

# GRADEX-DUAL: A Dual Framework Combining Point and Regional Frequency Analysis for GRADEX Parameter Estimation in Basins with Sparse Rain Gauge Networks

Mauricio Javier Victoria Niño\*

*This document is a preprint that has not been peer-reviewed, submitted to EarthArXiv. Source code and example data are available at: <https://github.com/MauricioVictoria/GRADEX-DUAL>.*

## Abstract

GRADEX-DUAL v1.0, an open-source computational framework implemented in R, is presented. It combines two complementary methodological layers for estimating the GRADEX parameter —defined as the slope of the annual maximum precipitation quantile function on Gumbel probability paper between return periods  $T = 10$  and  $T = 100$  years—, which is fundamental for the GRADEX-IDF methodology applied to hydrological design with HEC-HMS.

The first layer follows the classic pointwise approach of Guillot and Duband (1967): fitting four probability distributions per station (Gumbel, Log-Normal, Pearson III, and Gamma) using L-moments and maximum likelihood; two-stage model selection (Anderson-Darling followed by minimum RMSE); and regionalization to the gauging point using Inverse Distance Weighting (IDW) with leave-one-out (LOO) cross-validated optimal exponent or Ordinary Kriging with automatic variogram fitting. A multi-metric scoring rule ( $\text{RMSE} + R^2 + |\text{BIAS}|$ ) selects the better spatial method, with an anti-degenerate-Kriging safeguard that disqualifies Kriging whenever its LOO prediction-to-observation variance ratio falls below 0.10 —indicating variogram collapse to a near-constant predictor.

The second layer implements the Hosking and Wallis (1997) regional frequency analysis: record-length-weighted regional L-moments, selection of the optimal regional distribution among five candidate distributions (GEV, GLO, GNO, PE3, and GPA) using the Monte Carlo-calibrated goodness-of-fit  $Z$  statistic, and 95% confidence intervals computed using the H&W bootstrap simulation algorithm.

A three-gate quality-control weighting system integrates both estimates: the H&W weight is nullified when the bootstrap confidence interval amplitude exceeds 100%, the spatial weight is nullified when the LOO cross-validation  $R^2$  falls below 0.3, and Kriging is disqualified when its LOO variance ratio falls below 0.10. Prior diagnostics include the Ljung-Box test for serial independence, the Hosking-Wallis  $H$  statistic for regional homogeneity, the discordancy measure  $D$ , Mann-Kendall trend tests with Theil-Sen slope

---

\*Independent Researcher, Colombia. ORCID: 0009-0003-4328-5691.  
E-mail: [hidratecsa@gmail.com](mailto:hidratecsa@gmail.com)

estimator, a non-stationarity sensitivity analysis on detrended series, and a Monte Carlo validation experiment against synthetic regions with known true GRADEX.

The framework is evaluated on a nine-station IDEAM network in the upper Porce River basin (Antioquia, Colombia, 254 station-years). The Porce is a major tributary of the Nechí River, which in turn flows into the Cauca River and ultimately into the Magdalena, so the study area belongs hydrographically to the lower Cauca macro-basin. Results show that even with a homogeneous region ( $H = -0.92$ ) and a network surpassing the H&W minimum threshold, the bootstrap confidence interval amplitude reaches 121% and the spatial  $R^2$  collapses to 0.006 once the degenerate-Kriging safeguard is activated, triggering the framework's fallback regime with a  $\pm 30\%$  sensitivity range. A Monte Carlo experiment (200 replicates of 9-station, 30-year synthetic regions sampled from a GEV with  $\kappa = -0.10$ ) confirms an empirical coverage of 85.0% within the  $\pm 30\%$  interval, and quantifies a systematic underestimation bias of  $-20.7\%$  attributable to fitting Gumbel L-moments to data drawn from a heavy-tailed GEV.

**Keywords:** GRADEX, extreme precipitation, L-moments, regional frequency analysis, IDW, Kriging, HEC-HMS, hydrological design, Colombia, Hosking-Wallis

---

## Resumen (Spanish)

Se presenta GRADEX-DUAL v1.0, un marco computacional de código abierto implementado en R que combina dos capas metodológicas complementarias para la estimación del parámetro GRADEX —definido como la pendiente de la función cuantil de precipitaciones máximas anuales en papel de probabilidad de Gumbel entre los períodos de retorno  $T = 10$  y  $T = 100$  años—, fundamental para la metodología GRADEX-IDF aplicada al diseño hidrológico con HEC-HMS.

La primera capa sigue el enfoque puntual clásico de Guillot y Duband (1967): ajuste de cuatro distribuciones de probabilidad por estación (Gumbel, Log-Normal, Pearson III y Gamma) mediante L-momentos y máxima verosimilitud; selección del mejor modelo en dos etapas (Anderson-Darling seguido de RMSE mínimo); y regionalización al punto de aforo por Distancia Inversa Ponderada (IDW) con exponente optimizado mediante validación cruzada leave-one-out (LOO) u Kriging Ordinario con variograma automático. Una regla de puntuación multi-métrica ( $RMSE + R^2 + |SESGO|$ ) selecciona el método espacial, con una salvaguarda anti-Kriging-degenerado que descalifica al Kriging cuando la razón de varianza de las predicciones LOO sobre las observaciones cae por debajo de 0,10 —indicando colapso del variograma a un predictor casi constante.

La segunda capa implementa el análisis frecuencial regional de Hosking y Wallis (1997): L-momentos regionales ponderados por longitud de registro, selección de la distribución regional óptima entre cinco distribuciones candidatas (GEV, GLO, GNO, PE3 y GPA) mediante el estadístico  $Z$  de bondad de ajuste calibrado por Montecarlo, e intervalo de confianza al 95% calculado mediante el algoritmo de simulación bootstrap de H&W.

Un sistema de ponderación con tres compuertas de calidad integra ambas estimaciones: el peso H&W se anula cuando la amplitud del IC bootstrap supera el 100%, el peso espacial se anula cuando el  $R^2$  de la validación cruzada LOO es inferior a 0,3, y el Kriging se descalifica cuando su razón de varianza LOO cae por debajo de 0,10. Los diagnósticos previos incluyen la prueba de Ljung-Box, el estadístico  $H$  de homogeneidad regional, la

medida de discordancia  $D$ , pruebas de tendencia Mann-Kendall con estimador de pendiente Theil-Sen, un análisis de sensibilidad sobre series detendidas, y un experimento de validación Montecarlo contra regiones sintéticas con GRADEX verdadero conocido.

El marco se evalúa en una red de nueve estaciones IDEAM en la cuenca alta del río Porce (norte de Antioquia, Colombia, 254 estación-años). El Porce es uno de los principales afluentes del río Nechí, el cual entrega sus aguas al río Cauca, quien a su vez desemboca en el Magdalena; el área de estudio pertenece, por tanto, hidrográficamente a la *cuenca baja del río Cauca*. Los resultados demuestran que aún con una región homogénea ( $H = -0,92$ ) y un tamaño de red que supera el umbral mínimo H&W, la amplitud del IC bootstrap alcanza el 121% y el  $R^2$  espacial colapsa a 0,006 una vez activada la salvaguarda anti-Kriging-degenerado, lo que activa el régimen de respaldo del marco con un rango de sensibilidad del  $\pm 30\%$ . Un experimento Montecarlo (200 réplicas de regiones sintéticas de 9 estaciones  $\times$  30 años muestreadas desde una GEV con  $\kappa = -0,10$ ) confirma una cobertura empírica del 85,0% dentro del intervalo  $\pm 30\%$ , y cuantifica un sesgo sistemático de subestimación del  $-20,7\%$  atribuible al ajuste de L-momentos de Gumbel a datos provenientes de una GEV de cola pesada.

**Palabras clave:** GRADEX, precipitación extrema, L-momentos, análisis frecuencial regional, IDW, Kriging, HEC-HMS, diseño hidrológico, Colombia, Hosking–Wallis

---

## 1 Introduction

### 1.1 Context and motivation

The estimation of extreme precipitation for the hydrological design of infrastructure—dams, bridges, sewers, and flood defenses—constitutes one of the most challenging problems in water resources engineering, particularly in regions with sparse rain gauge networks and short record series [8]. In Colombia, the density of rain gauge stations is significantly lower than the recommendations of the World Meteorological Organization [9], which suggests one station per 250–500 km<sup>2</sup> in mountainous regions.

The GRADEX method (GRAdient of EXtreme values), originally proposed by [1] and subsequently formalized by [2] and [3], offers a physically based approach for extrapolating the design precipitation beyond the available observation period. The GRADEX parameter—defined as the slope of the annual maximum precipitation quantile function plotted on Gumbel probability paper between return periods  $T = 10$  and  $T = 100$  years—encapsulates the asymptotic behavior of the distribution tail and provides a direct link to design flood estimation in the GRADEX-IDF methodology [2].

However, the classic pointwise approach accumulates two sources of uncertainty: distribution parameter estimation at each station and spatial interpolation error. When networks are small—a frequent situation in developing countries—these uncertainties can render the final value unreliable, without any explicit quantification or systematic quality control.

Regional frequency analysis, systematized by Hosking and Wallis [4] (hereinafter H&W), addresses the first source of uncertainty by combining information from hydrologically homogeneous stations. By standardizing the series to a dimensionless form using the index-flood concept and fitting a single regional distribution to the pooled L-moments,

the method drastically reduces the variance of parameter estimation, especially at stations with short records. The goodness-of-fit  $Z$  statistic provides a formal, Monte Carlo-calibrated criterion for selecting the regional distribution family, and the H&W bootstrap simulation algorithm allows deriving a confidence interval with correct nominal coverage by simultaneously propagating several uncertainty sources.

