https://doi.org/10.5194/egusphere-2025-5543
Preprint. Discussion started: 9 February 2026

Characterizing Climate-Driven Shifts in Chilean Rainfall Regimes with a Hybrid Hidden Markov–Copula Framework

Mauricio Herrera-Marín[1], Francisca Kleisinger[1], Diego Rivera[1], Andrés Wilson[1], and Alex Godoy-Faundez[1]

[1]Research Center on Sustainability and Strategic Resource Management (CISGER). Faculty of Engineering, Universidad del Desarrollo (UDD). Avda. Plaza 680. Santiago. Chile

Correspondence: Mauricio Herrera-Marín (mherrera@udd.cl)

Abstract. Chile's hydroclimate exhibits pronounced meridional gradients and strong interannual variability, posing persistent challenges for regime-aware, probabilistic rainfall prediction. We introduce a hierarchical framework that explicitly separates large-scale regime dynamics from local spatial dependence. The approach integrates: (i) a covariate-driven non-homogeneous Hidden Markov Model (nHMM) to learn synoptic precipitation regimes and their transitions; (ii) Dynamic Time Warping (DTW) clustering to delineate precipitation-coherent climatic zones; and (iii) state-conditional Regular Vine copulas with Generalized Pareto (GPD) tails to model residual spatial dependence and extremes. The analysis employs the 0.05° daily CR2MET precipitation product over continental Chile (462 grid points, May–August 1980–2021) together with large-scale atmospheric covariates including the Southern Oscillation Index (SOI), the Oceanic Niño Index (ONI), and Global Mean Sea-Surface Temperature (GMSST).

Five physically consistent rainfall regimes emerge, spanning from an anticyclonic dry state to a cyclonic wet state, confirmed by composites of mean sea-level pressure, 850-hPa winds, and 500-hPa geopotential height. Mixed-effects inference on the transition matrix reveals a statistically significant decline in wet-state persistence of $\sim$0.34 % yr^{-1} ($\approx$14.5 % over 1980–2022), coincident with rising GMSST. Out-of-sample ensembles for 2022 (100 daily members conditioned on Viterbi states) are well calibrated: central 90 % prediction intervals achieve near-nominal coverage, low asymmetry, and widths increasing southward with climatological variance.

By disentangling regime timing and drivers from residual spatial co-variability and extremes, the proposed nHMM – DTW – vine – GPD framework yields meteorologically coherent states, spatially consistent probabilistic simulations, and quantitatively validated forecasts. The method is computationally tractable and transferable, offering a principled pathway for regime-conditioned, uncertainty-aware precipitation prediction to support hydroclimate risk management in Chile and other topographically complex regions.

1 Introduction

Understanding precipitation variability and its long-term modulation is a central challenge in hydroclimatology, particularly in regions of pronounced climatic heterogeneity such as Chile (Sarricolea et al., 2017). The country spans an extreme meridional

https://doi.org/10.5194/egusphere-2025-5543
Preprint. Discussion started: 9 February 2026

gradient—from hyperarid deserts in the north (18–26°S) to temperate and subpolar climates in the south (44–54°S)—that produces sharp transitions in precipitation regimes and associated hydrological processes. Rainfall variability in Chile operates across multiple temporal scales, including seasonal (Montecinos and Aceituno, 2003; Garreaud et al., 2009), interannual (Garreaud et al., 2017; Montecinos et al., 2011), and decadal to multidecadal fluctuations (Garreaud, 2009; Latoja et al., 2024). Among the large-scale climate drivers, the El Niño–Southern Oscillation (ENSO) dominates interannual variability, exerting strong control on regional circulation and precipitation anomalies. Warm (El Niño) and cold (La Niña) phases modulate storm-track position, jet strength, and moisture transport toward subtropical South America (Trenberth, 2014; Robertson et al., 2004a; Zhang et al., 2021). These effects are typically represented through indices such as the Southern Oscillation Index (SOI), the Oceanic Niño Index (ONI), and the Global Mean Sea Surface Temperature (GMSST), which encapsulate large-scale forcing but often fail to explain the strong spatial gradients in rainfall induced by orography, mesoscale dynamics, and land–atmosphere coupling (Vrac et al., 2012; Bárdossy and Pegram, 2010).

The persistence of extreme drought conditions—particularly the decade-long "megadrought" affecting central Chile since 2010 (Garreaud et al., 2020a; Bozkurt et al., 2021)—underscores the urgency of developing analytical frameworks that integrate both large-scale climate drivers and localized hydrological responses. Annual precipitation deficits between 15 % and 45 % relative to the pre-2010 baseline (Boisier et al., 2018), combined with regional warming and snowpack decline (Burger et al., 2018; Barría et al., 2019), have exacerbated water scarcity and altered runoff regimes. These changes highlight the limitations of traditional statistical models, which often assume linearity or stationarity in the relationship between atmospheric forcing and precipitation (Charles et al., 2014; Wilks, 2011). Capturing hidden or emergent shifts in rainfall regimes thus requires probabilistic frameworks capable of representing temporal evolution, regime persistence, and non-linear teleconnections.

In this context, we employ a Non-Homogeneous Hidden Markov Model (nHMM) to identify latent rainfall regimes and characterize their temporal dynamics. Hidden Markov Models (HMMs) have been extensively used to represent atmospheric circulation regimes and daily rainfall sequences (Ailliot et al., 2009; Franzke and Osprey, 2014). The non-homogeneous extension allows transition probabilities to depend explicitly on time-varying covariates such as ENSO indices or large-scale circulation fields (Hughes et al., 1999; Bellone et al., 2000; Robertson et al., 2004b), thus enabling the detection of evolving regime behavior under changing climatic conditions. Beyond identifying discrete rainfall states, we assess their long-term persistence and transition probabilities using linear mixed models, quantifying statistically significant trends that reflect climate-induced shifts in regime stability.

However, a key limitation of conventional nHMM applications lies in their assumption of conditional spatial independence—i.e., that rainfall at different locations is independent once the hidden state is known. This simplification is rarely valid in complex terrains like Chile, where strong orographic forcing and mesoscale circulations generate residual spatial dependence even after conditioning on large-scale atmospheric states. To address this issue, we develop a hybrid *nHMM–Copula* framework that combines the temporal structure and climatic interpretability of nHMMs with explicit spatial modeling via multivariate copulas. The approach begins by partitioning Chile into homogeneous precipitation zones using Dynamic Time Warping (DTW) clustering (Berndt and Clifford, 1994; Herrera et al., 2025), ensuring that grouped stations share coherent rainfall dynamics (Maharaj and Alonso, 2011; Montero and Vilar, 2015). Within each zone, we estimate state-specific cop-

https://doi.org/10.5194/egusphere-2025-5543
Preprint. Discussion started: 9 February 2026

ula models to reconstruct spatial dependence, thereby relaxing the restrictive independence assumption while preserving the probabilistic and physically interpretable structure of the HMM (Bárdossy and Pegram, 2010; Vrac et al., 2012).

Previous studies have independently applied HMMs and copula models to precipitation (Ailliot et al., 2009; Vrac et al., 2012), yet few have integrated spatial dependence within a covariate-driven nHMM framework. The novelty of this study lies in bridging these two perspectives through a hybrid regime-based and dependence-aware modeling system. By explicitly linking hidden rainfall regimes to large-scale atmospheric drivers while reconstructing state-conditioned spatial dependencies, our framework enables spatially coherent probabilistic simulations of precipitation fields that are consistent with both physical mechanisms and statistical constraints. This approach provides an interpretable, transferable, and uncertainty-aware basis for rainfall regime analysis and probabilistic forecasting across Chile's diverse hydroclimatic zones.

The remainder of this paper is organized as follows. Section 2 describes the data and methods, including the CR2MET precipitation dataset, the nHMM formulation, the DTW-based climatic zoning, and the copula construction. Section 3 presents the resulting rainfall regimes, their atmospheric interpretation, temporal trends, and spatially coherent probabilistic simulations. Finally, Section 4 discusses the methodological advances, implications for hydroclimatic predictability, and potential extensions for coupled hydroclimate–impact modeling.

2 Data and Methods

2.1 Data

Daily precipitation in Chile is monitored by official agencies (CR2, 2024), yet many ground stations exhibit discontinuities, missing data, or record lengths shorter than 30 years, limiting their suitability for robust climate diagnostics. To ensure spatial completeness and temporal consistency, we employed the *CR2MET* dataset (Boisier, 2023; Boisier et al., 2018), a high-resolution ($0.05°$) gridded product developed specifically for continental Chile and rigorously validated against dense observational networks. CR2MET integrates local station observations with global reanalyses—primarily ERA5 (Hersbach et al., 2023)—using statistical downscaling techniques to better capture precipitation over complex orography. While ERA5 provides a physically consistent global depiction of atmospheric conditions, the regional refinement in CR2MET significantly improves the realism of daily rainfall estimates across Chile, particularly at the spatial scales relevant for hydrological and climate-impact studies (Mayer et al., 2021; Cavalleri et al., 2024).

From CR2MET, we extracted 462 grid points collocated with or proximal to operational rain gauges exhibiting uninterrupted daily records during 1980–2021. Each grid point contributes $123 \times 42 = 5166$ daily winter observations, where the austral winter season (May–August; MJJA) was isolated for analysis. This period was selected because frontal systems during MJJA account for approximately 70–90 % of annual rainfall between $30°$S and $44°$S (Pica-Téllez et al., 2020), dominating streamflow generation, groundwater recharge, and snow accumulation in the Andes, and therefore conditioning water availability during spring and summer (Valdés-Pineda et al., 2016).

The fine spatial resolution ($0.05°$, $\sim$5 km) ensures that the 462 selected grid cells are distinct and adequately resolve sub-regional precipitation variability. This granularity is crucial in a country characterized by strong climatic and orographic gra-

https://doi.org/10.5194/egusphere-2025-5543
Preprint. Discussion started: 9 February 2026

dients, where coarse global reanalyses often underestimate variability and spatial dependence. In contrast, CR2MET captures both the magnitude and spatial organization of precipitation across coastal, valley, and mountain sectors, providing a robust basis for identifying rainfall regimes and their extremes.

Climatological controls on winter precipitation are consistent with established large-scale dynamics over the southeastern Pacific. Seasonal totals are modulated primarily by the South Pacific Subtropical High and the Southern Hemisphere storm track (Hoskins and Hodges, 2005; Barrett and Hameed, 2017). Orographic enhancement along the Andes intensifies rainfall from frontal systems (Garreaud et al., 2009; Arias et al., 2021), while persistent blocking or reduced frontal incursions can produce substantial deficits (Latoja et al., 2024), even during ENSO-neutral conditions (Montecinos et al., 2011). These processes are accurately represented in CR2MET, which thus provides a consistent observational foundation for both the temporal regime identification via the nHMM and the spatial dependence analysis through copula modeling.

To represent large-scale atmospheric variability influencing daily precipitation, we incorporated multiple climatic covariates into the nHMM transition matrix, including geopotential height at 500 hPa, mean sea-level pressure, sea surface temperature, evapotranspiration, maximum and mean near-surface temperature, and zonal–meridional wind components (u, v). These fields were extracted from ERA5 reanalysis (Hersbach et al., 2023) at $0.25° \times 0.25°$ resolution, covering both continental Chile and adjacent oceanic regions (approximately $15°$N–$60°$S, $120°$W–$60°$W). The inclusion of oceanic domains is essential given the strong influence of the South Pacific High, storm-track dynamics, and sea–atmosphere interactions on winter precipitation over central Chile. To reduce dimensionality and multicollinearity among predictors, we applied Principal Component Analysis (PCA) to each gridded covariate field and retained the leading components explaining more than 80 % of total variance. These principal components were then used as exogenous predictors in the nHMM transition model.

2.2 Non-Homogeneous Hidden Markov Model (nHMM)

The Non-Homogeneous Hidden Markov Model (nHMM) (Khalil et al., 2010; Kirshner, 2005; Cioffi et al., 2020; Hughes et al., 1999; Robertson et al., 2004b) provides a probabilistic framework to identify latent (hidden) climate regimes and model their temporal evolution in daily rainfall sequences. The model represents precipitation as a stochastic process governed by unobserved states that capture large-scale atmospheric configurations. Transitions between these states follow the Markov property, where the probability of a state on a given day depends solely on the state of the preceding day. The nHMM extends the conventional (homogeneous) HMM by allowing transition probabilities to vary in time through dependence on external covariates, thereby linking local precipitation to evolving large-scale climatic drivers.

2.2.1 Model structure.

Let $R_{t_i}^{r_j}$ denote the daily precipitation at location $r_j = (\text{long}_j, \text{lat}_j)$, where long and lat are longitude and latitude, respectively, and let $i = 1, \ldots, n$ and $j = 1, \ldots, q$. In this study, $n = 5166$ corresponds to 42 winters (1980–2021) with 123 days each (May–August; MJJA), and $q = 462$ spatial locations. The year 2022 is reserved for model evaluation. We define $R_{t_i} = (R_{t_i}^{r_1}, \ldots, R_{t_i}^{r_q})$ as the q-dimensional precipitation vector at time t_i, and $R_{1:n} = (R_{t_1}, \ldots, R_{t_n})$ as the sequence of daily ob-

https://doi.org/10.5194/egusphere-2025-5543
Preprint. Discussion started: 9 February 2026

servations. The temporal dependence is governed by a latent (unobserved) Markov chain $S_t = (S_1, \ldots, S_n)$ taking values in $\{1, \ldots, K\}$, where K is the optimal number of hidden states estimated from data.

