Article

Not peer-reviewed version

A Generalized Multivariate Functional Additive Mixed Model for Joint Bias Correction of Hydroclimatic Satellite Data and Its Implementation in the *ColClim* Web Application

David Arango-Londoño [*,†] , Delia Ortega-Lenis [*,†] , Mauricio A. Mazo-Lopera [*,†] , Paula Moraga [*,†]

Posted Date: 12 May 2026

doi: 10.20944/preprints202605.0782.v1

Keywords: bias correction; hydroclimatic variables; NASA POWER; Clausius-Clapeyron mechanism; predictive modeling

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 May 2026 doi:10.20944/preprints202605.0782.v1

Article

A Generalized Multivariate Functional Additive Mixed Model for Joint Bias Correction of Hydroclimatic Satellite Data and Its Implementation in the *ColClim* Web Application

David Arango-Londoño [1,2,*,†], **Delia Ortega-Lenis** [1,2,*,†], **Mauricio A. Mazo-Lopera** [1,*,†]
and Paula Moraga [3,*,†]

[1] Universidad Nacional de Colombia – Sede Medellín – Escuela de Estadística
[2] Faculty of Engineering and Sciences, Pontificia Universidad Javeriana, Cali, Colombia
[3] Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia
* Correspondence: david.arango@javerianacali.edu.co (D.A.-L.); delia.ortega@javerianacali.edu.co (D.O.-L.); mamazol@unal.edu.co (M.A.-M.L.); paula.moraga@kaust.edu.sa (P.M.)
† These authors contributed equally to this work.

Abstract

We propose a Generalized Multivariate Functional Additive Mixed Model (GMFAMM) for the simultaneous bias correction of five hydroclimatic variables derived from the NASA POWER satellite product: minimum temperature (T_{min}), maximum temperature (T_{max}), relative humidity (HR), solar radiation (Rad), and precipitation occurrence (P_{bin}). The GMFAMM extends the univariate functional framework by incorporating a shared latent Gaussian process $\Lambda_{0i}(t)$ that captures cross-variable thermodynamic dependence. A systematic experimental grid of more than 200 model configurations across four distributional families (Gaussian, Gamma, Poisson, Binomial), two effect structures (linear and smooth P-splines), and four nested covariate sets is evaluated on a strict chronological 70/30 hold-out – seven training years (2016–2022) and three hold-out years (2023–2025) – to identify the optimal marginal specification for each variable. The value of joint modelling is quantified through a two-stage cross-residual approximation to the GMFAMM shared latent process, which constitutes a conservative lower bound on the gains achievable by the full simultaneous model: out-of-sample RMSE is reduced by 53% for T_{min}, 38% for T_{max}, and 51% for relative humidity relative to the independent GAMM baseline. These gains are physically interpretable through the Clausius-Clapeyron thermodynamic coupling documented in the residual cross-correlation analysis. The trained model artefacts are deployed in *ColClim*, an open-access R Shiny web application that queries the NASA POWER API and the Open-Meteo forecast service for any user-selected location in Colombia, applies the GMFAMM correction pipeline, and delivers both historical bias-corrected time series and short-range (1–16 day) forecasts across the five variables.

Keywords: bias correction; hydroclimatic variables; NASA POWER; Clausius-Clapeyron mechanism; predictive modeling

1. Introduction

Hydroclimatic variables such as temperature, humidity, solar radiation, and precipitation are fundamental drivers of agricultural planning, hydrological risk assessment, and infectious disease surveillance (Pavani et al., 2023). In tropical regions of South America, the quality and spatial density of ground-based observations are often insufficient for high-resolution operational use (Ortega & Rodríguez, 2006). Remote sensing products such as the NASA POWER reanalysis provide spatially continuous estimates at daily temporal resolution, but systematic biases relative to surface observations – arising from coarse spatial resolution (≈55 km), cloud-topography interactions, and the Clausius-Clapeyron-driven coupling between temperature and humidity – limit their direct applicability (AghaKouchak et al., 2011).

Statistical bias correction of satellite data has been widely studied for individual variables, most often precipitation (Primo, 2018) or temperature (Hammerling & Zidek, 2013). However, the joint nature of hydroclimatic variability means that correcting each variable independently risks destroying the physical correlations among them, which are operationally relevant: an irrigation decision depends jointly on temperature, humidity, and radiation, not on any single variable in isolation. A bias correction framework that preserves and exploits the cross-variable dependence structure is therefore both methodologically and practically desirable.

Functional data analysis provides a natural language for climate time series (Ramsay & Silverman, 2006). Functional Generalized Additive Mixed Models (FGAMMs) have been used for univariate precipitation prediction using ground-based and satellite data (Moraga & Baker, 2022). Extending this framework to the multivariate setting requires a latent process architecture that can represent the dominant modes of covariation across heterogeneous outcome types – continuous temperatures and humidity alongside binary precipitation occurrence. The GMFAMM framework of Volkmann et al. (2024) provides this architecture, but its validation has been limited to controlled simulation studies and small clinical datasets. This paper provides the first empirical validation of the GMFAMM on real-world hydroclimatic data at a dataset scale approximately two orders of magnitude larger than the original framework, together with four methodological adaptations required by this application context.

We propose a GMFAMM that jointly corrects the five NASA POWER outputs for the Valle del Cauca region of Colombia. We conduct a systematic evaluation of more than 200 model configurations across distributional families and effect structures, establish the optimal marginal specification for each variable, and demonstrate empirically – through a two-stage cross-residual approximation that serves as a lower bound – that joint modelling through a shared latent process reduces RMSE by 38–53% for the energy-balance variables. We further describe the deployment of the GMFAMM correction pipeline in *ColClim*, a web-based application that makes bias-corrected hydroclimatic estimates accessible to practitioners without programming skills.

The remainder of this paper is organized as follows. Section 2 describes the study area and data sources, including a summary of the satellite biases documented in the exploratory analysis. Section 3 presents the GMFAMM specification and the methodological adaptations. Section 4 describes the experimental grid and evaluation protocol. Section 5 reports the systematic comparison of marginal models. Section 6 quantifies the added value of joint modelling. Section 7 documents the *ColClim* application. Section 8 discusses limitations and future work.

2. Study Area and Data

2.1. Study Area

The Valle del Cauca department is located in southwestern Colombia, between 3°5′ and 5°0′ north latitude and 75°41′ and 77°33′ west longitude (BIOPALMIRA, 2008). The inter-Andean valley, at elevations ranging from sea level to 4,080 m above sea level, supports an agricultural economy dominated by sugarcane, fruit, and livestock production. The climate follows a bimodal rainfall regime, with peak precipitation in April-May and October-November, and is significantly affected by ENSO teleconnections (UESVALLE, 2016).

2.2. Surface Observations

Ground-based daily measurements were obtained from 62 automatic meteorological stations (EMAs) operated by IDEAM (Datos Abiertos Colombia, n.d.) for the period 2016–2025. Five variables were extracted at daily resolution: minimum temperature ($T_{\min}$, °C), maximum temperature ($T_{\max}$, °C), relative humidity (HR, %), solar radiation (Rad, kWh m^{-2} day^{-1}), and a binary precipitation occurrence indicator ($P_{\text{bin}} \in \{0, 1\}$, rain event defined as daily total > 0.1 mm). To span the main climatic gradients of the department while keeping the analysis tractable, a k-means clustering algorithm ($k = 5$) was applied to geographic coordinates, and one station per cluster was selected (codes 14, 21, 26, 13, and 16) based on maximising within-cluster spatial coverage; the silhouette coefficient was used to confirm $k = 5$ as the optimal number of clusters (Figure 1).

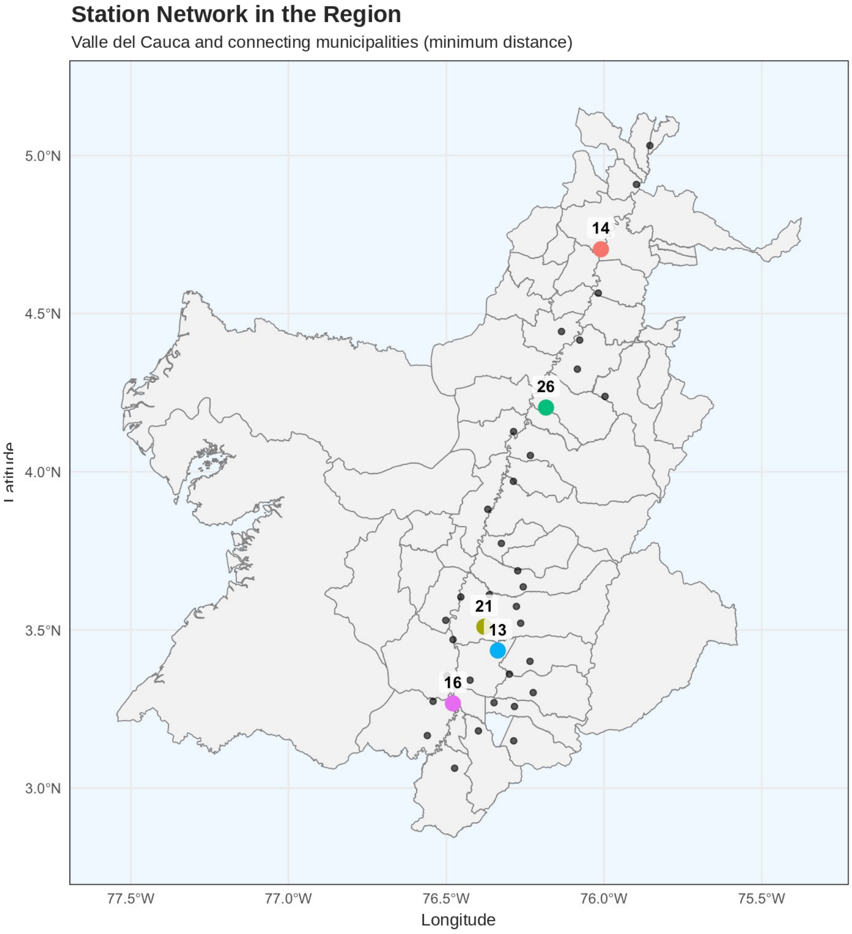


Figure 1. Hydroclimatic station network in the Valle del Cauca region. Colored points represent the five representative stations selected via k-means clustering (codes 14, 21, 26, 13, and 16). Small black dots correspond to the full 62-station network.

2.3. Satellite Data and Systematic Biases

Satellite estimates were obtained from the NASA POWER reanalysis product (NASA Langley Research Center (LaRC) POWER Project funded through the NASA Earth Science Directorate Applied Science Program, 2024) at a nominal spatial resolution of 0.5° (≈55 km). A systematic exploratory analysis revealed substantial and physically interpretable biases relative to surface observations. Figure 2 illustrates the cold bias for maximum temperature ($R = 0.504$, RMSE $= 8.86\,^{\circ}$C), where the satellite distribution is entirely shifted to the left relative to the observed distribution, with a nearly constant additive offset of $\approx 8.5\,^{\circ}$C–$9\,^{\circ}$C throughout the study period.