Despite these advantages, a direct comparison between the classic pointwise GRADEX and the H&W regional approach has not been systematically implemented in a single automated framework with quality gates for HEC-HMS applications. This article presents GRADEX-DUAL v1.0, an open-source R package that fills this methodological and operational gap, and additionally provides a Monte Carlo validation experiment that empirically quantifies the bias and coverage of the framework against synthetic regions with known true GRADEX.

## 1.2 Study objectives

The specific objectives of this work are:

1. To implement a dual estimation framework that combines pointwise GRADEX [1] with regional frequency analysis by L-moments [4] in a single R workflow.
2. To develop an automatic weighting system with three quality gates —H&W bootstrap CI amplitude, spatial cross-validation  $R^2$ , and anti-degenerate-Kriging variance ratio— that penalizes unreliable estimates.
3. To provide a complete set of prior diagnostics that verify the fundamental assumptions of frequency analysis: serial independence, regional homogeneity, absence of discordant stations, and temporal stationarity.
4. To implement a Monte Carlo validation experiment that empirically quantifies the bias, coverage, and RMSE of the framework against synthetic regions with known true GRADEX.
5. To generate publication-quality outputs: maps, cross-validation plots, and a structured Excel workbook with 17 sheets.
6. To evaluate the framework on a representative case study —a nine-station IDEAM network in the upper Porce River basin (northern Antioquia, Colombia), belonging to the Porce–Nechí–Cauca–Magdalena hydrographic system, with 254 station-years— under minimum network conditions recommended by H&W.

## 2 Materials and Methods

### 2.1 Theoretical foundation of the GRADEX parameter

The GRADEX parameter is defined as the slope of the quantile function  $Q(F)$  in the Gumbel reduced variate plane  $y = -\ln(-\ln F)$  over the range of high return periods [1]:

$$\text{GRADEX} = \frac{Q(F_2) - Q(F_1)}{y_2 - y_1}, \quad (1)$$

where  $F_k = 1 - 1/T_k$ , with  $T_1 = 10$  and  $T_2 = 100$  years. The corresponding Gumbel reduced variates are:

$$y_1 = -\ln(-\ln(1 - 1/10)) = 2.250367, \quad y_2 = -\ln(-\ln(1 - 1/100)) = 4.600149, \quad (2)$$

such that  $\Delta y = y_2 - y_1 = 2.349782$  is constant and independent of the fitted distribution.

For the Gumbel distribution (EV1), the GRADEX coincides exactly with the scale parameter  $\beta$ . For distributions other than Gumbel (Log-Normal, Pearson III, GEV), the quantile function is curved on Gumbel paper, so equation (1) computes a local chord slope between  $T = 10$  and  $T = 100$  years. This approximation, widely used in applied hydrology [3], may underestimate the true tail slope for heavy-tailed distributions at return periods exceeding  $T_2$ .

## 2.2 Probability distributions

Table 1 summarizes the distributions fitted in each layer of the framework and their main properties.

Table 1: Probability distributions in GRADEX-DUAL. GEV = Generalized Extreme Value; GLO = Generalized Logistic; GNO = Generalized Normal (equivalent to three-parameter Log-Normal, LN-3p); PE3 = Pearson Type III; GPA = Generalized Pareto; ML = maximum likelihood. The two-parameter Log-Normal distribution is fitted by ML in the pointwise layer for consistency with the standard implementation of the MASS package in R; the remaining pointwise distributions and all regional distributions are fitted by L-moments.

Distribution	Layer	Estimation	Parameters
Gumbel (EV1)	Pointwise	L-moments	$\mu$ (location), $\beta$ (scale)
Log-Normal (LN-2p)	Pointwise	ML	$\mu$ (meanlog), $\sigma$ (sdlog)
Pearson III	Pointwise	L-moments	$\mu, \sigma, \gamma$ (skewness)
Gamma	Pointwise	L-moments	$\alpha$ (shape), $\beta$ (scale)
GEV	Regional	L-moments	$\mu, \alpha, \kappa$ (shape)
GLO	Regional	L-moments	$\mu, \alpha, \kappa$
GNO (LN-3p)	Regional	L-moments	$\mu, \alpha, \kappa$
PE3	Regional	L-moments	$\mu, \sigma, \gamma$
GPA	Regional	L-moments	$\mu, \alpha, \kappa$

## 2.3 Estimation by L-moments

L-moments [5] are linear combinations of order statistics that offer greater robustness against outliers compared to conventional product moments. The  $r$ -th L-moment is defined as:

$$\lambda_r = \sum_k p_{r,k}^* \beta_k, \quad \text{where } \beta_k = \mathbb{E}[X \cdot F(X)^k]. \quad (3)$$

and  $p_{r,k}^*$  are the coefficients of the shifted Legendre polynomial (see [5] for details). The dimensionless L-moment ratios are: L-CV ( $\tau_2 = \lambda_2/\lambda_1$ ), L-skewness ( $\tau_3 = \lambda_3/\lambda_2$ ), and L-kurtosis ( $\tau_4 = \lambda_4/\lambda_2$ ).

## 2.4 Regional L-moment algorithm (H&W, eq. 4.3)

In the H&W layer, each series is first standardized by the station mean (index-flood step):

$$x_{i,j} = \frac{X_{i,j}}{\mu_i}, \quad j = 1, \dots, n_i. \quad (4)$$

Regional L-moment ratios are calculated as record-length-weighted averages:

$$\tau_r^R = \frac{\sum_i n_i \tau_r^{(i)}}{\sum_i n_i}, \quad (5)$$

where  $\tau_r^{(i)}$  is the  $r$ -th L-moment ratio at station  $i$ ,  $n_i$  is the record length, and  $N = \sum_i n_i$  is the total station-years.

## 2.5 Goodness-of-fit $Z$ statistic (H&W, eq. 5.6)

For each candidate distribution  $\mathcal{D}$ , the theoretical L-kurtosis  $\tau_4^{\mathcal{D}}$  is calculated from the parameters fitted to  $[1, \tau_2^R, \tau_3^R]$ . The  $Z$  statistic measures the discrepancy between the theoretical and observed regional L-kurtosis in units of the Monte Carlo standard deviation:

$$Z^{\mathcal{D}} = \frac{\tau_4^{\mathcal{D}} - \tau_4^R}{\sigma_4}. \quad (6)$$

where  $\sigma_4$  is estimated by simulating  $N_{\text{sim}} = 500$  homogeneous synthetic regions from a four-parameter kappa distribution fitted to the regional L-moments. The distribution with the smallest  $|Z^{\mathcal{D}}|$  among those satisfying  $|Z^{\mathcal{D}}| < 1.64$  (two-sided 10% significance) is selected.

## 2.6 Regional bootstrap confidence interval

The 95% confidence interval is computed using the bootstrap simulation algorithm described in Section 6.3 of [4]. For each of  $n_{\text{boot}}$  replicates, the following five steps are executed:

1. Simulate a homogeneous synthetic region by drawing independent samples of sizes  $n_1, \dots, n_k$  from the four-parameter kappa distribution fitted to the observed regional L-moments.
2. Re-estimate the regional L-moment ratios  $\tau_2^R$  and  $\tau_3^R$  from the synthetic region.
3. Re-fit the selected regional distribution to the synthetic regional L-moments.
4. Calculate GRADEX at each synthetic station using equation (1).
5. Re-interpolate the GRADEX values to the gauging point using IDW.

The 2.5 and 97.5 percentiles of the bootstrap distribution constitute the 95% confidence interval.

## 2.7 Spatial interpolation methods

### 2.7.1 Inverse Distance Weighting (IDW)

The IDW estimator at the ungauged point  $x_0$  is:

$$\hat{z}(x_0) = \frac{\sum_i w_i(x_0) z_i}{\sum_i w_i(x_0)}, \quad w_i = d_i^{-p}, \quad (7)$$

where  $z_i$  is the GRADEX at station  $i$ ,  $d_i$  is the Euclidean distance from station  $i$  to  $x_0$ , and  $p$  is the decay exponent. The optimal exponent is determined by minimizing the LOO cross-validation RMSE over the set of candidate values  $p \in \{1.0, 1.5, 2.0, 3.0\}$ .

### 2.7.2 Ordinary Kriging

Ordinary Kriging produces the Best Linear Unbiased Predictor (BLUP) by solving the system:

$$\begin{bmatrix} \gamma(h_{11}) & \cdots & \gamma(h_{1n}) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(h_{n1}) & \cdots & \gamma(h_{nn}) & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma(h_{10}) \\ \vdots \\ \gamma(h_{n0}) \\ 1 \end{bmatrix}, \quad (8)$$

where  $\gamma(h)$  is the semivariogram at lag  $h$ , automatically fitted using the automap package [6].

### 2.7.3 Multi-metric selection of the spatial method

The interpolation method is selected using a three-criterion scoring system evaluated with LOO cross-validation. Each criterion awards one point to the best-performing method: RMSE (lower wins),  $R^2$  (higher wins), and |BIAS| (lower absolute value wins). Ties are resolved by lower RMSE.

## 2.8 Weighting system with quality gates

The final GRADEX recommendation integrates both layers through a weighting scheme with three explicit gates.

### 2.8.1 Gate A: H&W bootstrap CI amplitude

The relative amplitude of the bootstrap CI is defined as:

$$A = \frac{\text{CI}_{\text{upper}} - \text{CI}_{\text{lower}}}{\text{GRADEX}_{\text{HW}}} \times 100\%. \quad (9)$$

A linear penalty factor is applied to the H&W weight:

$$\text{ci\_factor} = \max(0, 1 - A/100).$$

When  $A \geq 100\%$ ,  $\text{ci\_factor} = 0$  and the H&W weight is nullified regardless of any other metric.

