

Response to Referee Comments

Hydrology and Earth System Sciences
Interactive Public Discussion

Manuscript:

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

Authors:

Mauricio Herrera-Marín, Francisca Kleisinger, Diego Rivera,
Andrés Wilson, Alex Godoy-Faundez

Affiliation:

CISGER, Faculty of Engineering, Universidad del Desarrollo (UDD), Santiago, Chile

Corresponding author: mherrera@udd.cl

Format note. This document follows the Copernicus/HESS interactive public discussion convention. Each referee comment is quoted in a shaded box; the authors’ response follows in plain text; any corresponding change to the manuscript is indicated in a green box with the exact section or table reference.

Letter to the Editor

Dear Editor,

We thank the Editor and both Referees for their thorough and constructive evaluation of our manuscript. The reviews were detailed and fair; we believe the revised manuscript is substantially stronger as a result.

Below we summarise the principal changes made in response to the referees’ comments. Individual point-by-point responses to each referee follow in Sections 2 and 3.

Summary of major revisions

- 1. Structural reorganisation** (*Referee 1, General remarks; Referee 2, General remarks*). The Methods section was entirely restructured as a sequential, step-by-step pipeline. A paragraph-map at the opening of Section 2 guides the reader through each sub-section. The former “Experimental Setup” section, which duplicated material already presented in Sections 2.2–2.4, was eliminated; its content was absorbed into the relevant sub-sections. All overlap between Sections 2 and 3 has been removed.
- 2. Results moved from Supplement to main text** (*Referee 1, General remark 2*). Seven tables containing the core quantitative results are now in the main manuscript: Table 1 (DTW CVI consensus); Table 2 (GPD parameter estimates with 95 % bootstrap CIs); Table 3 (vine structure summary, Zone 2); Table 4 (model selection grid with

N_{obs}/p ratio); Table 5 (transition matrix, 1980–2022); Table 6 (interval uncertainty metrics); and Table 7 (full verification summary). The Supplement now contains genuinely auxiliary material.

3. **Parameter count and identifiability** (*Referee 1, General remark 4*). The ratio $N_{\text{obs}}/p \approx 82$ (29,284 parameters from 2,386,692 observations) is now stated in the abstract, Table 4, and Sect. 2.5.
4. **Confidence intervals and p -values** (*Referee 1, General remark 5*). Bootstrap 95 % CIs for GPD shape parameters are in Table 2; bootstrap CIs for POC are in Table 6; the 95 % CI for the $S_4 \rightarrow S_4$ persistence trend is reported in Sect. 3.3; BH-FDR correction is applied over all 25 transition pairs.
5. **Moisture-related atmospheric variables** (*Referee 1, General remark 6*). Vertically integrated moisture flux (VIMF) and its divergence (VIMFD) were added to the covariate set, with physical justification in Sect. 2.1.
6. **Decision criteria for statistical tests** (*Referee 1, General remark 3*). Each validation sub-section now contains an explicit *Retention criterion* or *Decision criterion* paragraph. A new paragraph in Sect. 2.4.2 explains why p -values of zero on KS/CvM tests of Rosenblatt components do not imply operational failure of the vine at large training sample sizes.
7. **Zone 4 characterisation corrected** (*identified during revision*). The erroneous description of Zone 4 as “hyper-arid” was removed. Zone 4 covers the semi-arid Mediterranean transition of central Chile ($\approx 33^\circ\text{S}$ – 38°S). The inapplicability of zone-level areal CRPS in Zones 2 and 4 is explained through two mechanistically distinct causes in Sect. 3.5.1; CRPS entries for these zones in Table 7 are replaced by n/a with a footnote.
8. **Figure–Table coherence**. Table 6 was updated to use values from the validated production run ($M = 100$) that generated Figure 6, ensuring full numerical consistency.
9. **Zone 7 clarified**. Zone 7 covers the austral fjord region south of $\approx 45^\circ\text{S}$. The AUC entry in Table 7 is n/a because 25 mm exceedance events were insufficient in the 2022 evaluation season.

We believe the manuscript now addresses all substantive concerns raised by both referees.

Sincerely,

Mauricio Herrera-Marín (on behalf of all co-authors)

Response to Referee 1

We thank Referee 1 for a thorough and expert review. All page and section references below are to the **revised manuscript**.

Referee comment: *General assessment.* “This is a puzzling manuscript: on one side it delivers evidence that the chosen approach ... is a scientifically fruitful approach ... But the presentation of the results is the ‘black’ side of the medal.”