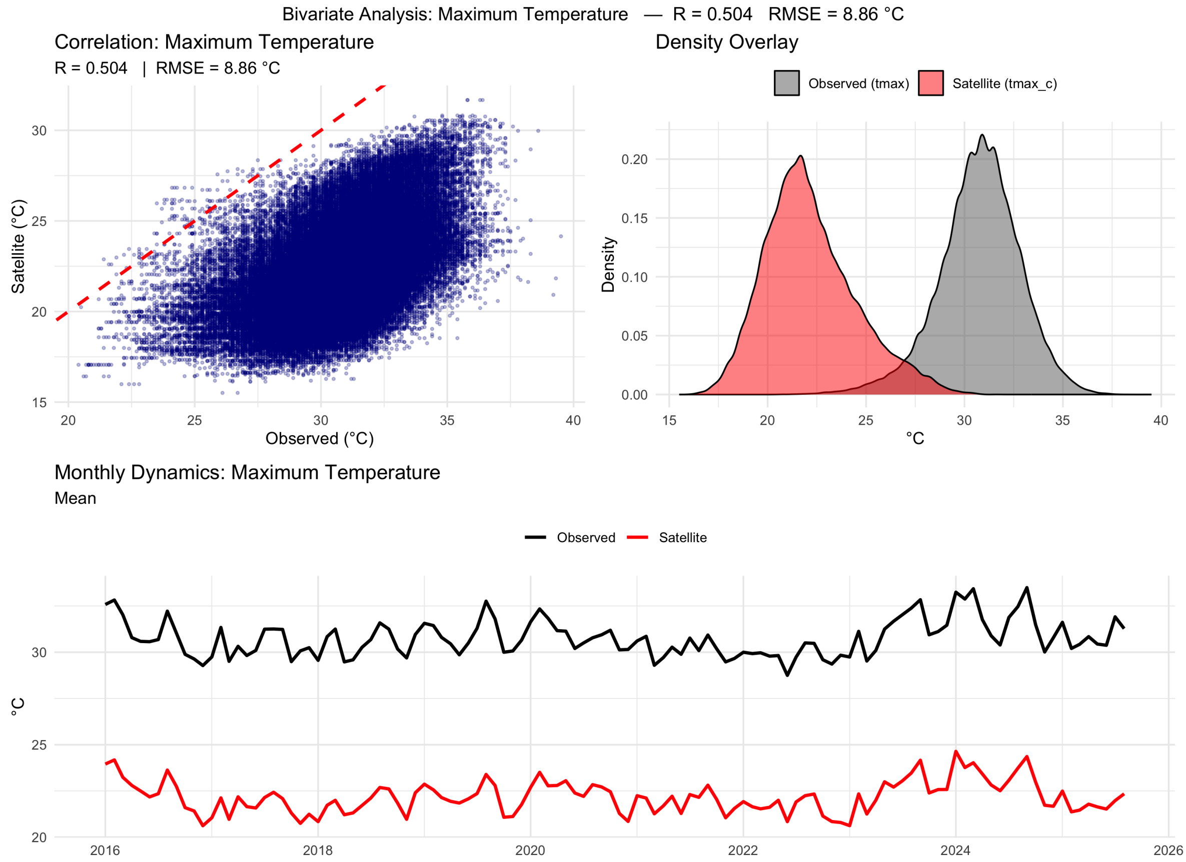


Figure 2. Bivariate analysis of Maximum Temperature (°C): $R = 0.504$, RMSE $= 8.86\,°$C. *Top left:* scatter plot with identity line (dashed red). *Top right:* overlaid density functions (gray: observed $T_{\max}$; red: satellite). *Bottom:* monthly average dynamics (black: observed; red: satellite), confirming the systematic cold bias of $\approx 8.5\,°$C–$9\,°$C.

Figure 3 shows the corresponding relative humidity analysis ($R = 0.551$, RMSE $= 7.31\%$). The satellite overestimates HR by $\approx 4.5\%$ on average – a direct consequence of the cold bias operating through the Clausius-Clapeyron relation – and acts as a low-pass filter that suppresses extreme dry days below 66%.

The multi-scale correlation analysis (Figure 4) reveals that satellite–surface agreement improves substantially with temporal aggregation for all variables: for minimum temperature, $r = 0.51$ (daily) $\rightarrow$ $r = 0.84$ (monthly) $\rightarrow r = 0.96$ (annual). This pattern confirms that the satellite captures low-frequency hydroclimatic variability with sufficient fidelity to serve as a functional covariate, and justifies the statistical calibration strategy adopted in this paper.

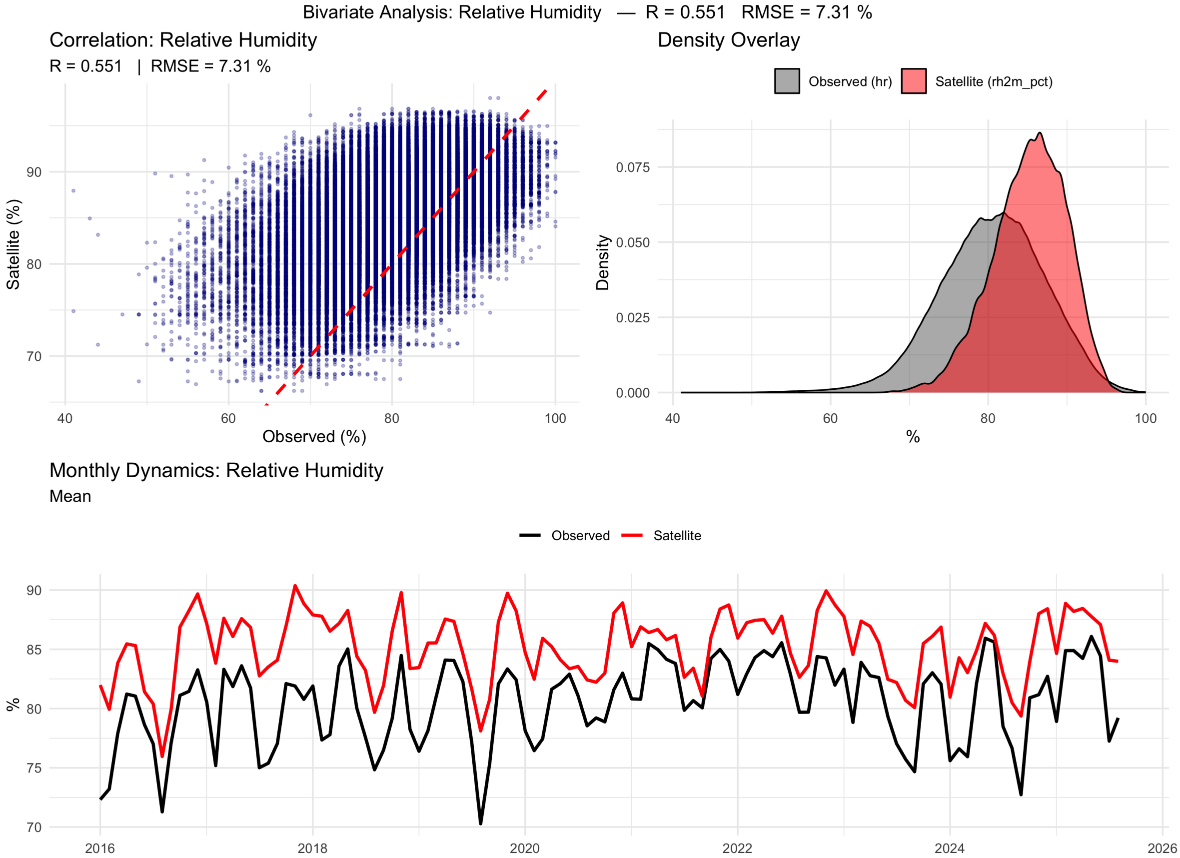


Figure 3. Bivariate analysis of Relative Humidity (%): $R = 0.551$, RMSE $= 7.31\%$. The overestimation visible in the scatter plot (point cloud predominantly above the identity line) is physically caused by the cold bias through the Clausius-Clapeyron relation: a satellite that underestimates temperature must overestimate the saturation-relative humidity ratio.

The key statistics for all five variables are summarized in Table 1, showing the magnitude of the biases that motivate the GMFAMM correction.

Table 1. Summary of satellite biases relative to surface observations (62 EMAs, 2016–2025). Bias = Satellite − Observed. The cold bias in temperature propagates into humidity overestimation through the Clausius-Clapeyron relation.

Variable	**Obs. Mean**	**Sat. Mean**	**Mean Bias**	**RMSE**	**r (daily)**
T_{min} (°C)	18.93	13.85	-5.08	5.67	0.213
T_{max} (°C)	30.61	22.17	-8.44	8.86	0.504
HR (%)	80.70	85.20	$+4.50$	7.31	0.551
Rad (kWh/m^2/day)	4.10	3.87	-0.23	1.38	0.621
P_{bin} (occ. rate)	52.2%	64.2%	$+12.0$ pp	—	0.309

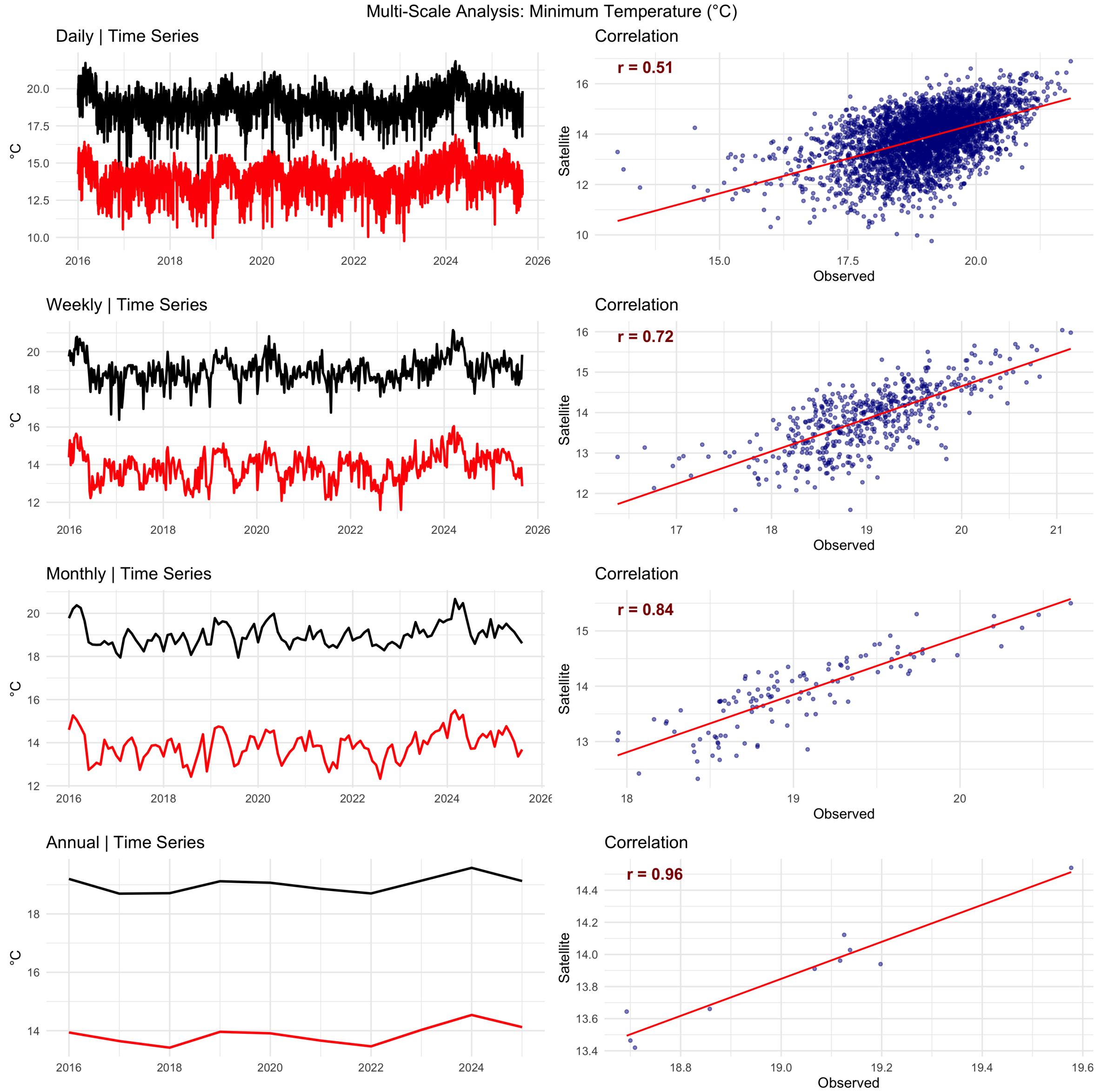


Figure 4. Multi-scale analysis of Minimum Temperature (°C). Each row corresponds to an aggregation scale (daily, weekly, monthly, annual). Left column: time series (black: observed; red: satellite). Right column: scatter plot with Pearson coefficient r. The systematic improvement in r with aggregation scale ($0.51 \to 0.72 \to 0.84 \to 0.96$) confirms that the satellite captures low-frequency variability reliably while showing high daily noise.

3. The Generalized Multivariate Functional Additive Mixed Model

3.1. Outcome Space and Notation

Let $\mathbf{Y}_i(t) = \left(Y_i^{(1)}(t), \ldots, Y_i^{(K)}(t)\right)^\top$ be the K-dimensional daily hydroclimatic response at station i on day t, partitioned as

$$\mathbf{Y}_i(t) = \Big(\underbrace{Y_i^{(1)}, \ldots, Y_i^{(K_c)}}_{\text{continuous}}, \underbrace{Y_i^{(K_c+1)}, \ldots, Y_i^{(K)}}_{\text{binary}}\Big)^\top, \tag{1}$$

where $K_c = 4$ ($T_{\min}$, $T_{\max}$, HR, Rad) and $K - K_c = 1$ (P_{bin}). The outcome space is the mixed product $\mathcal{Y} = \mathbb{R}^{K_c} \times \{0,1\}^{K-K_c}$.

