

Responses to reviewer 1 comments

We appreciate the reviewer's efforts, time, and suggestions for improving the quality of our work. The comments and suggestions will increase our confidence and motivation to produce more scientific research. In the following sections, we have answered and responded to each reviewer's comment with a detailed explanation, and the corresponding text is modified in the revised manuscript. We believe these changes will answer the reviewers' concerns and improve our work quality. The response to each comment and suggestion is given below.

General Assessment

Dear Editor, thank you for providing me a chance to review the above-mentioned manuscript draft. I have critically reviewed it and provide constructive comments to authors, which I believe significantly improves the manuscript's quality and increases reader's and scientific community interest. This manuscript presents a comprehensive analysis of hydroclimatic responses to multiple Stratospheric Aerosol Injection (SAI) and solar radiation modification (SRM) experiments using CESM2-WACCM simulations under the GeoMIP framework. The study focuses on regional hydroclimatic changes over Asia, including the Tibetan Plateau, Central Asia, and monsoon-dominated regions, and introduces sulfur-normalized efficiency metrics and "available water" (AW) diagnostics.

The topic is relevant to current SRM research and climate intervention assessment. The use of multiple GeoMIP-style experiments (G6solar, G6sulfur, Geo-SAI, and G6-1.5K-SAI) strengthen this draft. However, the manuscript in its current form suffers from **methodological inconsistencies, insufficient justification of derived metrics, and overstated physical interpretations of statistical diagnostics**. Furthermore, its novelty relative to existing GeoMIP literature is not clearly established. I recommend **major revision** before the manuscript can be considered for publication.

Major Comments

Comment 1: Novelty and Positioning Within Existing SRM Literature

The manuscript does not sufficiently distinguish its contribution from prior SRM and GeoMIP studies, particularly:

- Simpson et al. (2019, 2020): hydroclimatic responses to SRM
- Vioni et al. (2020–2024): G6sulfur and G6-1.5K-SAI experimental design
- Rezaei et al. (2023–2025): CESM2-based hydrological responses under SAI
- MacMartin, Kravitz, and Tilmes group studies on GeoMIP SRM response patterns

The analysis framework used here (CESM2-WACCM, G6 experiments, precipitation–evapotranspiration diagnostics, and regional hydroclimatic partitioning) is **largely consistent with existing literature**, with incremental rather than transformative novelty.

Required revision:

The authors must clearly articulate:

- What is fundamentally new (not just regional application)

- Whether sulfur-normalized “efficiency” is a novel diagnostic or reinterpretation
- How this study extends beyond existing GeoMIP ensemble analyses

Without this, the manuscript risks being perceived as a **regional re-analysis of established GeoMIP outputs** rather than a novel contribution.

Response:

We thank the reviewer for this important comment. We agree that the original manuscript did not sufficiently clarify how this study extends beyond previous SRM and GeoMIP analyses. In the revised manuscript, we have substantially strengthened the novelty and positioning by explicitly distinguishing our framework from previous studies focusing mainly on global-scale or limited-variable assessments.

Specifically, we clarified that this study does not introduce a new SRM strategy, but provides an integrated regional assessment of hydroclimatic responses across four contrasting IPCC climate regions (WCA, ECA, TIB, and SAS) using nine hydroclimatic and land-surface variables within a single CESM2-WACCM6 framework. We further emphasized that, beyond previous analyses of mean-state changes, our study examines annual peaks, seasonal amplitude, peak timing, ensemble-member uncertainty, and raw versus deseasonalized statistical associations with climatic water balance (CWB).

We also revised the description of sulfur-normalized responses to clarify that they represent a descriptive scaling diagnostic rather than a new physical efficiency metric. The revised manuscript now explicitly states that sulfur-normalized values should not be interpreted as intrinsic efficiency, sensitivity, or a ranking of SRM strategies.

These revisions have been incorporated in the Introduction (Lines 315–372), where we clarify the research gaps, contributions, and interpretation of sulfur-normalized responses. We also revised the Model simulations and scenarios section (Lines 445–483 and Lines 601–655) to emphasize that differences among SRM experiments arise from their distinct GHG backgrounds, control objectives, injection strategies, and aerosol processes rather than representing a comparison of intervention superiority.

Comment 2: Definition and Interpretation of “Available Water (AW)”

The manuscript defines AW as: $AW = \text{Precipitation} - \text{Evapotranspiration}$

This is problematic because:

- It does not represent true hydrological availability
- It excludes groundwater, snow storage, glacier melt, and anthropogenic water use
- It is a **climatic water balance proxy**, not a water resource metric

Yet the manuscript repeatedly interprets AW as:

- “water availability”

- “hydrological recovery”
- “water resource response under SRM”

Required revision:

The authors must:

- Reframe AW strictly as a **climate diagnostic index**
- Remove or soften all implications of water resource sufficiency
- Clearly distinguish atmospheric fluxes from hydrological storage systems

Failure to do so leads to **systematic over-interpretation of model output**.

