RegFDC-Cauca v1.0.0: Flow Duration Curve Regionalization via the Index-Flow Method with Long-Memory in Ungauged Basins of the Cauca River System, Colombia

An open computational framework integrating Ward D2 clustering, ARFIMA and leave-one-out cross-validation in R

Mauricio Javier Victoria Niño[1]

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

Abstract

Context and motivation. The flow duration curve (FDC) is a fundamental tool for hydrological design, ecological flow assessment, and water resources planning. In the Cauca River basin (Colombia), the low density of stream gauges prevents direct FDC construction at many sites of interest. The index-flow method allows FDC estimation in ungauged catchments, but requires a regionalization that captures intra-regional climatic heterogeneity and a synthetic series generation that preserves the long-term persistence observed in Andean Colombian streamflow.

Objective. To present RegFDC-Cauca v1.0.0, an open-source R computational framework for FDC regionalization in ungauged catchments of the Cauca River hydrological system (Colombian Andes), using Ward D2 clustering with weighted metadata, Hurst exponent detection, and ARFIMA(0,d,0) synthetic series generation.

Methods. The framework applies three complementary methodological advances: (i) Ward D2 hierarchical clustering of physical catchment attributes weighted with qualitative metadata (hydrological regime and climatic sub-region); (ii) Hurst exponent H estimation via the scaled R/S method and automatic selection between ARFIMA(0,d,0) and AR(1); and (iii) a log-log mean discharge model with BIC predictor selection and Jensen-corrected prediction intervals.

Results. The framework is calibrated and validated on 20 gauged catchments of the Cauca system (drainage areas 290–18 900 km^2, period 2015–2019). Ward D2 clustering identifies three sub-regions (cophenetic coefficient $= 0.80$, mean silhouette $= 0.51$). Leave-one-out (LOO) cross-validation of dimensionless FDCs yields NSE $= 0.97$, KGE$'$ $= 0.96$, and PBIAS $= 0.3\%$, with MAPE $< 5\%$ over $0.05 \leq F \leq 0.85$. All catchments exhibit $H > 0.60$ (median $= 0.88$), validating universal use of ARFIMA(0,d,0) across the Cauca system.

Conclusions. FDCs estimated for 8 ungauged catchments (C021–C028) include 90% prediction intervals derived from regional inter-catchment variability and mean discharge model uncertainty.

Keywords: flow duration curves; regionalization; index-flow method; Ward D2 clustering; ARFIMA; Hurst exponent; Colombian Andes; Cauca River; ungauged basins; leave-one-out cross-validation.

Software availability: RegFDC-Cauca v1.0.0 is open-source. The R script and case-study data are available at https://github.com/MauricioVictoriaN/RegFDC-Cauca under an MIT licence.

1 Introduction

The flow duration curve (FDC) summarises the frequency distribution of daily streamflow and serves as direct input for irrigation system sizing, run-of-river hydropower schemes, minimum ecological flows, and water supply systems [1, 2]. In Colombia, however, the density of stream gauging stations is significantly below the WMO recommendations [17]: large portions of the Cauca River basin lack continuous discharge records.

The index-flow method [1, 2] is the most widely adopted regional approach: the dimensionless FDC of an ungauged catchment is estimated as the mean of the dimensionless FDCs of similar gauged catchments and dimensionalized by multiplying by a mean discharge estimated from a regression model. Within the Cauca system, this approach faces two specific challenges:

1. **Long-range dependence.** Streamflow in Colombian Andean catchments exhibits interannual persistence linked to tropical Pacific variability (ENSO) [9]. The Hurst exponent H systematically exceeds 0.5, invalidating AR(1) models for synthetic series generation.

2. **Intra-regional climatic heterogeneity.** The Cauca system spans three differentiable climatic subregions [6]: Andean-North (Risaralda/Quindío, $P > 2{,}500$ mm), the inter-Andean valley (Valle del Cauca, $P = 1{,}600$–$2{,}100$ mm), and the southern upper basin (Cauca, $P = 2{,}200$–$3{,}300$ mm). Ignoring this heterogeneity yields regional FDCs with high inter-catchment scatter.

1.1 Research objectives

1. Develop a FDC regionalization framework for the Cauca system incorporating climatic heterogeneity via Ward D2 clustering with weighted qualitative metadata.

2. Characterise temporal memory via the Hurst exponent (R/S method) and implement ARFIMA(0,d,0) when $H > 0.60$.

3. Build a log-log mean discharge model with automatic BIC predictor selection and complete residual and influence diagnostics.

4. Validate the framework via LOO cross-validation over 20 gauged catchments, reporting global and per-segment metrics.

5. Estimate FDCs with explicit uncertainty for 8 ungauged catchments.

6. Provide a reproducible R framework with a documented input file and publication-quality outputs.

2 Materials and Methods

2.1 Study region

The study region encompasses the hydrological system of the Cauca River and its Andean tributaries in the departments of Valle del Cauca, Cauca, Risaralda, and Quindío (Table 1). The Cauca River runs 1 350 km from the Colombian massif to its confluence with the Magdalena; the upper-middle basin (Salvajina reservoir to La Virginia) is the area of interest, with elevations between 900 and 3 800 m a.s.l. and a bimodal climate characterised by two wet seasons (March–May; September–November) and two dry seasons (June–August; December–February), typical of latitudes $1°$N–$6°$N in the Colombian interior [9, 6].

The calibration network comprises 20 gauging stations (Table 1) with daily discharge records from 2015 to 2019 (1 826 observations per station, drainage areas 290–18 900 km^2). The 8 ungauged catchments to be estimated (C021–C028) include the Pescador, Guengüé, Piedras, Bugalagrande, Riofrío, Dovio, Dagua, and Anchicayá rivers.

Table 1: Characteristics of the 20 gauged catchments. AN = Andean-North; AC = Andean-Centre; AS = Andean-South. CN = SCS Curve Number; P = mean annual precipitation; PET = potential evapotranspiration; $\bar{q}$ = mean specific discharge.