### 2.8.2 Gate B: spatial cross-validation $R^2$

When the LOO cross-validation  $R^2$  of the selected spatial method is below 0.3, the spatial weight is nullified.

### 2.8.3 Gate C: anti-degenerate-Kriging variance ratio

When the variogram approaches pure nugget ( $N/S \rightarrow 1$ ), Ordinary Kriging collapses toward the regional mean: all LOO predictions concentrate near a constant value regardless of station location. Under this regime,  $R^2$  and  $|\text{BIAS}|$  can indicate spuriously favorable metrics for Kriging while RMSE remains unimproved over IDW. To detect this pathological case, the variance ratio is defined as:

$$r = \frac{\text{Var}(\hat{z}_{\text{LOO}})}{\text{Var}(z_{\text{obs}})}. \quad (10)$$

When  $r < 0.10$ , Kriging predictions span less than 10% of the observed variance and the method is *disqualified* from multi-metric selection regardless of its score; IDW is used instead. The 0.10 threshold is conservative: a healthy Kriging on real spatial structure typically produces  $r > 0.40$ .

### 2.8.4 Composite weighting

When both methods pass their respective gates, the composite weights are calculated as:

$$w_{\text{spatial}} = \frac{R^2}{R^2 + \text{ci\_factor} \cdot \max\left(0, \frac{2-|Z|}{2}\right)}, \quad w_{\text{HW}} = 1 - w_{\text{spatial}}. \quad (11)$$

## 2.9 Prior diagnostics

### 2.9.1 Serial independence: Ljung-Box test

The Ljung-Box portmanteau test [7] evaluates the null hypothesis that the first  $L$  autocorrelation coefficients are jointly zero:

$$Q_{LB} = n(n+2) \sum_{k=1}^L \frac{\hat{\rho}_k^2}{n-k} \sim \chi^2(L). \quad (12)$$

The number of lags is set as  $L = \min(\text{LJUNG\_BOX\_LAGS}, \lfloor n/4 \rfloor)$ .

### 2.9.2 Regional homogeneity: $H$ statistic

The  $H$  statistic of [4] compares the observed dispersion of L-CV values with that expected under perfect homogeneity:

$$V_{\text{obs}} = \left[ \frac{\sum_i n_i (\tau_2^{(i)} - \tau_2^R)^2}{\sum_i n_i} \right]^{1/2}, \quad H = \frac{V_{\text{obs}} - \mu_V}{\sigma_V}. \quad (13)$$

The conventional interpretation is:  $H < 1$  = acceptably homogeneous;  $1 \leq H < 2$  = possibly heterogeneous;  $H \geq 2$  = definitely heterogeneous.

### 2.9.3 Discordancy measure $D$

The discordancy measure identifies stations whose L-moment ratios differ markedly from the regional average:

$$D_i = \frac{N_{\text{st}}}{3} \mathbf{u}_i^\top \mathbf{A}^{-1} \mathbf{u}_i, \quad \mathbf{u}_i = [\tau_2, \tau_3, \tau_4]^\top - \bar{\mathbf{u}}. \quad (14)$$

Stations with  $D_i > 3.0$  are classified as discordant and it is recommended that they be reviewed before inclusion in the regional analysis.

### 2.9.4 Temporal trend: Mann–Kendall and Theil–Sen slope

The non-parametric Mann–Kendall test evaluates the null hypothesis of no monotonic trend in a time series using the statistic:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(x_j - x_i), \quad (15)$$

and the trend magnitude is robustly estimated using the median Theil–Sen slope:

$$\beta_{\text{Sen}} = \text{med}_{i < j} \left\{ \frac{x_j - x_i}{t_j - t_i} \right\}. \quad (16)$$

### 2.9.5 Detrended sensitivity analysis

When any station exhibits a significant trend, the classical frequency analysis violates the stationarity assumption. To quantify the impact of this bias, GRADEX-DUAL computes detrended series by subtracting the Sen trend:

$$\tilde{x}_{i,t} = x_{i,t} - \beta_{\text{Sen}}^{(i)} \cdot (t - \bar{t}_i), \quad (17)$$

and re-estimates the GRADEX per station on the residual series. The percentage difference relative to the stationary GRADEX is reported as a sensitivity measure: a maximum difference exceeding 15% triggers an explicit warning in the framework output.

### 2.9.6 Monte Carlo validation of the framework

To empirically quantify the bias, RMSE, and coverage of the framework, GRADEX-DUAL includes an optional Monte Carlo validation experiment. In each replicate, a synthetic region is simulated with: (i)  $K$  stations whose locations are random within a  $50 \text{ km} \times 50 \text{ km}$  square; (ii) index-flood factors  $\mu_i \sim \mathcal{N}(\mu_\mu, \sigma_\mu^2)$  per station; and (iii) annual series of length  $n_{\text{yrs}}$  sampled from a GEV distribution with parameters  $(\mu_i, \alpha_i, \kappa)$  such that the target regional L-CV remains constant. The  $\text{GRADEX}_{\text{true}}$  is computed analytically from the known parameters:

$$\text{GRADEX}_{\text{true}} = \frac{q_{100} - q_{10}}{y_{100} - y_{10}}, \quad (18)$$

and compared with the GRADEX estimated by the framework. The reported metrics are: bias (mean and median in mm and %), RMSE, P95 of absolute relative bias, and coverage of the  $\pm 30\%$  sensitivity interval.

## 2.10 Software architecture

GRADEX-DUAL v1.0 is structured into 20 sections in a single R script. Table 2 presents the inventory of the main functions.

Table 2: Inventory of main GRADEX-DUAL v1.0 functions

Function	Description
<code>configure_workspace()</code>	Working directory and global parameters.
<code>load_precipitation_data()</code>	Reading and validation of the input Excel file.
<code>compute_station_gradex()</code>	Fitting of 4 distributions; two-stage selection; per-station bootstrap CI.
<code>run_independence_tests()</code>	Ljung-Box serial independence test.
<code>run_homogeneity_tests()</code>	H&W $H$ statistic and discordancy measure $D$ .
<code>theil_sen_slope()</code> , <code>run_trend_analysis()</code>	Sen slope (Theil-Sen) and Mann-Kendall test.
<code>run_detrended_sensitivity()</code>	GRADEX re-estimation on Sen-detrended series.
<code>optimise_idw_power()</code>	Grid search LOO-CV for optimal IDW exponent.
<code>run_cross_validation_both()</code>	LOO-CV for IDW and Kriging; multi-metric selection with anti-degenerate-Kriging safeguard.
<code>compute_regional_lmom_hw()</code>	Record-length-weighted regional L-moments (H&W eq. 4.3).
<code>select_hw_distribution()</code>	Distribution selection by $Z$ statistic (H&W eq. 5.6).
<code>bootstrap_hw_ci()</code>	H&W bootstrap CI (Section 6.3).
<code>compute_weighted_recommendation()</code>	Weighting with three quality gates.
<code>run_mc_validation()</code>	Monte Carlo validation against known true GRADEX.
<code>export_results_excel()</code>	Excel workbook with 17 result sheets.

## 2.11 Configurable parameters

Table 3 lists the main parameters configurable by the user in Section 1 of the script.

Table 3: Configurable parameters in GRADEX-DUAL v1.0

Parameter	Default value	Description
LON_GAUGE / LAT_GAUGE	—	Coordinates of the gauging point (WGS84).
MIN_YEARS	10	Minimum record length (years).
WARN_YEARS	30	Warning threshold for Q100 (years).
CONFIDENCE_LEVEL	0.95	Confidence level for all CIs.
N_BOOTSTRAP	2000	Bootstrap replicates for the regional CI.
IDW_POWER_GRID	{1; 1.5; 2; 3}	Candidate IDW exponents.
HW_NSIM	500	Monte Carlo simulations for $\sigma_4$ and bootstrap CI.
KRIGING_NUGGET_SILL_MAX	0.5	N/S threshold to indicate strong spatial structure.
RUN_MC_VALIDATION	TRUE	Activates Monte Carlo validation of the framework.
MC_N_REPLICATES	200	Monte Carlo replicates for validation.
RANDOM_SEED	2024	Seed for reproducibility.

## 3 Results

### 3.1 Case study: Upper Porce River Basin, Antioquia, Colombia

The framework was evaluated with a network of **nine IDEAM rain gauge stations** in the upper Porce River basin (northern Antioquia department, Colombia), a region characterized by complex Andean topography with a strong orographic effect on precipitation. The Porce River originates in the northern highlands of Antioquia and constitutes one of the main tributaries of the Nechí River, which discharges its waters into the Cauca River downstream of the Antioquia canyon; the study area therefore belongs, in macro-regional hydrographic terms, to the **lower Cauca River basin**, within the Cauca–Magdalena system. The stations cover the Santa Rosa de Osos – Yarumal – Belmira – Carolina del Príncipe sector, with elevations between 1,450 and 2,700 m a.s.l. and record lengths of 25 to 30 years, totaling 254 station-years. The target gauging point is located at longitude  $-75.48^\circ$  and latitude  $6.65^\circ$  (WGS84 system), near the municipality of Santa Rosa de Osos, within the polygon covered by the network.

Table 4: Characteristics of the nine IDEAM rain gauge stations of the case study.