The nHMM defines a joint probability distribution over the observed rainfall sequence $R_{1:n}$ and the hidden state sequence $S_{1:n}$ through a set of conditional independence assumptions:

Assumption 1 (Markovian dynamics). The hidden state sequence $S_{1:n}$ is a first-order Markov process:

$$P(S_1, S_2, \ldots, S_n) = p(S_1) \prod_{t=2}^{n} p(S_t \mid S_{t-1}),$$

where $p(S_1)$ is the initial state probability and $p(S_t \mid S_{t-1})$ defines the transition matrix between states.

Assumption 2 (Conditional dependence of rainfall on the hidden state). The precipitation vector on day t, R_t, depends only on the hidden state on that day:

$$p(R_{1:n}, S_{1:n}) = p(S_1) \prod_{t=2}^{n} p(S_t \mid S_{t-1}) \prod_{t=1}^{n} p(R_t \mid S_t).$$

Assumption 3 (Conditional spatial independence). Conditional on the hidden state S_t, precipitation across stations is assumed independent:

$$p(R_t \mid S_t) = \prod_{m=1}^{q} p(R_t^m \mid S_t).$$

Each term $p(R_t^m \mid S_t = i)$ is modeled as a mixed distribution that accounts for dry and wet days:

$$p(r_t \mid S_t = i) = \prod_{m=1}^{q} \delta_{im},$$

with

$$\delta_{im} = \begin{cases} p_{im0}, & r_t^m = 0, \\ \sum_{c=1}^{C-1} \dfrac{p_{im1}\, \phi_{imc}^{\omega_{imc}} (r_t^m)^{\omega_{imc}-1} e^{-\phi_{imc} r_t^m}}{\Gamma(\omega_{imc})}, & r_t^m > 0, \end{cases}$$

where p_{im0} is the probability of no precipitation for state i at station m, $p_{im1} = 1 - p_{im0}$ is the complementary probability of rainfall, and $(\omega_{imc}, \phi_{imc})$ are the shape and rate parameters of the Gamma distribution for the c-th mixture component ($C-1$ components in total). This zero-inflated mixture model flexibly captures both dry-day frequency and rainfall intensity across stations.

Assumption 4 (Non-homogeneity and climatic covariates). To account for nonstationary behavior, transition probabilities between hidden states are allowed to vary as a function of exogenous atmospheric variables. Let $X_t = (X_t^1, X_t^2, \ldots, X_t^h)$ denote a vector of h climatic covariates at time t (e.g., geopotential height, mean sea-level pressure, wind components, temperature, and SST indices). The non-homogeneous formulation links large-scale atmospheric dynamics to local rainfall through:

$$p(R_t \mid S_{1:n}, R_{t-1}, X_{1:n}) = p(R_t \mid S_t)\, p(S_t \mid S_{t-1}, X_t),$$

https://doi.org/10.5194/egusphere-2025-5543
Preprint. Discussion started: 9 February 2026

where the second term, $p(S_t \mid S_{t-1}, X_t)$, introduces time-varying transition probabilities modeled via a multinomial logistic regression:

$$p(S_t = i \mid S_{t-1} = j, X_t = x) = \frac{\exp(\alpha_{j,i} + \beta_i^T x)}{\sum_{k=1}^{K} \exp(\alpha_{j,k} + \beta_k^T x)}.$$

Here, $\alpha_{j,i}$ represents the baseline log-odds of transitioning from state j to state i, while β_i is a vector of regression coefficients quantifying the influence of the covariates X_t on that transition. This formulation enables the model to capture climate-driven shifts in the frequency and persistence of rainfall regimes.

2.2.2 Estimation and inference.

Direct computation of the full joint likelihood over all possible hidden state sequences is computationally infeasible. Instead, parameter estimation proceeds via the forward–backward (Baum–Welch) algorithm, a variant of the Expectation–Maximization (EM) procedure (McLachlan and Krishnan, 2008). This iterative algorithm alternates between (i) estimating the posterior probabilities of hidden states given current parameters and data (E-step), and (ii) updating parameters to maximize the expected complete-data log-likelihood (M-step). We employ the implementation proposed by Kirshner (2005) and further adapted by Robertson et al. (2006), which efficiently handles covariate-dependent transitions in non-homogeneous HMMs.

Overall, the nHMM provides a flexible probabilistic representation of rainfall regimes, capturing both the stochastic nature of daily precipitation and the modulation of regime transitions by evolving large-scale atmospheric drivers.

2.3 Hybrid nHMM–Copula Framework with DTW–Based Zoning

The non-homogeneous Hidden Markov Model (nHMM) described in Sect. 2.2 provides a coherent mechanism to extract hidden rainfall regimes $\{S_t\}$ and link their transitions to large-scale covariates $\{X_t\}$. However, Assumption 3 (conditional spatial independence given S_t) may be too restrictive in regions with strong orographic gradients and mesoscale circulations such as Chile, where residual spatial dependence often persists even after conditioning on the regime (Becker and Schmitz, 2003; Garreaud, 2009; Pablo Sarricolea and Óliver Meseguer-Ruiz, 2017). We therefore propose a hierarchical framework in which the nHMM governs temporal regime dynamics, while state-conditioned spatial dependence is reconstructed via multivariate copulas within climatically homogeneous zones.

2.3.1 Zoning via DTW

Let $R_t = (R_t^{r_1}, \ldots, R_t^{r_q})$ denote the daily precipitation vector at $q = 462$ locations and $S_t \in \{1, \ldots, K\}$ the nHMM regime at time t. A direct q-variate copula is computationally prohibitive. We partition the network into $k^* = 7$ zones $\mathcal{Z} = \{1, \ldots, 7\}$, each with d_z stations, using hierarchical clustering on pairwise Dynamic Time Warping (DTW) dissimilarities computed from standardized MJJA daily series (Berndt and Clifford, 1994; Giorgino, 2009; Montero and Vilar, 2015).

For two stations i and j, DTW defines the alignment cost

$$\mathrm{DTW}(i,j) = \min_{\pi \in \Pi_{ij}} \sum_{(\tau,\kappa)\in\pi} \big\| y_i(\tau) - y_j(\kappa) \big\|,$$

https://doi.org/10.5194/egusphere-2025-5543
Preprint. Discussion started: 9 February 2026

where Π_{ij} is the set of monotone warping paths mapping the time indices of y_i to those of y_j. Ward's linkage (Ward, 1963) on the resulting distance matrix yields compact, interpretable clusters. The final partition (Fig. 6A) aligns with established climatic subregions (Pablo Sarricolea and Óliver Meseguer-Ruiz, 2017; Garreaud et al., 2009) and constrains copula modeling to tractable dimensions $d_z \in [50, 90]$.

2.3.2 State–conditional marginal specification

Fix a zone $z \in \mathcal{Z}$ and a regime $s \in \{1, \dots, K\}$. For station $m \in z$, let

$$F_{m,s}(x) \;=\; \mathbb{P}(R_t^m \le x \mid S_t = s)\,, \qquad f_{m,s}(x) \;=\; \frac{\mathrm{d}}{\mathrm{d}x} F_{m,s}(x),$$

be the state-conditioned CDF and density, estimated from the subset $\{R_t^m : S_t = s\}$.[1] To respect the nHMM emissions and explicitly address extremes, we use a hybrid marginal:

1. A point mass at zero with probability $p_{m,s}^{(0)} = \mathbb{P}(R_t^m = 0 \mid S_t = s)$.
2. A $(C-1)$-component Gamma mixture for positive amounts, with weights $w_{m,s,c}$, shapes $\omega_{m,s,c}$ and rates $\phi_{m,s,c}$, $c = 1, \dots, C-1$:

$$g_{m,s}(x) = \sum_{c=1}^{C-1} w_{m,s,c} \, \frac{\phi_{m,s,c}^{\omega_{m,s,c}} \, x^{\omega_{m,s,c}-1} e^{-\phi_{m,s,c} x}}{\Gamma(\omega_{m,s,c})}, \quad x > 0.$$

3. A Generalized Pareto Distribution (GPD) for the upper tail, fitted to exceedances $Y = x - u_{m,s}$ above the $u_{m,s}$-quantile ($u_{m,s}$ chosen at the empirical 0.90 level), with shape $\xi_{m,s}$ and scale $\sigma_{m,s} > 0$:

$$\mathbb{P}(Y \le y) = 1 - \left(1 + \frac{\xi_{m,s} y}{\sigma_{m,s}}\right)_+^{-1/\xi_{m,s}}, \quad y \ge 0.$$

We enforce continuity of $F_{m,s}$ at the threshold $u_{m,s}$ by stitching the body and tail with a probability-splicing construction:

$$F_{m,s}(x) = \begin{cases} p_{m,s}^{(0)} + \big(1 - p_{m,s}^{(0)}\big) \displaystyle\int_0^x g_{m,s}(y)\,\mathrm{d}y, & 0 \le x \le u_{m,s}, \\ F_{m,s}(u_{m,s}) + \big(1 - F_{m,s}(u_{m,s})\big)\, G_{m,s}^{\mathrm{GPD}}(x - u_{m,s}), & x > u_{m,s}, \end{cases}$$

where $G_{m,s}^{\mathrm{GPD}}$ is the GPD CDF. This hybrid specification (Hughes et al., 1999; Khalil et al., 2010; Zhang et al., 2021; Bárdossy and Pegram, 2010; Vrac et al., 2012) preserves the nHMM emission structure for the bulk while representing state-specific heavy tails relevant to hydroclimatic risk in central Chile (Garreaud et al., 2020a; Salazar et al., 2024). Tail parameter stability is checked via threshold sensitivity (0.85–0.95 quantiles) and mean-excess diagnostics (see Sect. 2.4).

[1] In practice we use the Viterbi-decoded states for simulation; Section 2.4 reports robustness under smoothed state probabilities.

https://doi.org/10.5194/egusphere-2025-5543
Preprint. Discussion started: 9 February 2026

2.3.3 Transformation to copula space

Define pseudo-observations (state-conditional PITs) for zone z and regime s:

$$U_t^{(m)} = F_{m,s}\big(R_t^m\big), \qquad m = 1, \ldots, d_z,$$

and $\mathbf{U}_t^z = \big(U_t^{(1)}, \ldots, U_t^{(d_z)}\big) \in [0,1]^{d_z}$. By Sklar's theorem (Nelsen, 2006; Genest and MacKay, 1986), the conditional joint law of $\mathbf{R}_t^z$ factorizes as

$$F_{z,s}(\mathbf{r}) = C_{z,s}\Big(F_{1,s}(r_1), \ldots, F_{d_z,s}(r_{d_z})\Big), \qquad f_{z,s}(\mathbf{r}) = c_{z,s}(\mathbf{u}) \prod_{m=1}^{d_z} f_{m,s}(r_m),$$

where $C_{z,s}$ is the copula of $(R_t^{(1)}, \ldots, R_t^{(d_z)})$ given $S_t = s$ and $c_{z,s}$ its density. Thus, marginal behavior is inherited from Sect. 2.3.2 and residual spatial dependence is entirely encoded by $c_{z,s}$.

2.3.4 Regular vine construction, simplifying assumption, and model selection

For d_z up to ∼90, we represent $c_{z,s}$ by a Regular Vine (R-vine) (Aas et al., 2009), decomposing the d_z-variate density into bivariate pair-copulas along a sequence of linked trees:

$$c_{z,s}(u_1, \ldots, u_{d_z}) = \prod_{m=1}^{d_z - 1} \prod_{e \in E_m} c_e\Big(u_{i|D}, u_{j|D}; \theta_e\Big),$$

where E_m denotes edges at tree level m, $c_e(\cdot, \cdot; \theta_e)$ is a bivariate copula family with parameter θ_e, and $u_{i|D}$ denotes conditional distribution functions given the conditioning set D determined by the vine structure. We adopt the simplifying assumption (pair-copula parameters do not depend on the values of conditioning variables), standard in high-dimensional applications and empirically adequate in our context (diagnostics in Sect. 2.4).

We allow a candidate family set $\mathcal{F} = \{\text{Gaussian}, \text{Student-}t, \text{Clayton}, \text{Gumbel}, \text{Frank}\}$, covering elliptical and Archimedean types with different tail properties. Structure selection and family assignment are performed edge-wise using likelihood-based criteria (AIC/BIC) as implemented by `RVineStructureSelect` in the `VineCopula R` package (Schepsmeier et al., 2015). Parameters are estimated by maximum likelihood. Multiple random initializations and, when beneficial, truncated vines are considered to balance fit and parsimony. In practice, lower-level trees capture the strongest spatial links (e.g., nearby stations), where Student-t or asymmetric Archimedean families may be selected; higher trees often favor Gaussian pairs due to attenuation of conditional dependence beyond ∼5–10 variables, consistent with Bárdossy and Pegram (2010).

2.3.5 State–aware simulation and back–transformation

Given a Viterbi-decoded regime $\widehat{S}_t$ for day t and zone z, we generate an ensemble member as follows:

1. Draw $\mathbf{U}_t^z \sim C_{z,\widehat{S}_t}$ using `RVineSim`.

https://doi.org/10.5194/egusphere-2025-5543
Preprint. Discussion started: 9 February 2026

2. Map back to precipitation using the stitched marginal inverse:

$$R_t^{(m)} = \begin{cases} F^{-1}_{m,\widehat{S}_t}\left(U_t^{(m)}\right), & U_t^{(m)} \le u^*_{m,\widehat{S}_t}, \\ F^{-1}_{m,\widehat{S}_t}\left(u^*_{m,\widehat{S}_t}\right) + Q^{-1}_{\mathrm{GPD},m}\left(\dfrac{U_t^{(m)} - u^*_{m,\widehat{S}_t}}{1 - u^*_{m,\widehat{S}_t}}\right), & U_t^{(m)} > u^*_{m,\widehat{S}_t}, \end{cases}$$

where $u^*_{m,\widehat{S}_t}$ is the tail threshold quantile and $Q^{-1}_{\mathrm{GPD},m}$ the GPD quantile function for station m. This preserves both central and tail behavior (Trenberth, 2014; Franzke and Osprey, 2014).