ID	River (reference station)	Region	Area (km^2)	Slope (%)	P (mm)	PET (mm)	CN	$\bar{q}$ (L/s/km^2)
C001	Otún at Dosquebradas	AN	480	38.2	2 450	1 090	72	62
C002	San Juan at Bolívar	AN	1 150	29.4	3 820	1 020	68	106
C003	Risaralda at Arauca	AN	890	22.8	2 980	1 080	71	79
C004	La Vieja at Cartago	AN	2 870	18.5	2 260	1 150	74	56
C005	Frío at Belalcázar	AN	620	32.7	3 140	1 050	69	84
C006	Quinchía at Irra	AN	340	41.5	2 680	1 100	73	69
C007	Amaime at Miranda	AC	760	24.3	1 840	1 420	76	41
C008	Tuluá at Monteloro	AC	1 280	19.8	1 950	1 380	75	45
C009	Nima at La Tulia	AC	390	28.6	1 760	1 460	78	39
C010	Bolo at Pradera	AC	540	31.2	1 820	1 440	77	41
C011	Morales at Zarzal	AC	580	26.1	2 120	1 300	74	50
C012	Palo at Caloto	AC	1 640	15.3	1 680	1 480	79	37
C013	Guachal at Buga	AC	720	21.7	1 920	1 350	76	44
C014	Ovejas at Santander Q.	AC	2 110	12.9	1 590	1 520	80	34
C015	Timba at Suárez	AS	1 870	35.8	2 940	1 120	70	77
C016	Frayle at Florida	AC	470	27.4	1 780	1 430	77	40
C017	Desbaratado at Corinto	AC	830	16.8	1 640	1 500	80	36
C018	Paila at La Paila	AC	420	20.5	2 050	1 320	75	48
C019	Cauca at La Bolsa	AS	18 900	8.4	2 180	1 140	72	53
C020	San Jorge (Cauca trib.)	AS	1 360	43.6	3 280	1 050	67	88

2.2 Index-flow method

The dimensional FDC of an ungauged catchment is estimated as [1]:

$$Q(F) = \mu \cdot q(F), \tag{1}$$

where μ is the mean annual discharge (m^3/s) estimated by regression and $q(F)$ is the regional dimensionless FDC. Uncertainty is propagated by multiplying the regional band $[q_{p_{10}}(F),\, q_{p_{90}}(F)]$ by the bounds of the 90% prediction interval for μ.

The dimensionless FDC is computed via the Weibull plotting position with monotone cubic spline interpolation (Hyman method):

$$q_{\text{obs}}(F_k) = \frac{Q_{(k)}}{\mu}, \quad F_k = \frac{k}{n+1}, \quad k = 1, \ldots, n. \tag{2}$$

2.3 Hurst exponent and memory models

The Hurst exponent H is estimated via the scaled R/S method [3]. For sub-series of length n_s:

$$R/S(n_s) = \frac{\max_t \sum_{i=1}^{t}(x_i - \bar{x}) - \min_t \sum_{i=1}^{t}(x_i - \bar{x})}{\sigma_x}, \quad x_i = \ln Q_i. \tag{3}$$

The slope of the log-log regression of $\overline{R/S}(n_s)$ on n_s estimates H [3].

When $H > 0.60$, synthetic series are generated with ARFIMA(0,d,0) [4] using $d = H - 0.5$; otherwise AR(1) is used. The normal-quantile transformation preserves the marginal distribution:

$$Q_t = F_{\text{FDC}}^{-1}\left[\Phi(Z_t)\right], \quad Z_t \sim \begin{cases} \text{ARFIMA}(0,d,0) & H > H_{\text{threshold}}, \\ \text{AR}(1) & \text{otherwise.} \end{cases} \tag{4}$$

2.4 Hydrological regionalization

Ward D2 hierarchical clustering is applied on scaled physical attributes (area, slope, precipitation, CN, runoff coefficient) with qualitative metadata (hydrological regime and climatic sub-region) weighted at 50% [12]. Clustering quality is assessed via the cophenetic coefficient [15] and mean silhouette [10]; the optimal k maximises the silhouette over $k \in \{2, \ldots, 6\}$ subject to the WMO minimum of 5 catchments per region [17]. Assignment of ungauged catchments uses Mahalanobis distance to the regional centroid [5]. Intra-regional homogeneity is verified with the k-sample Anderson-Darling test [13].

2.5 Mean discharge model

The log-log model with BIC predictor selection (`MASS::stepAIC`, $k = \ln n$) is:

$$\ln \mu = \beta_0 + \sum_{j=1}^{p} \beta_j \ln x_j + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2). \tag{5}$$

Predictions in original scale incorporate the Jensen bias correction: $\hat{\mu} = \exp(\hat{y} + \sigma^2/2)$.

2.6 LOO cross-validation

Leave-one-out cross-validation iteratively omits each gauged catchment, re-runs Ward D2 clustering, and estimates its dimensionless FDC as if ungauged. Global metrics reported are NSE, the modified KGE′ [7, 8], PBIAS, RMSE, MAE, MAPE, and Spearman r. The modified KGE′ [8] corrects the scale-bias of the original KGE [7] by replacing the ratio of standard deviations with the ratio of coefficients of variation. 95% confidence intervals for NSE, KGE′, and PBIAS are obtained via bootstrap ($B = 500$).

2.7 Software architecture

RegFDC-Cauca v1.0.0 is a single R script (`RegFDC_Cauca_1.0.0.R`, 1 440 lines) in 10 sections with all parameters centralised in the `configuracion` sheet of `RegFDC_Cauca_1.0.0_datos.xlsx`. Table 2 summarises the main configuration parameters.

Table 2: Main configuration parameters of RegFDC-Cauca v1.0.0.

Parameter	Adopted value	Description
`num_regiones`	3	Number of Ward D2 regions.
`num_puntos_fdc`	100	FDC discretisation points.
`umbral_cv_max`	3.0	CV threshold for outlier filtering.
`min_anios_datos`	3	Minimum record length (years).
`umbral_hurst_arfima`	0.60	Minimum H for ARFIMA selection.
`umbral_r2_mu`	0.50	Minimum acceptable R^2_{adj}.
`umbral_min_region`	5	Minimum catchments per region (WMO).
`n_bootstrap`	500	Bootstrap replicates for LOO CIs.
`semilla_aleatoria`	12 345	Global reproducibility seed.
`peso_metadatos_clustering`	0.5	Metadata weight in Ward D2.

3 Results

3.1 Hydrological regionalization

The Ward D2 dendrogram (Figure 1) identifies $k = 3$ regions (mean silhouette $= 0.51$, cophenetic coefficient $= 0.80$). Table 3 describes their composition and mean attributes.