Response: We thank the reviewer for this important clarification. We agree that precipitation minus evapotranspiration (P-ET) should not be interpreted as a direct measure of hydrological or water-resource availability because it does not explicitly account for groundwater storage, snow and glacier storage and melt, surface-water storage, or anthropogenic water withdrawals and management. In the revised manuscript, we have therefore removed the term “available water (AW)” and replaced it throughout with “climatic water balance (P-ET)”. We now explicitly describe P-ET as a climate diagnostic representing the balance between atmospheric water input through precipitation and water loss through evapotranspiration, rather than as a measure of water-resource sufficiency. Statements referring to “water availability,” “hydrological recovery,” or “water-resource response” based solely on P-ET have been revised or softened accordingly. We have also clarified that P-ET characterizes atmospheric-land surface flux balance and does not directly represent hydrological storage or managed water resources. These revisions avoid over-interpretation of the modeled P-ET response. The revised text now reads:

Climatic water balance (P-ET) was calculated as precipitation minus actual evapotranspiration. This diagnostic represents the balance between atmospheric water input and evapotranspiration loss and is used here as an indicator of hydroclimatic moisture balance. It should not be interpreted as direct water-resource availability because it does not explicitly account for groundwater, surface-water and cryospheric storage, glacier melt, or anthropogenic water use and management.

Comment 3: Sulfur-Normalized “Efficiency” Metric is Insufficiently Justified

The manuscript introduces sulfur-normalized response metrics to evaluate SRM “efficiency.” However:

- The mathematical formulation is not clearly defined in the main text
- The assumption of linear scaling between SO₂ injection and hydrological response is not justified
- Differences in injection strategy (latitude, control algorithm, baseline scenario) are not accounted for

This raises concerns about **physical interpretability and comparability across experiments**.

Required revision:

The authors must:

- Explicitly define normalization equations
- Clarify that this is a **diagnostic scaling metric, not a causal efficiency measure**
- Discuss nonlinearity and model dependence of aerosol–radiation–hydrology responses

Response: We thank the reviewer for this important comment. We agree that the original manuscript did not sufficiently define the sulfur-normalized metric. We have therefore revised the Methods, Results, and Discussion to clarify both its mathematical formulation and its interpretation. For each hydroclimatic variable, region, and SAI experiment, the sulfur-normalized response is now explicitly defined as the simulated Sen-slope trend divided by the corresponding mean SO₂ injection rate over the analysis period. Thus, the metric is a descriptive ratio used to place responses from experiments with different sulfur burdens on a common scaling basis. Importantly, it is not interpreted as a causal efficiency, linear sensitivity, or derivative of hydroclimatic response with respect to sulfur injection. No linear relationship between SO₂ injection and hydroclimatic response is assumed, and the normalized values are not used to extrapolate responses to untested injection amounts.

We have also clarified that normalization by sulfur amount alone does not account for important differences among the experiments, including injection latitude and altitude, spatial distribution of injections, feedback-control strategy, background GHG scenario, aerosol microphysics and size evolution, stratospheric transport, and circulation responses. These factors can produce nonlinear and experiment-dependent aerosol–radiation–hydrology responses. Consequently, differences in the sulfur-normalized values are interpreted only as differences in the simulated response magnitude per unit mean injection rate under each experiment's realized configuration, rather than as intrinsic differences in physical efficiency. Because the analysis is based on CESM2-WACCM6, we further note that the magnitude of these ratios is model dependent. In response to the reviewer, we have replaced the term “efficiency” with “sulfur-normalized response” throughout the manuscript and added an explicit methodological equation and discussion of these limitations. The newly added text in the method section now read as:

To facilitate descriptive scaling among the aerosol-based SAI experiments with different sulfur burdens, the regional hydroclimatic trends were additionally normalized by the corresponding mean SO₂ injection rate. For variable v , region r , and SAI experiment s , the sulfur-normalized response was calculated as:

$$R_{v,r,s}^{norm} = \frac{\beta_{v,r,s}}{\bar{I}_s}, \quad \text{Eq. (1)}$$

Where, $\beta_{v,r,s}$ is the Sen-slope trend of variable v in region r under experiment s , and $\bar{I}_s$ is the mean SO₂ injection rate over 2055–2084 for that experiment over the analysis period (TgSO₂ yr⁻¹). The resulting quantity represents the simulated trend magnitude scaled by the mean sulfur injection rate.

This normalization is used only as a descriptive scaling diagnostic and is not interpreted as an efficiency metric, causal sensitivity, or derivative of climate response with respect to sulfur injection. Thus,

$$R_{v,r,s}^{norm} \neq \frac{\partial R_{v,r,s}}{\partial I_{SO_2}}.$$

Eq. (2)

Hydroclimatic responses to SAI may scale nonlinearly with sulfur injection because aerosol microphysics, particle growth, stratospheric transport, injection latitude and altitude, background climate state, circulation adjustments, and feedback-control strategies differ among experiments. Accordingly, sulfur normalization accounts only for differences in the magnitude of sulfur injection and does not remove differences arising from experimental design and model-dependent aerosol–climate responses.

Comment 4: Inconsistency in Baseline Scenarios Across Experiments

A critical methodological issue arises from the use of inconsistent baselines:

- G6-1.5K-SAI is compared against SSP2-4.5
- G6solar, G6sulfur, and Geo-SAI are compared against SSP5-8.5

Despite acknowledgment, the manuscript still:

- Interprets cross-experiment differences as physically meaningful efficiency contrasts
- Draws implicit ranking conclusions across SRM strategies

This introduces structural bias.

Required revision:

The authors must either:

- Recompute all diagnostics relative to a unified baseline, **or**
- Explicitly avoid cross-experiment efficiency ranking and restrict interpretation to within-scenario contrasts

Response: We thank the reviewer for highlighting this important methodological issue. We agree that the different SRM experiments considered here do not share a common GHG background or control objective and therefore should not be directly ranked in terms of physical effectiveness or efficiency. Rather than recomputing all diagnostics relative to a single baseline, which would not remove the underlying differences in experimental design and would address a different comparison question, we have adopted the reviewer’s second suggested approach and revised the manuscript to restrict interpretation primarily to within-experiment contrasts relative to the corresponding no-SRM GHG background.