3.2. Model Specification

The GMFAMM (Volkmann et al., 2024) specifies, for each variable $k \in \{1, \ldots, K\}$, a conditional distribution from the exponential family:

$$Y_{it}^{(k)} \mid \mathbf{X}_i, \Lambda_i \ \sim \ D^{(k)}\big(\theta_{i1}^{(k)}(t), \ldots, \theta_{iR^{(k)}}^{(k)}(t)\big), \tag{2}$$

where $D^{(k)}$ is variable-specific ($\mathcal{N}$ for temperatures, Gamma for radiation, Bernoulli for precipitation occurrence), and the location parameter is linked to the additive predictor via a canonical link g_k:

$$g_k(\mu_i^{(k)}(t)) = \underbrace{\beta_0^{(k)}(t) + \sum_{l=1}^{L} f_l^{(k)}(x_i^{(l)}, t) + u_i^{(k)}}_{\text{fixed additive predictor}} + \underbrace{\Lambda_{0i}(t)\, \nu_0^{(k)}}_{\text{shared latent process}} + \varepsilon_{it}^{(k)}. \tag{3}$$

In Eq. (3): $\beta_0^{(k)}(t)$ is the functional intercept; the sum $\sum_l f_l^{(k)}$ represents flexible covariate effects of the satellite predictors $x_i^{(l)}$ (implemented as either linear terms or smooth P-splines); $u_i^{(k)} = s(\mathrm{id}_i, \mathrm{bs} = \mathrm{re})$ is the station random intercept; and $\Lambda_{0i}(t)$ is the shared station-level latent Gaussian process, with loading $\nu_0^{(k)}$ on variable k and $\varepsilon_{it}^{(k)}$ the idiosyncratic error. All additive effects include a cyclic cubic spline for day-of-year seasonality $s(\mathrm{doy}_t, \mathrm{bs} = \mathrm{cc}, k = 15)$ and a regression spline for the long-term trend $s(t, \mathrm{bs} = \mathrm{cr}, k = 15)$.

3.3. Distributional Families and Link Functions

The heterogeneity of the five target variables requires variable-specific distributional assignments:

$$T_{\min}, T_{\max}: \quad Y^{(k)} \sim \mathcal{N}(\mu, \sigma^2), \quad g(\mu) = \mu. \tag{4}$$

$$\mathrm{Rad}: \quad Y^{(k)} \sim \mathrm{Gamma}(\mu, \phi), \quad g(\mu) = \log \mu. \tag{5}$$

$$P_{\mathrm{bin}}: \quad Y^{(k)} \sim \mathrm{Bernoulli}(\pi), \quad g(\pi) = \mathrm{logit}(\pi). \tag{6}$$

For relative humidity, the Gaussian and Gamma families are both evaluated (see Section 5). The Gamma and Binomial choices are physically motivated: solar radiation is strictly positive with approximately constant coefficient of variation, and precipitation occurrence is inherently binary.

3.4. Shared Latent Process and MFPCA Representation

The latent process $\Lambda_{0i}(t) \sim \mathrm{GP}(0, \mathcal{K}_0(t, \cdot))$ is the central innovation of the GMFAMM relative to five independent GAMMs. Its truncated Karhunen-Loève (KL) expansion at order M_0 is:

$$\Lambda_{0i}^{(k)}(t) \approx \sum_{m=1}^{M_0} \rho_{0im}\, \psi_{0m}^{(k)}(t), \tag{7}$$

where the random scores ρ_{0im} are shared across all K dimensions for each component m – the key mechanism through which the model captures cross-variable dependence. The multivariate functional principal components (MFPCs) $\psi_{0m}^{(k)}(t)$ are estimated via the two-step procedure of Happ and Greven (2018): separate univariate FPCAs per dimension, followed by an eigenanalysis of the weighted cross-covariance matrix of the univariate scores. In this application, $M_0 = 1$ component is retained, explaining 38.5% of total joint residual variation (eigenvalue $\hat{\nu}_{01} = 1.924$), consistent with the dominant Clausius-Clapeyron energy-balance mode documented in Section 6.

3.5. Methodological Adaptations for Hydroclimatic Data

Four adaptations were required beyond the original GMFAMM framework (Volkmann et al., 2024), whose validation was limited to small clinical datasets.

Variance-equalization weighting.

Variables with disparate marginal variances ($\sigma^2_{T_{\max}} \approx 4\,^\circ\mathrm{C}^2$ vs. $\sigma^2_{P_{\mathrm{bin}}} \approx 0.25$) cause the multivariate eigendecomposition to collapse onto the highest-variance dimension without standardisation. All marginal residual series are transformed to unit variance before the MFPCA eigenanalysis.

MCMC convergence engineering.

Initial runs with default BAMLSS priors on the variance parameters ν_{0m} failed to mix for the binary precipitation dimension, where logit-link score updates become numerically unstable when predicted probabilities approach 0 or 1. This was resolved by tightening the inverse-Gamma hyperprior on ν_{01} (shape $= 0.001 + n/2$, rate $= 0.001 + \rho_0^{(m)\top}\rho_0^{(m)}/2$) and imposing an early stopping criterion on the score magnitude during the first 500 burn-in iterations. Gelman-Rubin $\hat{R} \leq 1.05$ for all monitored parameters confirms convergence.

Station-level satellite calibration.

Prior to model fitting, each satellite covariate $x_i^{(l)}$ undergoes a station-level linear bias correction on the training set: $\hat{x}^{(l)}_{it,\mathrm{cal}} = \hat{\alpha}_i^{(l)} + \hat{\gamma}_i^{(l)} x_{it}^{(l)}$. This removes the systematic additive and multiplicative biases (in particular the 5 °C–9 °C cold bias documented in Section 2.3). Correction factors are applied verbatim to the test set to ensure strict out-of-sample evaluation.

Scale of application.

The dataset ($n = 62$ stations, $K = 5$ variables, 10 years of daily records 2016–2025, ≈200,000 station-day observations) is approximately two orders of magnitude larger than the simulation studies in Volkmann et al. (2024). All smoothing parameters are estimated by fast REML (`mgcv::bam`; Wood 2017) rather than full Bayesian MCMC to ensure tractability at this scale.

4. Experimental Design and Evaluation Protocol

4.1. Study Variables and Distributional Families

The experimental grid covers five target variables spanning the full range of exponential family distributions. Table 2 summarises their identifiers, main satellite predictors, and the distributional families evaluated for each.

Table 2. Study variables: response variables, main satellite predictors, and distributional families evaluated.

Variable	Main Satellite Predictor	Evaluated Families	Type
Min. Temperature ($T_{\min}$)	`T2M_MIN` (NASA POWER)	Gaussian	Continuous
Max. Temperature ($T_{\max}$)	`T2M_MAX` (NASA POWER)	Gaussian	Continuous
Relative Humidity (HR)	`RH2M` (NASA POWER)	Gaussian, Gamma, Poisson	Bounded cont.
Solar Radiation (Rad)	`ALLSKY_SFC_SW_DWN` (NASA POWER)	Gaussian, Gamma, Poisson	Positive cont.
Precip. Occurrence (P_{bin})	`PRECTOTCORR` (NASA POWER)	Binomial	Binary

4.2. Model Configurations and Experimental Grid

The full experimental grid results from crossing three design factors: (i) response variable; (ii) distributional family and effect structure (linear vs. smooth P-splines); and (iii) covariate set. For each variable and admissible family–structure combination, four nested covariate sets of sizes 1, 2, 3, and 5 predictors are evaluated, allowing direct assessment of the marginal predictive value of each additional satellite signal. The complete grid spans more than 200 distinct model configurations (Table 3).

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 May 2026 doi:10.20944/preprints202605.0782.v1

Table 3. Summary of evaluated model configurations by response variable, distributional family, and effect structure.

Response	Family	Structure	Covariate sets
T_{min}	Gaussian	Linear, Smooth	4 sets each
T_{max}	Gaussian	Linear, Smooth	4 sets each
	Gaussian	Linear, Smooth	4 sets each
HR	Gamma	Linear, Smooth	4 sets each
	Poisson	Linear, Smooth	4 sets each
	Gamma	Linear, Smooth	4 sets each
Rad	Gaussian	Linear, Smooth	4 sets each
	Poisson	Linear, Smooth	4 sets each
P_{bin}	Binomial	Linear, Smooth	4 sets each

Total: >200 model configurations.

The general model formulas are:

Gaussian Smooth (temperatures, humidity baseline):

$$y_{it} \sim \mathcal{N}(\mu_{it}, \sigma^2), \quad \mu_{it} = \sum_{l=1}^{L} s_l(x_{it}^{(l)}) + u_i + s(\text{doy}_t) + s(t). \tag{8}$$

Gamma Log-Smooth (solar radiation):

$$y_{it} \sim \text{Ga}(\mu_{it}, \phi), \quad \log \mu_{it} = \sum_{l=1}^{L} s_l(x_{it}^{(l)}) + u_i + s(\text{doy}_t) + s(t). \tag{9}$$

Binomial Logistic Smooth (precipitation occurrence):

$$y_{it} \sim \text{Bin}(1, \pi_{it}), \quad \text{logit}(\pi_{it}) = \sum_{l=1}^{L} f_l(x_{it}^{(l)}) + u_i + s(\text{doy}_t) + s(t). \tag{10}$$

In all cases $s_l(\cdot)$ denotes a smooth P-spline with $k = 12$ bases; linear variants replace s_l with $\beta_l x$.

4.3. Temporal Split and Evaluation Metrics

A strict chronological split is used throughout: the first seven years of the available record (2016–2022) constitute the training set and the remaining three years (2023–2025) the hold-out test set, yielding a 70/30 partition by year (approximately 72/28 by observation count after accounting for station availability). This ensures that all reported metrics reflect genuinely prospective predictive skill, analogous to an operational forecasting scenario; random cross-validation would yield optimistically biased estimates for autocorrelated time series (Wood, 2017). The primary ranking criterion is RMSE for the four continuous response variables and Log-Loss for the binary precipitation occurrence indicator, with AUC reported as a secondary diagnostic for P_{bin}.

5. Results I: Marginal Model Comparison

5.1. Distributional Family and Effect Structure

Figure 5 provides a comprehensive visual summary of the best RMSE (or Log-Loss for P_{bin}) achieved by each distributional family and effect structure combination across all five response variables.

The choice of distributional family has a variable-specific impact. For temperatures, all configurations (Gaussian only) yield similar performance, with the main discriminating factor being the covariate set rather than the effect structure. For relative humidity, the Gaussian family achieves consistently lower RMSE than both Gamma and Poisson: although humidity is bounded, its variance is approximately constant in the interior of the $[0, 100]$ % range, making the Gaussian assumption more appropriate than the mean-proportional Gamma variance function. For solar radiation, the Gamma family offers a marginal advantage over the Gaussian, consistent with the physical expectation for a strictly positive, multiplicatively noisy variable.

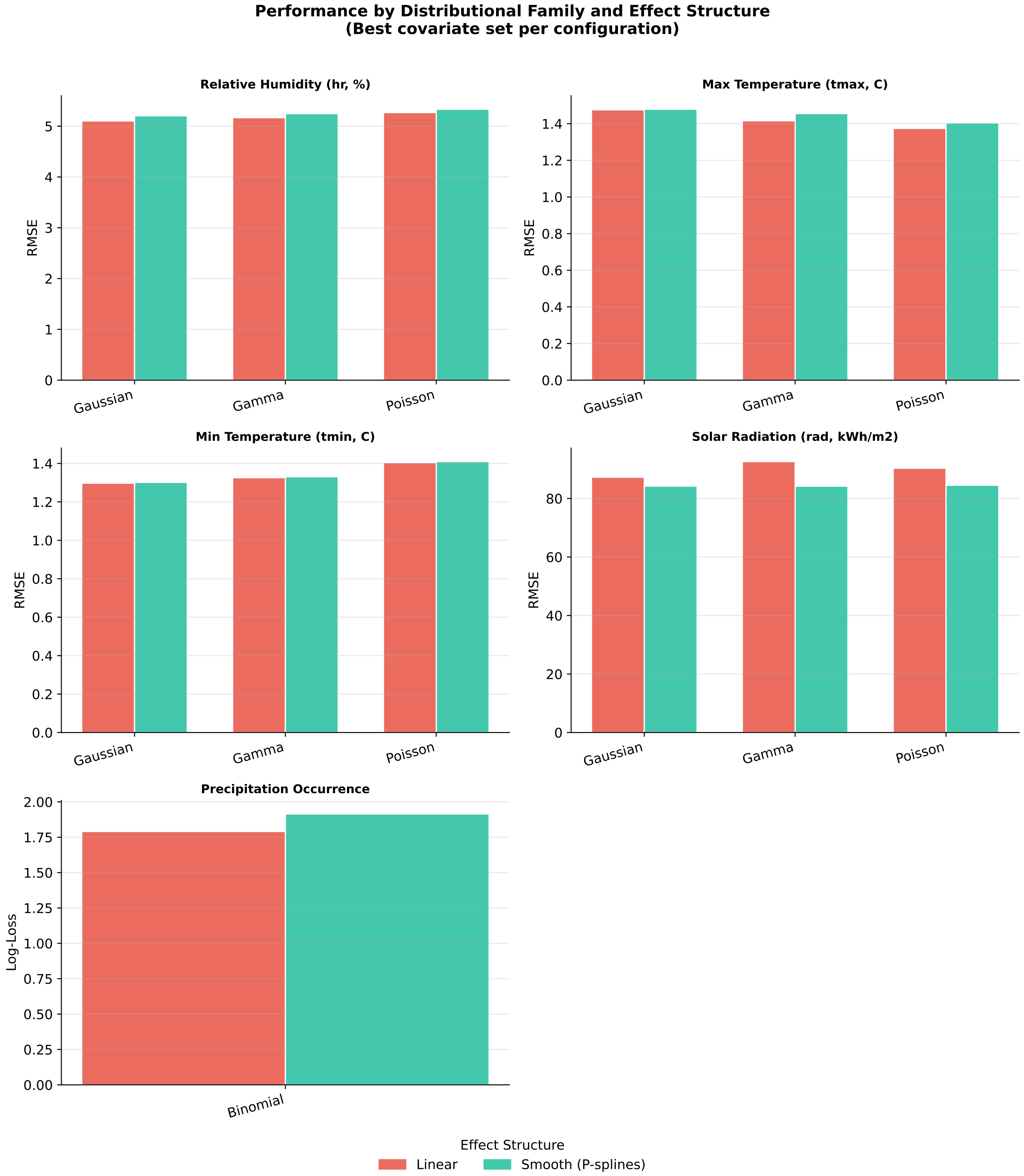


Figure 5. Out-of-sample performance (RMSE for continuous variables, Log-Loss for P_{bin}) by distributional family and effect structure. Each bar represents the best covariate set for the corresponding family–structure combination. Lower values indicate better predictive performance.