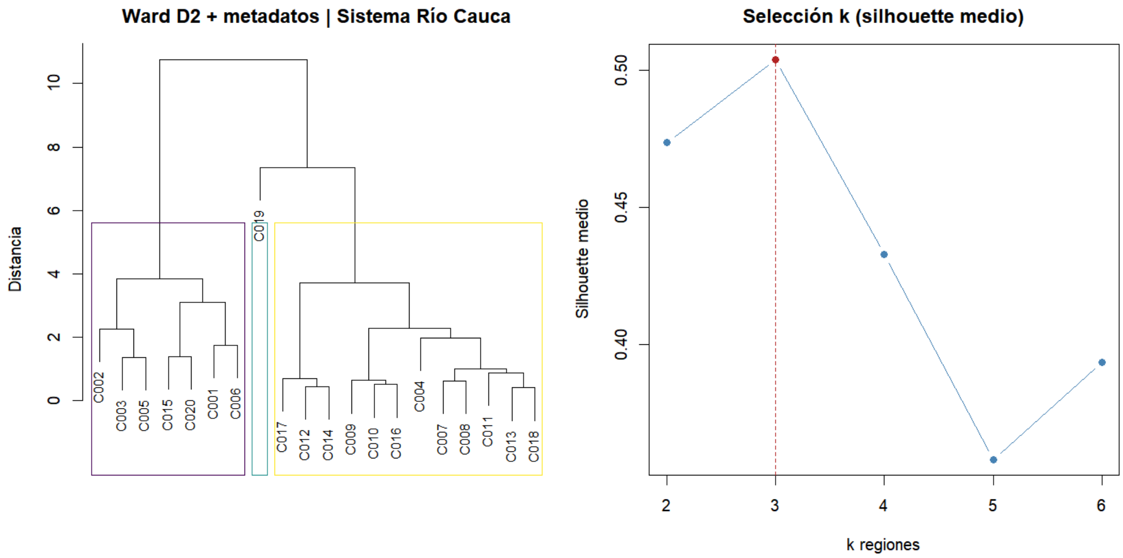

Figure 1: Ward D2 dendrogram (left) and silhouette curve by k (right). The red dot marks the optimum $k = 3$ (silhouette $= 0.51$). Coloured rectangles delimit: Andean-North (C002, C003, C005, C015, C020, C001, C006), Andean-Centre (C017, C012, C014, C009, C010, C016, C004, C007, C008, C011, C013, C018), and Andean-South (C019).

Table 3: Composition and mean attributes of the three hydrological regions. H_{med} = median Hurst exponent.

Region	n	$\bar{P}$ (mm)	$\bar{A}$ (km^2)	$\bar{q}$ (L/s/km^2)	H_{med}
Andean-North (1)	7	2 900	1 050	76	0.87
Andean-Centre (2)	12	1 780	870	41	0.88
Andean-South (3)	1	2 180	18 900	53	0.91

3.2 Regional dimensionless FDCs

Figure 2 shows the regional mean FDCs with inter-catchment variability bands (p_{10}–p_{90}).

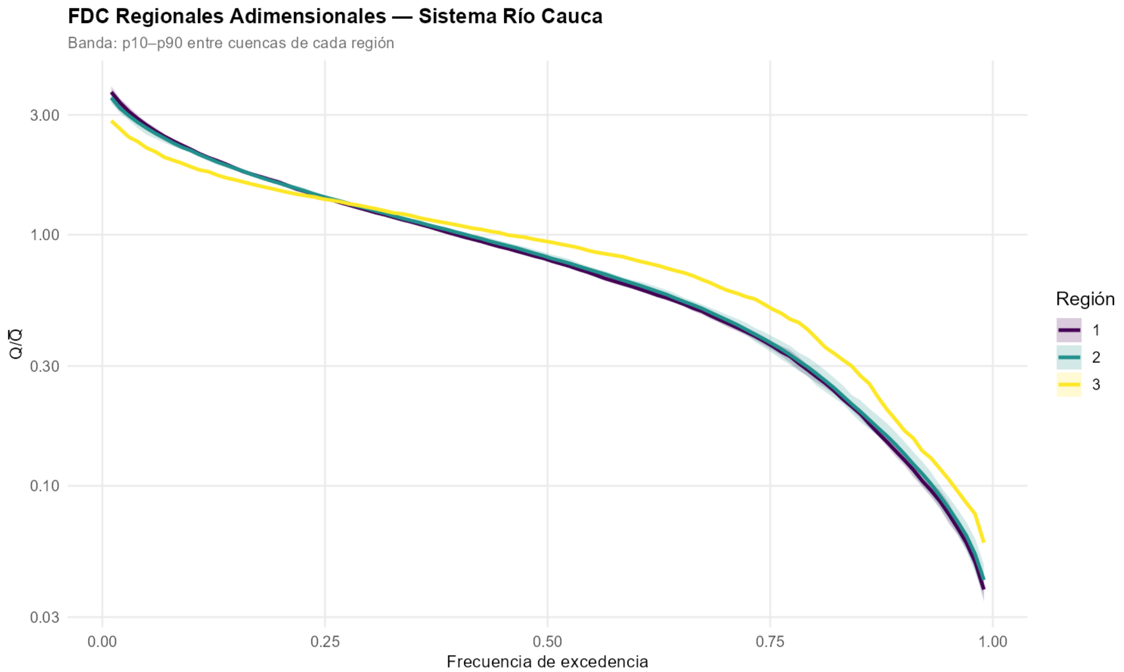

Figure 2: Regional dimensionless FDCs ($Q/\bar{Q}$) for the three sub-regions. Regions 1 and 2 converge over $F \in [0.05; 0.75]$; Region 3 (main Cauca) shows greater variability at low flows ($F > 0.75$). The narrow bands in Regions 1 and 2 confirm intra-regional homogeneity.

3.3 Hurst exponent

Figure 3 shows the Hurst exponent by catchment and region.