2.3.6 Computational aspects and reproducibility

The full system entails 7 zones $\times$ K=5 regimes = 35 copula models $\{C_{z,s}\}$, each with state-conditioned marginals $\{F_{m,s}\}$ for d_z stations. Estimation and simulation are parallelized across (z, s) using `parallel` and `doParallel` in `R`. Vines are fitted with `VineCopula` (Schepsmeier et al., 2015); clustering uses `TSclust` (Montero and Vilar, 2015); GPD fitting uses `fExtremes` and `evd`. The training set comprises $n = 5166$ MJJA days (1980–2021) and the 2022 season is reserved for out-of-sample evaluation.

2.3.7 Uncertainty quantification and summary metrics

From M ensemble members, define for each day t the central 90% prediction interval $[\widehat{L}_t, \widehat{U}_t]$ (empirical 5th and 95th percentiles) and the ensemble mean $\widehat{X}_t$. Let X_t^{obs} be the observation. We report:

$$\mathrm{AW} := \frac{1}{T}\sum_{t=1}^{T}\left(\widehat{U}_t - \widehat{L}_t\right), \qquad \text{(Average Width; sharpness)}$$

$$\mathrm{POC} := \frac{1}{T}\sum_{t=1}^{T}\mathbf{1}\left\{X_t^{\mathrm{obs}} \in [\widehat{L}_t, \widehat{U}_t]\right\}, \qquad \text{(Prediction-interval coverage; calibration)}$$

$$\mathrm{AAD}_t := \begin{cases} \dfrac{\widehat{L}_t - X_t^{\mathrm{obs}}}{\widehat{U}_t - \widehat{L}_t}, & X_t^{\mathrm{obs}} < \widehat{L}_t, \\ \dfrac{X_t^{\mathrm{obs}} - \widehat{U}_t}{\widehat{U}_t - \widehat{L}_t}, & X_t^{\mathrm{obs}} > \widehat{U}_t, \\ 0, & X_t^{\mathrm{obs}} \in [\widehat{L}_t, \widehat{U}_t], \end{cases} \qquad \mathrm{AAD} := \frac{1}{T}\sum_{t=1}^{T}\mathrm{AAD}_t, \qquad \text{(Average Asymmetry Degree).}$$

Proper scoring rules (CRPS, energy and variogram scores) and dependence diagnostics (Rosenblatt transform, tail-coincidence rates) are detailed in Sect. 2.4.

2.4 Model validation: state dynamics, residual spatial dependence, and predictive skill

We assess model adequacy at two complementary levels: (i) the temporal regime structure encoded by the non-homogeneous HMM (nHMM), which links $\{S_t\}$ to large-scale covariates and defines one-step-ahead predictive laws at each site; and (ii) the residual *state-conditioned* spatial dependence within DTW-defined zones, represented by vine copulas. We then verify out-of-sample probabilistic skill using strictly proper scores. Diagnostics, tests and scoring rules are described below.

https://doi.org/10.5194/egusphere-2025-5543
Preprint. Discussion started: 9 February 2026

2.4.1 nHMM fit and calibration

Let R_t^m denote daily precipitation at site $m \in \{1,\dots,q\}$ and $S_t \in \{1,\dots,K\}$ the latent regime. The nHMM yields the one-step predictive distribution

$$\mathcal{L}(R_t^m \mid \mathcal{F}_{t-1}) \;=\; \sum_{s=1}^{K} \Pr(S_t = s \mid \mathcal{F}_{t-1})\, \mathcal{L}(R_t^m \mid S_t = s),$$

where $\mathcal{F}_{t-1} = \sigma\{R_{1:t-1}, X_{1:t}\}$ and the transition law depends on X_t via the multinomial logit (Sect. 2.2). Denote $F_{t,m}$ the corresponding predictive CDF. Probability–integral–transform (PIT) residuals $U_{t,m} = F_{t,m}(R_t^m)$ should be i.i.d. $\mathrm{Unif}(0,1)$ under correct specification (Dawid, 1984; Gneiting and Raftery, 2007). We examine PIT histograms and apply uniformity tests (Kolmogorov–Smirnov, Cramér–von Mises), together with residual autocorrelation (Ljung–Box) to detect temporal under/overdispersion (Supplement S1).

Because emissions are mixed (point mass at zero, Gamma mixtures for the body, GPD for the upper tail), we validate occurrence and intensity separately. For occurrence, we verify probabilistic wet/dry forecasts using Brier scores and reliability diagrams (in-sample and rolling-origin). For positive amounts, we use state-specific QQ-plots and distributional tests at each site; tail diagnostics include mean-excess plots and confidence intervals for the GPD shape above the threshold u^*. As a state-aware check, we compute smoothed regime probabilities $\Pr(S_t = s \mid R_{1:n}, X_{1:n})$ and form conditional PITs using $\sum_s \Pr(S_t = s \mid R_{1:n}, X_{1:n})\, F_{m,s}(\cdot)$, which isolates misspecification of $\mathcal{L}(R_t^m \mid S_t = s)$ from transition errors (Zucchini et al., 2016). Model selection across K relies on maximized log-likelihood and AIC/BIC with stability checks across random initializations.

2.4.2 Vine–copula fit: residual spatial dependence

Within zone z and regime s, let $U_t^{(m)} = F_{m,s}(R_t^m)$ and $\mathbf{U}_t^z = (U_t^{(1)}, \dots, U_t^{(d_z)})$. The fitted vine $C_{z,s}$ implies density $c_{z,s}$ and a Rosenblatt transform $\mathcal{R}_{z,s}$ mapping $\mathbf{U}_t^z$ to $\mathbf{V}_t^z = \mathcal{R}_{z,s}(\mathbf{U}_t^z)$, which should be i.i.d. $\mathrm{Unif}(0,1)$ components under correct specification (Rosenblatt, 1952). We test marginal uniformity (KS/CvM), serial independence (runs, Ljung–Box), and joint adequacy via energy-distance tests against $\mathrm{Unif}([0,1]^{d_z})$ (Székely and Rizzo, 2013). At the pair level, we apply parametric-bootstrap goodness-of-fit for edge copulas c_e (Cramér–von Mises statistics for the empirical copula process) and compare empirical vs model-implied Kendall's τ and tail-dependence coefficients (Genest and Rémillard, 2008; Genest et al., 2009).

To diagnose spatial coherence, we compare empirical and model-implied dependence summaries: (i) tail-coincidence rates $\hat{\lambda}_U(q) = \Pr(U_i > q \mid U_j > q)$ at high q (e.g. 0.95), particularly along low-order edges where proximity is strongest; and (ii) variogram functionals $\gamma_\alpha(\mathbf{u}) = \sum_{i<j} w_{ij} |u_i - u_j|^\alpha$ with $\alpha \in (0,2]$, contrasted with their model expectations under $C_{z,s}$. Competing vine structures (C-, D-, R-vines), truncation levels, and family sets are compared by AIC/BIC; when fits are close, Vuong's test assesses relative performance (Vuong, 1989). Where dependence in higher-order trees is negligible, truncated vines are preferred for parsimony (Sect. 2.3).

https://doi.org/10.5194/egusphere-2025-5543
Preprint. Discussion started: 9 February 2026

2.4.3 Out-of-sample probabilistic verification

Because the goal is probabilistic forecasting, we use strictly proper scoring rules (Gneiting and Raftery, 2007). For a univariate predictive CDF F and realization y, the continuous ranked probability score (CRPS) is

$$\mathrm{CRPS}(F, y) = \int_{-\infty}^{\infty} \big(F(x) - \mathbf{1}\{x \geq y\}\big)^2 \, \mathrm{d}x,$$

estimated from an ensemble $\{Y^{(m)}\}_{m=1}^{M}$ via

$$\widehat{\mathrm{CRPS}} = \frac{1}{M} \sum_{m=1}^{M} |Y^{(m)} - y| - \frac{1}{2M^2} \sum_{m=1}^{M} \sum_{m'=1}^{M} |Y^{(m)} - Y^{(m')}|.$$

Exceedance events $\{R_t^m > r^\star\}$ are evaluated with Brier scores and reliability diagrams. At the multivariate (zone) level, we compute the energy score (ES) and variogram score (VS) with order $\alpha \in (0, 2]$ and weights w_{ij}:

$$\mathrm{ES}(F, \mathbf{y}) = \mathbb{E}\|\mathbf{X} - \mathbf{y}\| - \tfrac{1}{2}\mathbb{E}\|\mathbf{X} - \mathbf{X}'\|, \quad \mathrm{VS}_\alpha(F, \mathbf{y}) = \sum_{i<j} w_{ij} \Big(|y_i - y_j|^\alpha - \mathbb{E}|X_i - X_j|^\alpha \Big)^2,$$

with ensemble approximations. ES is sensitive to overall multivariate location/scale, whereas VS is explicitly diagnostic of *dependence* misspecification (Scheuerer and Hamill, 2015). We report station-wise CRPS and zone-wise (ES, VS) in rolling-origin blocks (e.g., train 1980–2010, validate 2011–2015; then train 1980–2015, validate 2016–2021), respecting temporal dependence (Bergmeir and Benítez, 2012). Differences vs a baseline without copulas (conditionally independent emissions) are tested with Diebold–Mariano statistics using HAC standard errors adapted to block designs (Diebold and Mariano, 1995).

2.4.4 Sensitivity and robustness

We probe robustness to design choices: number of DTW zones k^*, vine structure and truncation, admissible pair families, tail threshold u^*, and (if used) state-weighting in copula fitting. We track the response of CRPS/ES/VS and interval diagnostics under controlled perturbations. The vine's simplifying assumption is examined by checking the stability of selected pair families and parameter estimates across conditioning values on low-order edges; systematic violations prompt restricted family sets or stronger truncation. Together, these analyses provide converging evidence that the nHMM emissions and the state-conditioned copula layer jointly yield a calibrated and sharp forecasting system, capturing both regime dynamics and residual spatial dependence.

2.4.5 Conceptual justification

A legitimate conceptual concern is the apparent tension between (i) assuming conditional spatial independence to estimate the nHMM and (ii) subsequently introducing spatial dependence via copulas. Our *hierarchical* design resolves this by separating **temporal-regime inference** (governed by large-scale drivers) from **state-conditioned spatial dependence** (governed by mesoscale/orographic controls):

https://doi.org/10.5194/egusphere-2025-5543
Preprint. Discussion started: 9 February 2026

1. **Scale separation and identifiability.** Regime transitions are primarily driven by synoptic–planetary variability (ENSO, storm-track, South Pacific High), whereas subdaily–mesoscale organization shapes spatial correlation given a regime. Estimating regimes without a high-dimensional spatial layer avoids confounding temporal state dynamics with local dependence and stabilizes identification of K, transition coefficients, and emission parameters.

2. **Modular likelihood and efficiency.** The nHMM likelihood exploits conditional factorization to recover regime dynamics and marginal emissions at scale ($q = 462$). The copula is then estimated on state-conditioned PITs, which, under correct marginal specification, provide valid inputs for dependence modeling. This two-stage strategy attains near–full-model performance with orders-of-magnitude lower computational cost than embedding a d_z-variate copula inside EM.

3. **Statistical adequacy ensured by diagnostics.** Any potential mismatch introduced by the modular approach is explicitly probed by: (i) PIT-based and tail diagnostics for emissions; (ii) Rosenblatt-transform tests, variogram/tail-coincidence checks, and VS for dependence; and (iii) strictly proper scores (CRPS/ES/VS) in time-blocked out-of-sample verification. Systematic deviations trigger model refinements (thresholds, vine truncation/families, zoning), preserving inferential transparency.

4. **Physical interpretability and transferability.** Regimes retain their climatological meaning (circulation-linked rainfall states), while the copula layer restores the spatial coherence consistent with orography and coastal–Andean contrasts. This decoupling facilitates transferring the framework across Chile's hydroclimatic zones and to other mountainous regions without reengineering the temporal core.

Empirically, the copula-adjusted simulations reproduce observed spatial coherence and extreme-rainfall behavior (Sect. 3), which is essential for hydrological risk assessment in central Chile (Garreaud et al., 2020a; Bozkurt et al., 2021). Thus, the proposed hierarchy is not a mere convenience but a principled, testable, and computationally tractable approximation to a fully joint spatiotemporal model, with demonstrable gains in calibration and sharpness over conditionally independent baselines.

2.5 Experimental setup, model selection, and validation (overview)

All analyses use daily CR2MET MJJA precipitation at $q = 462$ locations (grid points co-located with operational stations) for 1980–2021, yielding $n = 42 \times 123 = 5166$ days and $N_{\text{obs}} = n\,q = 2{,}386{,}692$ scalar observations (Sect. 2). Let $R_t = (R_t^{r_1}, \ldots, R_t^{r_q})$ denote the q-variate precipitation on day t, and $S_t \in \{1, \ldots, K\}$ the latent regime (Sect. 2.2). Transitions $p(S_t \mid S_{t-1}, X_t)$ depend on large-scale covariates X_t via the multinomial logit; emissions $p(R_t \mid S_t)$ follow the Bernoulli–mixture-of-Gamma model with an explicit GPD tail refinement (Sects. 2.2, 2.3).

