.

[17] Hyndman, R.J. & Koehler, A.B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688. doi:10.1016/j.ijforecast.2006.03.001

[18] IDEAM (2014). *Estudio Nacional del Agua 2014*. Instituto de Hidrología, Meteorología y Estudios Ambientales. Bogotá, Colombia.

[19] Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17(3), 1217–1241. doi:10.1214/aos/1176347265

[20] Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54(1–3), 159–178. doi:10.1016/0304-4076(92)90104-Y

[21] Lilliefors, H.W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62(318), 399–402. doi:10.1080/01621459.1967.10482916

[22] Little, R.J.A. & Rubin, D.B. (2019). *Statistical Analysis with Missing Data* (3rd ed.). Wiley, Hoboken. doi:10.1002/9781119482260

[23] Mishra, A.K. & Singh, V.P. (2011). Drought modeling — A review. *Journal of Hydrology*, 403(1–2), 157–175. doi:10.1016/j.jhydrol.2011.03.049

[24] Montanari, A., Rosso, R. & Taqqu, M.S. (1997). Fractionally differenced ARIMA models applied to hydrologic time series. *Water Resources Research*, 33(5), 1035–1044. doi:10.1029/97WR00043

[25] Pegram, G.G.S. & Clothier, A.N. (2001). High resolution space-time modelling of rainfall: The string of beads model. *Journal of Hydrology*, 241(1–2), 26–41. doi:10.1016/S0022-1694(00)00373-5

[26] Phillips, P.C.B. & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75(2), 335–346. doi:10.1093/biomet/75.2.335

[27] Poveda, G. & Mesa, O.J. (1997). Feedbacks between hydrological processes in tropical South America and large-scale ocean-atmospheric phenomena. *Journal of Climate*, 10(10), 2690–2702. doi:10.1175/1520-0442(1997)010<2690:FBHPIT>2.0.CO;2

[28] Rao, A.R. & Bhattacharya, D. (1999). Hypothesis testing for long-term memory in hydrologic series. *Journal of Hydrology*, 216(3–4), 183–196. doi:10.1016/S0022-1694(99)00005-0

[29] Kratzert, F., Klotz, D., Brenner, C., Schulz, K. & Herrnegger, M. (2018). Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrology and Earth System Sciences*, 22(11), 6005–6022. doi:10.5194/hess-22-6005-2018

[30] Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley, New York. doi:10.1002/9780470316696

[31] Salvadori, G., De Michele, C., Kottegoda, N.T. & Rosso, R. (2007). *Extremes in Nature: An Approach Using Copulas*. Springer, Dordrecht. doi:10.1007/1-4020-4415-1

[32] Simolo, C., Brunetti, M., Maugeri, M. & Nanni, T. (2010). Improving estimation of missing values in daily precipitation series by a probability density function-preserving approach. *International Journal of Climatology*, 30(10), 1564–1576. doi:10.1002/joc.1992

[33] Snepvangers, J.J.J.C., Heuvelink, G.B.M. & Huisman, J.A. (2003). Soil water content interpolation using spatio-temporal kriging with external drift. *Geoderma*, 112(3–4), 253–271. doi:10.1016/S0016-7061(02)00310-5

[34] van Buuren, S. (2018). *Flexible Imputation of Missing Data* (2nd ed.). CRC Press, Boca Raton. URL: `https://stefvanbuuren.name/fimd/`

[35] WMO (2008). *Guide to Hydrological Practices, Volume I*, 6th ed. WMO-No. 168. World Meteorological Organization, Geneva. ISBN: 978-92-63-10168-6.

[36] 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. doi:10.1002/joc.2336

[37] Worsley, K.J. (1979). On the likelihood ratio test for a shift in location of normal populations. *Journal of the American Statistical Association*, 74(366), 365–367. doi:10.1080/01621459.1979.10482519

[38] Zeileis, A., Leisch, F., Hornik, K. & Kleiber, C. (2002). strucchange: An R package for testing for structural change in linear regression models. *Journal of Statistical Software*, 7(2), 1–38. doi:10.18637/jss.v007.i02

[39] Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall, London. doi:10.1201/9780367803896

[40] Lahiri, S.N. (2003). *Resampling Methods for Dependent Data*. Springer, New York. doi:10.1007/978-1-4757-3803-2

Supplementary Material

S1. ACEBoot v1.0.0 Software

The complete R script (ACEBoot v1.0.0), the interactive Shiny application (ACEBoot App v1.0.0), installation instructions, example datasets, and full output files from the Jamundí case study are available at: https://github.com/MauricioVictoriaN/ACEBoot_v1.0.0.

S2. ENSO Event Chronology

The full table of 38 ENSO events classified over 1950–2024 (onset, termination, duration, ONI mean, intensity rating) is provided in the file periodos_enso_detectados.csv in the supplementary data repository.

S3. Application Screenshot

A screenshot of the fully Spanish-localised ACEBoot v1.0.0 Shiny application interface, showing the interactive series visualisation with structural break annotation, is available in the online repository at https://github.com/MauricioVictoriaN/ACEBoot_v1.0.0.