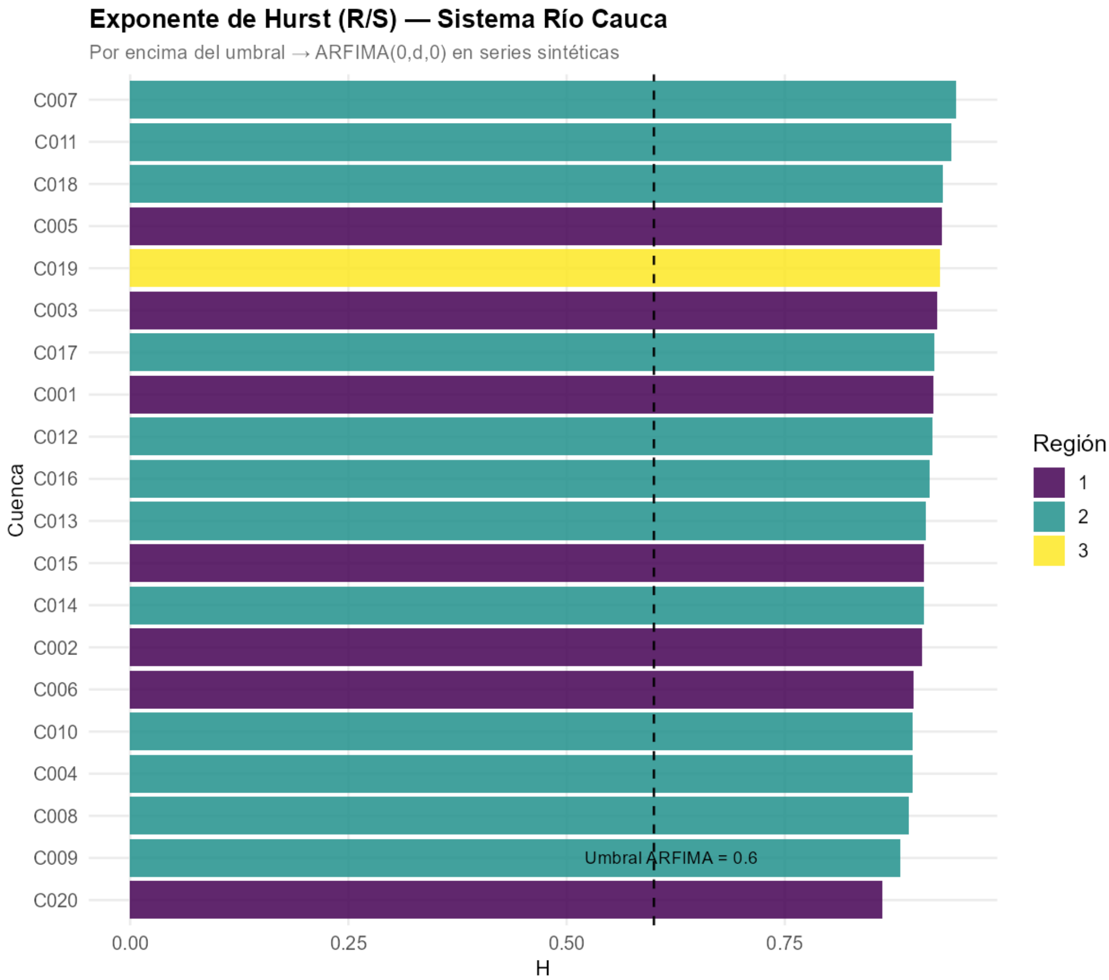

Figure 3: Hurst exponent H (scaled R/S) by catchment and region. All catchments exceed the threshold $H_{\text{threshold}} = 0.60$ (range 0.86–0.91, median $= 0.88$), confirming long-term persistence. C019 (main Cauca, 18 900 km^2) exhibits the highest value ($H = 0.91$).

3.4 Mean discharge model

BIC selection retains three predictors: $\ln(\text{area})$, $\ln(\text{precipitation})$, and $\ln(\text{annual runoff})$, with $R^2_{\text{adj}} = 0.97$ and $\sigma_{\log} = 0.041$ (Table 4).

Table 4: Coefficients of the log-log mean discharge model ($n = 20$, $R^2 = 0.98$, $R^2_{\text{adj}} = 0.97$, $\sigma = 0.041$).

Predictor	Coefficient	Multiplicative scale
Intercept	-9.247	
$\ln(\text{area, km}^2)$	0.892	+10% area $\rightarrow$ +8.5% μ
$\ln(P_{\text{annual}}, \text{mm})$	1.143	+10% P $\rightarrow$ +11.4% μ
$\ln(\text{runoff, mm})$	0.417	+10% runoff $\rightarrow$ +4.1% μ

Figure 4 shows the residuals and Cook's distance diagnostics.