2.5.1 Grid of nHMM configurations

We fitted a grid of non-homogeneous HMMs varying the number of regimes $K \in \{2, 3, 4, 5, 6\}$ and the number of Gamma components in the positive-amount mixture $C \in \{2, \ldots, 8\}$ (the point mass at zero is handled by the occurrence term). We

https://doi.org/10.5194/egusphere-2025-5543
Preprint. Discussion started: 9 February 2026

denote a configuration as $K\texttt{hs}C\texttt{dg}$. For each configuration, parameters θ maximize the incomplete-data likelihood

$$\ell(\theta) = \log p_\theta(R_{1:n}) = \log \sum_{S_{1:n}} p_\theta(S_{1:n}) \prod_{t=1}^{n} p_\theta(R_t \mid S_t),$$

via the Baum–Welch/EM algorithm (forward–backward E-step; M-step for emissions and transition logit; Sect. 2.2). Model selection relies on

$$\mathrm{AIC} = 2p - 2\ell(\hat{\theta}), \qquad \mathrm{BIC} = p \log N_{\mathrm{obs}} - 2\ell(\hat{\theta}),$$

where p counts regime priors, covariate coefficients in the transition kernel, and mixture parameters across sites and states. Table S1 reports $\ell(\hat{\theta})$, AIC, BIC, p, and a per-dimension log-score $\ell(\hat{\theta})/(nq)$. The BIC surface exhibits a clear elbow at $K = 5$: moving from $K = 4$ to $K = 5$ yields substantial likelihood gains that compensate the penalty, whereas $K = 6$ increases p sharply with negligible improvement and produces weakly separated regimes. Within $K = 5$, BIC attains its minimum at $C = 5$ (configuration `5hs5dg`, $\ell = -5{,}108{,}541.0$, AIC $= 10{,}275{,}650.1$, BIC $= 10{,}483{,}643.9$, $p = 29{,}284$). Larger C marginally improves ℓ but is penalized by BIC; smaller C underfits heavy tails. We therefore adopt $K = 5$, $C = 5$ for state interpretation and forecasting.

2.5.2 DTW-based zoning for dependence modeling

To represent subregional dependence at scale (Sect. 2.3), we partition the $q = 462$ series using Dynamic Time Warping (DTW) on standardized MJJA anomalies and Ward's hierarchical clustering. Candidate partitions use Ward.D and Ward.D^2 with $G \in \{5, \ldots, 10\}$ clusters. The number of zones is selected by a multi-criteria internal validation: Silhouette, Calinski–Harabasz, Davies–Bouldin (and DB^*), Dunn, and COP indices. Each index is z-scored across candidates and aggregated into an average rank. The best-performing setting is Ward.D with $G = 7$ clusters (entry `hierarchical_ward.D_7/config3_7`; Table S2), yielding spatially compact zones with $d_g \in [50, 90]$ sites that minimize cross-zone leakage while preserving sufficient dimensionality for robust copula estimation.

2.5.3 State-conditional vine copulas and simulation

Given the Viterbi path $\hat{S}_{1:n}$ from the selected nHMM, we model intra-zone dependence of daily totals *conditionally on regime* by Regular Vines (R-vines). For zone g with d_g sites, define state-conditional PIT variables

$$\mathbf{U}_t^{(g)} = \left(U_{t,1}^{(g)}, \ldots, U_{t,d_g}^{(g)}\right), \qquad U_{t,j}^{(g)} = F_{j,\hat{S}_t}\left(R_t^{(j)}\right),$$

where $F_{j,i}$ is the site–state marginal (Bernoulli–Gamma mixture for the body with GPD tail; Sect. 2.3, Supplement S3). For each pair (i, g) we select vine structure and pair families from $\{\mathrm{Gaussian}, \mathrm{Student}\text{-}t, \mathrm{Clayton}, \mathrm{Gumbel}, \mathrm{Frank}\}$ using `RVineStructureSelect` (AIC) and estimate parameters by MLE; computation is parallelized across (i, g). This yields regime-aware, spatially coherent ensembles respecting both $p(S_t \mid S_{t-1}, X_t)$ and the intra-zone dependence $c_{i,g}$.

https://doi.org/10.5194/egusphere-2025-5543
Preprint. Discussion started: 9 February 2026

2.5.4 Forecasting and uncertainty quantification

For each forecast origin, we: (i) propagate S_t using the fitted non-homogeneous transition kernel with observed/predicted X_t; (ii) simulate $\mathbf{U}_t^{(g)} \sim C_{g,\hat{S}_t}$; (iii) invert site–state marginals to precipitation space. Predictive uncertainty is summarized by empirical quantiles across M replicates. Let $[\hat{L}_t, \hat{U}_t]$ denote the central 90% interval and $\hat{X}_t$ the ensemble mean at time t. Calibration and sharpness are evaluated using coverage (POC), average width (AW), and the asymmetry degree (AAD), together with strictly proper scores: station-wise CRPS, and zone-wise energy (ES) and variogram (VS) scores (Sects. 2.4.2.4.3). Reliability diagnostics include PIT/rank histograms, QQ-plots for extremes, and state-conditional Kendall's τ/tail-dependence checks for low-order vine edges. Verification uses blocked (winter-wise) cross-validation and an out-of-sample holdout; results are reported by zone and regime in Tables S3–S6 and Figs. S4–S10. Across zones, 90% coverages are close to nominal with non-excessive widths; PIT histograms are acceptably flat; and copula-based τ reproduces observed inter-site ranks, supporting the adequacy of the nHMM–vine construction.

2.5.5 Rationale and coherence with the hierarchical design

The experimental design is aligned with the hierarchical modeling rationale established in Sects. 2.2–2.3:

1. **Separation of temporal and spatial scales.** The nHMM grid targets identifiability of regime dynamics and covariate effects at synoptic scales (choice of K and emission flexibility C), without contamination by high-dimensional spatial likelihoods. DTW zoning then confines copula learning to climatically coherent subdomains where residual dependence is most meaningful.

2. **Principled capacity control.** BIC-based selection over (K, C) controls temporal model complexity given N_{obs}, whereas the multi-criteria CVI selection of G controls spatial granularity. Vine truncation and AIC-based edge family selection further regularize dependence structure, ensuring parsimony where higher-tree conditional links are negligible.

3. **End-to-end verifiability.** The same quantities optimized at training (likelihood; marginal fits) are independently audited by PIT diagnostics, tail checks, Rosenblatt-transform tests, ES/VS, and event-based Brier scores under block validation, closing the loop between estimation and genuine predictive skill.

4. **Computational tractability without loss of interpretability.** The Viterbi-based conditioning allows decoupled, parallel estimation of $G \times K$ copulas, preserving the climatological interpretation of regimes while attaining multivariate coherence at the zone level, something infeasible within a single EM loop for $q = 462$.

Together, these choices yield a calibrated and sharp probabilistic forecaster that respects regime dynamics and reconstructs residual spatial dependence, providing a coherent bridge from methodology to the *Results* section.

https://doi.org/10.5194/egusphere-2025-5543
Preprint. Discussion started: 9 February 2026

3 Results

3.1 Model configuration (summary)

All results reported here use the non-homogeneous HMM with $K = 5$ latent states and $C = 5$ Gamma components per state (configuration `5hs5dg`), selected by BIC after fitting a $K \times C$ grid (Sect. 2.5.1); increasing to $K = 6$ produced redundant, weakly separated regimes with negligible likelihood gains relative to the BIC penalty (Table S1). Residual spatial dependence is modeled within $G = 7$ DTW-based zones (Ward clustering), chosen by a consensus of internal CVIs (Sect. 2.5.2; Table S2). For each zone and state, we estimate a state-conditional Regular Vine copula (candidate families: Gaussian/Clayton/Gumbel/Frank/Student-t; AIC selection) by MLE and use it for state-aware ensemble simulation (Sect. 2.5.3). Forecast verification employs central 90% prediction intervals (coverage and average width), AAD, and strictly proper scoring rules (CRPS for station-level medoids; ES/VS for multivariate zone fields) under blocked cross-validation and out-of-sample tests (Sects. 2.5.4, 2.4.3; Tables S3–S6, Figs. S4–S10).

3.2 Climatic organization of precipitation by Hidden Markov states

3.2.1 State-conditioned precipitation fields

Figure 1 synthesizes the austral-winter (MJJA) precipitation conditioned on the five nHMM regimes. In each column, the main map shows the conditional mean daily precipitation (mm d^{-1}) across the 462 locations for days assigned to that state by the Viterbi path; the inset displays anomalies (conditional minus MJJA climatology, 1980–2021). The joint depiction disentangles baseline meridional and orographic gradients from regime-dependent departures.

Climatologically, conditional means peak over southern Chile (40°S–55°S), consistent with vigorous mid-latitude cyclone activity and orographic enhancement along the Andes and coastal ranges. Northern Chile (north of 30°S) remains dry in all regimes due to persistent subtropical aridity. The anomaly insets reveal a coherent dry–to–wet progression across states. Two regimes ("0", "1") exhibit widespread negative anomalies over central–southern Chile (suppressed frontal activity and moisture transport). A transitional regime ("2") presents near-neutral, spatially patchy anomalies. Two regimes ("3", "4") display robust positive anomalies from ∼35°S to ∼45°S (strongest in "4"), with spatial coherence enhanced by windward ascent.

3.2.2 Dynamical interpretation from large-scale circulation

Figure 2 consolidates the large-scale circulation composites for each state: (A) MSLP anomalies (hPa) relative to the MJJA zonal climatology; (B) composite mean 850-hPa wind speed (m s^{-1}) with streamlines; (C) Z500 anomalies (m). The five regimes form a dynamically consistent sequence.

Dry states (e.g., "1"): a ridge–high couplet appears with +Z500 over the southeast Pacific and +MSLP centered over central Chile, displacing the storm track poleward and strengthening subsidence over 30–40°S. Low-level composites exhibit weakened westerlies (850 hPa) along the Pacific corridor and enhanced coastal southerlies, inhibiting onshore moisture flux and orographic triggering. *Wet states (e.g., "4"):* a deep −Z500 trough extends from the Amundsen–Bellingshausen sector

https://doi.org/10.5194/egusphere-2025-5543
Preprint. Discussion started: 9 February 2026

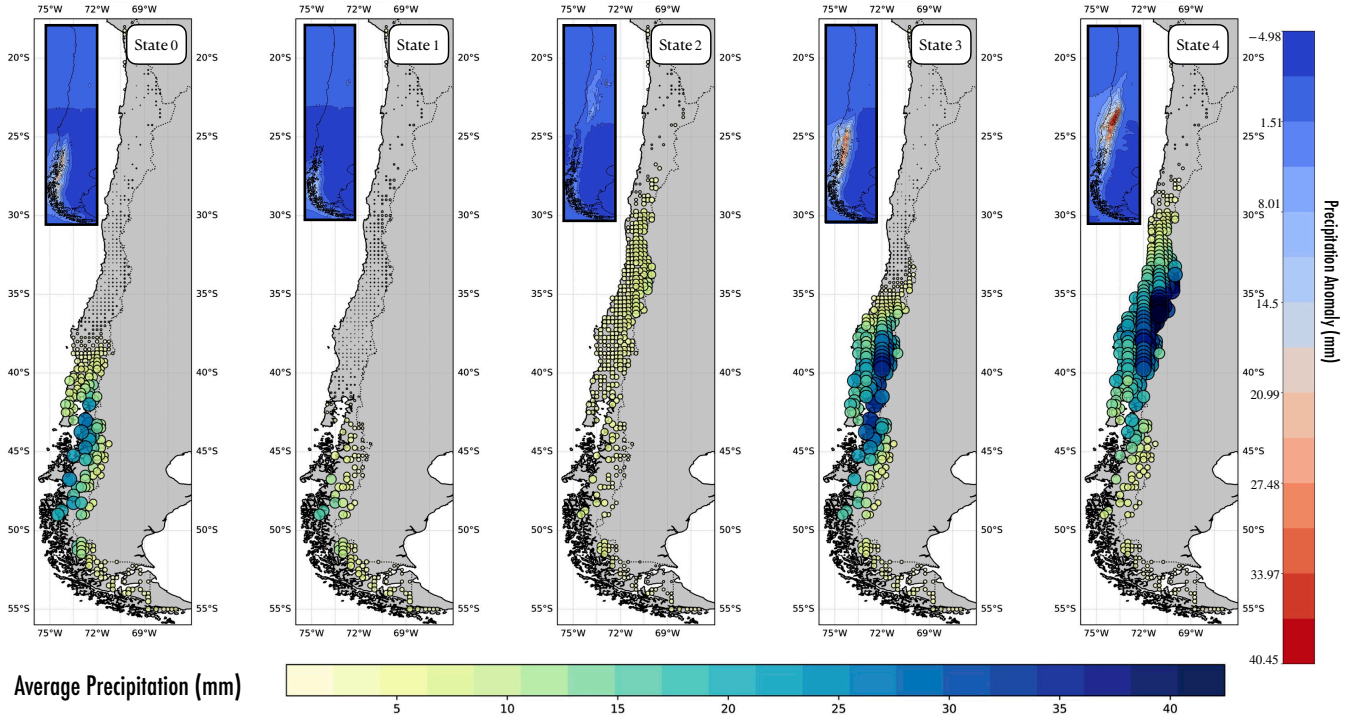

Figure 1. Mean precipitation and anomalies by nHMM state (MJJA). For each of the five latent states ("0"–"4"), the main panel maps conditional mean daily precipitation (mm d^{-1}) across the 462 locations (1980–2021). Insets: anomalies (mm d^{-1}) relative to the MJJA climatology. States "0"/"1": negative anomalies (dry regimes); "2": transitional; "3"/"4": positive anomalies, strongest over ∼35°S–45°S.

into the southeast Pacific with a −MSLP minimum south of ∼40°S, intensifying a more zonal low-level jet impinging on central–southern Chile; cyclonic curvature favors frontal genesis and baroclinic development. Transitional states ("0", "2", "3") show intermediate structures (dipolar pressure/height patterns, emerging westerly corridors), consistent with episodic storm incursions and mixed hydroclimatic outcomes. The columnwise correspondence between (A) and (C) evidences robust vertical coupling, while (B) underscores the central role of the state-dependent Pacific moisture corridor intersecting topography.