Authors’ response: We fully agree with this characterisation of the original submission. The revised manuscript has been restructured substantially to address each of the six specific presentation deficiencies identified in the General Remarks.

General Remark 1 — Repeated text in Sections 2 and 3

Referee comment: “In large parts, Section 2 and Section 3 contain repeated texts.”

Authors’ response: This repetition arose because the original Section 2.5 (“Experimental Setup”) largely duplicated material already presented in Sections 2.2–2.4. Section 2.5 has been eliminated in the revised manuscript. Its contents have been absorbed into the relevant sub-sections (model selection into Sect. 2.5 [now renamed “Model selection”]; DTW zoning criterion into Sect. 2.3.1; vine-copula setup into Sect. 2.3.4) or removed as genuinely redundant. The opening paragraph of Section 2 now provides a concise roadmap, replacing the need for the duplicated summary.

Manuscript change: Section 2.5 “Experimental Setup” removed. Content redistributed to Sects. 2.3.1, 2.3.4, and 2.5 (now “Model selection”). All overlap between Methods and Results eliminated.

General Remark 2 — Results displaced to Supplement; Section S8 unreadable

Referee comment: “A very large number of actual results is shifted into the supplement text S1–S10 (the total supplement covers 20 pages!) with the additional caveat that S8 is not readable.”

Authors’ response: We accept this criticism fully. The following results have been moved from the Supplement to the main text as numbered tables:

- Table 1 — DTW CVI consensus (top-ranked zoning candidates)
- Table 2 — GPD parameter estimates with 95 % bootstrap CIs
- Table 3 — Vine structure summary for Zone 2 (all states)
- Table 4 — Model selection grid (best per K , N_{obs}/p)
- Table 5 — Transition matrix (1980–2022)
- Table 6 — Interval uncertainty metrics (AW, POC with 95 % CI, AAD)
- Table 7 — Verification summary (CRPSS, CRPS, p_{DM} , pinball losses, AUC)

The Supplement has been rewritten and now contains genuinely auxiliary material: per-

zone PIT histograms (S1), full CVI table (S2), full GPD estimates for all zones (S3), vine structure for all zones (S4), ES/VS results (S5), areal-CRPS gains and Brier scores (S6), rank correlation robustness (S7), rolling-origin hindcast details (S8), and the full 30-configuration model selection grid (S9). Section S8 was rewritten from scratch and is fully readable in the revised version.

Manuscript change: Seven new tables added to main text. Supplement reduced and reorganised. Former Section S8 rewritten.

General Remark 3 — Tests without decision criteria

Referee comment: “A very large number of evaluation/test variables is introduced often only by name, and at no stage it is mentioned how this multitude of tests is used for inferences about the results and how the actual decision of choosing the appropriate model is influenced by those test variables.”

Authors’ response: Each validation sub-section now contains an explicit *Retention criterion* or *Decision criterion* paragraph. We quote the three key criteria added to the revised manuscript:

Sect. 2.4.1 (nHMM calibration): “Retention criterion: a configuration is retained if PIT histograms are visually flat, KS/CvM tests do not reject uniformity at the 5 % level for > 80 % of stations, and no significant Ljung–Box autocorrelation is detected. Regime count K is selected by BIC, with PIT-based adequacy as a secondary check.”

Sect. 2.4.2 (vine-copula fit): “Operative criterion for vine adequacy: (i) PIT histograms visually consistent with uniformity for verification states and zones; (ii) variogram scores lower for the vine model than for the independence baseline; and (iii) Diebold–Mariano tests confirming predictive skill gains out-of-sample.”

Sect. 2.4.3 (out-of-sample verification): “Decision criterion: the vine layer is deemed to improve over the independence baseline if the Diebold–Mariano test rejects equality at the 10 % level for station-wise CRPS or zone-wise VS in the majority of zones.”

Additionally, a new paragraph in Sect. 2.4.2 explains explicitly why KS/CvM p -values of zero on individual Rosenblatt components (as observed in Supplement S1) do not constitute operational failure of the vine: at training sample sizes of $n = 845$ – 1269 days per state, these tests have very high statistical power and will reliably detect even minor residual dependence that has no practical consequence for forecast calibration.

Manuscript change: *Retention criterion* / *Decision criterion* paragraphs added to Sects. 2.4.1, 2.4.2, and 2.4.3. Explanation of KS/CvM behaviour at large n added to Sect. 2.4.2.