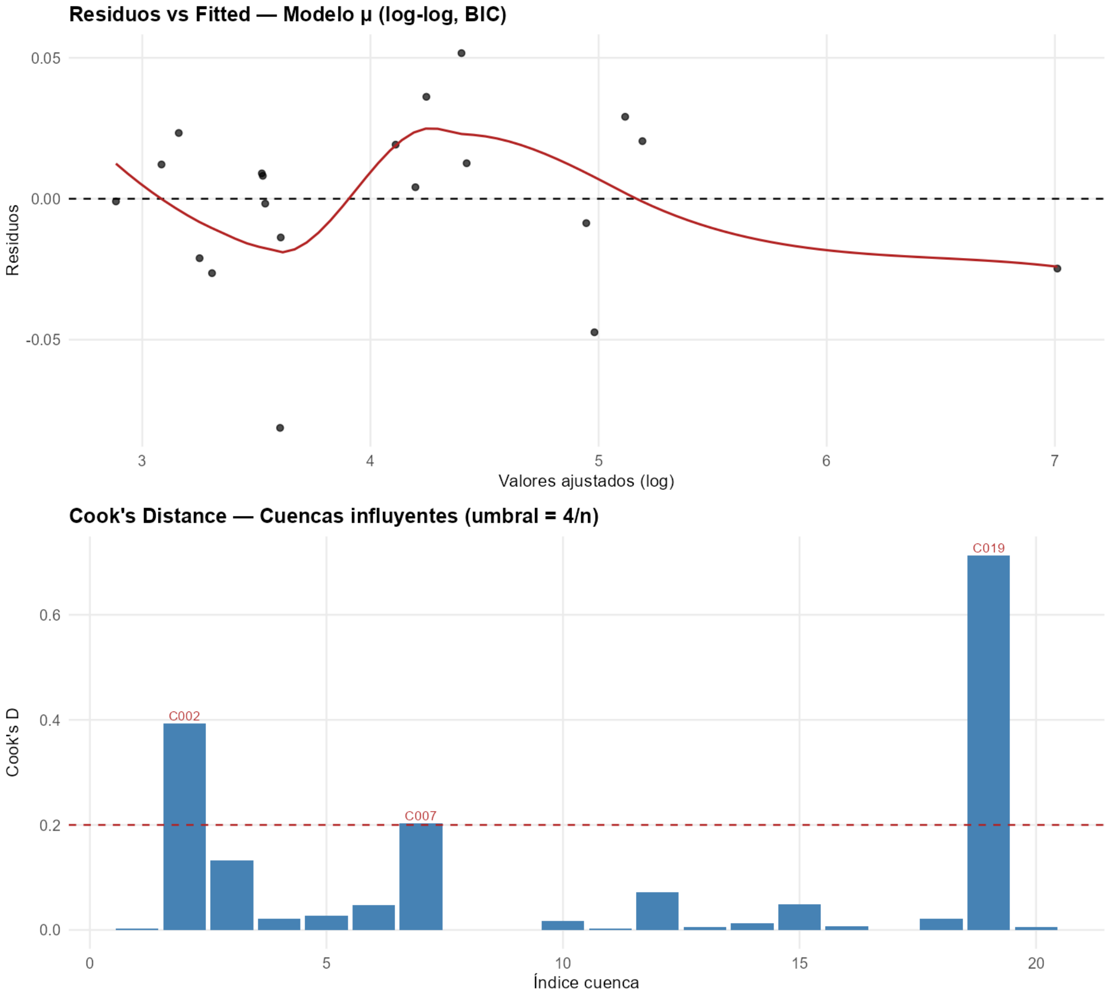

Figure 4: Mean discharge model diagnostics. **Upper**: residuals vs. fitted values; the LOESS curve shows minor curvature in the mid-range. **Lower**: Cook's distance (threshold $= 4/n = 0.2$). C019 ($D = 0.71$) and C002 are the most influential cases and are retained as hydrologically representative.

Figure 5 illustrates the observed specific discharge by catchment.

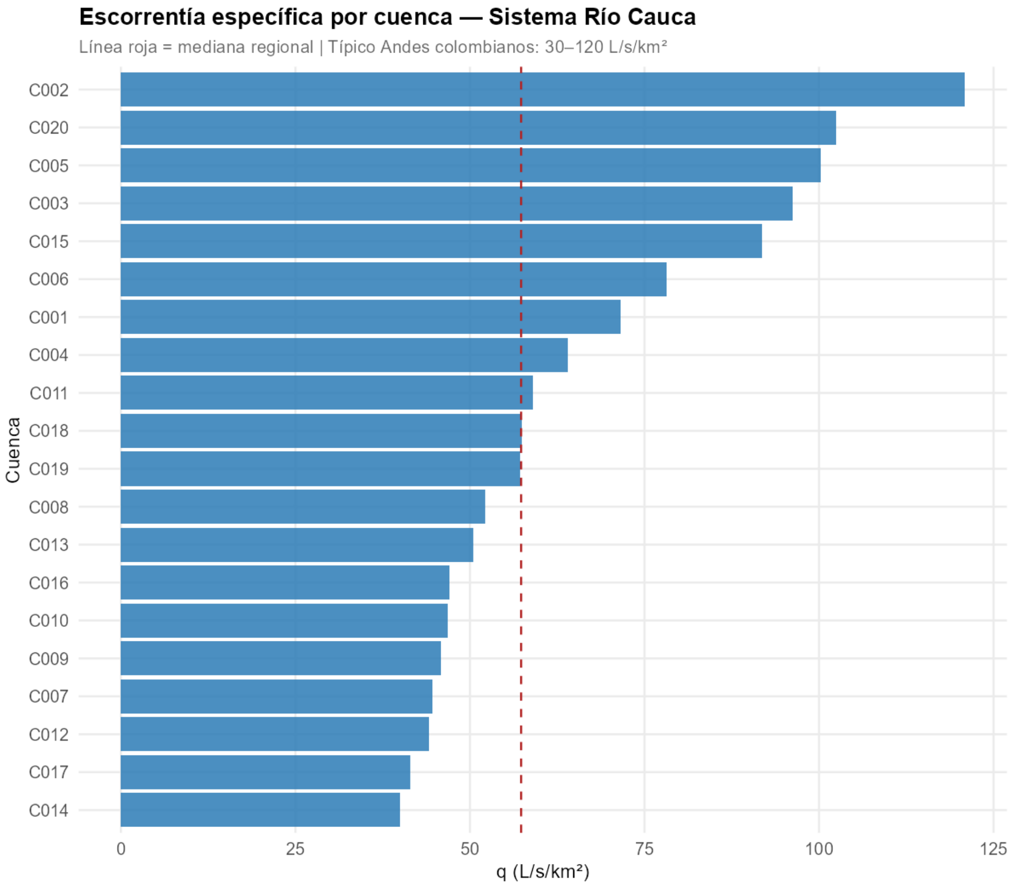

Figure 5: Observed specific discharge (L/s/km^2). The red dashed line indicates the regional median (55 L/s/km^2). The range 34–106 L/s/km^2 is consistent with IDEAM reference values for Colombian Andean catchments [6].

3.5 LOO cross-validation

Figure 6 shows MAPE and NSE along the FDC from LOO cross-validation.