3.3 Persistence and inter-regime transitions

Let $\widehat{S}_{1:n}$ denote the Viterbi state sequence for all MJJA days in 1980–2022. For $i, j \in \{0, \ldots, 4\}$ and year y, define annual transition counts

$$C_{ij}^{(y)} = \sum_{t \in \mathcal{T}_y} \mathbf{1}\{\widehat{S}_t = i,\ \widehat{S}_{t+1} = j\}, \qquad N_i^{(y)} = \sum_{t \in \mathcal{T}_y} \mathbf{1}\{\widehat{S}_t = i\},$$

and proportions $p_{ij}^{(y)} = C_{ij}^{(y)} / N_i^{(y)}$ $(N_i^{(y)} > 0)$. Aggregating over years yields $\widehat{p}_{ij} = \frac{\sum_y C_{ij}^{(y)}}{\sum_y N_i^{(y)}}$ and the sample transition matrix $\widehat{\mathbf{P}} = (\widehat{p}_{ij})$ (Table 1).

https://doi.org/10.5194/egusphere-2025-5543
Preprint. Discussion started: 9 February 2026

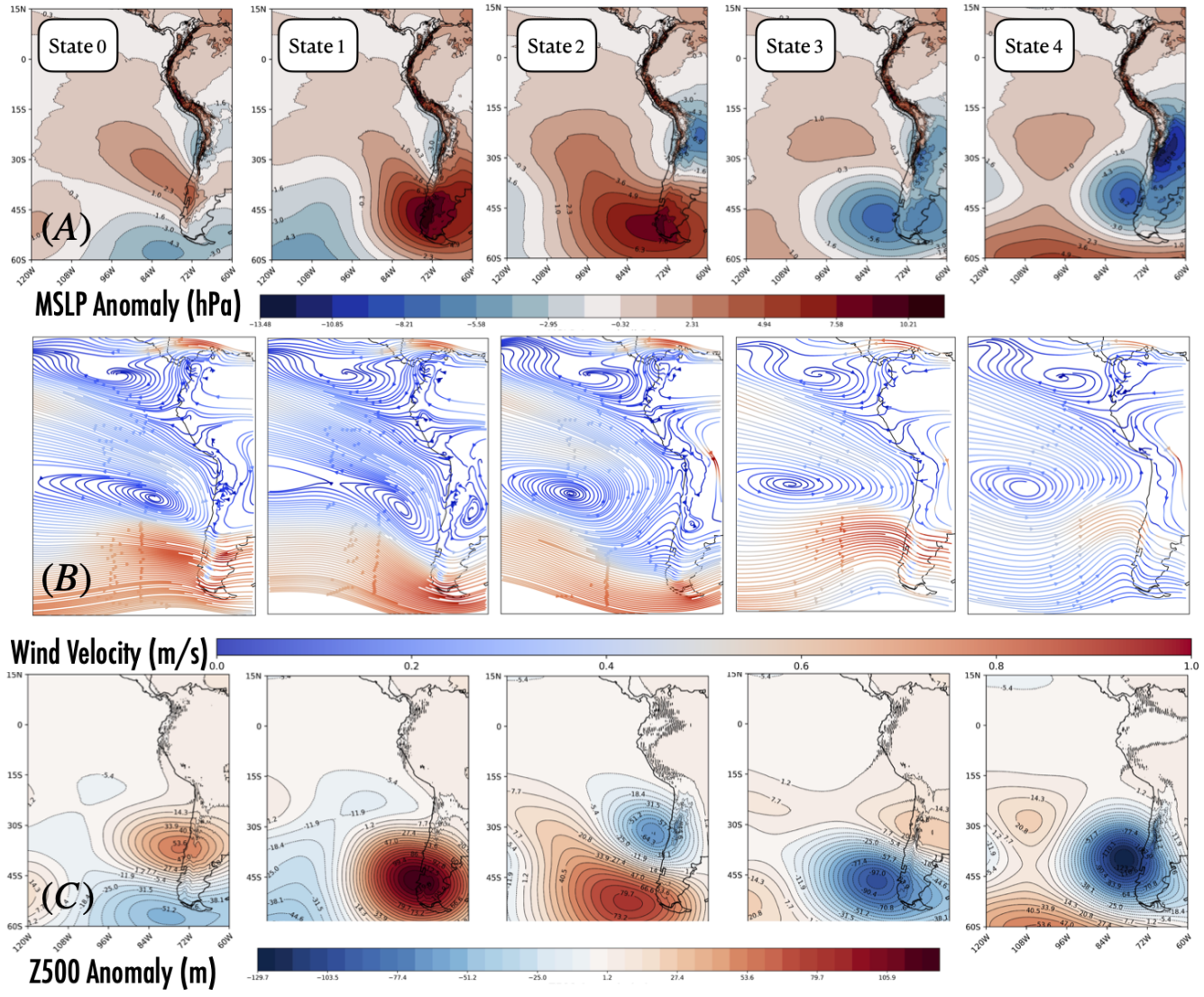

Figure 2. Circulation composites by nHMM state (MJJA, 1980–2021). Columns: states "0"–"4". (A) MSLP anomalies (hPa) vs MJJA zonal climatology; (B) composite mean 850-hPa wind speed with streamlines; (C) Z500 anomalies (m). Dry regimes (e.g., "1"): ridge–high couplet and weakened/deflected low-level westerlies. Wet regimes (e.g., "4"): deep trough–low couplet and intensified westerly jet with enhanced landfalling moisture flux.

435 $\widehat{\mathbf{P}}$ is diagonally dominant, evidencing persistence. The largest persistence is for the dry regime S_1 ($\widehat{p}_{11} = 0.56$), followed by the wet regime S_4 ($\widehat{p}_{44} = 0.43$); S_0, S_2, S_3 are moderately persistent (0.30–0.39), acting as transitional configurations. Direct jumps $S_1 \leftrightarrow S_4$ are rare ($\widehat{p}_{14} = 0.02$, $\widehat{p}_{41} = 0.03$), dynamically consistent with gradual synoptic reconfiguration from subtropical anticyclonic to mid-latitude cyclonic states. Transitions among S_0–S_1–S_2 are relatively likely, while $S_3 \leftrightarrow S_4$ is

https://doi.org/10.5194/egusphere-2025-5543
Preprint. Discussion started: 9 February 2026

Table 1. Average transition probabilities (1980–2022). Each entry is $\widehat{p}_{ij}$, the empirical probability of $i \to j$ given i.

From/to	$[,0]$	$[,1]$	$[,2]$	$[,3]$	$[,4]$
$[0,]$	39%	23%	6%	27%	5%
$[1,]$	22%	56%	13%	8%	2%
$[2,]$	18%	24%	32%	10%	16%
$[3,]$	20%	6%	13%	36%	25%
$[4,]$	9%	3%	27%	18%	43%

favored over links between wet and dry clusters, supporting S_3 as a preconditioning stage toward wetness and S_0–S_2 as egress from dryness.

To assess temporal drift, we analyze $\{p_{ij}^{(y)}\}_{y=1980}^{2022}$ with binomial GLMMs (or beta–binomial under overdispersion) when support is adequate (at least 25 nonzero years and $\sum_y C_{ij}^{(y)} \geq 100$), and use Mann–Kendall with Sen's slope for very rare transitions. Figure 3 shows vetted trends: $S_4 \to S_4$ persistence declines significantly, indicating shorter residence in the wet regime; $S_1 \to S_1$ is trendless. Bridging transitions display compensating behavior (e.g., modest increases in $0 \leftrightarrow 3$ with reduced $3 \to 4$), suggesting more frequent cycling through transitional states rather than persistent wet occupancy. About 73% of the variance in $\{p_{ij}^{(y)}\}$ arises from systematic differences among transitions, with comparatively weaker multidecadal modulation within a given $i \to j$.

3.4 Linkages between regimes and large-scale climate indices

For MJJA season y, the occupancy of state k is $f_k(y) = |T_y|^{-1} \sum_{t \in T_y} \mathbf{1}\{S_t = k\}$, $|T_y| = 123$. Pairwise correlations $\rho_{k,I} = \mathrm{corr}\big(f_k(y), I(y)\big)$ with SOI, Niño 3.4, ONI, and GMSST are compiled in Fig. 4 (rank correlations yield qualitatively identical patterns; Supplementary S7). Dry-state occupancy f_1 correlates positively with SOI and negatively with Niño/ONI; wet-state occupancy f_4 shows the opposite, in line with canonical ENSO forcing. Transitional states ($k \in \{0, 2, 3\}$) have weaker, sign-varying associations. Moderate $|\rho_{k,I}|$ values reflect nonlinearity and state dependence, which the hybrid framework captures via covariate-driven transitions and copulas.

Monthly MJJA relationships (1980–2022) clarify functional form (Fig. 5): ONI is negatively (positively) associated with f_1 (f_4), with strongest signal in ONI tails (El Niño/La Niña months). GMSST relates negatively to f_4, consistent with increased subtropical stability and a poleward baroclinicity shift. Dashed OLS lines are descriptive; inference in our approach relies on covariate-sensitive transitions (Sect. 2.2) and copulas that permit nonlinear/tail dependence (Sect. 2.3).

All results are robust to rank measures (Spearman, Kendall) with effective-sample adjustments for persistence and to stationary block bootstrap uncertainty; signs stay invariant under FDR control (Pyper and Peterman, 1998; Politis and Romano, 1994; Benjamini and Hochberg, 1995) (Supplementary S7).

https://doi.org/10.5194/egusphere-2025-5543
Preprint. Discussion started: 9 February 2026

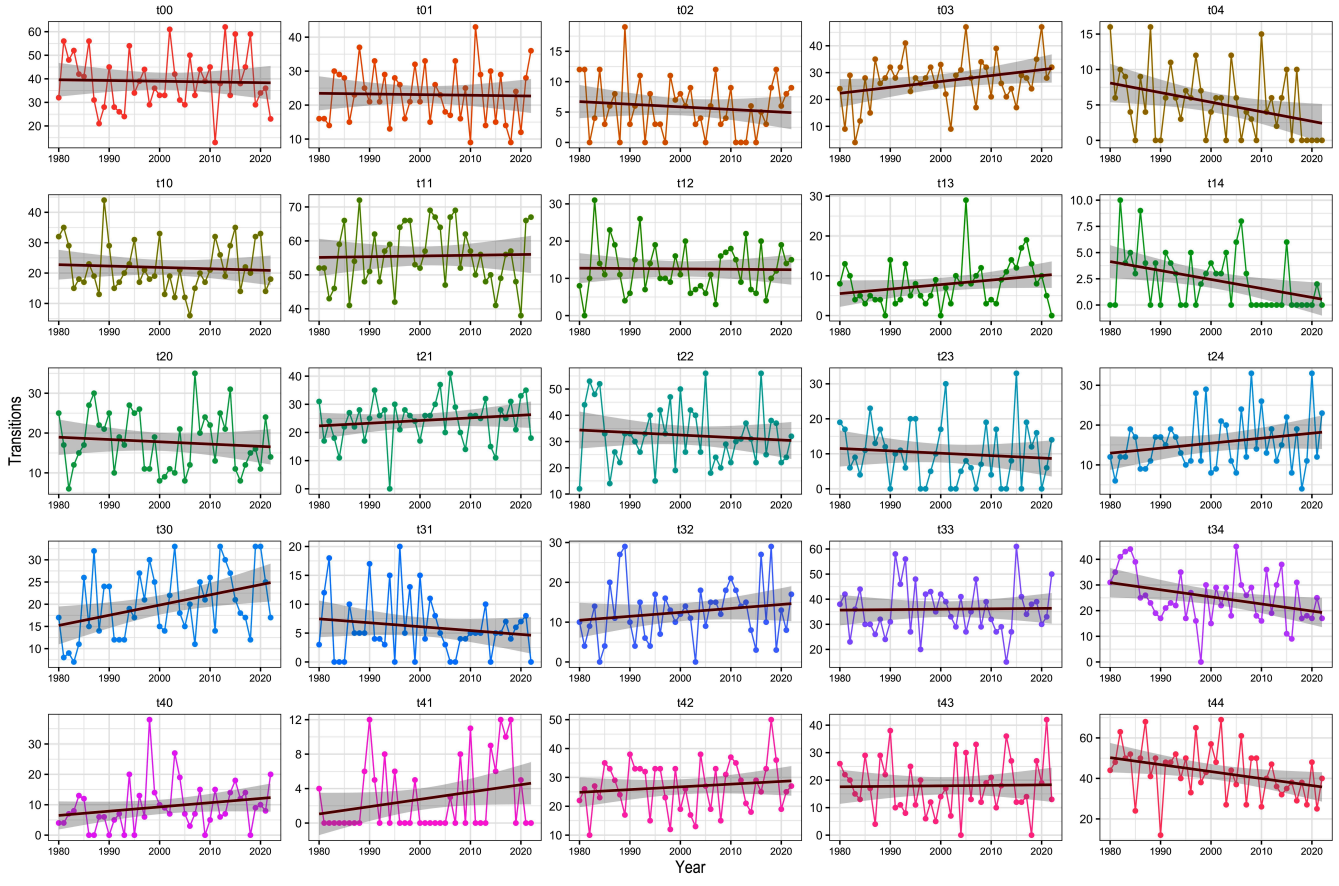

Figure 3. Annual transition proportions and trends. For each $i \to j$, points are $p_{ij}^{(y)} = C_{ij}^{(y)}/N_i^{(y)}$ (MJJA, 1980–2022). Solid lines: GLMM logit–linear trends where diagnostics support binomial/beta–binomial modeling; panels lacking support are shown without trend lines and excluded from inference.