General Remark 4 — Parameter count vs. observations not reported

Referee comment: “At no stage of the manuscript the actual number of estimated parameters vs. the actual number of sampling points in space and time used to perform the inference/estimation is mentioned.”

Authors’ response: The following information is now stated explicitly at three locations:

- *Abstract*: “The nHMM comprises $p = 29,284$ estimated parameters fit to $N_{\text{obs}} = 2,386,692$ daily observations ($N_{\text{obs}}/p \approx 82$), confirming adequate model identifiability.”
- *Table 4*: includes columns for p (free parameters) and N/p (observations-per-parameter ratio) for all three candidate model configurations.
- *Sect. 2.5*: “The ratio $N_{\text{obs}}/p \approx 82$ confirms adequate identifiability: there are roughly 82 observations per estimated parameter.”

Manuscript change: N_{obs}/p ratio added to abstract, Table 4, and Sect. 2.5.

General Remark 5 — GPD shape parameters without confidence intervals; bootstrap analysis required

Referee comment: “Very often uncertainty measures such as confidence intervals or p -values in case of test are not given or even mentioned. One example is table S8 in the supplements, here the estimated shape parameter of the Pareto fit of the medoid station is reported ... for each synoptic regime with one close to zero value and two positive and two negative values which would indicate either clear differences of the tail behavior or very weak properties of the underlying data set to estimate the notorious difficult shape value; this is a clear case to be studied in detail e.g. by bootstrapping across the stations around the medoid.”

Authors’ response: We thank the referee for this specific and constructive suggestion. We implemented a parametric bootstrap following exactly the spirit of the recommendation. For each (zone, state) combination, exceedances of the medoid station itself were resampled ($B = 500$ replicates), separating body observations and exceedances to preserve the exceedance proportion, and the GPD was re-fitted each time, producing empirical 95 % CIs for $\hat{\xi}$.

The results for Zone 5 (main Table 2) reveal a physically interpretable pattern:

- **States 0, 3, 4** (anticyclonic/transitional): $\hat{\xi} \leq 0$ (bounded tails); CIs are either entirely negative or straddle zero only narrowly.
- **States 1 and 2** (episodic frontal intrusions): $\hat{\xi} > 0$ (heavy, unbounded tails); CIs have predominantly positive bounds.

This sign contrast is statistically robust and physically interpretable: bounded tails arise under weak synoptic forcing (sparse, limited precipitation), while heavy tails arise under organised deep frontal systems (intense precipitation with no structural upper bound). Full per-zone results are in Supplement Table S3.

We also note that the original concern about “two positive and two negative values” in the previous submission arose partly because the table lacked confidence intervals and partly because it appeared to show the medoid of a different zone. The revised Table 2 explicitly identifies the Zone 5 medoid (southern Chile, storm-dominated) and provides bootstrap CIs throughout.

Manuscript change: Table 2 (main text) now reports $\hat{\xi}$, $\hat{\sigma}$, 95 % bootstrap CI for ξ , n_{exc} , and tail classification for all five states. Supplement Table S3 provides the same information for all six active zones. Physical interpretation of the dry/wet tail

contrast added to Sect. 2.3.2.

General Remark 6 — Manuscript reads like a project report; moisture variables absent from covariate set

Referee comment: “Except for the introduction the remaining manuscript ... reads like a split project report ... the selection of atmospheric variables lacks any moisture related variables like specific humidity or vertically integrated moisture transports which is more than the 850 (u, v) data and which can be easily extracted from C3S/ERA5 data sets.”

Authors’ response: We accept this criticism on both counts.

Regarding presentation: the restructuring described under Remarks (1) and (2) has substantially improved the manuscript’s coherence. The Methods section now reads as a continuous, sequentially organised narrative rather than a collection of sub-reports.

Regarding moisture variables: vertically integrated moisture flux (VIMF) and its divergence (VIMFD) have been added to the covariate set. The physical motivation is stated explicitly in Sect. 2.1: “The inclusion of VIMF and VIMFD is motivated by their direct link to landfalling moisture flux over Chile: u_{850} and v_{850} alone do not fully characterize atmospheric moisture delivery in the presence of orographic blocking and low-level jets.” Both fields were extracted from ERA5 at $0.25^\circ \times 0.25^\circ$ resolution over the same domain used for the other covariates.