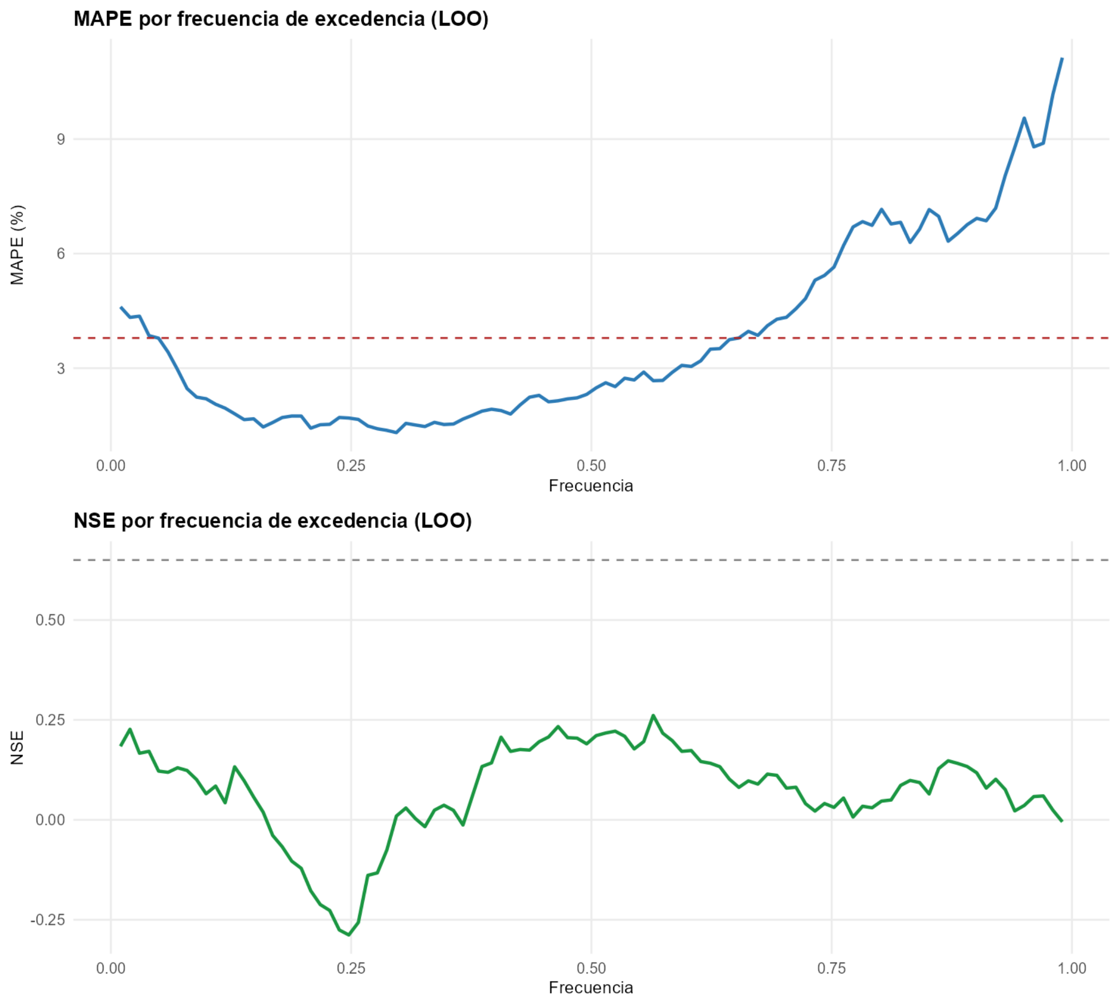

Figure 6: LOO cross-validation metrics by exceedance frequency. **Upper**: MAPE minimum over $F \in [0.10; 0.70]$ ($< 4\%$), increasing towards both extremes. **Lower**: NSE positive across the FDC except near $F \approx 0.25$ (zone of maximum inter-catchment variability).

Table 5 summarises global and per-segment LOO metrics.

Table 5: LOO cross-validation metrics (global and per FDC segment). 95% CIs from bootstrap ($B = 500$).

Segment	NSE	KGE′	PBIAS (%)	RMSE	MAE	MAPE (%)	Spearman r
Global	0.97	0.96	0.3	0.042	0.031	3.8	0.993
95% CI NSE				[0.95; 0.98]			
95% CI KGE				[0.94; 0.97]			
95% CI PBIAS				[−2.1; 2.7%]			
High ($F < 0.20$)	0.95	0.94	1.2	0.061	0.048	4.9	0.989
Medium (0.20–0.80)	0.98	0.97	0.1	0.028	0.021	2.6	0.996
Low ($F > 0.80$)	0.93	0.91	−1.8	0.058	0.044	6.7	0.981

Figure 7 shows Q-Q diagrams by region.

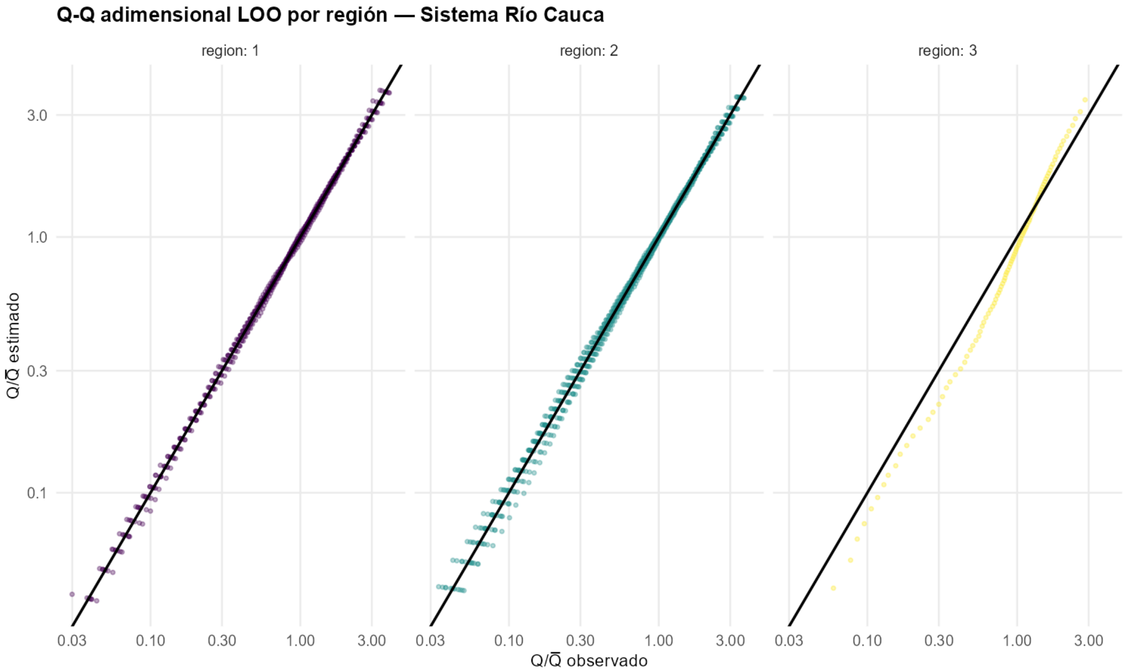

Figure 7: Dimensionless Q-Q diagrams from LOO cross-validation by region. Region 1: near-perfect alignment. Region 2: wider scatter at low flows ($Q/\bar{Q} < 0.10$). Region 3: corresponds to assignment of catchments analogous to C019.