Specifically, G6-1.5K-SAI is evaluated relative to SSP2-4.5, whereas G6solar, G6sulfur, and Geo-SAI are evaluated relative to SSP5-8.5. These pairings reflect the GHG background under which the corresponding SRM experiments are implemented. We now explicitly state that quantitative

differences across these experiment families should not be interpreted as rankings of SRM performance because they reflect differences in baseline forcing, temperature-control objectives, injection strategy, and sulfur burden.

This treatment is consistent with Vioni et al. (2023), who demonstrated that the choice of baseline can substantially influence the apparent climatic response to SAI and emphasized that different reference states answer different scientific questions. Accordingly, we have revised statements that implied that one SRM strategy is intrinsically “more effective,” “stronger,” or “more efficient” than another when those experiments use different baselines. Cross-experiment sulfur-normalized responses are now presented only as descriptive scaled responses and not as measures of comparative efficiency. The revised Methods, Results, and Discussion explicitly distinguish within-baseline intervention effects from descriptive differences among experiments and acknowledge that cross-experiment contrasts reflect the combined influence of baseline forcing and experimental design. For more clarification, the newly added text in the method section now read as below:

Because these SRM experiments differ in their underlying GHG pathways, control objectives, and intervention designs, their responses are not interpreted as a common-baseline ranking of SRM performance. Each experiment is evaluated primarily relative to its corresponding no-SRM GHG background: G6-1.5K-SAI relative to SSP2-4.5, and G6solar, G6sulfur, and Geo-SAI relative to SSP5-8.5. Where additional G6sulfur contrasts relative to SSP2-4.5 are reported, they are identified explicitly and treated as supplementary baseline-dependent comparisons rather than as a common reference for ranking the experiments. Cross-experiment contrasts are presented descriptively to characterize differences among experimental configurations rather than to infer that one SRM experiment is intrinsically stronger, more effective, or more efficient than another. This distinction is important because the choice of baseline can substantially influence the apparent magnitude and interpretation of SAI-induced climate responses (Vioni et al., 2023).

Comment 5: Overstatement of Causal Interpretation

Several statements in the manuscript imply causality beyond what is supported by the experimental design, such as:

- “SRM restores hydroclimate”
- “efficiency depends on sulfur loading”
- “regional water systems recover under SAI”

However, SRM impacts in CESM2-WACCM are:

- Emergent nonlinear responses
- Strongly modulated by internal variability and CO₂ physiological forcing
- Regionally heterogeneous and model-dependent

Required revision:

The discussion must:

- Replace deterministic language with probabilistic phrasing
- Avoid implying restoration or optimization
- Emphasize process uncertainty and structural model limitations

Response: We appreciate the reviewer’s important comment. We agree that some statements in the original manuscript could be interpreted as implying deterministic or causal relationships beyond what is supported by the experimental design. We have revised the manuscript by replacing terms such as “restore,” “recovery,” and “efficiency” with more cautious process-based descriptions, including “modify,” “partially offset,” and “descriptive scaled response.” We have also clarified that SRM responses are model-dependent, regionally heterogeneous, and influenced by background forcing, aerosol characteristics, and experimental configuration. These revisions are mainly reflected in the Abstract, Section 2.2 (scenario description), Discussion, and Conclusion sections.

We have also clarified that the simulated hydroclimatic responses emerge from coupled and potentially nonlinear interactions among radiative forcing, aerosol microphysics, atmospheric circulation, land–atmosphere feedbacks, internal climate variability, and persistent CO₂ physiological effects. Consequently, the magnitude and direction of the responses vary among regions and experiments and should not be interpreted as deterministic restoration or optimization of the hydroclimatic system.

In addition, we now explicitly acknowledge that the results are based on CESM2-WACCM6 and therefore remain subject to structural model uncertainty. The revised Discussion emphasizes that the reported responses represent model-simulated outcomes under the specific experimental configurations considered here, rather than universally applicable causal relationships. These changes also align with our revisions addressing the differing experiment baselines and the interpretation of sulfur-normalized responses.

Comment 6: Lack of Uncertainty Quantification

The manuscript does not adequately address:

- Ensemble spread across CESM2 members
- Internal climate variability
- Statistical significance of regional changes
- Robustness of seasonal shifts and extremes

Given SRM sensitivity to variability, this is a **major omission**.

Required revision:

Include:

- Ensemble mean vs spread analysis
- Confidence intervals for regional AW and precipitation changes
- Significance testing across ensemble members

Response: We appreciate the reviewer's important comment. To address the lack of uncertainty quantification, we have substantially revised the manuscript to explicitly incorporate ensemble spread, confidence intervals, and member-level significance testing. First, we revised Figure 2 so that solid lines represent ensemble-mean annual anomalies and shaded envelopes represent the spread across ensemble members, thereby illustrating the role of internal variability. Second, we added Table S7, which reports ensemble-mean percentage changes and associated 95% confidence intervals for regional precipitation and CWB. Third, we revised the seasonal-peak analysis by adding member-level significance testing and uncertainty estimates for peak hydroclimatic variables in Tables S6 and S8. We also clarified these procedures in the Methods section. Together, these revisions provide a more explicit assessment of ensemble spread, statistical significance, and the robustness of simulated hydroclimatic responses. The updated figures, tables and the corresponding text in the method section can now be read as:

2.5 Ensemble-member uncertainty and statistical analysis

To quantify internal climate variability, individual CESM2-WACCM6 ensemble members were retained separately prior to ensemble aggregation. For Figure 2, annual anomalies were calculated for each member relative to the corresponding 2015–2034 reference climatology, followed by a centered 5-year running mean. Solid lines represent the ensemble mean, while shaded envelopes denote ± 1 standard deviation across members. Linear trends over 2055–2084 were calculated for each member, and significance of the mean member trend was assessed using a one-sample t-test against zero ($p < 0.05$). For G6sulfur, additional contrasts were obtained by subtracting the contemporaneous SSP2-4.5 or SSP5-8.5 ensemble-mean anomaly from the G6sulfur anomaly. For regional precipitation and CWB, uncertainty was evaluated from ensemble-member period means. Ensemble-mean percentage changes and 95% confidence intervals were estimated using bootstrap resampling of ensemble members with 10,000 replicates, and future-versus-present differences were tested using Welch's two-sample t-test ($p < 0.05$) (Table S6). For annual-peak diagnostics, annual maxima were additionally calculated for each member and year, and significance was assessed across member period-mean peaks using Welch's test. Member-bootstrap 95% confidence intervals for peak percentage changes were likewise estimated using 10,000 replicates (Tables S5 and S7). Seasonal timing uncertainty was assessed separately using the member-year circular-statistical and bootstrap procedures described in Section 2.4. Because ensemble sizes are limited for some experiments, these uncertainty estimates characterize inter-member variability within CESM2-WACCM6 rather than structural model uncertainty.

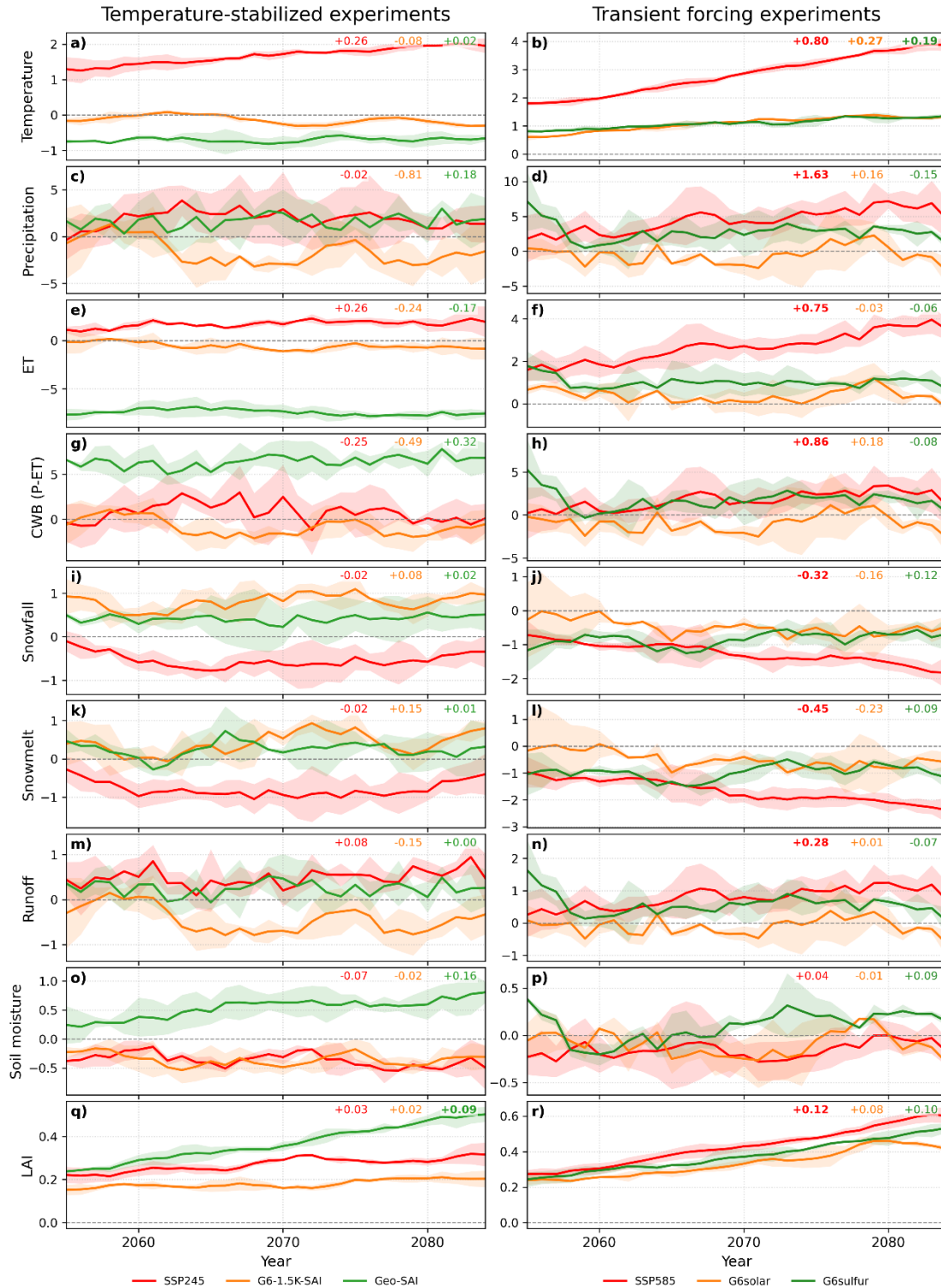


Figure 2. Annual mean anomalies of temperature, precipitation, ET, CWB, snowfall, snowmelt, runoff, soil moisture, and LAI during 2055–2084 relative to the corresponding 2015–2034 baseline under temperature-stabilized experiments (left column) and transient forcing experiments (right column). Solid lines denote ensemble-mean annual anomalies, and shaded envelopes indicate the spread across ensemble members after applying a 5-year running mean. Colored numbers indicate the mean linear trend over 2055–2084, and bold

values denote trends that are statistically significant at the 95% confidence level based on ensemble-member trend distributions.