3.5 Out-of-sample probabilistic simulations and verification

We evaluate the hybrid nHMM–Copula–GPD forecaster on MJJA 2022. The nHMM is trained on 1980–2021; the 2022 state sequence is inferred by Viterbi using observed covariates. Conditional on state and zone, a state–zone-specific R-vine couples sites; the marginal tail at each site is refined by a GPD above the 0.90 quantile. Unless stated otherwise, $M = 100$ Monte Carlo members per day form predictive distributions and aggregates.

Figure 6 summarizes design and skill. Panel (A): seven DTW zones (medoids marked). Panel (B): for six medoids (Zone 1 excluded due to negligible MJJA totals), monthly boxplots of 2022 predictive distributions (100 daily draws per day), with observed monthly means (orange). Insets: daily ensembles (semi-transparent) vs observations (orange). Ensembles reproduce seasonal cycle and intermittency; spreads are smaller in arid north, larger (and well-centered) in storm-dominated south.

https://doi.org/10.5194/egusphere-2025-5543
Preprint. Discussion started: 9 February 2026

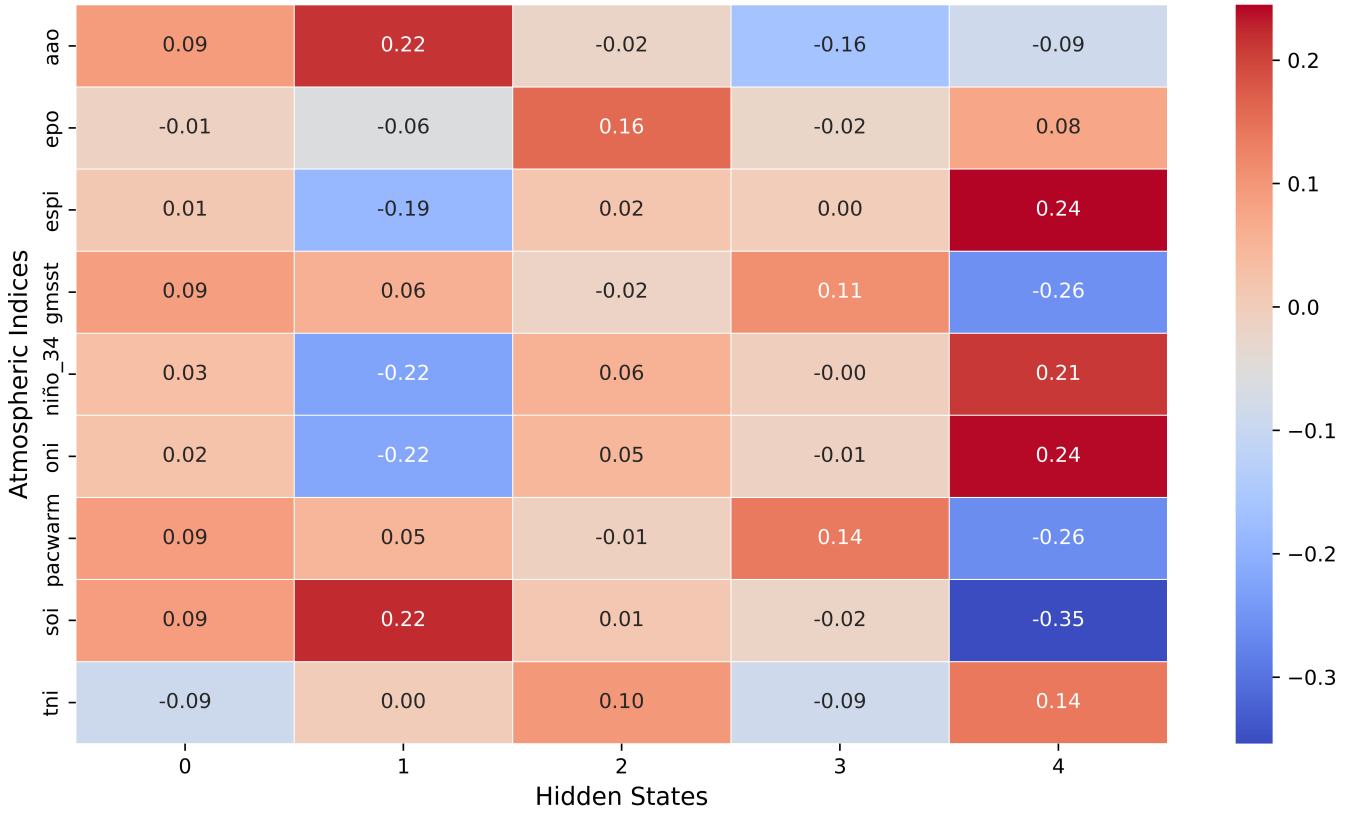

Figure 4. Correlation between HMM state frequencies and atmospheric indices (MJJA, 1980–2022). Correlations between seasonal state frequencies $f_k(y)$ (columns) and standard indices (rows). Signs agree with ENSO modulation (SOI vs Niño metrics). Magnitudes are moderate, consistent with teleconnections that modulate—not deterministically control—synoptic regimes.

Table 2 reports uncertainty metrics at medoids. Average Width (AW) increases meridionally, reflecting increasing variability; POC remains close to nominal (88–98%), indicating calibrated intervals; AAD is uniformly small (≤ 0.07), with mild right-skew in Zones 2 and 7 (slightly under-dispersed ensembles under arid conditions). These outcomes indicate informative (sharp yet reliable) uncertainty quantification across hydroclimatic regimes.

3.5.1 Proper scores and event-based skill

Interval diagnostics are complemented with strictly proper scores and exceedance skill. For day t and site observation x_t, the ensemble-based CRPS estimator is

$$\widehat{\mathrm{CRPS}}(\hat{F}_t, x_t) = \frac{1}{M}\sum_{m=1}^{M}|Y^{(m)} - x_t| - \frac{1}{2M^2}\sum_{m=1}^{M}\sum_{m'=1}^{M}|Y^{(m)} - Y^{(m')}|.$$

Skill relative to a state-agnostic MJJA climatology (1980–2021) is reported as $\mathrm{CRPSS} = 1 - \overline{\mathrm{CRPS}}/\overline{\mathrm{CRPS}}_{\mathrm{clim}}$. Median CRPSS is positive in all zones, peaking in Zones 5–6 where regime sequencing (MSLP/Z500/wind coupling) is most informa-

https://doi.org/10.5194/egusphere-2025-5543
Preprint. Discussion started: 9 February 2026

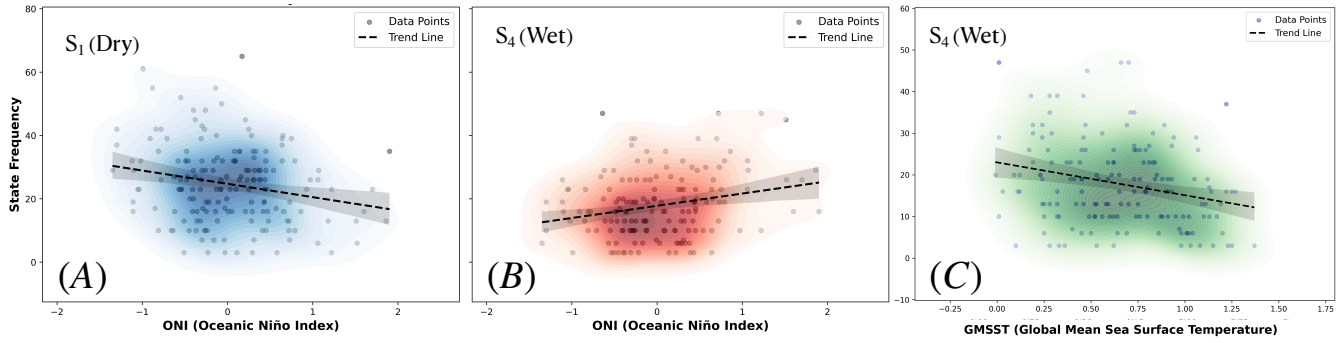

Figure 5. State–index relationships (MJJA, 1980–2022). (A) ONI vs monthly frequency of the dry state f_1; (B) ONI vs f_4; (C) GMSST vs f_4. Kernel-density contours emphasize high-probability regions; dashed lines: descriptive OLS fits (95% bands). Tail-enhanced associations motivate copula-based modeling.

Table 2. Uncertainty metrics at medoids for MJJA 2022 (100 daily draws per day). AW: Average Width (mm) of central 90% interval; POC: empirical coverage (%); AAD: Average Asymmetry Degree. Zone 1 excluded due to near-zero MJJA totals.

Zone	AW (mm)	POC (%)	AAD
Zone 2	6.873	96.56	0.070
Zone 3	18.117	97.67	0.016
Zone 4	26.644	96.90	0.016
Zone 5	32.521	97.45	0.008
Zone 6	44.020	93.33	0.012
Zone 7	11.560	88.83	0.029

tive. Reliability is examined with PIT histograms and KS tests (Supplementary S8): PITs are close to uniform in Zones 3–6; slight right-skew in Zones 2 and 7 matches the POC findings. For thresholds $u \in \{10, 25\}$ mm, Brier Scores, Brier Skill Scores and ROC AUC (from ensemble exceedance probabilities) indicate discrimination materially above chance (AUC $\gtrsim 0.7$), again peaking in Zones 5–6. Pinball losses at $\tau \in \{0.90, 0.95\}$ show no systematic bias at upper quantiles (Supplementary S8).

3.5.2 Spatial generalization and multi-year robustness

To test generalization beyond medoids, we verify 2022 forecasts at three additional non-medoid stations per zone. Distributions of POC and CRPSS mirror medoid behavior (Supplementary S8), indicating that regime–copula structure transfers within zones. A rolling-origin hindcast (2012–2021; train 1980–y–1, forecast year y) yields positive median CRPSS throughout, with interannual fluctuations tracking ENSO phase but preserving overall gains (Supplementary S8).

https://doi.org/10.5194/egusphere-2025-5543
Preprint. Discussion started: 9 February 2026

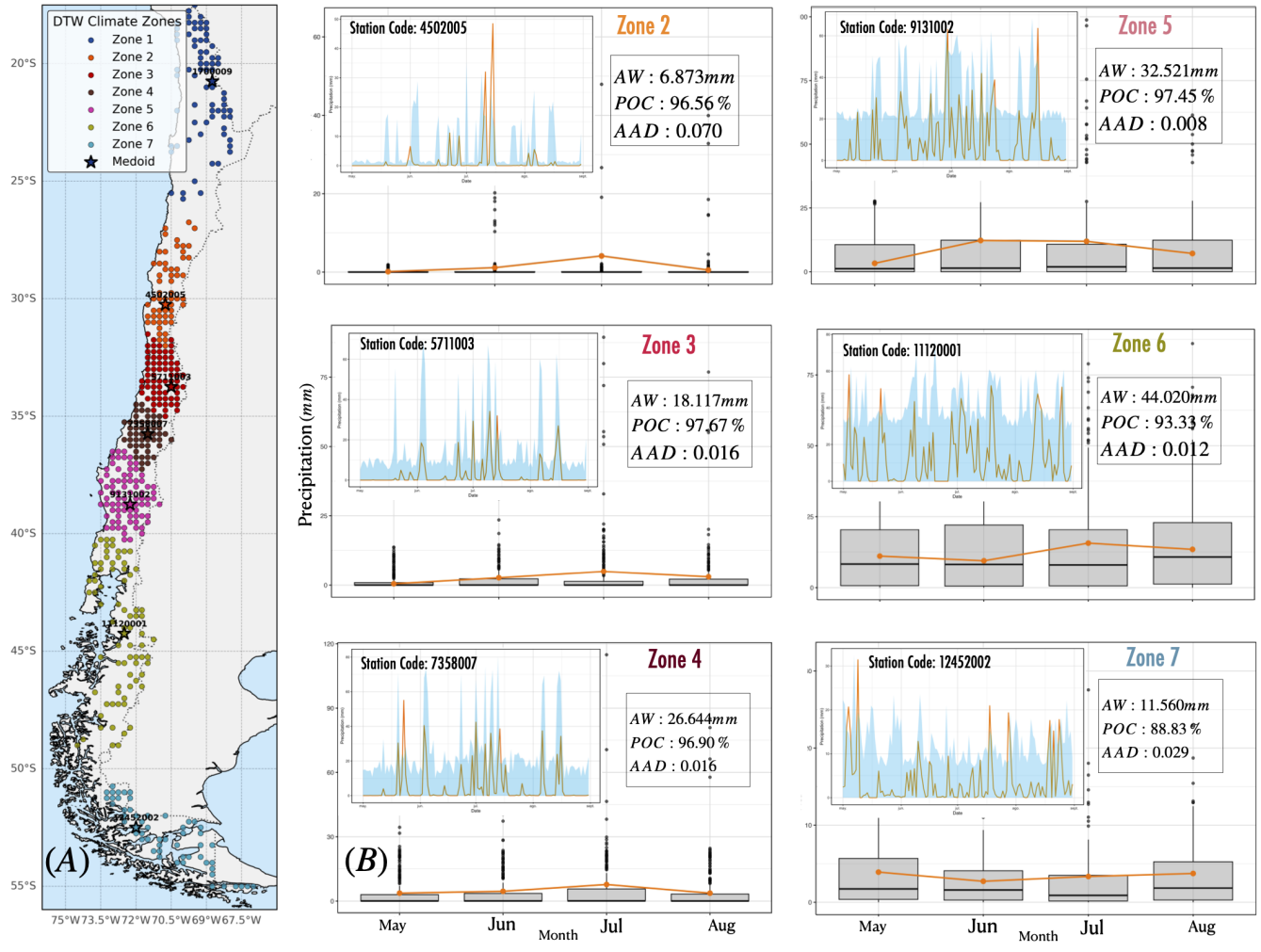

Figure 6. DTW climate zones and out-of-sample MJJA 2022 simulations. (A) Seven DTW-based climate zones (medoids: stars). (B) For six medoids, monthly boxplots of MJJA 2022 predictive distributions (100 daily draws per day) overlaid with observed monthly mean (orange). Insets: daily ensembles vs observations. Training: 1980–2021; the 2022 hidden-state sequence is inferred via Viterbi and conditions the copula–GPD simulator.