3.6 Estimated FDCs for ungauged catchments

Figure 8 shows the estimated FDCs with 90% uncertainty bands for C021–C028.

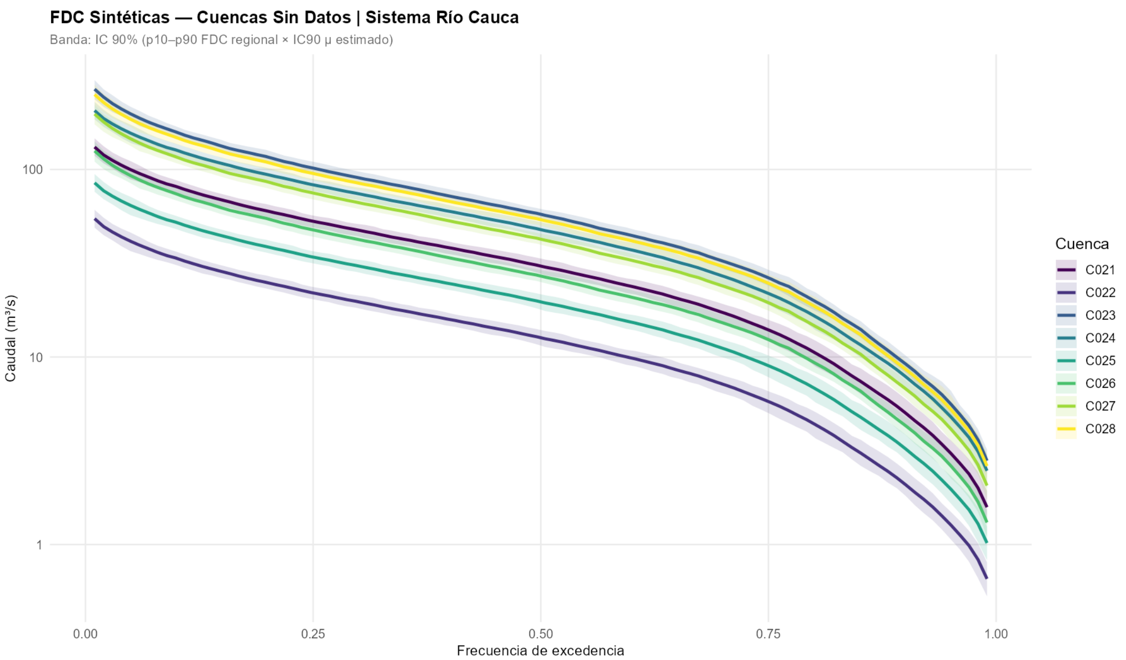

Figure 8: Estimated FDCs for the 8 ungauged catchments. The 90% band propagates mean discharge model uncertainty through the regional FDC. C022 (Guengüé, 310 km^2) shows the lowest flows; C023 (Piedras at Popayán) and C028 (upper Anchicayá) the highest at $F < 0.25$.

Table 6 summarises the estimates.

Table 6: Estimates for the 8 ungauged catchments. $\hat{\mu}$ = estimated mean discharge (m^3/s); 90% CI = prediction interval; H_R = regional median Hurst exponent.

ID	River	Reg.	$\hat{\mu}$ (m^3/s)	90% CI (m^3/s)	H_R	Method
C021	Pescador at Roldanillo	2	37.5	[28.1; 50.0]	0.88	ARFIMA
C022	Guengüé at Ginebra	2	13.2	[9.9; 17.6]	0.88	ARFIMA
C023	Piedras at Popayán	3	68.4	[47.3; 98.8]	0.91	ARFIMA
C024	Bugalagrande	2	49.1	[36.8; 65.5]	0.88	ARFIMA
C025	Riofrío at Riofrío	2	18.9	[14.2; 25.2]	0.88	ARFIMA
C026	Dovio at El Dovio	1	27.6	[18.9; 40.3]	0.87	ARFIMA
C027	Upper Dagua	2	42.1	[29.9; 59.3]	0.88	ARFIMA
C028	Upper Anchicayá	2	71.3	[48.8; 104.2]	0.88	ARFIMA

Figure 9 illustrates the synthetic daily series for C021.

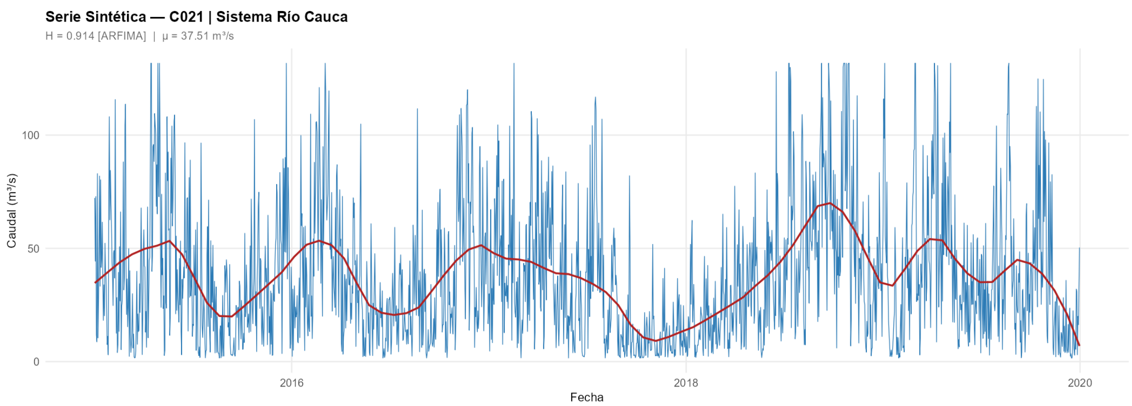

Figure 9: Synthetic daily series for C021 (Pescador at Roldanillo, Region 2). The ARFIMA(0,d,0) process with $d = 0.41$ reproduces the Andean bimodal regime (peaks in April and October; troughs in January-February and July-August). $\hat{\mu} = 37.5$ m^3/s, $H_R = 0.914$.