Table S7. Percent change (%) in peak hydroclimatic variables during the future period (2055–2084) relative to the present-day baseline (2015–2034) across WCA, ECA, TIB, and SAS. Positive (negative) values indicate increases (decreases) relative to the present-day climatology. Bold values denote statistically significant changes at the 95% level ($p < 0.05$) based on Welch’s tests applied to ensemble-member period-mean annual peaks. G6-1.5K-SAI is evaluated relative to SSP2-4.5, whereas G6solar and Geo-SAI are evaluated relative to SSP5-8.5; G6sulfur is evaluated relative to both SSP2-4.5 and SSP5-8.5, as indicated by the scenario labels. Corresponding member-bootstrap 95% confidence intervals are provided in Table S8.

region	scenario	TMP	PCP	ET	CWB	Snowfall	Snowmelt	Runoff	SM	LAI
WCA	SSP2-4.5	6.68	-1.43	4.29	-0.21	-9.29	-19.29	0.74	-2.79	28.70
	SSP5-8.5	11.43	4.49	4.33	7.04	-24.94	-28.59	2.68	-2.10	42.53
	G6-1.5K-SAI	0.08	3.26	-2.49	-9.03	11.63	7.10	-0.53	-1.96	21.59
	Geo-SAI	-1.18	10.51	-3.81	35.42	3.68	3.89	5.42	2.09	38.30
	G6solar	4.20	6.21	-1.83	13.73	-6.50	-13.86	0.02	-0.58	33.81
	G6sulfur-SSP2-4.5	5.95	2.80	-0.83	-24.14	-11.60	-24.09	-2.36	-4.70	30.11
	G6sulfur-SSP5-8.5	4.83	6.17	-0.68	7.59	-17.96	-22.81	3.09	-1.35	35.95
ECA	SSP2-4.5	7.58	1.57	0.95	7.54	7.01	-2.89	3.43	-0.77	10.47
	SSP5-8.5	13.35	3.23	2.17	22.80	4.51	-11.81	6.54	2.52	21.49
	G6-1.5K-SAI	-1.47	-0.58	-2.11	-49.19	17.84	19.48	-2.84	0.51	7.15
	Geo-SAI	-1.12	8.91	0.39	30.45	1.65	19.99	9.61	5.43	33.66
	G6solar	4.14	1.67	-0.84	37.28	10.64	1.52	4.88	3.02	24.39
	G6sulfur-SSP2-4.5	6.29	8.21	0.10	-31.47	6.19	-4.68	2.69	-0.40	24.33
	G6sulfur-SSP5-8.5	5.62	8.37	0.34	65.39	0.67	-3.44	8.38	2.21	27.63
TIB	SSP2-4.5	13.26	3.29	5.88	6.69	-6.03	-14.63	6.15	-3.67	19.06
	SSP5-8.5	24.17	13.02	2.55	18.76	-9.83	-26.27	12.38	-1.82	26.31
	G6-1.5K-SAI	1.13	-5.43	0.51	-4.17	18.71	5.91	-2.44	-3.36	7.44
	Geo-SAI	3.91	-0.69	-2.74	2.38	9.60	3.96	1.23	1.68	12.70
	G6solar	10.71	1.04	-2.68	3.74	-2.18	-9.93	0.99	0.17	16.46
	G6sulfur-SSP2-4.5	13.44	10.51	-0.94	19.06	-0.08	-9.48	14.59	-1.25	19.99
	G6sulfur-SSP5-8.5	12.32	9.75	-0.10	15.26	-0.47	-10.68	11.22	1.09	20.03
SAS	SSP2-4.5	5.94	2.07	6.52	5.96	-29.71	-25.03	9.86	0.39	16.29
	SSP5-8.5	9.89	7.53	4.74	9.44	-49.01	-44.31	12.08	0.12	32.53
	G6-1.5K-SAI	1.88	-3.80	2.35	-12.20	-4.74	-2.03	-1.52	-1.09	21.09
	Geo-SAI	1.57	4.00	-12.41	8.91	-5.98	-7.99	8.44	-0.47	53.73
	G6solar	4.93	2.17	2.61	2.85	-24.14	-21.68	5.35	-0.40	40.09
	G6sulfur-SSP2-4.5	3.13	6.76	20.29	-3.25	-24.12	-21.25	15.22	1.82	41.45
	G6sulfur-SSP5-8.5	4.81	9.24	2.65	12.81	-24.99	-19.82	15.55	0.77	45.06

Table S5. Ensemble-mean percentage changes and associated 95% confidence intervals for regional precipitation and CWB during 2055–2084 relative to 2015–2034 across WCA, ECA, TIB, and SAS under GHG and SRM scenarios. Values are reported as ensemble-mean percentage changes with 95% confidence intervals

in brackets, reflecting variability across ensemble members; asterisks indicate statistically significant changes at the 95% level ($p < 0.05$) based on ensemble-member distributions. G6-1.5K-SAI is evaluated relative to SSP2-4.5, whereas G6solar and Geo-SAI are evaluated relative to SSP5-8.5; G6sulfur is evaluated relative to both SSP2-4.5 and SSP5-8.5, as indicated by the scenario labels.