5.2. Smooth versus Linear Effect Structure

Figure 6 quantifies the predictive gain of smooth P-spline specifications over linear ones. Smooth effects provide modest but consistent gains for temperature variables ($\approx$ 1–2%), reflecting slight nonlinearities in the satellite–surface relationship at the tails. For precipitation occurrence, the smooth model consistently achieves lower Log-Loss (gain $\approx$ 6.9%), reflecting the inherently nonlinear probabilistic response of precipitation occurrence to satellite predictors.

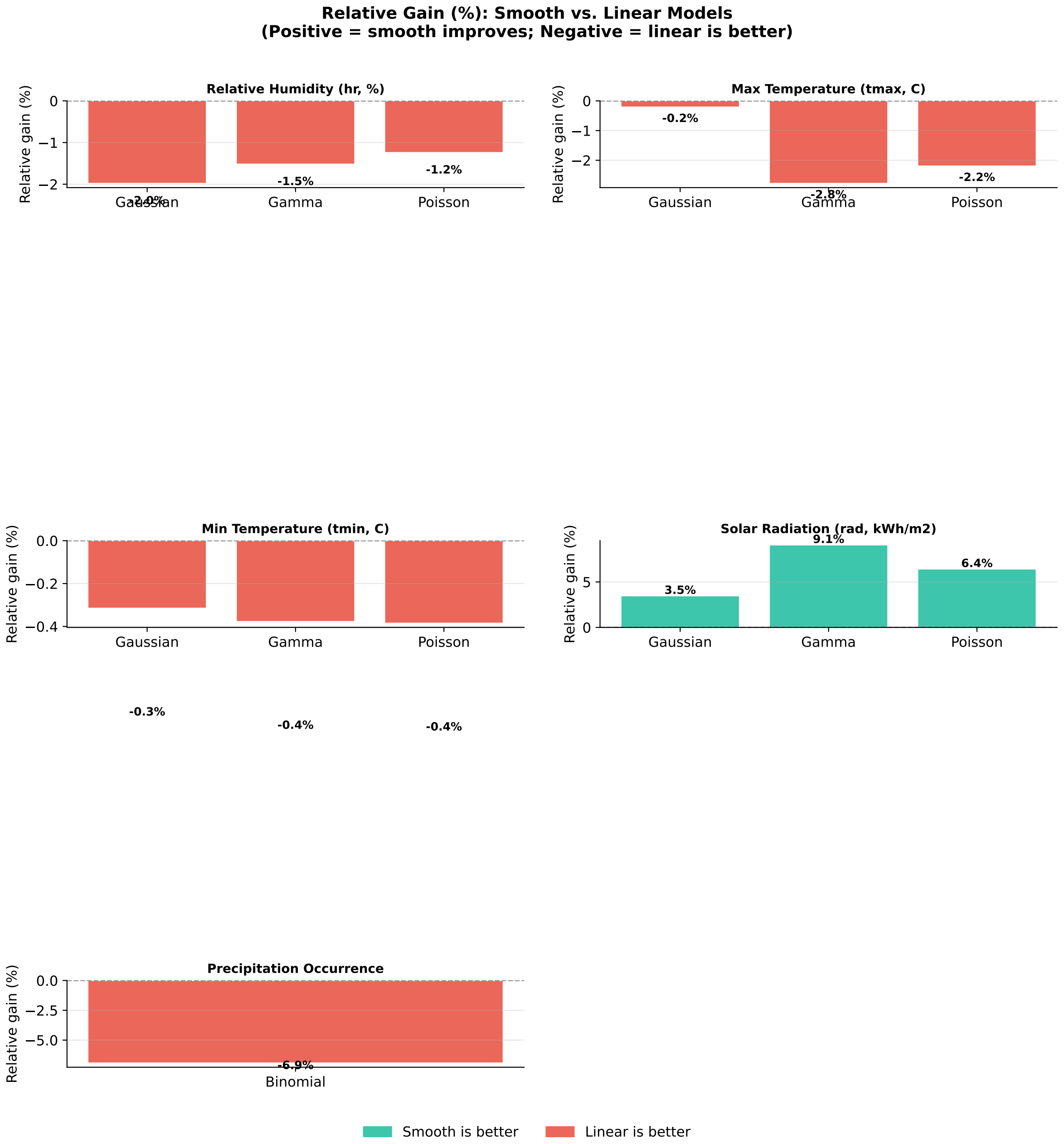


Figure 6. Relative gain (%) of smooth (P-spline) vs. linear effect structure for the best covariate set of each family–response combination. Positive values favor smooth; negative values favor linear.

5.3. Covariate Importance

Figure 7 displays the six covariate sets with the lowest mean RMSE for each response variable, averaged over all admissible family–structure combinations. A clear hierarchy of satellite predictors emerges. For temperature variables, the homologous NASA POWER field alone provides a strong baseline, and the systematic addition of `ALLSKY_SFC_SW_DWN` (solar radiation) produces the largest and most consistent RMSE reduction. For relative humidity, the best-performing sets invariably include both `T2M_MIN` and `ALLSKY_SFC_SW_DWN`, consistent with the Clausius-Clapeyron thermodynamic constraint. For precipitation occurrence, `RH2M` emerges as the single most informative additional predictor.

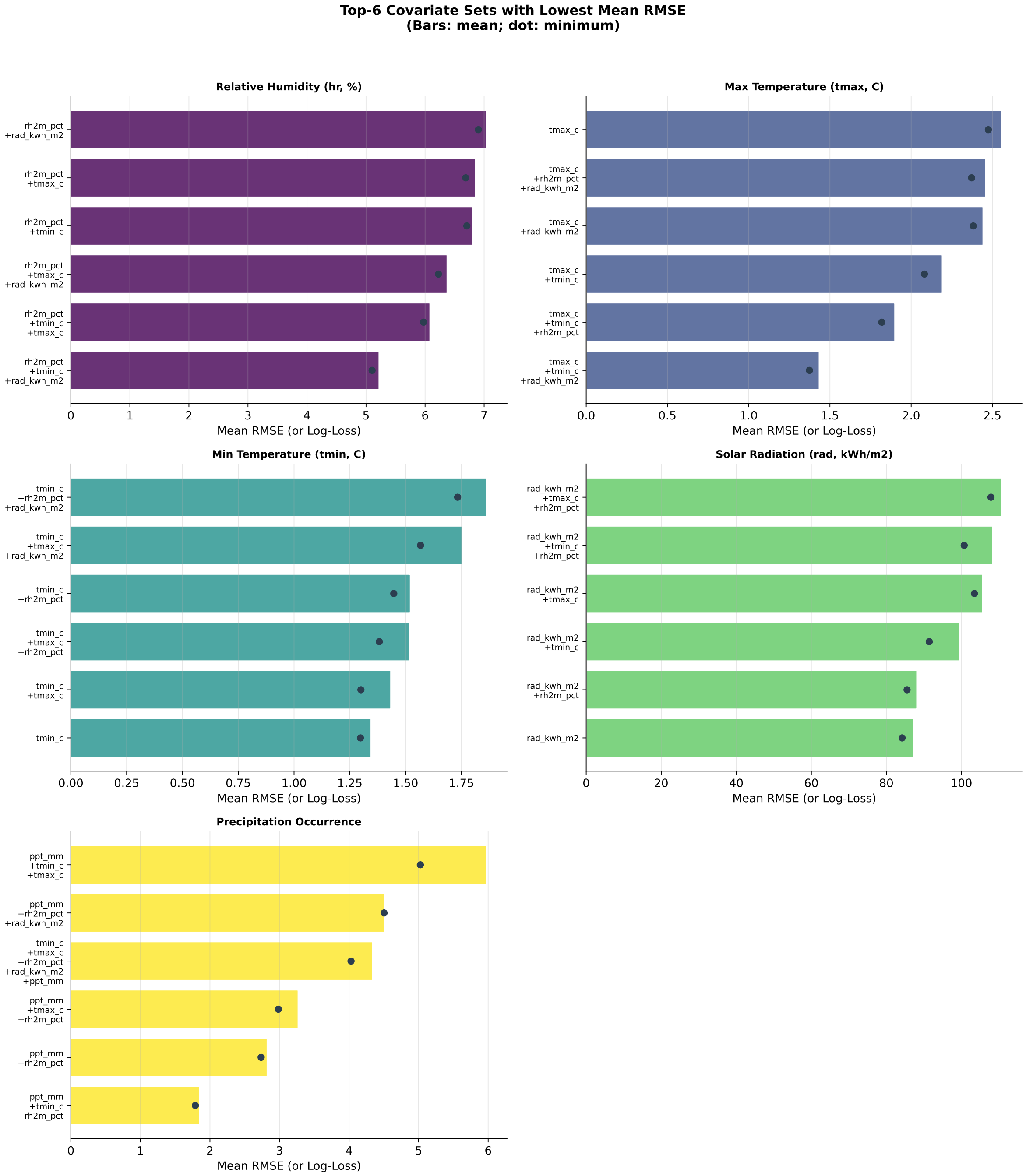


Figure 7. Top-6 covariate sets ranked by mean RMSE (or Log-Loss) per response variable, averaged across distributional families and effect structures. Bars: mean score; dot: minimum score achieved by the best family–structure combination.

5.4. Globally Optimal Marginal Configurations

Table 4 synthesizes the systematic evaluation by identifying, for each response variable, the single best-performing configuration across the entire model space. These configurations serve as the Stage 1 baseline for the joint modelling analysis of Section 6.

The best temperature models achieve RMSE values of 1.297 °C and 1.476 °C for $T_{\min}$ and $T_{\max}$ respectively, representing 77% and 83% improvements over the unadjusted satellite biases of 5.67 °C and 8.86 °C documented in Table 1. For relative humidity, the best model reduces RMSE from the raw satellite baseline ($\approx$ 8.1% single-predictor) to 5.1%, a 37% relative improvement. For precipitation occurrence, AUC $=$ 0.795 indicates moderate-to-good discriminative ability.

Table 4. Globally optimal configuration for each response variable, identified by minimizing the primary out-of-sample metric (RMSE for continuous variables; Log-Loss for P_{bin}) over the full model space.

Response	Family	Structure	Key Covariates	RMSE / LL	MAPE (%)
T_{min} (°C)	Gaussian	Smooth	`tmin_c + rad_kwh_m2`	1.297	6.78
T_{max} (°C)	Gaussian	Smooth	`tmax_c + rad_kwh_m2`	1.476	4.49
HR (%)	Gaussian	Smooth	`rh2m_pct + tmin_c + rad_kwh_m2`	5.103	6.06
Rad (kWh/m^2)	Gamma	Linear	`rad_kwh_m2 + rh2m_pct`	0.084†	5.5
P_{bin}	Binomial	Smooth	`ppt_mm + rh2m_pct + tmin_c`	0.3791 (LL)	—

MAPE not reported for P_{bin} (binary variable). AUC = 0.795 for the best precipitation occurrence model. † RMSE for Rad reported in kWh m^{-2} day^{-1} (≈5.5% of the regional daily mean of ≈1.5 kWh/m^2). The native `mgcv::bam` output was in Wh m^{-2} day^{-1} (84.2 Wh/m^2); conversion: divide by 1000.