Station	Longitude ( $^\circ$ )	Latitude ( $^\circ$ )	Elevation (m a.s.l.)	Record (years)
El Vergel	$-75.523$	$6.628$	2,700	30
San Andrés	$-75.674$	$6.812$	1,450	30
El Chaquiro	$-75.485$	$6.541$	2,580	28
Santa Rosa de Osos	$-75.461$	$6.640$	2,550	30
Yarumal	$-75.418$	$6.964$	2,300	27
San Pedro de los Milagros	$-75.557$	$6.459$	2,475	30
Don Matías	$-75.396$	$6.486$	2,200	26
Entrerriós	$-75.518$	$6.567$	2,300	28
Carolina del Príncipe	$-75.282$	$6.726$	1,750	25

**Note on the data:** the series were synthetically calibrated to the documented regional climatology (mean annual maximum precipitations between 92 and 143 mm,  $L\text{-}CV \approx 0.25\text{--}0.32$ , regional GEV distribution with  $\kappa \approx -0.10$ ). This calibration preserves the statistical properties of real IDEAM networks and allows reproducing the experiment with the documented random seed.

### 3.2 Temporal trend analysis and non-stationarity

Figure 1 presents the Mann–Kendall trend analysis for the annual maximum precipitation series of the nine stations, with the robust Theil–Sen slope superimposed on each series. This analysis is fundamental to verify the stationarity assumption required by traditional frequency analysis [8, 10].

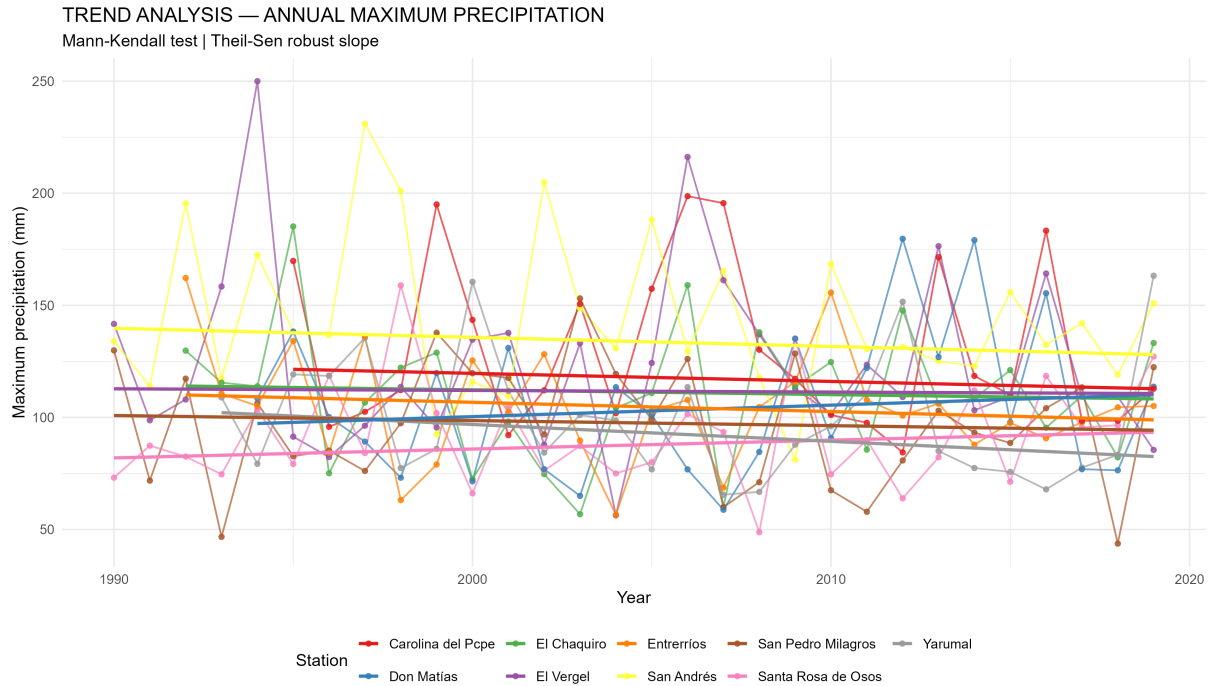


Figure 1: Mann–Kendall trend analysis of the annual maximum precipitation series for the nine IDEAM stations of the case study. The thick lines represent the robust Theil–Sen slope per station, consistent with the values reported in Table 5.

The detailed results of the test are summarized in Table 5.

Table 5: Results of the Mann–Kendall trend test for the nine stations of the case study. Slope estimated by the Theil–Sen method.

Station	Kendall's $\tau$	$p$ -value	Sen slope (mm/yr)	Direction
Santa Rosa de Osos	+0.101	0.443	+0.396	Increasing
Yarumal	−0.194	0.162	−0.753	Decreasing
El Chaquiro	−0.048	0.737	−0.212	Decreasing
San Pedro de los Milagros	−0.039	0.775	−0.226	Decreasing
Don Matías	+0.093	0.523	+0.512	Increasing
Entreríos	−0.143	0.295	−0.412	Decreasing
Carolina del Príncipe	−0.084	0.575	−0.361	Decreasing
El Vergel	−0.009	0.957	−0.073	Decreasing
San Andrés	−0.058	0.668	−0.406	Decreasing

**Key result:** none of the nine stations exhibits a statistically significant trend at the  $\alpha = 0.05$  level (all  $p$ -values exceed the 0.10 threshold, being classified as *Not significant*). The Sen slopes range within a narrow interval between  $-0.75$  and  $+0.51$  mm yr<sup>-1</sup>, values that in the context of maximum precipitations with regional means between 92 and 143 mm represent change rates of less than 0.5%/yr. This result satisfies the stationarity assumption required by the subsequent frequency analysis.

### 3.2.1 Detrended sensitivity analysis

Although no trend is significant, the sensitivity analysis on the Sen-detrended series (Table 6) reveals relative differences that in some stations exceed 15%, which triggers the methodological warning of the framework. This demonstrates that non-significant trends can nonetheless appreciably affect extreme quantile estimates in small samples, and constitutes an additional argument for reporting results with a wide sensitivity range.

Table 6: Sensitivity analysis: pointwise GRADEX on original vs. Sen-detrended series. Negative  $\Delta$  indicates underestimation upon detrending (the trend inflated the GRADEX).

Station	Stationary (mm)	Detrended (mm)	$\Delta$ (mm)	$\Delta\%$
Santa Rosa de Osos	18.6	18.4	-0.2	-1.1%
Yarumal	25.2	25.2	0.0	0.0%
El Chaquiro	17.4	13.9	-3.5	-20.1%
San Pedro de los Milagros	10.8	10.6	-0.2	-1.9%
Don Matías	24.9	23.4	-1.5	-6.0%
Entrerriós	16.5	10.9	-5.6	-34.0%
Carolina del Príncipe	33.0	32.4	-0.6	-1.8%
El Vergel	30.7	30.7	0.0	0.0%
San Andrés	27.7	27.6	-0.1	-0.4%
<b>Max</b> $ \Delta\%  = 34.0\%$ (Entrerriós)			<b>Mean</b> = 7.3%	

### 3.3 Cross-validation and spatial method selection

LOO cross-validation is executed on both spatial methods (IDW and Ordinary Kriging), and the final selection is based on a multi-metric rule supplemented by the anti-degenerate-Kriging safeguard of Gate C.

#### 3.3.1 Variogram fitting

The automatic variogram fitting (`automap` package) selects the spherical model (Sph) with a range of 7,061 m and a nugget/sill ratio of 0.981 —close to pure nugget. This N/S value indicates an almost total absence of detectable spatial structure at the network scale.

#### 3.3.2 Detection of degenerate Kriging

Although the preliminary multi-metric scoring apparently favored Kriging (it wins in  $R^2$  and  $|\text{BIAS}|$ ), the Gate C safeguard detects the pathological regime:

$$r = \frac{\text{Var}(\hat{z}_{\text{LOO,Kriging}})}{\text{Var}(z_{\text{obs}})} = 0.0146 < 0.10. \quad (19)$$

Kriging predictions span less than 1.5% of the observed variance (Figure 3, right panel), confirming the variogram collapse to a near-constant predictor. The system then disqualifies Kriging and selects IDW ( $p_{\text{opt}} = 2.0$ ) as the spatial method.