Variable	Scenario	WCA	ECA	TIB	SAS
Precipitation	SSP2-4.5	-0.16 [-2.59, 2.11]	6.96 [3.23, 10.44]*	5.30 [3.29, 7.25]*	2.71 [-1.16, 5.87]
	SSP5-8.5	6.19 [2.47, 10.62]*	12.38 [10.56, 14.37]*	11.45 [8.73, 14.27]*	5.36 [1.46, 8.95]
	G6-1.5K-SAI	0.21 [-2.60, 3.11]	-1.69 [-5.57, 1.87]	-2.92 [-4.16, -1.70]*	-3.56 [-6.21, -1.41]
	G6solar	-0.52 [-3.92, 3.69]	4.28 [1.03, 7.63]	-0.86 [-2.98, 1.51]	-1.46 [-7.03, 4.11]
	G6sulfur-SSP2-4.5	-4.04 [-5.71, -2.47]*	4.77 [0.63, 8.87]	8.19 [5.57, 10.79]	7.14 [3.26, 11.14]
	G6sulfur-SSP5-8.5	-0.67 [-3.60, 3.11]	4.29 [1.14, 7.54]	6.08 [2.97, 9.31]	6.15 [1.43, 10.95]
	Geo-SAI	4.09 [0.41, 8.39]	4.49 [3.05, 6.11]*	-0.98 [-2.89, 1.11]	3.63 [-0.82, 8.17]
CWB	SSP245	-554.30 [-8513.86, -92.68]	27.07 [-15.48, 92.88]	6.89 [4.46, 9.35]*	3.94 [-3.61, 10.50]
	SSP585	-8.99 [-30.06, 7.48]	45.10 [34.79, 57.06]*	15.20 [10.58, 19.88]*	5.41 [-5.86, 18.23]
	G6-1.5K-SAI	-613.86 [-5911.8, 766.7]*	27.62 [-4.09, 79.40]	-1.99 [-4.77, 0.40]	-10.77 [-17.20, -5.31]
	G6solar	-9.21 [-33.77, 10.60]	34.81 [21.90, 49.19]	0.41 [-2.81, 3.81]	-12.54 [-29.03, 5.59]
	G6sulfur-SSP245	-1937.16 [-20556.3, -833.05]*	46.46 [9.99, 102.89]	16.06 [11.26, 21.00]	-28.42 [-35.30, -20.88]*
	G6sulfur-SSP585	-16.68 [-38.27, 0.08]	31.10 [15.27, 48.73]	10.89 [5.79, 16.13]	13.34 [-0.84, 29.29]
	Geo-SAI	-139.21 [-152.21, -127.76]*	53.00 [42.88, 65.10]*	2.94 [-0.07, 6.19]	64.88 [45.97, 85.48]*

Table S6. Ensemble-member uncertainty in percentage changes in peak hydroclimatic variables during 2055–2084 relative to 2015–2034 across WCA, ECA, TIB, and SAS. Values are reported as ensemble-mean percentage changes with member-bootstrap 95% confidence intervals in brackets. Asterisks indicate statistically significant future-versus-present differences at the 95% level ($p < 0.05$) based on Welch’s tests applied to ensemble-member period-mean annual peaks. The 95% confidence intervals were estimated by bootstrap resampling of ensemble members using 10,000 replicates. G6-1.5K-SAI is evaluated relative to SSP2-4.5, whereas G6solar and Geo-SAI are evaluated relative to SSP5-8.5; G6sulfur is evaluated relative to both SSP2-4.5 and SSP5-8.5, as indicated by the scenario labels. The confidence intervals characterize finite inter-member variability within CESM2-WACCM6 and do not represent structural model uncertainty.