3.5.3 Sensitivity to modeling choices

Allowing Student-t pairs in vines, varying the GPD threshold within $[0.85, 0.95]$, and changing state count to $K_s \in \{4, 6\}$ leaves PIT shapes, coverage and event skill qualitatively unchanged (Supplementary S9). Student-t pairs marginally improve heavy-rain discrimination in Zones 5–6; lower thresholds slightly increase over-coverage, supporting the parsimonious main specification.

https://doi.org/10.5194/egusphere-2025-5543
Preprint. Discussion started: 9 February 2026

3.6 Synthesis and implications

The hybrid framework delivers: (i) an interpretable regime taxonomy with physically coherent dry ↔ wet progression and robust vertical coupling (MSLP/Z500/850 hPa), (ii) quantified persistence/connectivity with rare direct dry–wet jumps and a detected decline in wet-state residence, and (iii) calibrated, sharp probabilistic forecasts whose multivariate dependence matches observed spatial coherence. These advances arise from the principled separation of temporal regime dynamics (nHMM) and state-conditioned spatial dependence (vine copulas with extreme-aware marginals), enabling coherent, uncertainty-aware precipitation simulations across Chile's heterogeneous hydroclimate. The results provide a process-consistent basis for hydrologic risk assessment and for scenario analyses where regime frequencies are modulated by large-scale drivers.

4 Discussion and Conclusions

This work develops and validates a regime–based probabilistic framework for austral–winter (MJJA) precipitation in Chile that couples a covariate–driven non–homogeneous Hidden Markov Model (nHMM) with DTW–based climatic zoning and state–conditioned Regular Vine copulas augmented by Generalized Pareto (GPD) tails. The modeling architecture is explicitly designed to separate *temporal* regime dynamics—linked to large–scale atmospheric controls—from *spatial* residual dependence—organized by mesoscale/orographic processes—thereby overcoming the conditional–independence limitation in classical HMM applications to rainfall. Using CR2MET at $0.05°$ resolution ensures spatial completeness and downscaled realism over complex terrain while retaining consistency with station–based evaluations for Chile.

4.1 Summary of scientific contributions.

(i) We identify a physically interpretable taxonomy of five hidden regimes that organize MJJA precipitation over Chile along a coherent dry ↔ wet spectrum. State composites of MSLP, Z500 and 850–hPa winds demonstrate robust vertical coupling: dry regimes exhibit a ridge–high couplet with weakened/deflected low–level westerlies, whereas wet regimes show a trough–low couplet with an intensified zonal jet impinging on central–southern Chile. (ii) We quantify regime persistence and connectivity from 1980–2022, finding diagonally–dominant transitions, rare direct $S_1 \leftrightarrow S_4$ jumps, and a statistically significant decline in wet–state self–transitions ($S_4 \rightarrow S_4$), consistent with documented drying signatures in central Chile. (iii) We fuse regimes with a *state–aware* spatial simulator: DTW zoning yields climatically coherent subdomains, and state–conditional R–vines fitted to PIT–transformed series recover spatial dependence and tail behavior (via GPD) within each regime, enabling spatially coherent probabilistic simulations. (iv) We establish out–of–sample probabilistic skill in MJJA 2022: central 90% coverage (POC) remains close to nominal across zones (88–98%), intervals are sharp yet reliable (AW increases meridionally as expected), AAD is uniformly small (≤ 0.07), median CRPSS is positive in all zones (peaking where regime sequencing is most informative), and event–based scores (Brier, AUC) indicate discrimination above chance for moderate/heavy thresholds. Together, these results provide converging evidence that the proposed hierarchy is calibrated, sharp, and physically consistent.

https://doi.org/10.5194/egusphere-2025-5543
Preprint. Discussion started: 9 February 2026

4.2 Relation to prior work and novelty.

HMMs and nHMMs have been widely used to characterize atmospheric regimes and precipitation sequences, and copulas have been employed to represent spatial dependence and extremes. However, explicit *integration* of (i) covariate–driven regime transitions; (ii) DTW–based zoning that respects temporal asynchronies in rainfall; and (iii) *state–conditioned* high–dimensional R–vines with tail augmentation for daily totals has been largely unexplored. The present framework bridges these strands by treating regimes as the organizing temporal layer and learning residual spatial dependence conditional on regime and zone. This design preserves interpretability (regime physics) while achieving multivariate coherence and extreme–aware behavior at daily scales—capabilities difficult to obtain within a single monolithic spatiotemporal likelihood for $q = 462$ sites.

4.3 Hydroclimatic implications.

The detected decrease in $S_4 \to S_4$ persistence implies shorter residence times in wet, cyclone–dominated states during MJJA, coherently aligning with reduced winter accumulation and increasing hydro–climatic stress downstream (streamflow, snow storage, groundwater recharge). The regime classification translates large–scale circulation variability (e.g., ENSO phases and background SST anomalies) into state frequencies with spatially coherent precipitation responses. This enables state–contingent forecasting and scenario analysis: given seasonal predictors (indices/dynamical fields), the transition kernel provides regime probabilities, which—propagated through the copula–GPD layer—become spatially consistent precipitation ensembles suitable for risk–aware operations (e.g., reservoirs, drought management, flood preparedness).

4.4 Robustness and validation.

Multiple, complementary diagnostics support the adequacy of each model layer. For emissions, PIT histograms, wet/dry reliability, QQ plots and GPD threshold stability confirm calibration across regimes, with expected challenges in the hyper–arid north (mass at zero, data scarcity in the tail). For spatial dependence, Rosenblatt–transform tests, variogram/tail–coincidence diagnostics, and variogram scores validate the selected vines; truncated structures are preferred when higher–order conditional links attenuate, limiting over–parameterization. Proper scoring rules (CRPS/ES/VS) under block cross–validation and out–of–sample evaluation document genuine predictive skill beyond in–sample fit. Sensitivity checks (vine families, thresholds, $K \in \{4, 6\}$) leave conclusions qualitatively unchanged, with Student–t pairs offering marginal gains for heavy–rain discrimination in southern zones.

4.5 Limitations and scope.

First, Viterbi decoding induces state–assignment error; residual misclassification can propagate to the copula layer. In practice, effects are modest in our diagnostics, but a smoothed–probability simulation scheme could further attenuate classification noise. Second, within–state nonstationarity (e.g., slow structural changes in spatial dependence) is not explicitly modeled; regime–varying or covariate–varying vines, or mixture copulas, could capture such evolution. Third, tail modeling relies on a fixed exceedance threshold (here 0.90), which trades bias and variance; threshold ensembles or Bayesian threshold selection

https://doi.org/10.5194/egusphere-2025-5543
Preprint. Discussion started: 9 February 2026

could formalize this choice. Fourth, the nHMM assumes conditional independence during estimation; although our hierarchy repairs dependence post hoc in a principled manner, a fully joint spatiotemporal likelihood would be ideal but is computationally prohibitive at this dimensionality.

4.6 Pathways for future work.

Methodological extensions include (i) zero–inflated/hurdle emissions with regime–specific tails to improve arid–zone behavior; (ii) hierarchical Bayesian formulations that integrate regime, vine, and tail components with shared priors, enabling joint uncertainty quantification and model selection; (iii) regime–aware multivariate fields (precipitation–temperature–snow) for compound–risk analysis; and (iv) assimilation of sub–seasonal/seasonal dynamical predictors (e.g., MJO/jet diagnostics, RCM output) as covariates in the transition kernel to extend lead times. From an applications perspective, state–contingent ensembles support reservoir rule–curve testing, drought early–warning, and impact–model forcing with coherent spatial fields.

4.7 Concluding remarks.

By disentangling large–scale regime dynamics from residual spatial dependence and explicitly accounting for extremes, the nHMM–DTW–vine–GPD framework delivers (i) meteorologically coherent regimes, (ii) quantified and interpretable changes in regime persistence, and (iii) calibrated, sharp, spatially consistent probabilistic forecasts. The approach reconciles physical insight with statistical parsimony and computational tractability at continental scale, offering a transferable template for probabilistic hydroclimate modeling in Chile and other mountainous, hydroclimatically heterogeneous regions.

Code and data availability. All code and data required to reproduce the analyses and figures in this study are openly available at the public GitHub repository

https://github.com/mauricio-herrera/hybrid-nhmm-copula-precipitation-framework.

The repository contains: (i) the R script `Hybrid_nHMM_Copula_Simulation_Validation_2022.R`, (ii) the compressed dataset `mapa_estaciones_precipitaciones_clean.csv.zip`, which must be unzipped before running the script, and (iii) documentation files (`README.md`, `LICENSE`, `CITATION.cff`) describing data structure and execution steps. All analyses were performed in R (version 4.3 or later) using publicly available packages (`VineCopula`, `evd`, `scoringRules`, etc.).

Author contributions.

- Mauricio Herrera-Marín: Conceptualization, Methodology, Formal Analysis, Investigation, Writing – Original Draft, Visualization
- Francisca Kleisinger: Data Curation, Methodology, Software, Visualization
- Diego Rivera: Conceptualization, Writing – Review & Editing, Supervision
- Andrés Wilson: Data Curation, Methodology, Visualization
- Alex Godoy-Faúndez: Conceptualization, Writing – Review & Editing, Supervision.

https://doi.org/10.5194/egusphere-2025-5543
Preprint. Discussion started: 9 February 2026

Competing interests. The authors declare that they have no conflicts of interest or competing interests.

Acknowledgements. D. Rivera acknowledges financial support from FONDECYT 1230520 and ANID/FONDAP/1523A0001.

https://doi.org/10.5194/egusphere-2025-5543
Preprint. Discussion started: 9 February 2026

References

Aas, K., Czado, C., Frigessi, A., and Bakken, H.: Pair-copula constructions of multiple dependence, Insurance: Mathematics and Economics, 44, 182–198, 2009.

Ailliot, P., Thompson, C., and Thomson, P.: Space–Time Modeling of Precipitation Using a Hidden Markov Model and Censored Gaussian Distributions, Journal of Hydrology, 365, 56–73, https://doi.org/10.1016/j.jhydrol.2008.11.027, 2009.

Arias, P. A., Garreaud, R., Poveda, G., Espinoza, J. C., Molina-Carpio, J., Masiokas, M., Viale, M., Scaff, L., and Van Oevelen, P. J.: Hydroclimate of the Andes part II: Hydroclimate variability and sub-continental patterns, Frontiers in Earth Science, 8, 505 467, 2021.

Barrett, B. S. and Hameed, S.: Seasonal variability in precipitation in central and southern Chile: Modulation by the South Pacific high, Journal of Climate, 30, 55–69, 2017.

Barría, P., Rojas, M., Moraga, P., Muñoz, A., Bozkurt, D., and Alvarez-Garreton, C.: Anthropocene and streamflow: Long-term perspective of streamflow variability and water rights, Elem Sci Anth, 7, 2, 2019.

Becker, E. and Schmitz, G.: Climatological effects of orography and land–sea heating contrasts on the gravity wave–driven circulation of the mesosphere, J. Atmos. Sci., 60, 103–118, https://doi.org/10.1175/1520-0469(2003)060<0103:CEOOAL>2.0.CO;2, 2003.

Bellone, E., Hughes, J. P., and Guttorp, P.: A Hidden Markov Model for Downscaling Synoptic Atmospheric Patterns to Precipitation Amounts, Climate Research, 15, 1–12, https://doi.org/10.3354/cr015001, 2000.

Benjamini, Y. and Hochberg, Y.: Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, Journal of the Royal Statistical Society: Series B, 57, 289–300, 1995.

Bergmeir, C. and Benítez, J. M.: On the Use of Cross-Validation for Time Series Predictor Evaluation, Information Sciences, 191, 192–213, https://doi.org/10.1016/j.ins.2011.12.028, 2012.

Berndt, D. J. and Clifford, J.: Using dynamic time warping to find patterns in time series, KDD workshop, 10, 359–370, 1994.

Boisier, J. P.: CR2MET v2.5: High-resolution precipitation and temperature dataset for 1960–2021 in continental Chile, Dataset, https://doi.org/10.5281/zenodo.7529682, data set, 2023.

Boisier, J. P., Alvarez-Garretón, C., Cepeda, J., Osses, A., Vásquez, N., and Rondanelli, R.: CR2MET A high-resolution precipitation and temperature dataset for hydroclimatic research in Chile, 2018.