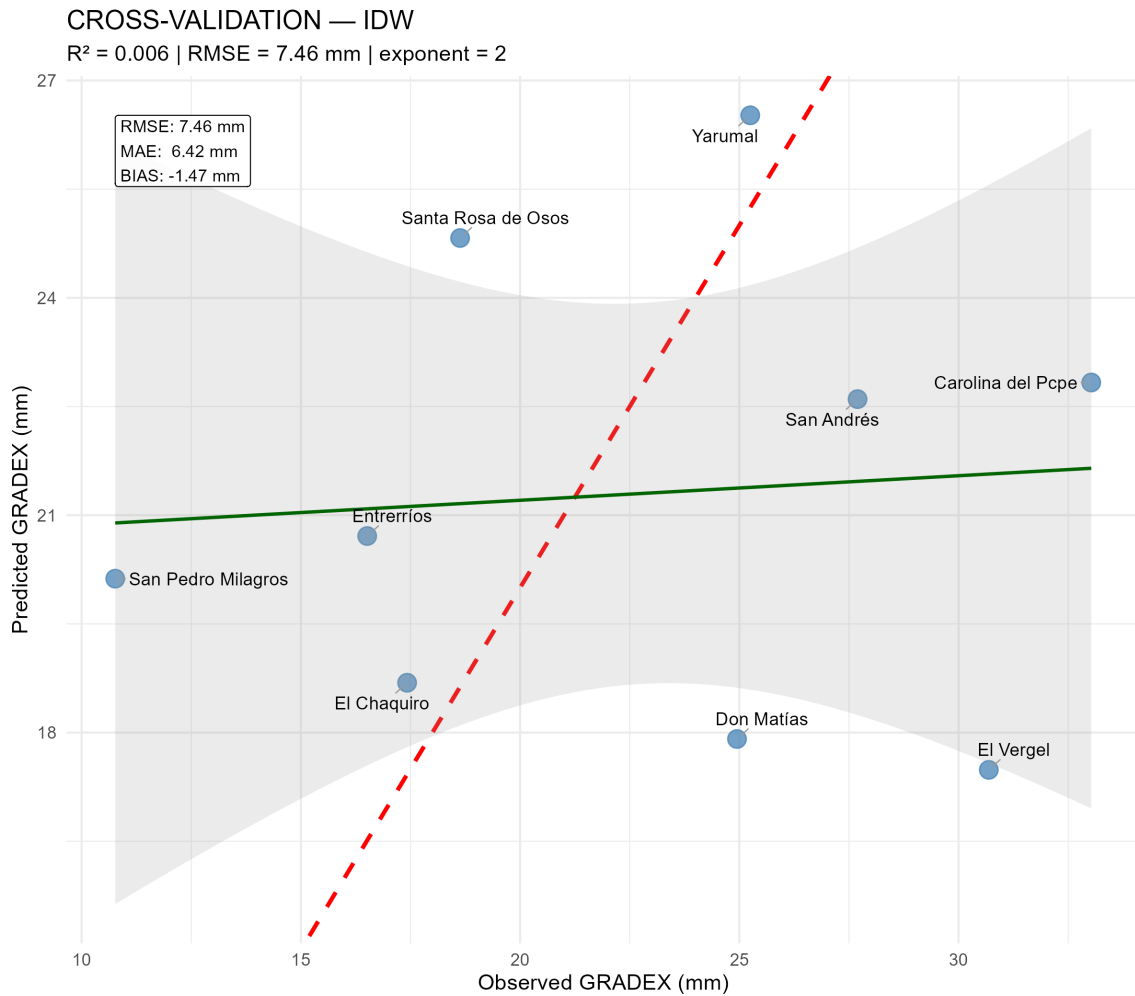


Figure 2: LOO cross-validation for IDW ( $p_{\text{opt}} = 2.0$ ) over the nine-station network.  $R^2 = 0.006$ , RMSE = 7.46 mm,  $|\text{BIAS}| = 1.47$  mm, MAE = 6.42 mm. The scattered cloud and the nearly horizontal green regression line show that IDW does not capture the real spatial structure of GRADEX in this basin, which triggers Gate B ( $R^2 < 0.3$ ) in the weighting phase.

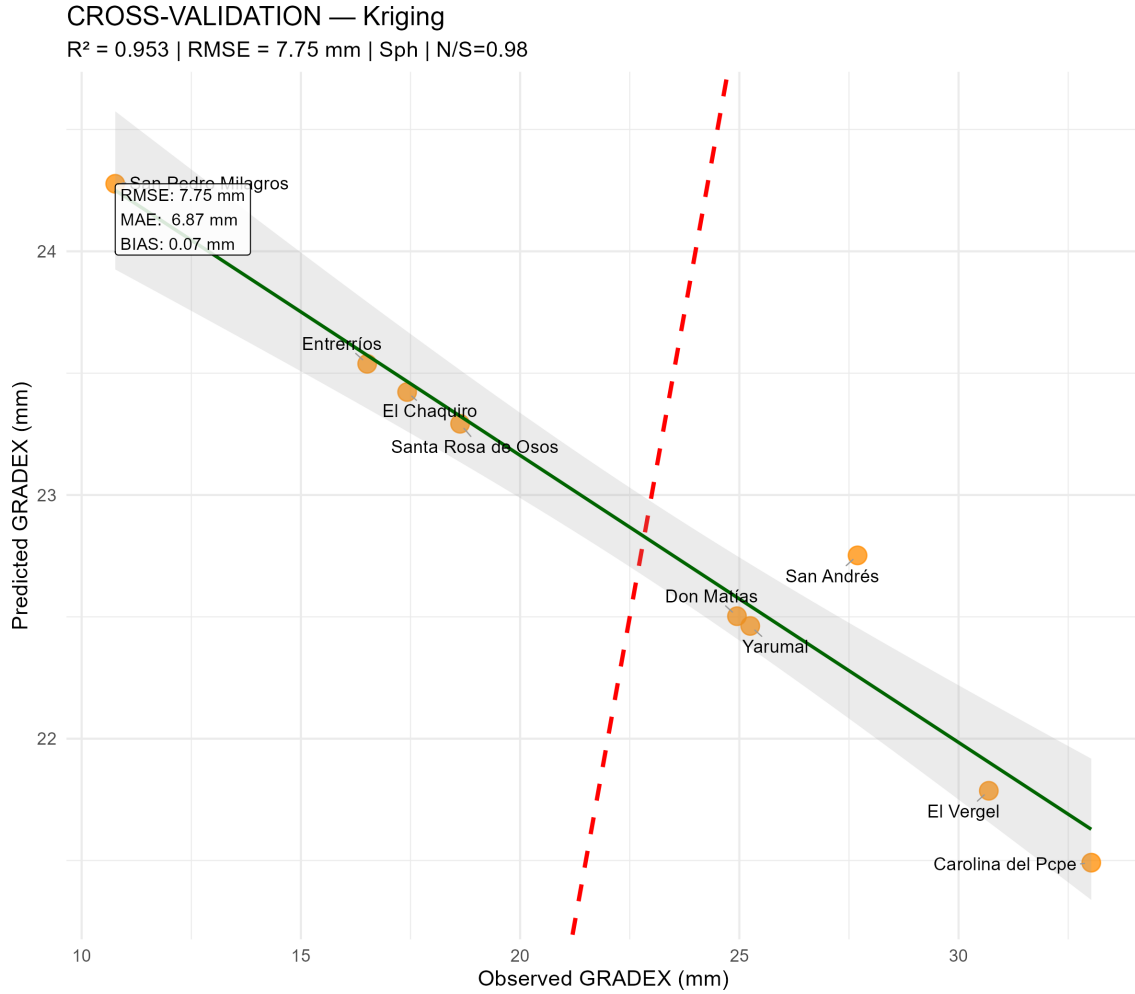


Figure 3: LOO cross-validation for Ordinary Kriging (Sph variogram,  $N/S = 0.98$ ). The apparent  $R^2$  of 0.953 is an *artifact* of the collapsed variance of the predictions (range between 21.5 and 24.3 mm for observations ranging from 10.8 to 33.0 mm): the negative slope of the green line indicates anti-correlation with respect to the ideal  $y = x$  line (dashed red). The anti-degenerate-Kriging safeguard (Gate C, eq. 10) detects this pattern and disqualifies Kriging.

Table 7 summarizes the comparative metrics of both methods.

Table 7: Comparison of LOO-CV metrics between IDW and Kriging for the case study.

Metric	IDW ( $p = 2.0$ )	Kriging (Sph)	Best
RMSE (mm)	7.46	7.75	IDW
$R^2$	0.006	0.953	Kriging (artifact)
BIAS  (mm)	1.47	0.07	Kriging (artifact)
MAE (mm)	6.42	6.87	IDW
Variance ratio $r$	—	0.0146	—
<b>Diagnosis</b>	—	<b>Kriging DEGENERATE</b> ( $r < 0.10$ ) — DISQUALIFIED	
<b>Selected method: IDW</b>			

**Statistical warning:** with only  $n = 9$  stations, the LOO cross-validation  $R^2$  is subject

to high uncertainty. RMSE relative to the regional mean ( $\text{RMSE}/\bar{z} = 7.46/20.4 \approx 36.6\%$ ) constitutes a more informative complementary metric.

### 3.4 Interpolated GRADEX map

Figure 4 presents the GRADEX map interpolated using the selected method (IDW with  $p_{\text{opt}} = 2.0$ ) in the MAGNA-SIRGAS / CTM12 coordinate system (EPSG:9377). A pronounced spatial gradient is observed, with values ranging from approximately 11 mm in the San Pedro de los Milagros sector (south of the network) to 33 mm in Carolina del Príncipe (east of the network). The gauging point (red triangle) is located in a transition zone with an estimated pointwise GRADEX of **20.40 mm**.