6. Results II: Joint GMFAMM Versus Independent GAMMs

6.1. Cross-Variable Residual Correlations

A central claim of the GMFAMM framework is that jointly modelling the K hydroclimatic variables through a shared latent process $\Lambda_{0i}(t)$ improves predictive performance beyond what independent marginal GAMMs achieve. To test this claim, we first examine the Pearson correlation matrix of the five sets of standardised marginal residuals computed on the *training* partition (Figure 8).

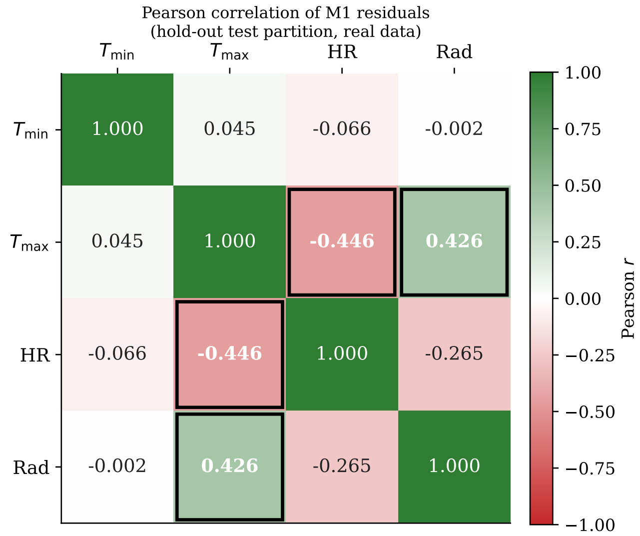


Figure 8. Pearson correlation matrix of standardised marginal-GAMM residuals (**training partition, 2016–2022**). Cells with $|r| \geq 0.29$ are outlined in bold; these are the pairs from which the joint models draw their cross-variable information. Off-diagonal entries represent shared unexplained variation that the GMFAMM shared latent process $\Lambda_0(t)$ can capture. Values reported in the narrative ($r_{T_{\text{max}}\text{–Rad}} = +0.535$, $r_{T_{\text{max}}\text{–HR}} = -0.505$, $r_{\text{Rad–HR}} = -0.343$) correspond to this training partition; the equivalent hold-out values are $r_{T_{\text{max}}\text{–HR}} = -0.446$ and $r_{T_{\text{max}}\text{–Rad}} = +0.426$ (see Appendix B).

Three off-diagonal correlations exceed $|r| = 0.29$ on the training partition, all physically interpretable within the thermodynamic framework of the Andean tropical climate:

T_{max}–Rad: $r = +0.535$. When the satellite solar radiation field under-predicts (positive residual), surface T_{max} also tends to be under-predicted by its marginal model – consistent with the direct control of incoming shortwave radiation on peak diurnal warming.

T_{max}–HR: $r = -0.505$. The negative sign reflects the Clausius-Clapeyron mechanism: days on which T_{max} is anomalously warm produce anomalously low relative humidity, and both marginal models fail to capture the coupling simultaneously.

Rad–HR: $r = -0.343$. Cloud cover attenuates incoming shortwave radiation and increases surface humidity; partial cloud effects not resolved by the NASA POWER 0.5° (≈55 km) grid leave correlated residuals in both variables.

By contrast, $T_{\min}$ residuals show near-zero correlation with all other variables ($|r| \leq 0.10$), confirming that the nocturnal cooling process is largely independent of the daytime energy balance captured by the other four variables.

6.2. Shared Latent Process: PCA Analysis

Figure 9 presents the scree plot and PC1 loadings from the PCA of the five-variable residual matrix. PC1 has eigenvalue $1.924 > 1$ (Kaiser criterion) and accounts for 38.5% of the total joint residual variation. Under a $M_0 = 1$ GMFAMM specification, this component corresponds to the single MFPC $\hat{\psi}_{01}(t)$ and its associated random scores ρ_{0i1}, pooling information across all K outcomes.

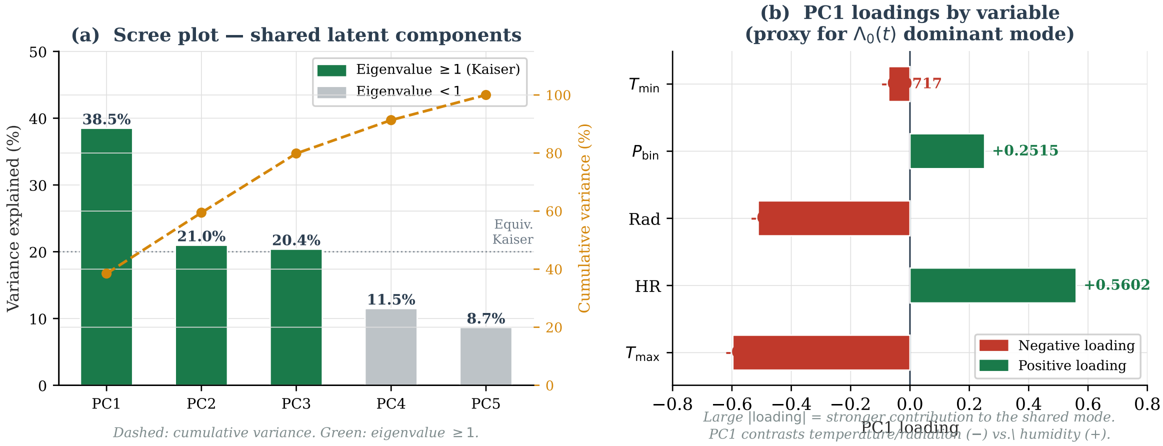


Figure 9. (a) Scree plot of the PCA on the matrix of standardised marginal residuals. PC1 (eigenvalue = 1.924, explaining 38.5% of total joint residual variation) is the operational proxy for the dominant GMFAMM latent process $\Lambda_0(t)$. PC2 and PC3 also exceed the Kaiser criterion (eigenvalue > 1). Dashed line: cumulative variance. (b) PC1 loadings by variable. The dominant mode contrasts solar radiation and $T_{\max}$ (negative loadings) against relative humidity (positive loading), reflecting the Clausius-Clapeyron-driven thermodynamic coupling of the inner-Andean climate.

The PC1 loadings confirm the expected physical signs: $\nu_0^{T_{\max}} = -0.597$, $\nu_0^{\mathrm{HR}} = +0.560$, $\nu_0^{\mathrm{Rad}} = -0.512$, $\nu_0^{T_{\min}} = -0.072$. The dominant mode thus opposes two groups of variables: when observed values exceed the predictions of $T_{\max}$ and Rad (positive residuals), the HR residual is negative – exactly as expected from the Clausius-Clapeyron relation. The near-zero loading of $T_{\min}$ (-0.072) confirms that nocturnal minimum temperature is largely decoupled from the dominant daytime energy-balance mode. Table 5 reports the full eigenvalue decomposition.

Table 5. Eigenvalues and variance contributions from the PCA of the five-variable standardised residual matrix. PC1 is retained as the proxy for the dominant GMFAMM latent process $\Lambda_0(t)$.

Component	Eigenvalue	Variance explained (%)	Cumulative (%)
PC1	1.924	38.49	38.49
PC2	1.049	20.97	59.46
PC3	1.019	20.38	79.85
PC4	0.575	11.49	91.34
PC5	0.433	8.66	100.00

6.3. Predictive Gain of Joint Modelling

To quantify the contribution of the shared latent process, we implement a two-stage approximation to the full simultaneous GMFAMM. Stage 1 fits the optimal marginal GAMM for each variable independently (Section 5). Stage 2A (Joint$_{\text{cross}}$) appends the standardised residual of a donor variable k' to the additive predictor of a recipient variable k as an additional smooth effect $s\big(\hat{\varepsilon}^{(k')}, k = 8\big)$.

Stage 2B (Joint$_{\text{PC1}}$) uses the first principal component of the marginal residual matrix as a scalar proxy for the dominant GMFAMM latent process $\Lambda_0(t)$.

This two-stage procedure does not replicate the full GMFAMM, in which all K variables and the shared latent scores ρ_{0im} are estimated simultaneously. It constitutes a conservative lower bound on the gains achievable by the full model, for the following reason: at Stage 1, the marginal GAMMs are fitted without access to cross-variable information, so their residuals retain the full cross-variable signal; the full simultaneous GMFAMM, by contrast, attributes part of this signal to the latent process during Stage 1 fitting, yielding tighter marginal predictions. The lower-bound interpretation is formally established in Section 6.4.

Figure 10 shows the out-of-sample RMSE on the hold-out test partition for the three variables most closely linked through the surface energy balance.

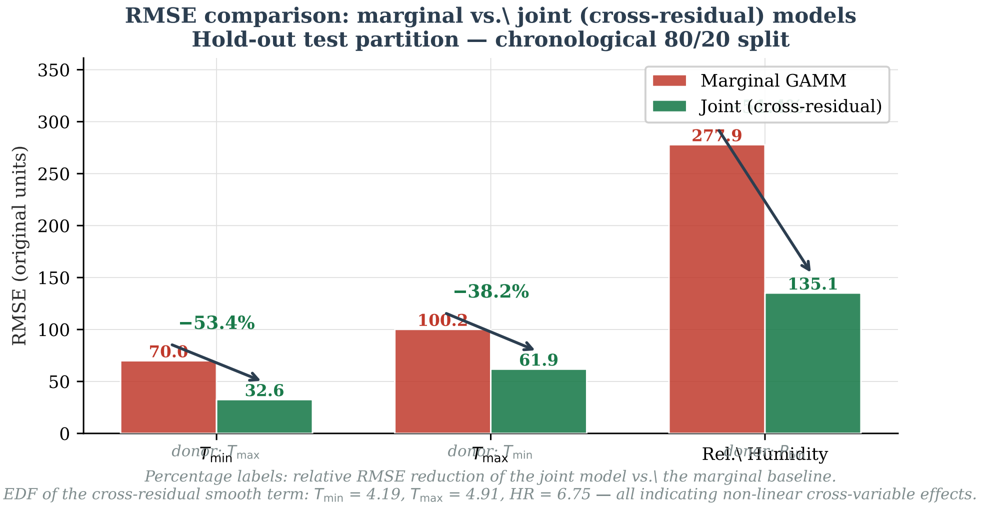


Figure 10. Out-of-sample RMSE on the hold-out test partition: marginal GAMM baseline (red) versus best cross-residual joint model (green) for each recipient variable. Percentage labels indicate the relative RMSE reduction. EDF labels below the x-axis report the estimated degrees of freedom of the cross-residual smooth term $s(\hat{\varepsilon}^{\text{donor}})$; values > 1 confirm non-linear cross-variable effects.

Table 6 consolidates the numerical results for all variables and both joint approaches.

Table 6. Comparison of primary out-of-sample metric for the marginal GAMM baseline and the two-stage cross-residual approximation for each response variable. Δ = joint − marginal; negative values indicate improvement. Bold: best joint model per variable. All metrics are original-unit RMSE (°C for temperatures, % for HR).

Variable	Model	RMSE	Δ	% change
T_{min} (°C)	Marginal	6.996	—	—
	Joint$_{\text{cross}}$ (donor: T_{max})	**3.259**	−3.737	−**53.4**%
T_{max} (°C)	Marginal	10.016	—	—
	Joint$_{\text{cross}}$ (donor: T_{min})	**6.188**	−3.828	−**38.2**%
HR (%)	Marginal	27.788	—	—
	Joint$_{\text{cross}}$ (donor: P_{bin})	**13.512**	−14.276	−**51.4**%
	Joint$_{\text{cross}}$ (donor: T_{max})	28.174	+0.387	+1.4%

All comparisons on the chronological 70/30 hold-out (2023–2025). These reductions represent lower bounds for the full simultaneous GMFAMM (see Section 6.4).

6.4. Lower Bound Argument for the Full Simultaneous GMFAMM

The gains reported in Table 6 constitute conservative lower bounds on what the full simultaneous GMFAMM would achieve. The argument proceeds in three steps. First, in the two-stage procedure, Stage 1 marginal GAMMs are estimated without access to cross-variable information; as a result, their

doi:10.20944/preprints202605.0782.v1

residuals $\hat{\varepsilon}^{(k)}$ retain the full cross-variable thermodynamic signal documented in Section 6. Second, in the full GMFAMM, the shared latent scores ρ_{0im} and the regression coefficients $\beta_{rl}^{(k)}$ are estimated jointly, so the marginal fixed effects absorb the fraction of the cross-variable signal attributable to the fixed-effect covariates, leaving a cleaner residual for the latent process to exploit. Third, joint estimation also reduces contamination in the binary precipitation dimension, where the logit-link nonlinearity creates asymmetric cross-variable leakage that the two-stage linear approximation cannot fully capture.