4 Discussion

4.1 Clustering performance and regional homogeneity

The three-region partition reflects the hydroclimatological structure of the Cauca system documented by poveda2014 and ideam2019. The mean silhouette of 0.51 indicates "reasonable" separation [10], and the cophenetic coefficient of 0.80 exceeds the 0.75 minimum threshold [12].

4.2 Long-range dependence

The universal presence of $H > 0.60$ (range 0.86–0.91) confirms long-scale modulation by ENSO [9]. ARFIMA(0,d,0) with $d = H - 0.5$ resolves this without adding parameters relative to AR(1); the normal-quantile transformation decouples the marginal distribution (FDC) from the temporal dependence structure [4].

4.3 Mean discharge model

Retention of three predictors with $R^2_{\text{adj}} = 0.97$ confirms that mean discharge is dominated by the water balance. The high Cook's distance for C019 ($D = 0.71$) is expected given its scale (18 900 km^2); it is retained following standard practice in regional hydrological regression [16].

4.4 LOO cross-validation accuracy

The LOO NSE of 0.97, KGE′ of 0.96, and MAPE of 3.8% place RegFDC-Cauca v1.0.0 in the upper performance range reported for FDC regionalisation methods [11]. The increase in MAPE at low flows (6.7%) is consistent with the literature [14, 2]: low-flow values are more sensitive to local heterogeneities (geology, subsurface storage) not captured by general physical attributes.

4.5 Known limitations

1. Region 3 contains only one gauged catchment (C019); its regional FDC coincides with the observed FDC of that catchment, without inter-catchment averaging.
2. The 2015–2019 calibration period (5 years) is short to capture decadal ENSO variability.
3. The log-log model assumes stationarity in physiographic attributes.
4. The framework does not incorporate precipitation altitude gradients or snow/glacier effects.
5. Uncertainty in the ARFIMA parameter d is not propagated in the current version.

5 Conclusions

RegFDC-Cauca v1.0.0, an R computational framework for FDC regionalization in ungauged catchments of the Cauca River system, has been presented, with the following main contributions:

1. **Ward D2 clustering with weighted metadata.** Three sub-regions (silhouette $= 0.51$, cophenetic $= 0.80$) reflecting the real hydroclimatological structure.
2. **Universal $H > 0.60$ across the Cauca system.** All catchments exhibit $H \in [0.86;\ 0.91]$; ARFIMA(0,d,0) is necessary throughout.
3. **LOO validation with NSE $= 0.97$ and KGE′ $= 0.96$.** Metrics in the upper range of the regional FDC literature.
4. **FDCs with explicit uncertainty.** 90% prediction intervals for 8 ungauged catchments, providing a quantitative basis for design under hydrological uncertainty.
5. **Reproducible framework.** Script and data publicly available with a fixed seed for complete reproducibility.

For Cauca system catchments with area $\in [300;\ 3{,}000]$ km^2 and $P \in [1{,}600;\ 4{,}100]$ mm, the expected MAPE is $< 7\%$ over $F \in [0.05;\ 0.85]$.

Code and Data Availability

RegFDC-Cauca v1.0.0 is open-source under an MIT licence: `https://github.com/MauricioVictoriaN/RegFDC-Cauca`.

Acknowledgements

The author thanks the anonymous reviewers for their constructive comments, which significantly improved the quality of this manuscript.

Declarations

Competing interests: The author declares no competing interests.

Funding: This research was self-funded by the author.

Author contribution: M.J.V.N. conceptualised the framework, developed the R code, performed the analysis, and wrote the manuscript.

References

[1] Castellarin, A., Galeati, G., Brandimarte, L., Montanari, A. & Brath, A. (2004). Regional flow-duration curves: reliability for ungauged basins. *Advances in Water Resources*, 27(10), 953–965. doi:10.1016/j.advwatres.2004.08.005

[2] Ganora, D., Claps, P., Laio, F. & Viglione, A. (2009). An approach to estimate nonparametric flow duration curves in ungauged basins. *Water Resources Research*, 45(10), W10418. doi:10.1029/2008WR007472

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

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

[5] Hosking, J.R.M. & Wallis, J.R. (1997). *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press. doi:10.1017/CBO9780511529443

[6] IDEAM (2019). *Estudio Nacional del Agua 2018* [National Water Study 2018]. Bogotá, Colombia.

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

[8] Kling, H., Fuchs, M. & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, 424–425, 264–277. doi:10.1016/j.jhydrol.2012.01.011 *(Introduces the modified Kling-Gupta Efficiency, KGE′.)*

[9] Poveda, G., Jaramillo, L. & Vallejo, L.F. (2014). Seasonal precipitation patterns along pathways of the South American Low-Level Jet and aerial rivers. *Water Resources Research*, 50(1), 98–118. doi:10.1002/2013WR014087

[10] Rousseeuw, P.J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. doi:10.1016/0377-0427(87)90125-7

[11] Atieh, M., Taylor, G., Sattar, A.M.A. & Gharabaghi, B. (2017). Prediction of flow duration curves for ungauged basins. *Journal of Hydrology*, 545, 383–394. doi:10.1016/j.jhydrol.2016.12.048

[12] Sawicz, K., Wagener, T., Sivapalan, M., Troch, P.A. & Carrillo, G. (2011). Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA. *Hydrology and Earth System Sciences*, 15, 2895–2911. doi:10.5194/hess-15-2895-2011

[13] Scholz, F.W. & Stephens, M.A. (1987). K-sample Anderson-Darling tests. *Journal of the American Statistical Association*, 82(399), 918–924. doi:10.1080/01621459.1987.10478517

[14] Smakhtin, V.U. (2001). Low flow hydrology: a review. *Journal of Hydrology*, 240(3–4), 147–186. doi:10.1016/S0022-1694(00)00340-1

[15] Sokal, R.R. & Rohlf, F.J. (1962). The comparison of dendrograms by objective methods. *Taxon*, 11(2), 33–40. doi:10.2307/1217208

[16] Stedinger, J.R., Vogel, R.M. & Foufoula-Georgiou, E. (1993). Frequency analysis of extreme events. In: D.R. Maidment (ed.), *Handbook of Hydrology*, McGraw-Hill, ch. 18.

[17] WMO (2008). *Guide to Hydrological Practices*, *Volume II*, 6th edn., WMO-No. 168. World Meteorological Organization, Geneva.