Manuscript change: VIMF and VIMFD added to the covariate set (Sect. 2.1). Physical justification for their inclusion stated. Manuscript reorganised as a coherent scientific narrative throughout.

Additional correction (identified during revision) — Zone 4 characterisation

During the revision process we identified an error in the previous version that described Zone 4 as “hyper-arid,” which is factually incorrect. Zone 4 covers the semi-arid Mediterranean transition of central Chile (approximately 33°S – 38°S), including the pre-cordillera of O’Higgins, Maule, and northern Bío-Bío, a region of significant but highly heterogeneous precipitation driven by strong orographic gradients between coastal and Andean stations.

The zone-level areal-CRPS for Zones 2 and 4 is reported as n/a in the revised Table 7, with a detailed explanation in Sect. 3.5.1 distinguishing two mechanistically distinct failure modes:

- **Zone 2** (desert transition, $\approx 25^\circ\text{S}$ – 30°S): dominance of zero-precipitation days causes the vine ensemble to occasionally generate positive draws on universally zero-observation days, inflating the areal mean and causing CRPS to diverge.
- **Zone 4** (semi-arid Mediterranean transition, $\approx 33^\circ\text{S}$ – 38°S): extreme within-zone orographic heterogeneity (coastal vs. Andean stations differ by up to an order of magnitude in annual totals) means that rare extreme draws at high-elevation stations — plausible given the GPD shape $\hat{\xi} = 0.921$ for State 1 at the Zone 4 medoid (the largest shape estimate in the dataset), fitted with only $n_{\text{exc}} = 34$ exceedances — can dominate the areal mean even when lower-elevation stations record near-zero totals. This is a case of unstable tail estimation propagating into the areal score, not a fundamental failure of the marginal or vine specification.

In both zones, interval-based diagnostics at the medoid level (Table 6: POC $\approx$ 96–97 %, AAD $\leq$ 0.07) confirm well-calibrated performance. The limitation lies in the choice of areal-CRPS as an evaluation metric for climatically heterogeneous zones. Recommendations for future work (orographically weighted areal averages, stratified sub-domain validation) are given in Sect. 4.4.

Manuscript change: “Hyper-arid” characterisation of Zone 4 removed throughout. Sect. 3.5.1 expanded with mechanistically distinct explanations for Zones 2 and 4. Table 7 CRPS/CRPSS entries for Zones 2 and 4 replaced by *n/a* with footnote. Sect. 4.4 updated with recommendations.

Response to Referee 2

We thank Referee 2 for a careful and constructive review. The comments focused primarily on methodological clarity and organisation; we have implemented all suggestions. All section references are to the **revised manuscript**.

Main Comment 1 — Methodology section hard to follow; step-by-step presentation requested

Referee comment: “Overall, I found the methodology section somewhat challenging to follow, likely due to the way it is currently organized. In particular, sections 2.3.2 through 2.4.2 were initially difficult to navigate, and I had to read them multiple times to fully grasp the workflow. My impression is that the clarity could be enhanced by presenting the methodology in a more step-by-step manner. Specifically, rather than interleaving statistical theory and its practical application multiple times, it might be beneficial to either present the statistical framework comprehensively at the beginning and then focus on the application, or to integrate theory and application in a more continuous, guided narrative ... For example, in section 2.3.4, the procedure was not immediately clear and only became understandable after reading subsequent sections.”

Authors’ response: We fully agree. The Methods section has been restructured as a sequential pipeline with four explicitly numbered and titled steps:

- **Step 1** — Zoning via DTW (Sect. 2.3.1)
- **Step 2** — State-conditional marginal specification (Sect. 2.3.2)
- **Step 3** — Transformation to copula space (Sect. 2.3.3)
- **Step 4** — Regular vine construction and model selection (Sect. 2.3.4)

A paragraph-map at the opening of Section 2 previews the complete pipeline, so the reader knows at every point where they are in the workflow. Within each step, statistical theory and its practical implementation are presented together in a continuous narrative rather than interleaved across separate sub-sections. The former “Experimental Setup” section — which broke the narrative by revisiting already-presented material in a different order — has been removed.

Regarding Section 2.3.4 specifically, which the referee found unclear on first reading: the revised version opens with a plain-language statement of what the vine achieves (“For d_z up to ~ 90 we represent $C_{z,s}$ by a Regular Vine (R-vine), decomposing the d_z -variate density into bivariate pair-copulas”), states the candidate family set and selection criterion, and then immediately provides Table 3 (vine structure summary for Zone 2) to ground the abstract formulation in a concrete example.