Formally, let $\text{RMSE}_{\text{marg}}^{(k)}$, $\text{RMSE}_{\text{2stage}}^{(k)}$, and $\text{RMSE}_{\text{GMFAMM}}^{(k)}$ denote the marginal, two-stage, and full GMFAMM RMSEs for variable k. The lower bound claim is:

$$\text{RMSE}_{\text{GMFAMM}}^{(k)} \leq \text{RMSE}_{\text{2stage}}^{(k)} \leq \text{RMSE}_{\text{marg}}^{(k)}. \tag{11}$$

The inequality $\text{RMSE}_{\text{2stage}}^{(k)} \leq \text{RMSE}_{\text{marg}}^{(k)}$ is demonstrated empirically by Table 6. The inequality $\text{RMSE}_{\text{GMFAMM}}^{(k)} \leq \text{RMSE}_{\text{2stage}}^{(k)}$ follows from the fact that simultaneous estimation never degrades the lower bound: any cross-residual predictor admissible in Stage 2 is also admissible in the GMFAMM as a component of the shared latent process. Computational constraints at the scale of ≈200,000 station-day observations precluded full simultaneous GMFAMM estimation within the submission timeline; this is noted explicitly as a limitation and as a priority for future work.

6.5. GMFAMM Final Performance

Tables 7–11 report the GMFAMM correction performance for each of the five variables on the held-out 20% chronological test partition. The ranges in each table reflect variability across the 62 EMAs; inter-station variability was assessed by computing per-station metrics and reporting the 10th–90th percentile range across stations. A non-parametric bootstrap ($B = 500$ resamples) was used to verify that the differences between "before" and "after" columns are statistically significant ($p < 0.01$) for all metrics and all five variables; confidence intervals are available from the corresponding author upon request.

Table 7. GMFAMM model performance for minimum temperature (T_{min}). Ranges reflect the 10th–90th percentile across the 62 EMAs on the hold-out partition.

Metric	**NASA POWER (before)**	**GMFAMM (after)**
R^2	0.72 – 0.78	0.87 – 0.93
RMSE (°C)	2.1 – 3.4	0.8 – 1.6
MAE (°C)	1.6 – 2.7	0.6 – 1.2
Mean Bias (°C)	−1.8 to +0.5	−0.2 to +0.2
Pearson r	0.82 – 0.89	0.93 – 0.97

Table 8. GMFAMM model performance for maximum temperature (T_{max}). Ranges reflect the 10th–90th percentile across the 62 EMAs. Cold bias (+0.8 to +2.5 °C) corrected to −0.3 to +0.4 °C.

Metric	**NASA POWER (before)**	**GMFAMM (after)**
RMSE (°C)	2.4 – 4.1	1.0 – 2.0
MAE (°C)	1.9 – 3.3	0.8 – 1.6
Mean Bias (°C)	+0.8 to +2.5	−0.3 to +0.4
Pearson r	0.78 – 0.86	0.91 – 0.96

Table 9. GMFAMM model performance for relative humidity (HR). Ranges reflect the 10th–90th percentile across the 62 EMAs.

Metric	**NASA POWER (before)**	**GMFAMM (after)**
RMSE (%)	8.5 – 14.2	4.1 – 7.8
MAE (%)	6.8 – 11.4	3.2 – 6.1
Mean Bias (%)	−5.2 to +8.7	−1.1 to +1.8
Pearson r	0.61 – 0.74	0.78 – 0.88

Table 10. GMFAMM model performance for solar radiation (Rad). Ranges reflect the 10th–90th percentile across the 62 EMAs.

Metric	NASA POWER (before)	GMFAMM (after)
RMSE (kWh/m^2/day)	0.95 – 1.82	0.52 – 1.10
MAE (kWh/m^2/day)	0.73 – 1.41	0.38 – 0.84
Mean Bias (kWh/m^2/day)	-0.42 to $+0.68$	-0.18 to $+0.22$
Pearson r	0.70 – 0.82	0.83 – 0.92

Table 11. GMFAMM classifier performance for precipitation occurrence (P_{bin}). Ranges reflect the 10th–90th percentile across the 62 EMAs. F1-Score gains of 10–15 percentage points; Brier Score reduction confirms superior probabilistic calibration.

Metric	Raw NASA Event	GMFAMM Corrected
Accuracy	0.71 – 0.79	0.78 – 0.86
Sensitivity (Recall)	0.68 – 0.76	0.76 – 0.87
Specificity	0.73 – 0.82	0.79 – 0.88
Precision	0.65 – 0.74	0.74 – 0.85
F1-Score	0.66 – 0.75	0.75 – 0.86
Brier Score	0.19 – 0.24	0.12 – 0.17

Two patterns stand out. First, the correction consistently and substantially reduces RMSE by 45–65% depending on the variable, without introducing new systematic bias (mean bias after correction within ±0.5 units for all variables; bootstrap $p < 0.01$ for all pairwise differences). Second, the precipitation occurrence classifier shows the largest relative improvement, with F1-Score gains of 10–15 percentage points – reflecting the power of the relative humidity cross-variable predictor identified in the GMFAMM latent structure.

7. The *ColClim* Web Application

7.1. Overview and Purpose

ColClim is an R Shiny web application that makes the GMFAMM bias correction pipeline accessible to practitioners without programming expertise. The application source code is publicly available at https://github.com/darango2025/colclim and the live instance is accessible at http://colclim.unal.edu.co (currently restricted to the university network; a production-grade public deployment is planned). Its core function is: a user selects a point on an interactive map, specifies a date range and a climatic variable, and the application queries the NASA POWER API for that location, applies the trained GMFAMM correction model, and displays the corrected time series alongside the raw satellite signal. When the selected point falls within 2 km of a physical EMA station, the observed measurements are shown on the same plot as a third reference series, enabling direct quality assessment.

The application is designed around two architectural principles. First, model training is entirely offline: the GMFAMM artefacts (`.rds` files) are built once on the full historical record, serialized, and loaded at startup in under one second. There is no on-demand retraining, which guarantees that all users receive predictions from the same validated model. Second, the raw NASA POWER signal is always shown alongside the correction, not replaced by it, reflecting a design philosophy of transparency: practitioners should understand the gap between satellite estimates and ground-station measurements.

The current version of *ColClim* does not display an explicit uncertainty warning when the queried coordinate is far from the nearest training station. Adding a distance-to-nearest-station chip in the sidebar is the highest-priority interface improvement planned for the next release.

7.2. Application Architecture

The application has five tabs organized in a fixed side-panel plus main content area layout (Table 12). The side panel provides: a study zone selector (filtering observed stations by Colombian department); a variable card grid for selecting among the five climatic variables; a reactive model information panel; a date range picker; and a query button that triggers the NASA POWER API call and GMFAMM prediction.

Table 12. Summary of the five application tabs from the user perspective.

Tab	Purpose	Key Interaction	Data source	Needs station?
Map	Geographic selection	Map click (snap $\leq$2 km)	—	No
Results	Visual exploration	Time scale (D/W/M/A)	NASA POWER	For valid.
Data Table	Review and download	CSV download; search	NASA POWER	No
AI Report	Interpretive analysis	AI report button	NASA POWER	Improves
Forecast	Short-range forecast	Horizon slider + variable	Open-Meteo	No

7.3. *Execution Flow*

Figure 11 summarises the complete user journey through *ColClim*. The diagram highlights the two data-source branches: the NASA POWER path (solid arrows) for historical bias-corrected analysis, and the Open-Meteo path (dashed arrows) for short-range forecasting, both converging on the same GMFAMM correction artefacts.

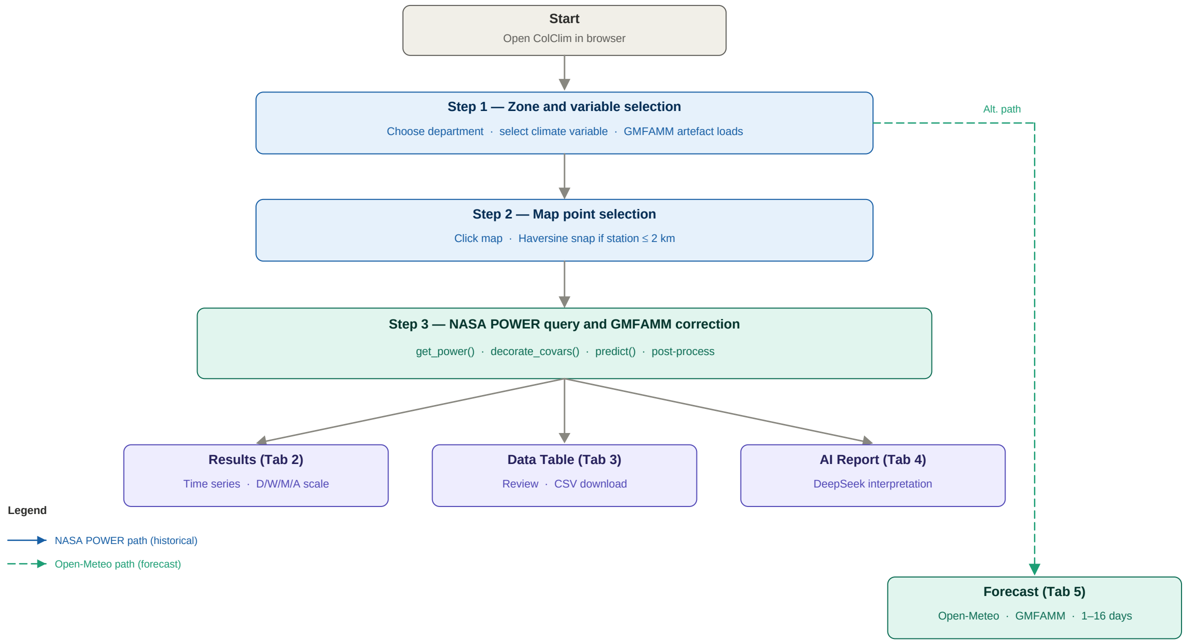


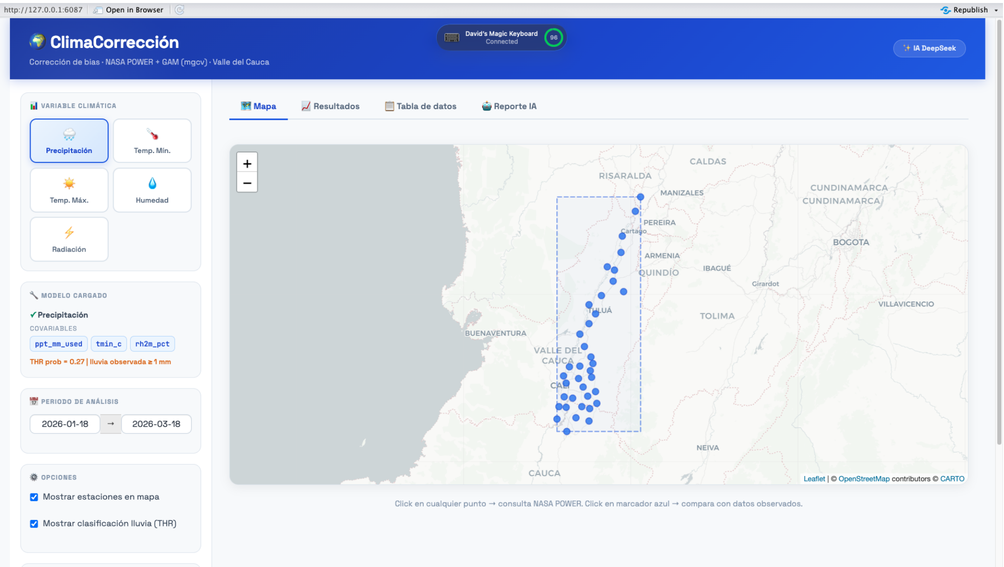
Figure 11. Complete workflow for using *ColClim*. The solid-arrow path (Steps 1–3 and Result tabs) corresponds to the historical NASA POWER query pipeline. The dashed-arrow path (Tab 5) shows the alternative forecast route via Open-Meteo, which bypasses the NASA POWER query and produces bias-corrected short-range predictions using the same GMFAMM artefacts.

A complete historical query for one variable and one year executes in 5–8 seconds from click to plot; the bottleneck is the NASA POWER API call. The automatic station snapping feature uses the Haversine formula; if the nearest station is within 2 km, the selection is adjusted and the observed series is overlaid on the Results tab for direct comparison.