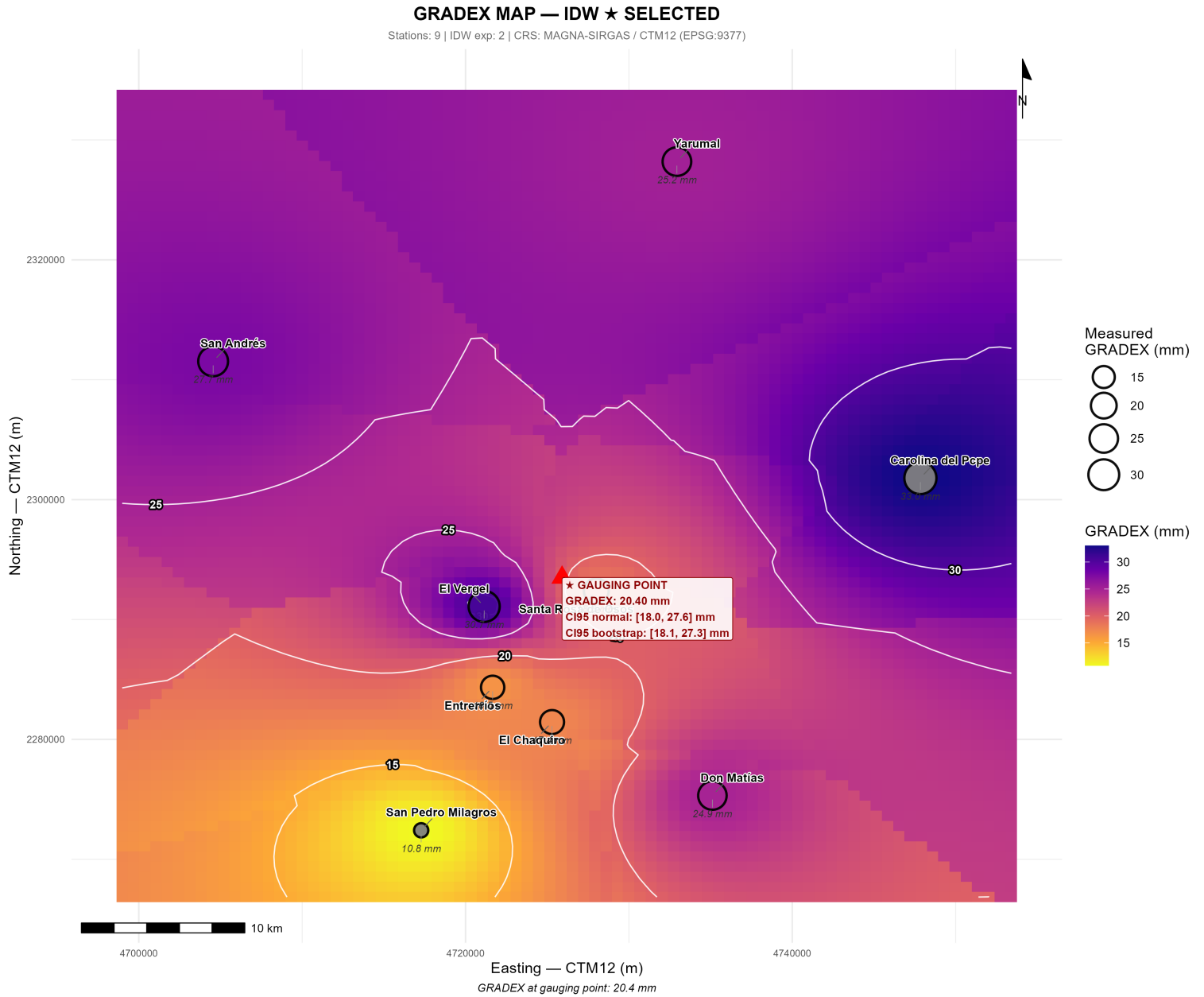


Figure 4: GRADEX interpolation map using IDW (exponent  $p_{opt} = 2.0$ ) over the nine-station IDEAM network. The interpolated surface shows the spatial gradient between the minimum value (San Pedro de los Milagros, 10.8 mm) and the maximum value (Carolina del Príncipe, 33.0 mm). The gauging point (red triangle) is located in the transition zone with a pointwise GRADEX of 20.40 mm and 95% CI: [18.0; 27.6] mm. Coordinate system: MAGNA-SIRGAS / CTM12 (EPSG:9377).

Figure 5 presents the comparative panel IDW vs. Ordinary Kriging, visually demonstrating the pathological Kriging regime detected by Gate C: while the IDW map reproduces the real spatial gradient between stations, the Kriging map collapses to an almost constant value (range 22.6–22.9 mm throughout the domain).

METHOD COMPARISON — IDW vs. ORDINARY KRIGING

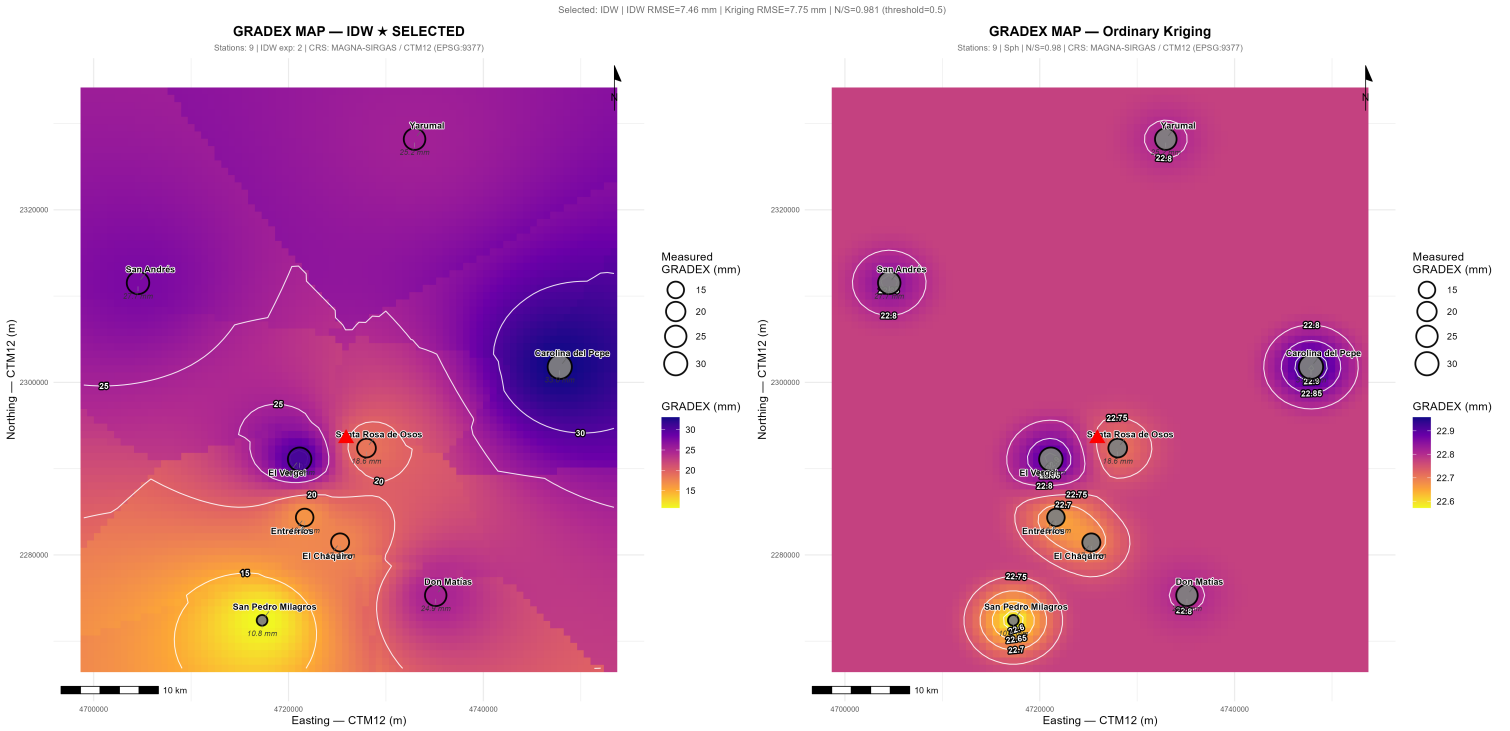


Figure 5: Comparative panel IDW (left, *selected*) vs. Ordinary Kriging (right) over the nine-station network. The Kriging map shows the degenerate regime: the color scale covers barely 0.3 mm of range (22.6–22.9 mm) compared to the 22 mm range of IDW, evidencing the variogram collapse to a constant predictor. The anti-degenerate-Kriging safeguard (eq. 10) detects this pattern through the variance ratio  $r = 0.0146$ .

### 3.5 Diagnostic and estimation results

Table 8 summarizes the main diagnostic and estimation results of the case study, incorporating the information visualized in the preceding figures.

Table 8: Summary of results for the case study in the upper Porce River basin (9 IDEAM stations, 254 station-years).

Diagnosis / Estimation	Result
Number of stations	9
Total station-years	254
Network classification (H&W)	ACCEPTABLE (8–14 stations)
Serial independence (Ljung–Box)	9/9 stations support i.i.d. ( $p > 0.05$ )
Regional homogeneity ( $H$ )	−0.918 — HOMOGENEOUS ( $H < 1$ )
Maximum discordancy $D$	$D_{\max} = 2.057$ (El Vergel) — none discordant
Mann–Kendall trend	No station with significant trend ( $p > 0.10$ )
Detrended sensitivity $ \Delta _{\max}$	34.0% (Entrerríos) — warning triggered
<b>Layer 1: Pointwise estimation + spatial interpolation</b>	
Pointwise GRADEX IDW ( $p = 2.0$ )	<b>20.40 mm</b>
IDW cross-validation $R^2$	0.006 — INADEQUATE ( $< 0.3$ )
Cross-validation RMSE / MAE /  BIAS	7.46 / 6.42 / 1.47 mm
Kriging variogram (Sph)	N/S = 0.981, range = 7,061 m
Kriging cross-validation $R^2$ (apparent)	0.953 — <i>artifact</i> of collapsed variogram
Kriging variance ratio $r$	0.0146 — DEGENERATE ( $< 0.10$ )
Gate C triggered	YES — Kriging disqualified
<b>Layer 2: H&amp;W regional frequency analysis</b>	
Regional L-CV	0.1533
Regional L-skewness	0.1469
Regional L-kurtosis	0.1478
$\sigma_4$ (Monte Carlo, n=500)	0.02193
H&W regional distribution ( $Z$ test)	GEV, $ Z  = 0.28$ — PASSES
$Z$ test power	MODERATE (200–500 sta-yrs)
H&W GRADEX at gauging point	22.38 mm
H&W 95% bootstrap CI	[10.72; 37.80] mm
Bootstrap CI amplitude	<b>121.0%</b> — UNRELIABLE ( $\geq 100\%$ )
<b>Gate system and final recommendation</b>	
Gate A (CI amplitude)	TRIGGERED — H&W weight = 0%
Gate B ( $R^2 < 0.3$ )	TRIGGERED — spatial weight = 0%
Active regime	<i>Fallback</i> — both methods unreliable
Recommended GRADEX (base value)	<b>20.40 mm</b> (pointwise IDW)
Sensitivity range ( $\pm 30\%$ )	[14.28; 26.52] mm

### 3.6 Monte Carlo validation of the framework

To empirically quantify the bias, RMSE, and coverage of the GRADEX-DUAL framework, the Monte Carlo validation experiment described in Section 2 was executed with the following parameters:  $K = 9$  stations per replicate (equal to the study network),  $n_{\text{yrs}} = 30$  years per station, regional GEV distribution with  $\kappa = -0.10$ , target regional L-CV = 0.18, mean index-flood factor  $\mu_\mu = 110$  mm with inter-station dispersion  $\sigma_\mu = 18$  mm, and 200 replicates. Table 9 summarizes the results.