Boisier, J. P., Alvarez-Garreton, C., Cordero, R. R., Damiani, A., Gallardo, L., Garreaud, R. D., Lambert, F., Ramallo, C., Rojas, M., and Rondanelli, R.: Anthropogenic drying in central-southern Chile evidenced by long-term observations and climate model simulations, Elementa, 6, https://doi.org/10.1525/elementa.328, 2018.

Bozkurt, D., Rondanelli, R., Gonzalez-Reyes, A., Garreaud, R. D., Muñoz, A., Flores-Aqueveque, V., Aguilera, D., Farias-Barahona, D., Lambert, F., and Ramallo, C.: Climate change impacts on water availability and agriculture in Central Chile, International Journal of Climatology, 41, 1–21, https://doi.org/10.1002/joc.6789, 2021.

Burger, F., Brock, B., and Montecinos, A.: Seasonal and elevational contrasts in temperature trends in Central Chile between 1979 and 2015, Global and Planetary Change, 162, 136–147, 2018.

Bárdossy, A. and Pegram, G.: Copula-Based Multivariate Distribution Functions for Hydrological Modeling, Water Resources Research, 46, W07 536, https://doi.org/10.1029/2009WR007857, 2010.

Cavalleri, F., Lussana, C., Viterbo, F., Brunetti, M., Bonanno, R., Manara, V., Lacavalla, M., Sperati, S., Raffa, M., Capecchi, V., Cesari, D., Giordani, A., Cerenzia, I. M. L., and Maugeri, M.: Multi-scale assessment of high-resolution reanalysis precipitation fields over Italy, Atmospheric Research, 312, 107 734, https://doi.org/https://doi.org/10.1016/j.atmosres.2024.107734, 2024.

https://doi.org/10.5194/egusphere-2025-5543
Preprint. Discussion started: 9 February 2026

Charles, S. P., Timbal, B., and Nguyen, K.: Statistical Downscaling for Assessing Simulated Precipitation over Western Victoria, Australia, International Journal of Climatology, 34, 719–732, https://doi.org/10.1002/joc.3727, 2014.

Cioffi, F., Conticello, F. R., and Lall, U.: Stochastic Scenarios for 21st Century Rainfall Seasonality, Daily Frequency, and Intensity in South Florida, Journal of Water Resources Planning and Management, 146, 04020 058, https://doi.org/10.1061/(ASCE)WR.1943-5452.0001250, 2020.

CR2: Centro de Ciencia del Clima y la Resiliencia. Datos de Precipitación, https://www.cr2.cl/datos-de-precipitacion/, accessed: 2024-05-29, 2024.

Dawid, A. P.: Present Position and Potential Developments: Some Personal Views: Statistical Theory: The Prequential Approach, Journal of the Royal Statistical Society. Series A, 147, 278–292, https://doi.org/10.2307/2981683, 1984.

Diebold, F. X. and Mariano, R. S.: Comparing Predictive Accuracy, Journal of Business & Economic Statistics, 13, 253–263, https://doi.org/10.1080/07350015.1995.10524599, 1995.

Franzke, C. L. E. and Osprey, S.: Climate variability and extremes: Statistical Erameworks and Applications, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 470, 20130 267, https://doi.org/10.1098/rspa.2013.0267, 2014.

Garreaud, R., Alvarez-Garreton, C., Barichivich, J., Boisier, J., Christie, D., Galleguillos, M., LeQuesne, C., McPhee, J., and Zambrano-Bigiarini, M.: The 2010–2015 megadrought in central Chile: impacts on regional hydroclimate and vegetation, Hydrology and Earth System Sciences, 21, 6307–6327, https://doi.org/10.5194/hess-21-6307-2017, 2017.

Garreaud, R., Alvarez-Garreton, C., Barichivich, J., and et al.: The 2010–2018 Megadrought in Central Chile: Causes, Impacts and Lessons, Journal of Climate, 33, 5765–5780, 2020a.

Garreaud, R. D.: The Andes climate and weather, Advances in geosciences, 22, 3–11, 2009.

Garreaud, R. D., Vuille, M., Compagnucci, R., and Marengo, J.: Present-day South American climate, Palaeogeography, Palaeoclimatology, Palaeoecology, 281, 180–195, 2009.

Garreaud, R. D., Boisier, J. P., Rondanelli, R., Montecinos, A., Sepúlveda, H. H., and Veloso-Aguila, D.: The Central Chile Mega Drought (2010–2018): A Climate Dynamics Perspective, International Journal of Climatology, 40, 421–439, https://doi.org/10.1002/joc.6219, 2020b.

Genest, C. and MacKay, J.: A primer on the copula approach to modeling dependence between random variables, Statistical Science, 7, 113–122, 1986.

Genest, C. and Rémillard, B.: Validity of the Parametric Bootstrap for Goodness-of-Fit Testing in Semiparametric Models, Annales de l'Institut Henri Poincaré, Probabilités et Statistiques, 44, 1096–1127, https://doi.org/10.1214/07-AIHP148, 2008.

Genest, C., Rémillard, B., and Beaudoin, D.: Goodness-of-Fit Tests for Copulas: A Review and a Power Study, Insurance: Mathematics and Economics, 44, 199–213, https://doi.org/10.1016/j.insmatheco.2007.10.005, 2009.

Giorgino, T.: Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package, Journal of Statistical Software, 31, 1–24, https://doi.org/10.18637/jss.v031.i07, 2009.

Gneiting, T. and Raftery, A. E.: Strictly Proper Scoring Rules, Prediction, and Estimation, Journal of the American Statistical Association, 102, 359–378, https://doi.org/10.1198/016214506000001437, 2007.

Herrera, M., Kleisinger, F., and Wilsón, A.: Spatiotemporal Forecasting in Climate Data Using EOFs and Machine Learning Models: A Case Study in Chile, https://arxiv.org/abs/2502.17495, 2025.

Hersbach, H., Bell, B., and et. al.: ERA5 hourly data on single levels from 1940 to present., Copernicus Climate Change Service (C3S) Climate Data Store (CDS), https://doi.org/10.24381/cds.adbb2d47, 2023.

https://doi.org/10.5194/egusphere-2025-5543
Preprint. Discussion started: 9 February 2026

Hoskins, B. J. and Hodges, K. I.: A new perspective on Southern Hemisphere storm tracks, Journal of Climate, 18, 4108–4129, 2005.

Hughes, J. P., Guttorp, P., and Charles, S. P.: A non-homogeneous hidden Markov model for precipitation occurrence, Journal of the Royal Statistical Society: Series C (Applied Statistics), 48, 15–30, 1999.

Khalil, A. F., Kwon, H.-H., Lall, U., and Kaheil, Y. H.: Predictive Downscaling Based on Non-Homogeneous Hidden Markov Models, Hydrological Sciences Journal, 55, 333–350, https://doi.org/10.1080/02626661003780342, 2010.

Kirshner, S.: Modeling of Multivariate Time Series Using Hidden Markov Models, Ph.d. thesis, University of California, Information and Computer Sciences, 2005.

Latoja, D., Lillo-Saavedra, M., Gonzalo-Martin, C., Godoy-Faúndez, A., Somos-Valenzuela, M., and Rivera, D.: Decadal Variability of Dry Days in Central Chile, Earth Systems and Environment, pp. 1–18, 2024.

Maharaj, E. A. and Alonso, A. M.: Discriminant analysis of multivariate time series: Application to diagnosis based on ECG signals, Computational Statistics & Data Analysis, 55, 973–988, https://doi.org/10.1016/j.csda.2010.07.015, 2011.

Mayer, J., Mayer, M., and Haimberger, L.: Consistency and Homogeneity of Atmospheric Energy, Moisture, and Mass Budgets in ERA5, Journal of Climate, 34, 3955–3974, https://doi.org/10.1175/JCLI-D-20-0676.1, publisher Copyright: © 2021 American Meteorological Society., 2021.

McLachlan, G. and Krishnan, T.: The EM Algorithm and Extensions, vol. 382, John Wiley & Sons, Hoboken, NJ, 2nd edn., 2008.

Montecinos, A. and Aceituno, P.: Seasonality of the ENSO-related rainfall variability in central Chile and associated circulation anomalies, Journal of climate, 16, 281–296, 2003.

Montecinos, A., Kurgansky, M. V., Muñoz, C., and Takahashi, K.: Non-ENSO interannual rainfall variability in central Chile during austral winter, Theoretical and Applied Climatology, 106, 557–568, https://doi.org/10.1007/s00704-011-0457-1, 2011.

Montero, P. and Vilar, J. A.: TSclust: An R package for time series clustering, Journal of Statistical Software, 62, 1–43, https://doi.org/10.18637/jss.v062.i01, 2015.

Nelsen, R. B.: An Introduction to Copulas, Springer Series in Statistics, 2 edn., https://doi.org/10.1007/0-387-28678-0, 2006.

Pablo Sarricolea, M. H.-O. and Óliver Meseguer-Ruiz: Climatic regionalisation of continental Chile, Journal of Maps, 13, 66–73, https://doi.org/10.1080/17445647.2016.1259592, 2017.

Pica-Téllez, A., Garreaud, R., Meza, F., Bustos, S., Falvey, M., Ibarra, M., Duarte, K., Ormazábal, R., Dittborn, R., and Silva, I.: Informe proyecto arclim: Atlas de riesgos climáticos para chile, Tech. rep., Centro de Ciencia del Clima y la Resiliencia, Centro de Cambio Global UC y Meteodata para el Ministerio del Medio Ambiente a través de la Deutsche Gesellschaft für Internationale Zusammenarbeit (giz), Santiago, Chile, 2020.

Politis, D. N. and Romano, J. P.: The Stationary Bootstrap, Journal of the American Statistical Association, 89, 1303–1313, https://doi.org/10.1080/01621459.1994.10476870, 1994.

Pyper, B. J. and Peterman, R. M.: Comparison of methods to account for autocorrelation in correlation analyses of fish data, Canadian Journal of Fisheries and Aquatic Sciences, 55, 2127–2140, https://doi.org/10.1139/f98-104, 1998.

Robertson, A. W., Kirshner, S., and Smyth, P.: Downscaling of Daily Rainfall Using a Hidden Markov Model, Journal of Climate, 17, 4407–4429, https://doi.org/10.1175/3190.1, 2004a.

Robertson, A. W., Kirshner, S., and Smyth, P.: Nonstationary hidden Markov models for modeling daily rainfall occurrence over Brazil, Journal of Climate, 17, 4407–4424, 2004b.

Robertson, A. W., Kirshner, S., Smyth, P., Charles, S. P., and Bates, B. C.: Subseasonal-to-interdecadal variability of the Australian monsoon over North Queensland, Quarterly Journal of the Royal Meteorological Society, 132, 519–542, 2006.

https://doi.org/10.5194/egusphere-2025-5543
Preprint. Discussion started: 9 February 2026

Rosenblatt, M.: Remarks on a Multivariate Transformation, Annals of Mathematical Statistics, 23, 470–472, https://doi.org/10.1214/aoms/1177729394, 1952.

Salazar, A., Thatcher, M., Goubanova, K., et al.: CMIP6 precipitation and temperature projections for Chile, Climate Dynamics, 62, 2475–2498, https://doi.org/10.1007/s00382-023-07034-9, 2024.

Sarricolea, P., Herrera-Ossandon, M., and Meseguer-Ruiz, Ó.: Climatic regionalisation of continental Chile, Journal of Maps, 13, 66–73, 2017.

Schepsmeier, U., Stöber, J., Brechmann, E. C., Graeler, B., Nagler, T., Erhardt, T., Almeida, C., Min, A., and Ruckdeschel, P.: Package 'VineCopula', R Package Version 1.6.0, 2015.

Scheuerer, M. and Hamill, T. M.: Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities, Monthly Weather Review, 143, 1321–1334, https://doi.org/10.1175/MWR-D-14-00269.1, 2015.

Székely, G. J. and Rizzo, M. L.: Energy Statistics: A Class of Statistics Based on Distances, Statistical Science, 28, 558–575, https://doi.org/10.1214/13-STS437, 2013.

Trenberth, K. E.: Changes in Precipitation with Climate Change, Climate Research, 47, 123–138, 2014.

Valdés-Pineda, R., Valdes, J. B., Diaz, H. F., and Pizarro-Tapia, R.: Analysis of spatio-temporal changes in annual and seasonal precipitation variability in South America-Chile and related ocean–atmosphere circulation patterns, International Journal of Climatology, 36, 2979–3001, 2016.

Vrac, M., Naveau, P., Li, Z., Noblet-Ducoudré, N., Drobinski, P., Somot, S., Cattiaux, J., and Yiou, P.: Bridging Climate Gaps: A Stochastic Downscaling Method to Link Large-Scale Variables with Local Climate, Journal of Climate, 25, 6447–6464, https://doi.org/10.1175/JCLI-D-11-00628.1, 2012.

Vuong, Q. H.: Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses, Econometrica, 57, 307–333, https://doi.org/10.2307/1912557, 1989.

Ward, J. H.: Hierarchical Grouping to Optimize an Objective Function, Journal of the American Statistical Association, 58, 236–244, https://doi.org/10.1080/01621459.1963.10500845, 1963.

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, Academic Press, 3rd edn., 2011.

Zhang, L., Chen, X., and Li, S.: Application of Non-Homogeneous Hidden Markov Model to Simulate Rainfall Sequences in a Changing Climate, Hydrology and Earth System Sciences, 25, 3925–3941, https://doi.org/10.5194/hess-25-3925-2021, 2021.

Zucchini, W., MacDonald, I. L., and Langrock, R.: Hidden Markov Models for Time Series: An Introduction Using R, Chapman & Hall/CRC, 2nd edn., 2016.