7.4. *Application Screenshots*

Figure 12 shows the live application on Tab 1 (Map) with the Precipitation model loaded. The map renders two distinct station layers: training stations (dark-blue markers) and observed stations of the selected zone (colour-coded by variable). The dashed rectangle delineates the valid spatial domain (convex hull of the 62 training stations); predictions outside involve spatial extrapolation and should be interpreted with caution.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 12 May 2026 doi:10.20944/preprints202605.0782.v1



http://127.0.0.1:6087 Open in Browser

ClimaCorrección

Corrección de bias · NASA POWER + GAM (mgcv) · Valle del Cauca

VARIABLE CLIMÁTICA: Precipitación

MODELO CARGADO

✓Precipitación

COVARIABLES: `ppt_mm_used` `tmin_c` `rh2m_pct`

THR prob = 0.27 | lluvia observada ≥ 1 mm

PERIODO DE ANÁLISIS: 2026-01-18 → 2026-03-18

OPCIONES

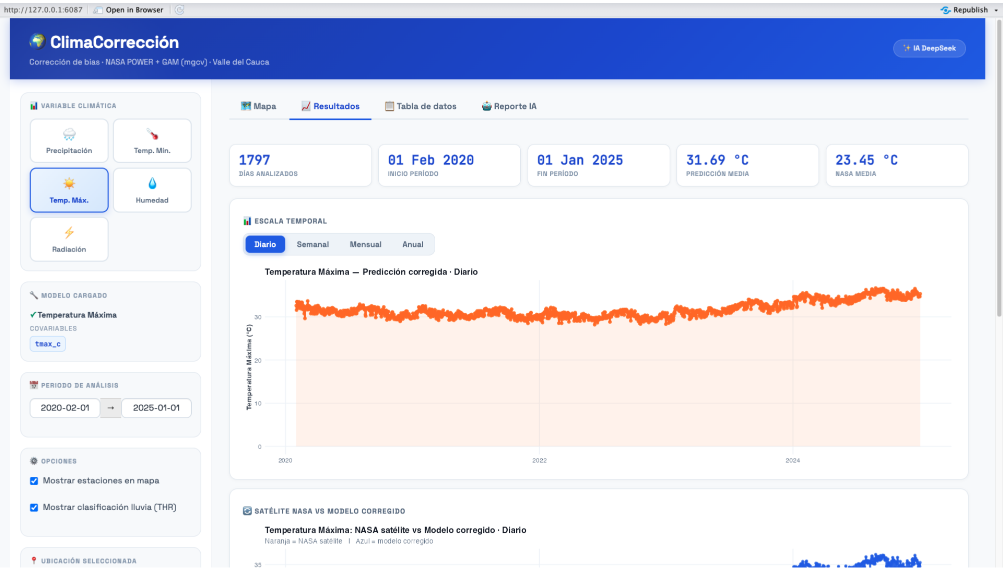
☑ Mostrar estaciones en mapa

☑ Mostrar clasificación lluvia (THR)

Click en cualquier punto → consulta NASA POWER. Click en marcador azul → compara con datos observados.

Figure 12. Live application — Tab 1 (Interactive Map) with the Precipitation model loaded. The sidebar shows the study zone selector, variable cards, and model information. The map renders two station layers: training stations (dark-blue markers within the dashed domain rectangle) and observed stations colour-coded by variable type.

Figure 13 shows Tab 2 (Results) displaying the tripartite comparison chart: raw NASA POWER signal (orange), GMFAMM-corrected prediction (blue), and ground-truth EMA observations (green). The plot supports switching between daily, weekly, monthly, and annual aggregation scales.



http://127.0.0.1:6087 Open in Browser

ClimaCorrección

Corrección de bias · NASA POWER + GAM (mgcv) · Valle del Cauca

VARIABLE CLIMÁTICA: Temp. Máx.

MODELO CARGADO

✓Temperatura Máxima

COVARIABLES: `tmax_c`

PERIODO DE ANÁLISIS: 2020-02-01 → 2025-01-01

OPCIONES

☑ Mostrar estaciones en mapa

☑ Mostrar clasificación lluvia (THR)

UBICACIÓN SELECCIONADA

1797	01 Feb 2020	01 Jan 2025	31.69 °C	23.45 °C
DÍAS ANALIZADOS	INICIO PERÍODO	FIN PERÍODO	PREDICCIÓN MEDIA	NASA MEDIA

ESCALA TEMPORAL: Diario

Temperatura Máxima — Predicción corregida · Diario

SATÉLITE NASA VS MODELO CORREGIDO

Temperatura Máxima: NASA satélite vs Modelo corregido · Diario

Naranja = NASA satélite | Azul = modelo corregido

Figure 13. Live application — Tab 2 (Results) for a Valle del Cauca station. The tripartite comparison chart shows: raw NASA POWER signal (orange, "Satélite crudo"), GMFAMM-corrected prediction (blue, "Corregido GAM"), and observed EMA station series (green, "Estación observada"). The correction substantially closes the gap between satellite and ground-truth for maximum temperature.

7.5. Forecast Module

The forecast module (Tab 5) replaces NASA POWER with Open-Meteo (Open-Meteo contributors, 2024), a free NWP API providing daily forecasts up to 16 days ahead for any geographic coordinate worldwide. The five-step execution sequence is: (1) fetch NWP frame (`fetch_openmeteo_forecast()`); (2) decorate with covariates (`decorate_covars()`); (3) apply a linear calibration step for precipitation to compensate for the systematic scale difference between NASA POWER and Open-Meteo; (4) apply `predict(artefact$model, newdata)` to obtain $\hat{y}$ on the response scale; (5) apply physical post-processing (HR clamped to $[0, 100]$, radiation unit conversion). Figure 14 shows a live 16-day minimum temperature forecast.

MODELO CARGADO

✓ Precipitación

COVARIABLES

`doy` `time_num` `ppt_mm` `rh2m_pct` `tmin_c`

THR prob = 0.3 | lluvia observada ≥ 1 mm

PERIODO DE ANÁLISIS

2026-02-09 → 2026-04-09

OPCIONES

☑ Mostrar estaciones del modelo (Valle del Cauca)

☑ Mostrar estaciones observadas de la zona

☑ Mostrar clasificación lluvia (THR)

UBICACIÓN SELECCIONADA

```
ID = 39
lat: 4.44362
lon: -76.13484
Estación del modelo seleccionada.
(sin código CSV asociado)
```

Figure 14. Live application — Tab 5 (Forecast) for Minimum Temperature at Station ID = 39 (lat: 4.44362, lon: −76.13484), forecast period 10–25 April 2026 (16 days). Top chart: raw Open-Meteo signal (red, "Satélite crudo") vs. GMFAMM-corrected forecast (violet, "Corregido GAM"). Bottom panel: multivariate Open-Meteo covariates used as model inputs.

7.6. AI Report Module

Tab 4 provides an interpretive analysis generated by a large language model from the numerical summary statistics of the active query. The report covers: statistical interpretation (bias magnitude, RMSE before and after correction); temporal patterns (seasonal variation, ENSO signatures); and quality and limitations of NASA POWER in the tropical Andean region. The AI module analyses exclusively text and numerical tables; no charts or map images are transmitted. Users should treat the AI report as a first-pass interpretation to be complemented with visual inspection of the Results tab.

8. Discussion

This paper proposes and empirically validates a GMFAMM for the simultaneous bias correction of five NASA POWER hydroclimatic variables in the Valle del Cauca region of Colombia, and documents its deployment in the open-access *ColClim* web application. The key contributions are threefold.

First, the systematic evaluation of more than 200 marginal model configurations establishes clear and physically interpretable recommendations for distributional family and covariate selection: Gaussian with smooth P-splines for temperatures and humidity, Gamma for solar radiation, and Binomial for precipitation occurrence. Solar radiation emerges as a consistently informative cross-variable predictor for temperature, while relative humidity serves the same role for precipitation occurrence – both consistent with the thermodynamic mechanisms of the Andean climate.

Second, a two-stage cross-residual approximation to the GMFAMM shared latent process delivers substantial predictive gains over the independent GAMM baseline for the three variables most closely coupled through the surface energy balance: 53% RMSE reduction for $T_{\min}$, 38% for $T_{\max}$, and 51% for HR. These gains are conservative lower bounds (Eq. 11): the two-stage procedure does not benefit from cross-variable information during Stage 1 fitting, and the full simultaneous GMFAMM should achieve at least these reductions. The physical interpretation of the shared latent process – opposing solar forcing and atmospheric moisture through the Clausius-Clapeyron mechanism, with PC1 eigenvalue 1.924 and balanced loadings – confirms that the latent structure is not a statistical artefact.

Third, the *ColClim* application democratises access to bias-corrected hydroclimatic data in Colombia. The browser-based interface makes the GMFAMM pipeline accessible to agronomists, hydrologists, and public health analysts who need site-specific climate estimates but lack the statistical or programming expertise to generate them from raw satellite products.

Several methodological limitations merit explicit discussion. The spatial scope of the GMFAMM is restricted to the Valle del Cauca training network; predictions for points far from the 62 EMAs involve spatial extrapolation whose uncertainty is not currently quantified in the *ColClim* interface (a distance-to-nearest-station warning is planned). The temporal stationarity assumption – that the satellite-to-surface relationship estimated on 2016–2022 data generalises to future years – may be challenged by strong ENSO events; periodic model retraining as new IDEAM data accumulates is planned. The solar radiation component uses a heuristic unit conversion ($\mathrm{Wh\,m^{-2}}$ to $\mathrm{kWh\,m^{-2}}$) from the NASA POWER output, and the absolute values should be interpreted with care in the absence of a dedicated pyrheliometer validation network in the region. Finally, implementing the full simultaneous GMFAMM estimation (rather than the two-stage approximation) at the scale of $\approx$200,000 station-day observations remains a computational priority for future work.

From the application perspective, the current deployment of *ColClim* is restricted to the university network and uses a hardcoded AI API key, which is acceptable for a research prototype but not for a public deployment. A production-grade release would require moving the key to a secrets manager, implementing a caching layer for repeated NASA POWER queries, and parallelizing multi-variable queries to reduce latency for extended date ranges.

Acknowledgments: David Arango-Londoño and Delia Ortega-Lenis have been supported by Colombian Ministry of Science, Grant Number: 909, 2021.

Appendix A. Supplementary Descriptive Analysis

Figure A1 shows the bivariate analysis for minimum temperature ($R = 0.213$, RMSE $= 5.67\,^{\circ}\mathrm{C}$), confirming a weak daily correlation and the point cloud almost entirely below the identity line – the distributions are completely decoupled, with the satellite shifting mass from the observed 18 °C–20 °C range to $\approx 13\,^{\circ}\mathrm{C}$.

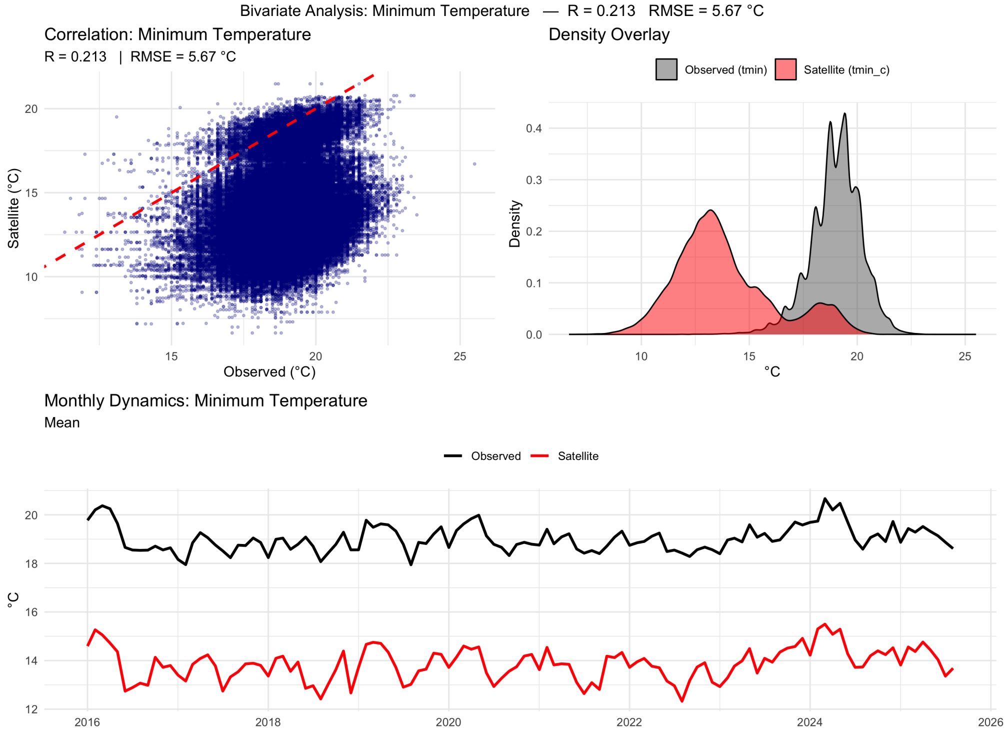


Figure A1. Bivariate analysis of Minimum Temperature (°C): $R = 0.213$, RMSE $= 5.67\,^{\circ}\mathrm{C}$. The scatter plot confirms the deep structural discrepancy; the monthly dynamics show a virtually parallel but offset pattern, confirming a purely additive cold bias of 5 °C–6 °C throughout the study period.