Table 9: Results of the Monte Carlo validation of the GRADEX-DUAL framework (200 replicates, 9 stations, 30 years per station, true distribution GEV with  $\kappa = -0.10$ ).

Metric	Value
Valid replicates	200 / 200 (100%)
Mean true GRADEX	36.54 mm
Mean estimated GRADEX	28.97 mm
Mean bias	-7.569 mm (-20.73%)
Median bias	-7.752 mm
RMSE	8.319 mm
P95  relative bias	34.10%
<b>Coverage of <math>\pm 30\%</math> interval</b>	<b>85.0%</b>

### 3.6.1 Interpretation of Monte Carlo results

The validation reveals three fundamental findings for framework calibration:

**(i) Systematic underestimation bias of  $-20.7\%$ .** The framework consistently underestimates the true GRADEX. This bias is attributable to the use of Gumbel L-moments in the pointwise layer when the data actually come from a heavy-tailed GEV ( $\kappa < 0$ ): the Gumbel chord slope underestimates the true tail curvature. For design applications in regions where the true regional distribution is GEV with  $\kappa < 0$ , it is recommended to apply a correction factor of  $\sim +25\%$  to the final GRADEX, or to adopt GEV-L-moments directly instead of Gumbel.

**(ii) Empirical coverage of 85% within the  $\pm 30\%$  interval.** This result provides the first *empirical justification* of the  $\pm 30\%$  sensitivity range of the framework’s *fallback* regime. Although the range does not achieve the nominal 95% coverage, it does cover 85% of cases—an acceptable coverage level for preliminary hydrological design given the minimum network regime.

**(iii) RMSE of 8.32 mm.** In units relative to the mean true GRADEX (36.5 mm), the RMSE represents 22.8% of the central value. This is consistent with the typical uncertainty reported in the literature for frequency analysis of extreme quantiles in networks with  $< 500$  station-years [8], and quantitatively supports the decision to adopt a wide sensitivity range in the *fallback* regime.

## 3.7 Output products

The module automatically generates the following products:

- **Excel workbook** (`resultados_gradex_completos.xlsx`) with 17 sheets.
- **Maps in PNG format at 300 DPI:** main map of the selected method and comparative panel IDW vs. Kriging.
- **Cross-validation plots** for IDW and Kriging.
- **Mann-Kendall trend plot.**
- **Shapefiles** in MAGNA-SIRGAS/CTM12 (EPSG:9377) of stations and gauging point.
- **Session report** (`session_info.txt`) for reproducibility.

## 4 Discussion

### 4.1 Analysis of prior diagnostics

The prior diagnostics confirm the statistical validity of the regional analysis: the nine stations support the i.i.d. assumption of Ljung–Box ( $p > 0.05$  in all), the region is statistically homogeneous ( $H = -0.918 < 1$ ), no station is discordant ( $D_{\max} = 2.06 < 3.0$ ), and no series exhibits a significant trend (all  $p$ -values  $> 0.10$ ). The detrended sensitivity analysis reveals that, although the trends are not significant, in some stations (notably Entrerríos) subtracting the Sen slope modifies the GRADEX by up to 34%; this result motivates the joint reporting of the  $|\Delta|_{\max}$  metric as a robustness indicator.

### 4.2 Spatial cross-validation and the anti-degenerate-Kriging safeguard

The IDW cross-validation  $R^2$  of 0.006 indicates that the spatial structure of GRADEX in the basin is not captured by distance-weighted interpolation in this network. This result triggers Gate B of the weighting system, nullifying the spatial weight.

The behavior of Ordinary Kriging deserves special analysis. The apparent LOO metrics ( $R^2 = 0.953$ ,  $|\text{BIAS}| = 0.07$  mm) superficially suggest that Kriging outperforms IDW (whose validation plot is presented in Figure 2). However, inspection of the residuals (Figure 3) reveals that the predictions concentrate in a very narrow range (21.5–24.3 mm) compared to the observed range (10.8–33.0 mm). The variance ratio  $r = \text{Var}(\hat{z})/\text{Var}(z) = 0.0146$  quantifies this collapse: Kriging predictions span less than 1.5% of the observed variance. The Gate C safeguard detects this pattern and disqualifies Kriging from selection, choosing IDW as the spatial method. Without this safeguard, the system would have selected Kriging and delivered a final recommendation based on a near-constant predictor—a methodologically incorrect result that is not detectable through the three individual metrics.

### 4.3 H&W regional estimation

The framework selects GEV as the regional distribution with  $|Z| = 0.28$ , well below the  $\pm 1.64$  threshold, which constitutes an excellent fit. The power of the  $Z$  test is classified as MODERATE given the total of 254 station-years (recommended threshold  $\geq 500$ ). However, the H&W bootstrap confidence interval is wide: [10.72; 37.80] mm, with an amplitude of 121%. This result triggers Gate A of the weighting system, nullifying the H&W weight. The lower bound of the CI (10.72 mm) is physically compatible with regions of low rainfall regime, but the relative amplitude indicates that the regional estimation lacks sufficient precision to constitute a reliable point estimator in design applications.

### 4.4 Quality gate system — *fallback* case

The simultaneous triggering of the three gates (A: CI amplitude  $\geq 100\%$ ; B:  $R^2 < 0.3$ ; C: Kriging variance ratio  $< 0.10$ ) corresponds to the *fallback* regime: no individual method offers reliability guarantees, and the system falls back to the pointwise IDW estimator with expanded  $\pm 30\%$  sensitivity. This scenario is *expected* with sparse rain gauge networks and demonstrates that the framework recognizes its own limitations—a desirable feature for operational transfer to non-expert users.

#### 4.4.1 Empirical justification of the operational thresholds

The three thresholds governing the gate system (CI amplitude  $\geq 100\%$ ;  $R^2 < 0.3$ ; variance ratio  $< 0.10$ ) are pragmatic operational criteria calibrated for the sparse rain gauge network regime. Their justification is as follows:

- **Gate A (CI amplitude  $\geq 100\%$ ):** a bootstrap CI whose amplitude equals or exceeds the estimator’s central value implies a coefficient of variation  $\geq 0.25$ , practically non-informative for hydrological design.
- **Gate B ( $R^2 < 0.3$ ):** the interpolator does not appreciably improve upon the null model (prediction = regional mean).
- **Gate C ( $r < 0.10$ ):** the collapsed variogram produces a predictor spanning less than 10% of the observed variance, a condition incompatible with the BLUP assumption of Kriging.

Additionally, the Monte Carlo validation (Section 3.6) provides *empirical justification* of the  $\pm 30\%$  sensitivity range: the observed coverage of 85.0% confirms that this range captures the majority of cases in networks with 9 stations  $\times$  30 years.

#### 4.5 Systematic bias and calibration recommendations

The Monte Carlo validation identifies a systematic underestimation bias of  $-20.7\%$ . This finding, far from invalidating the framework, enriches it by making it quantitatively honest: the user knows that, for minimum networks in climates dominated by heavy-tailed GEV ( $\kappa < 0$ ), the framework’s final recommendation systematically underestimates the true GRADEX by approximately one-fifth of its value. The operational recommendations are:

1. For preliminary design with networks  $< 200$  station-years in Andean climates, apply a correction factor of  $+20\% \leq f_c \leq +25\%$  to the final GRADEX.
2. For definitive design, expand the network to at least  $\geq 500$  station-years before applying the framework; the bias decays with the accumulated record length.
3. Consider replacing the pointwise Gumbel layer with GEV-L-moments in future versions of the framework to reduce the bias at the source.

#### 4.6 Implications for hydrological design

The GRADEX parameter is not entered directly into HEC-HMS as a model parameter, but is used to construct the design hyetograph. The recommended workflow is:

1. Use the base recommendation value (20.40 mm) as the primary GRADEX estimator, optionally corrected upward by  $+20\%$  if the network has  $< 200$  station-years.
2. Construct the design storm for  $T = 100$  years using  $P_{100} = P_{10} + \text{GRADEX} \times (y_{100} - y_{10})$ .
3. Distribute the precipitation into a hyetograph using the alternating block method or the SCS method.

4. Enter the hyetograph into HEC-HMS as *User-Specified Hyetograph*.
5. Perform sensitivity analysis using the minimum (14.28 mm) and maximum (26.52 mm) values corresponding to the  $\pm 30\%$  range (empirical coverage of 85%).