region	scenario	TMP	PCP	ET	CWB	Snowfall	Snowmelt	Runoff	SM	LAI
WCA	SSP2-4.5	6.68 [5.77, 7.61]*	-1.43 [-7.63, 6.01]	4.29 [3.39, 5.48]*	-0.21 [-20.10, 18.06]	-9.29 [-19.19, -1.06]*	-19.29 [-21.33, -15.60]*	0.74 [-3.74, 4.14]	-2.79 [-4.29, -1.12]*	28.70 [24.72, 31.77]*
	SSP5-8.5	11.43 [10.94, 11.90]*	4.49 [-1.65, 13.28]	4.33 [2.56, 6.20]*	7.04 [-11.09, 25.77]	-24.94 [-32.93, -16.45]*	-28.59 [-31.20, -25.94]*	2.68 [-1.98, 6.16]	-2.10 [-3.76, -0.15]	42.53 [38.07, 47.31]*
	G6-1.5K-SAI	0.08 [-0.58, 0.71]	3.26 [-4.61, 10.77]	-2.49 [-3.50, 0.07]	-9.03 [-27.06, 10.15]	11.63 [0.91, 24.41]	7.10 [4.40, 12.90]*	-0.53 [-5.46, 1.73]*	-1.96 [-3.43, -0.44]*	21.59 [18.24, 24.60]*
	Geo-SAI	-1.18 [-1.64, -0.30]	10.51 [1.88, 22.54]	-3.81 [-4.67, -1.73]*	35.42 [10.14, 59.30]*	3.68 [-8.01, 11.76]	3.89 [0.53, 8.36]*	5.42 [1.11, 13.80]	2.09 [0.05, 4.30]	38.30 [34.53, 42.66]*
	G6solar	4.20 [3.90, 4.69]*	6.21 [-1.36, 15.47]	-1.83 [-3.29, -0.28]	13.73 [-9.13, 30.43]	-6.50 [-17.11, 3.22]	-13.86 [-17.04, -6.41]	0.02 [-4.11, 4.84]	-0.58 [-2.26, 1.45]	33.81 [29.96, 37.82]*
	G6sulfur-SSP2-4.5	5.95 [5.02, 7.14]*	2.80 [-5.21, 13.90]	-0.83 [-1.32, -0.39]*	-24.14 [-36.12, -5.01]*	-11.60 [-22.39, 0.76]*	-24.09 [-25.99, -17.34]*	-2.36 [-7.29, 1.02]	-4.70 [-6.12, -2.99]*	30.11 [25.70, 34.18]*
ECA	G6sulfur-SSP5-8.5	4.83 [4.11, 5.80]	6.17 [-1.38, 16.99]	-0.68 [-1.57, 0.61]	7.59 [-14.04, 32.60]	-17.96 [-26.69, -6.49]*	-22.81 [-25.48, -16.87]*	3.09 [-1.20, 6.67]	-1.35 [-3.02, 1.12]	35.95 [31.20, 41.10]*
	SSP2-4.5	7.58 [6.78, 8.38]*	1.57 [-3.21, 7.54]	0.95 [-0.27, 3.01]	7.54 [-9.13, 30.11]	7.01 [-2.38, 18.86]*	-2.89 [-7.56, 4.29]	3.43 [-2.98, 11.90]	-0.77 [-1.77, 0.74]	10.47 [7.58, 13.36]*
	SSP5-8.5	13.35 [12.69, 14.12]*	3.23 [-1.51, 10.58]*	2.17 [1.00, 3.38]*	22.80 [-3.66, 56.74]*	4.51 [-6.17, 20.60]	-11.81 [-19.15, -7.39]*	6.54 [0.61, 13.25]*	2.52 [0.80, 3.89]*	21.49 [18.86, 24.22]*
	G6-1.5K-SAI	-1.47 [-2.06, -0.83]*	-0.58 [-5.55, 7.35]	-2.11 [-3.65, -0.27]	-49.19 [-55.32, -27.77]*	17.84 [6.98, 31.58]*	19.48 [11.36, 27.81]*	-2.84 [-9.04, 4.42]	0.51 [-1.14, 1.92]	7.15 [3.29, 11.65]
	Geo-SAI	-1.12 [-1.90, -0.20]	8.91 [2.88, 17.80]*	0.39 [-0.46, 1.21]	30.45 [0.19, 80.18]*	1.65 [-8.41, 17.28]	19.99 [12.07, 24.81]*	9.61 [2.48, 20.63]	5.43 [3.88, 6.32]*	33.66 [30.90, 35.60]*
	G6solar	4.14 [3.66, 4.79]*	1.67 [-3.86, 10.25]	-0.84 [-2.28, 0.86]	37.28 [0.76, 89.31]	10.64 [-2.05, 22.07]	1.52 [-6.04, 9.40]	4.88 [-1.55, 11.07]	3.02 [1.55, 4.57]*	24.39 [20.84, 29.03]
TIB	G6sulfur-SSP2-4.5	6.29 [5.65, 6.81]*	8.21 [2.51, 19.93]	0.10 [-0.94, 1.74]	-31.47 [-40.99, -6.16]*	6.19 [-4.24, 17.18]	-4.68 [-9.97, 0.34]*	2.69 [-3.73, 13.25]	-0.40 [-1.68, 1.62]	24.33 [19.83, 28.59]*
	G6sulfur-SSP5-8.5	5.62 [5.08, 6.24]*	8.37 [2.59, 19.37]	0.34 [-0.61, 2.22]	65.39 [21.39, 106.63]*	0.67 [-10.87, 13.93]	-3.44 [-9.88, 0.37]*	8.38 [1.73, 19.16]	2.21 [0.74, 4.32]*	27.63 [23.71, 31.49]
	SSP2-4.5	13.26 [12.12, 14.34]*	3.29 [-0.02, 7.26]	5.88 [4.99, 6.76]*	6.69 [2.09, 13.37]*	-6.03 [-11.97, 0.57]*	-14.63 [-17.24, -11.22]*	6.15 [1.20, 12.25]*	-3.67 [-4.38, -1.89]*	19.06 [18.45, 19.96]*
	SSP5-8.5	24.17 [23.29, 25.17]*	13.02 [8.78, 20.43]*	2.55 [1.37, 3.44]*	18.76 [12.16, 30.09]*	-9.83 [-16.35, -2.24]*	-26.27 [-30.33, -20.85]*	12.38 [7.02, 21.33]*	-1.82 [-2.60, -0.01]	26.31 [25.12, 27.72]*
	G6-1.5K-SAI	1.13 [0.21, 2.03]	-5.43 [-8.18, -1.72]*	0.51 [-0.34, 1.23]	-4.17 [-8.76, 1.48]	18.71 [10.32, 27.53]*	5.91 [2.62, 9.77]	-2.44 [-6.99, 1.82]	-3.36 [-4.46, -2.22]*	7.44 [7.27, 7.66]*
	Geo-SAI	3.91 [2.93, 4.96]*	-0.69 [-2.88, 2.67]	-2.74 [-3.93, -1.58]*	2.38 [-1.31, 8.62]	9.60 [2.12, 20.71]*	3.96 [-2.59, 10.67]	1.23 [-1.45, 5.44]	1.68 [0.68, 2.76]	12.70 [11.22, 14.66]*
SAS	G6solar	10.71 [9.61, 12.01]*	1.04 [-1.46, 8.42]	-2.68 [-3.91, -1.97]*	3.74 [-0.16, 15.22]	-2.18 [-8.85, 4.60]	-9.93 [-14.75, -3.87]*	0.99 [-1.70, 9.35]	0.17 [-0.82, 1.88]	16.46 [14.92, 18.13]*
	G6sulfur-SSP2-4.5	13.44 [12.43, 14.35]*	10.51 [6.93, 16.61]	-0.94 [-1.77, 0.14]	19.06 [11.76, 28.02]	-0.08 [-8.34, 4.39]*	-9.48 [-11.42, -6.30]*	14.59 [8.06, 20.88]	-1.25 [-2.36, 0.76]	19.99 [19.39, 20.54]*
	G6sulfur-SSP5-8.5	12.32 [11.54, 13.21]*	9.75 [6.55, 16.20]	-0.10 [-1.34, 1.20]	15.26 [10.05, 25.77]	-0.47 [-7.26, 4.41]	-10.68 [-15.08, -4.76]*	11.22 [7.89, 18.75]	1.09 [0.08, 3.18]	20.03 [18.82, 21.35]*
	SSP2-4.5	5.94 [5.19, 6.67]*	2.07 [-2.36, 6.16]	6.52 [3.85, 9.34]*	5.96 [-1.52, 10.00]	-29.71 [-35.84, -23.56]*	-25.03 [-28.45, -20.49]*	9.86 [5.72, 17.05]*	0.39 [-1.23, 1.86]	16.29 [13.01, 18.48]*
	SSP5-8.5	9.89 [9.15, 10.42]*	7.53 [5.00, 12.66]*	4.74 [2.83, 5.90]*	9.44 [5.85, 16.38]*	-49.01 [-53.60, -38.43]*	-44.31 [-48.50, -39.08]*	12.08 [8.54, 17.22]*	0.12 [-0.74, 1.17]	32.53 [29.53, 33.84]*
	G6-1.5K-SAI	1.88 [1.32, 2.31]*	-3.80 [-7.24, 1.14]	2.35 [-0.00, 4.85]	-12.20 [-17.16, -5.93]	-4.74 [-12.78, 7.50]	-2.03 [-6.40, 4.57]	-1.52 [-5.60, 3.29]	-1.09 [-2.64, 0.39]	21.09 [17.72, 23.24]*
SAS	Geo-SAI	1.57 [0.95, 2.22]*	4.00 [0.67, 8.91]	-12.41 [-14.05, -9.28]*	8.91 [4.32, 15.44]	-5.98 [-17.08, 9.08]	-7.99 [-15.22, -0.22]	8.44 [3.12, 13.03]	-0.47 [-1.35, 1.19]	53.73 [50.36, 55.07]*
	G6solar	4.93 [4.29, 5.47]*	2.17 [-0.49, 10.21]	2.61 [0.61, 4.48]	2.85 [-0.83, 14.17]	-24.14 [-31.95, -11.68]*	-21.68 [-27.78, -15.46]*	5.35 [0.82, 14.20]	-0.40 [-1.51, 1.18]	40.09 [36.77, 41.85]*
	G6sulfur-SSP2-4.5	3.13 [2.56, 3.81]*	6.76 [2.95, 11.33]*	20.29 [16.23, 23.24]*	-3.25 [-8.17, 2.52]	-24.12 [-31.23, -17.16]*	-21.25 [-24.77, -16.04]*	15.22 [10.45, 20.74]*	1.82 [-0.37, 3.64]	41.45 [36.47, 42.82]*
	G6sulfur-SSP5-8.5	4.81 [4.12, 5.54]*	9.24 [6.40, 13.91]*	2.65 [0.51, 5.24]	12.81 [8.78, 20.47]*	-24.99 [-31.16, -15.92]*	-19.82 [-25.85, -14.22]*	15.55 [11.07, 21.38]*	0.77 [-0.36, 2.28]	45.06 [41.81, 46.25]*