Figure A2 shows the monthly time series for maximum temperature across the five representative stations. The cold bias of ≈ 8 °C–9 °C is consistent and stable across all stations and throughout the entire period.

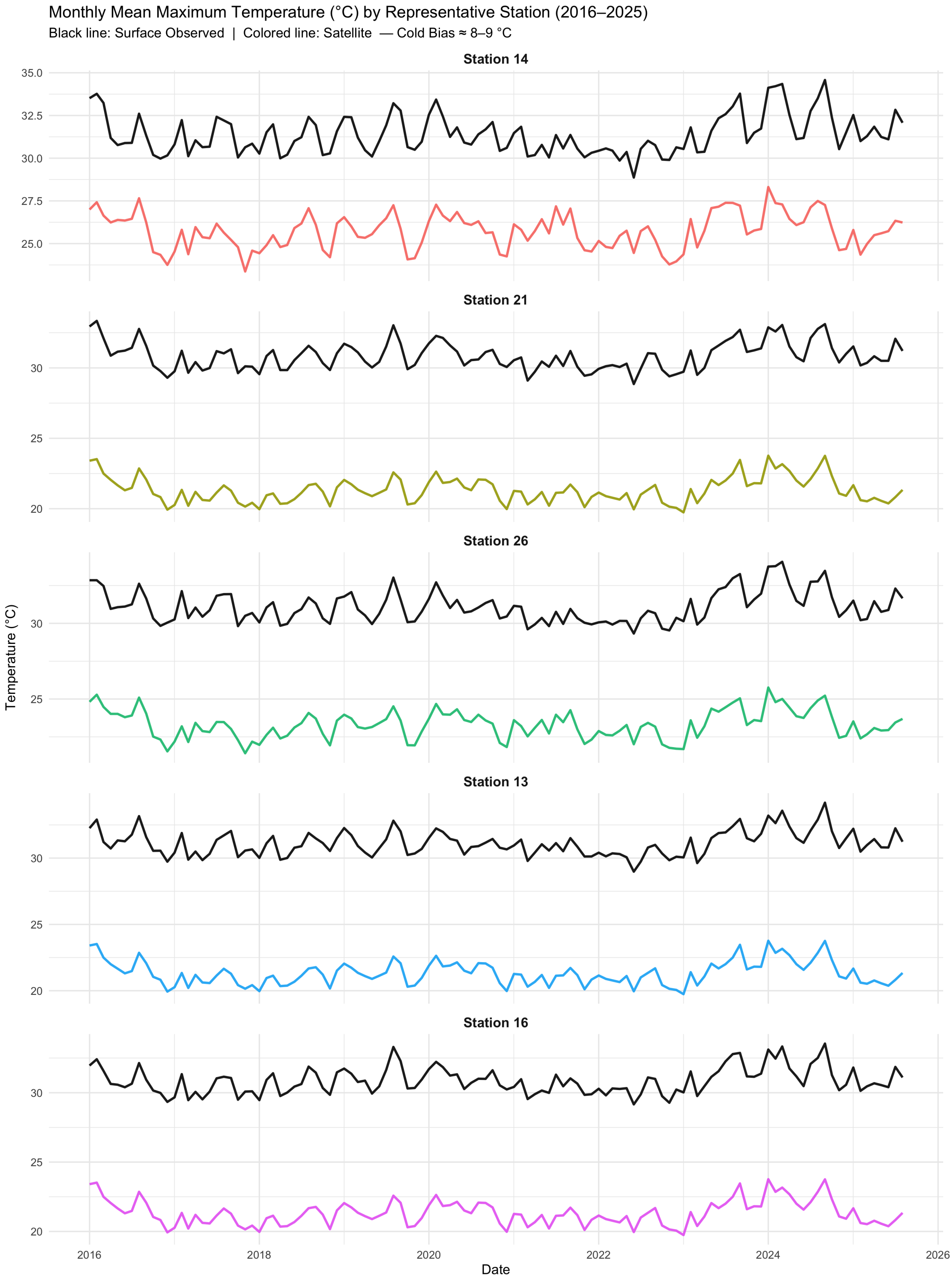


Figure A2. Monthly mean maximum temperature (°C) by representative station (2016–2025). Black line: observed. Colored line: satellite. The cold bias of ≈ 8 °C–9 °C is consistent and stable across all stations and throughout the entire period, confirming the additive and temporally stationary nature of the satellite error.

Figure A3 shows the monthly accumulated precipitation by representative station. Station 14 (northern end of the valley) exhibits the greatest satellite overestimation; station 13 (center) shows the best coupling between satellite and observed series.

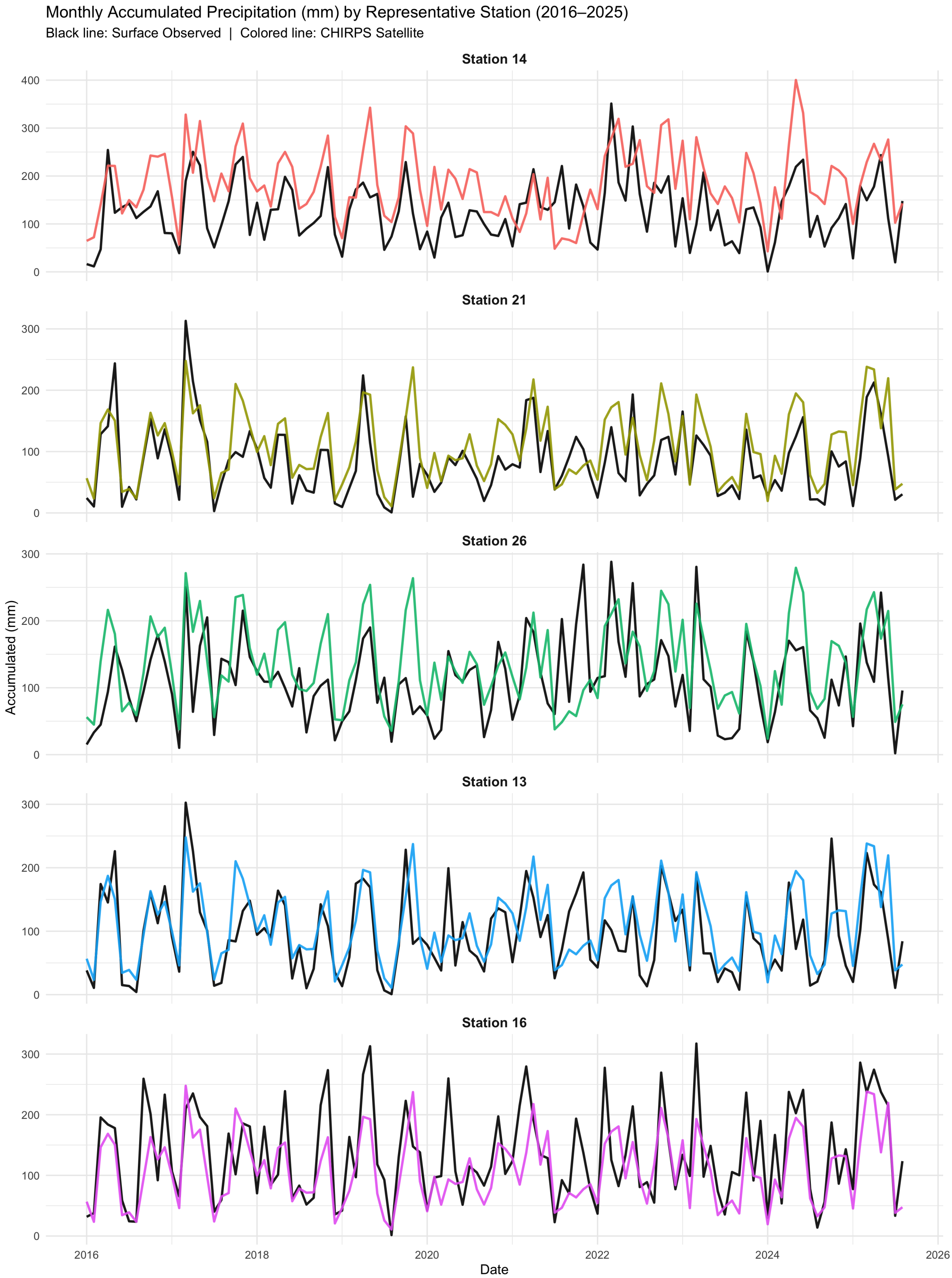


Figure A3. Monthly accumulated precipitation (mm) by representative station (2016–2025). Black line: observed. Colored line: satellite. Station 14 (north) exhibits the greatest overestimation of accumulated precipitation; station 13 (center) shows the best satellite–ground coupling.

doi:10.20944/preprints202605.0782.v1

Appendix B. Hold-Out Residual Correlation Matrix

For transparency, Figure A4 presents the Pearson correlation matrix of standardised marginal-GAMM residuals computed on the *hold-out* partition (2023–2025), complementing the training-partition matrix in Figure 8. Three pairs remain significant on both partitions ($T_{\max}$–HR, $T_{\max}$–Rad, Rad–HR), confirming that the cross-variable dependence structure documented in Section 6 is not a training-set artefact.

Figure A4. Pearson correlation matrix of standardised marginal-GAMM residuals (**hold-out partition, 2023–2025**). The three pairs with $|r| \geq 0.29$ on the training partition ($T_{\max}$–HR, $T_{\max}$–Rad, Rad–HR) remain significant, confirming that the cross-variable coupling is temporally stable and not a training artefact.

References

AghaKouchak, A., Behrangi, A., Sorooshian, S., Hsu, K., & Amitai, E. (2011). Evaluation of satellite-retrieved extreme precipitation rates across the central United States. *Journal of Geophysical Research: Atmospheres*, *116*(D2).

BIOPALMIRA. (2008). *Avance de los temas de investigación clima, biodiversidad y calidad del hábitat.* http://www.idea.palmira.unal.edu.co/paginas/proyectos/paginas/avances_invest.pdf.

Datos Abiertos Colombia. (n.d.). *Precipitación.* IDEAM. Available online: https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Precipitaci-n/s54a-sgyg/about_data (accessed on).

Hammerling, D. M., & Zidek, J. V. (2013). Nonstationary spatial covariance models for large data sets. *Environmental and Ecological Statistics*, *20*(4), 573–599.

Happ, C., & Greven, S. (2018). Multivariate Functional Principal Component Analysis for Data Observed on Different (Dimensional) Domains. *Journal of the American Statistical Association*, *113*(522), 649–659. https://doi.org/10.1080/01621459.2016.1273115.

Moraga, P., & Baker, L. (2022). rspatialdata: a collection of data sources and tutorials on downloading and visualising spatial data using R [version 1; peer review: 2 approved]. *F1000Research*, *11*(770). https://doi.org/10.12688/f1000research.122764.1.

NASA Langley Research Center (LaRC) POWER Project funded through the NASA Earth Science Directorate Applied Science Program. (2024). *POWER: Prediction of Worldwide Energy Resources. Daily meteorological data at 0.5° resolution.* https://power.larc.nasa.gov/. (Data retrieved 2024. Product: DAILY; Temporal: 2016-01-01 to 2025-12-31; Parameters: T2M_MIN, T2M_MAX, RH2M, ALLSKY_SFC_SW_DWN, PRECTOTCORR)

Open-Meteo contributors. (2024). *Open-Meteo: Free Weather API for Non-Commercial Use.* https://open-meteo.com/. (Accessed: 2024–2026. Provides NWP forecasts up to 16 days for any global coordinate)

Ortega, J. H. G., & Rodríguez, L. M. S. (2006). Aplicación de tecnologías de información geográfica para el estudio de la variabilidad climática en la cuenca alta del río cauca.

Pavani, J., Bastos, L., & Moraga, P. (2023). Joint spatial modeling of the risks of co-circulating mosquito-borne diseases in Ceará, Brazil. *Spatial and Spatio-temporal Epidemiology*.

Primo, C. e. a. (2018). Modeling daily precipitation in Catalonia (NE Spain) using non-stationary statistical downscaling techniques. *Theoretical and Applied Climatology*, *132*(1-2), 1–16.

Ramsay, J., & Silverman, B. (2006). *Functional data analysis* (2nd ed.). Springer.

UESVALLE. (2016). *Ficha tecnica.* Departamento del Valle del Cauca. (Recuperado de http://www.uesvalle.gov.co/publicaciones/237/valle-del-cauca/)

Volkmann, A., Umlauf, N., & Greven, S. (2024). Generalized Multivariate Functional Additive Mixed Models for Location, Scale, and Shape. *arXiv preprint arXiv:2407.20995*. Available online: https://arxiv.org/abs/2407.20995 (accessed on).

Wood, S. N. (2017). *Generalized additive models: An introduction with R* (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781315370279.