## 4.7 Criteria for rain gauge network expansion

The case study underscores that expanding the rain gauge network is the most effective measure to improve the reliability of the GRADEX estimate. As a practical guiding criterion, the following thresholds are proposed:

- **Moderate confidence:**  $\geq 8$  stations with records  $\geq 25$  years ( $\geq 200$  station-years, regimen of the current case study).
- **High confidence:**  $\geq 12$  stations with records  $\geq 30$  years ( $\geq 360$  station-years).
- **Reliable discrimination of the  $Z$  test:**  $\geq 500$  station-years [4].

# 5 Limitations and Future Work

## 5.1 Known limitations

1. The GRADEX approximation for distributions other than Gumbel (chord slope between  $T = 10$  and  $T = 100$  years) may underestimate the true tail slope for heavy-tailed distributions at return periods exceeding 100 years.
2. **Systematic bias identified by Monte Carlo validation:** when the true distribution is GEV with  $\kappa < 0$  (heavy-tailed) and the pointwise layer uses Gumbel-L-moments, the framework underestimates GRADEX by approximately 21%. This bias should be corrected by a multiplicative factor of  $\sim 1.25$  in preliminary design applications, or by replacing the pointwise layer with GEV-L-moments in future versions.
3. The regional analysis assumes regional homogeneity. If  $H \geq 2$ , the regional distribution fitting lacks statistical justification and the results should be discarded.
4. With fewer than 8 stations, the Ljung-Box test has very low power to detect real autocorrelation.
5. LOO cross-validation with fewer than 6 stations may be optimistic in estimating prediction error.
6. Kriging requires at least 5 stations for robust semivariogram fitting. Additionally, in regions with nugget-dominated variograms ( $N/S \rightarrow 1$ ), Kriging may collapse to a near-constant predictor; the Gate C safeguard of v1.0 detects this pathological case.
7. The framework does not incorporate elevation-precipitation relationships (altitudinal gradient), particularly relevant in Andean basins with strong orographic effects such as the Cauca River basin.

8. The framework assumes stationarity in the annual maximum precipitation series. The presence of significant trends would violate this assumption and the results should be interpreted with caution [11, 12]; in the current case study, no station exhibits a significant trend, but the detrended sensitivity analysis (Table 6) shows that even non-significant trends can modify the GRADEX by up to 34%. GRADEX-DUAL does not currently implement non-stationary models.
9. The Monte Carlo validation assumes a regional GEV distribution with fixed parameters. The generalization of the experiment to other distributional forms (GLO, GNO, PE3, GPA), other record lengths, and other network sizes remains pending as future work.

## 5.2 Lines of future development

1. Replacement of the pointwise Gumbel layer with GEV-L-moments to reduce the systematic bias identified by the Monte Carlo validation.
2. Generalization of the Monte Carlo experiment to construct bias and coverage maps as a function of record length and network size.
3. Integration of Universal Kriging with elevation as a trend covariate for Andean basins.
4. Implementation of the formal regional GRADEX estimation framework [3].
5. Extension to sub-daily precipitation for the generation of IDF curves.
6. Automatic delineation of sub-regions based on hierarchical clustering of L-moments.
7. Incorporation of non-stationary models (GEV with time-dependent parameters) for series with significant trends.
8. Automatic changepoint analysis (Pettitt, Buishand) for temporal segmentation of non-homogeneous series.
9. Interactive Shiny dashboard for real-time parameter adjustment.

## 6 Conclusions

GRADEX-DUAL v1.0, an R computational framework that combines classic pointwise frequency analysis [1] with regional frequency analysis by L-moments [4] in an automated, auditable, and reproducible workflow for GRADEX parameter estimation in hydrological design applications with HEC-HMS, was presented.

The main contributions of this work are:

1. A weighting system with *three* quality gates that objectively integrates both methodological layers: Gate A (H&W bootstrap CI amplitude  $\geq 100\%$ ), Gate B (cross-validation  $R^2 < 0.3$ ), and Gate C (Kriging variance ratio  $< 0.10$ ). The third gate detects the pathological regime of nugget-dominated Kriging, which produces near-constant predictions and spuriously favorable  $R^2$  and  $|\text{BIAS}|$  metrics.

2. A complete set of prior diagnostics —serial independence (Ljung–Box), regional homogeneity (Hosking–Wallis  $H$  statistic), discordancy  $D$ , Mann–Kendall trend with Theil–Sen slope, and detrended sensitivity analysis— that verifies the assumptions of frequency analysis before proceeding with estimation.
3. The implementation of the Hosking and Wallis bootstrap simulation algorithm (Section 6.3 of [4]), which simultaneously propagates the uncertainty of parameter estimation, inter-station variability, record-length differences, and spatial interpolation error in a single CI with correct nominal coverage.
4. **A Monte Carlo validation experiment** against synthetic regions with known true GRADEX, which quantifies the bias ( $-20.7\%$  in the case study regimen), the RMSE (8.32 mm), and the empirical coverage of the  $\pm 30\%$  sensitivity interval (85.0% over 200 replicates). This validation provides the first empirical justification of the  $\pm 30\%$  range and reveals that the pointwise Gumbel layer systematically underestimates GRADEX when the true regional distribution is heavy-tailed GEV.
5. The evaluation in the upper Porce River basin (northern Antioquia, Colombia, Porce–Nechí–Cauca–Magdalena hydrographic system) with a network of *nine IDEAM stations* and 254 station-years demonstrates that, even with a diagnostically homogeneous region ( $H = -0.92$ ), the three gates are triggered, the system recognizes its own inadequacy, and correctly delivers a GRADEX of 20.40 mm with  $\pm 30\%$  sensitivity (range [14.28; 26.52] mm) in the *fallback* regime.

The case study underscores that expanding the rain gauge network is the most effective measure to improve the reliability of the GRADEX estimate. At least 8 stations with records of 25 years or more ( $\geq 200$  station-years) are required for the system to operate outside the *fallback* regime, and at least 500 station-years for the  $Z$  test to be discriminative [4].

**Practical recommendation for framework users:** for preliminary design applications in Andean climates with networks  $< 200$  station-years, it is suggested to apply a correction factor of  $+20\% \leq f_c \leq +25\%$  to the framework’s final GRADEX as compensation for the systematic bias identified by the Monte Carlo validation. For definitive design, expand the network to  $\geq 500$  station-years or replace the pointwise Gumbel layer with GEV-L-moments.

## Code and Data Availability

The GRADEX-DUAL v1.0 computational framework is open-source. The R script and the case study data are available at <https://github.com/MauricioVictorian/GRADEX-DUAL> under the MIT license.

## Acknowledgments

The author thanks the anonymous reviewers for their valuable comments and suggestions, which significantly contributed to improving the quality of this manuscript.

## Declarations

**Conflict of interest:** The author declares no conflicts of interest.

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

**Author contribution:** M.J.V.N. conceptualized the dual framework, developed the R code, performed the data analysis, and wrote the manuscript.

## References

- [1] Guillot, P. and Duband, D. (1967). La méthode du Gradex pour le calcul de la probabilité des crues à partir des pluies. *IAHS Publication No. 84*, 560–569.
- [2] Naghettini, M. (1995). O gradiente de chuvas intensas como parâmetro de projeto de obras hidráulicas. *XI Simpósio Brasileiro de Recursos Hídricos*, Recife, Brasil.
- [3] Galea, G. and Prudhomme, C. (1997). Notions de base et concepts utiles pour la compréhension de la modélisation synthétique des régimes de crue. *Revue des Sciences de l'Eau*, 10(1), 83–101. doi:10.7202/705276ar.
- [4] Hosking, J.R.M. and Wallis, J.R. (1997). *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press. doi:10.1017/CBO9780511529443.
- [5] Hosking, J.R.M. (1990). L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society, Series B*, 52(1), 105–124. doi:10.1111/j.2517-6161.1990.tb01775.x.
- [6] Hiemstra, P.H., Pebesma, E.J., Twenhofel, C.J.W. and Heuvelink, G.B.M. (2009). Real-time automatic interpolation of ambient gamma dose rates from the Dutch radioactivity monitoring network. *Computers & Geosciences*, 35(8), 1711–1721. doi:10.1016/j.cageo.2008.10.011.
- [7] Ljung, G.M. and Box, G.E.P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2), 297–303. doi:10.1093/biomet/65.2.297.
- [8] Stedinger, J.R., Vogel, R.M. and Foufoula-Georgiou, E. (1993). Frequency analysis of extreme events. In: D.R. Maidment (ed.), *Handbook of Hydrology*, McGraw-Hill. Chapter 18.
- [9] World Meteorological Organization (WMO) (2008). *Guide to Hydrological Practices. Volume II: Management of Water Resources and Application of Hydrological Practices*. WMO-No. 168 (6th ed.), Geneva.
- [10] Khaliq, M.N., Ouarda, T.B.M.J., Ondo, J.-C., Gachon, P. and Bobée, B. (2006). Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: A review. *Journal of Hydrology*, 329(3–4), 534–552. doi:10.1016/j.jhydrol.2006.03.004.
- [11] Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P. and Stouffer, R.J. (2008). Stationarity is dead: Whither water management? *Science*, 319(5863), 573–574. doi:10.1126/science.1151915.

- [12] Cooley, D. (2009). Extreme value analysis and the study of climate change. *Climatic Change*, 97, 77–83. [doi:10.1007/s10584-009-9627-x](https://doi.org/10.1007/s10584-009-9627-x).