Manuscript change: Section 2 entirely reorganised with numbered Steps 1–4 and opening paragraph-map. Former Section 2.5 removed. Theory and application integrated in a continuous narrative throughout. Sect. 2.3.4 rewritten with Table 3 as a grounding concrete example.

Main Comment 2 — Selection of seven zones not justified early enough

Referee comment: “Additionally, regarding the methodology itself, I found it unclear why seven ‘zones’ were selected using the clustering approach. If a specific selection criterion was applied, it would be helpful to state it explicitly, ideally early in the methodology section, perhaps starting in section 2.3.1. This would provide the reader with a clear understanding of the rationale behind the choice and help contextualize subsequent analyses.”

Authors’ response: The selection criterion is now stated explicitly at the very beginning of Sect. 2.3.1, before the mathematical formulation. The revised text reads:

“The optimal number of zones is $k^* = 7$, selected via a consensus of five internal cluster validity indices (CVI): Silhouette, Calinski–Harabasz, Davies–Bouldin, Dunn, and COP (Table 1; full table in Supplement S2). The $G = 7$ solution achieves the best consensus rank across all five CVIs, aligns with established climatic subregions of continental Chile, and yields zones with $d_z \in [50, 90]$ stations and at least 200 training days per state, providing sufficient data for robust copula estimation.”

Table 1 (moved from the Supplement to the main text) shows the top five candidate configurations with their CVI scores, making the selection transparent and immediately verifiable by the reader.

Manuscript change: Zoning selection criterion (five-CVI consensus plus data sufficiency check) stated at the start of Sect. 2.3.1 before the mathematical formulation. Table 1 (CVI consensus table) moved from Supplement S2 to the main text.

Acknowledgement of positive assessment

The referee noted: *“Despite these points, I find this paper highly engaging and scientifically valuable. The study presents a good perspective on the topic, and the results are particularly insightful. The approach combines sophisticated statistical modeling with practical application, offering meaningful contributions. . . . With minor adjustments in clarity and organization, the manuscript would become even more accessible and impactful to readers.”*

We are grateful for this assessment and hope the structural improvements in the revised manuscript allow the scientific contribution to be more readily appreciated by all readers.

Complete Cross-Reference of Comments and Changes

The table below provides a complete cross-reference between each referee comment and the corresponding change in the revised manuscript, for the convenience of the Editor and both Referees.

Comment	Change in revised manuscript
Referee 1, GR (1) — Text repetition	Former Sect. 2.5 removed; roadmap paragraph added to Sect. 2; all overlapping passages between Sects. 2 and 3 eliminated.
Referee 1, GR (2) — Results in Supplement; S8 unreadable	Seven tables moved to main text; Supplement rewritten; former S8 replaced.
Referee 1, GR (3) — Tests without decision criteria	Retention/Decision criteria added to Sects. 2.4.1, 2.4.2, 2.4.3; KS/CvM behaviour at large n explained in Sect. 2.4.2.
Referee 1, GR (4) — N_{obs} vs. p not reported	$N_{\text{obs}}/p \approx 82$ in abstract, Table 4, Sect. 2.5.
Referee 1, GR (5) — GPD shape without CIs; bootstrap needed	95 % bootstrap CIs for $\hat{\xi}$ in Table 2 and Supp. Table S3; dry/wet tail contrast explained in Sect. 2.3.2.
Referee 1, GR (6) — No moisture variables; project-report style	VIMF and VIMFD added (Sect. 2.1); manuscript reorganised as coherent narrative.
Referee 1, Additional — Zone 4 incorrectly hyper-arid	Zone 4 correctly described as semi-arid/Mediterranean (Sect. 3.5.1); two distinct CRPS failure mechanisms explained; Table 7 entries replaced by n/a with footnote.
Referee 1, Additional — Figure/Table coherence	Table 6 updated to match validated production-run values consistent with Figure 6.
Referee 2, MC (1) — Methodology hard to follow; step-by-step needed	Sect. 2 reorganised with Steps 1–4 and paragraph-map; Sect. 2.3.4 rewritten with Table 3 as grounding example.
Referee 2, MC (2) — Seven zones: selection criterion not explicit	Five-CVI consensus criterion stated at start of Sect. 2.3.1; Table 1 moved from Supplement to main text.