Comment 7: Over-Interpretation of Regression (MLR) Results

The multiple linear regression (MLR) framework shows:

- Extremely high R^2 values (>0.99)
- Weak justification of lag selection (1–4 months)
- Potential multicollinearity issues

This suggests the model may be:

- Overfitting deterministic seasonal structure
- Capturing correlation rather than physical causality

Required revision:

- Provide residual diagnostics
- Test alternative lag structures systematically
- Clarify physical interpretation limitations of regression outputs

Response: We thank the reviewer for this important and detailed comment. We agree that the original MLR analysis required a more rigorous treatment of the part–whole relationship between CWB and its defining components, multicollinearity, temporal dependence, seasonality, model selection, and statistical interpretation. We have therefore substantially revised the entire MLR workflow.

First, precipitation and evapotranspiration were removed from the interpretive MLR models because they are the defining components of CWB and their inclusion produced a strong part–whole relationship that contributed to the previously very high R^2 values. Their relationship with CWB is now treated separately as a component-reconstruction diagnostic rather than as evidence of independent explanatory importance. The revised raw LMG figure consequently includes only the retained interpretive predictors—temperature, snowfall, snowmelt, runoff, soil moisture, and LAI—and explicitly states that LMG contributions represent shares of explained variance rather than causal effects.

Second, we replaced the previous manually chosen lag structure with a systematic model-selection procedure. For each retained predictor, temporal positions from 0 to 4 months were evaluated, with only one temporal position allowed for each predictor in a candidate model. All possible lag combinations were compared using rolling-origin time-series cross-validation. Models with a maximum variance inflation factor (VIF) greater than 5 were excluded, and the VIF-acceptable model with the minimum cross-validated RMSE was selected. Where the complete predictor set could not satisfy the predefined VIF threshold, scenario–region-specific predictor refinement was performed rather than relaxing the VIF criterion. The complete selected lag structures and associated R^2 , adjusted R^2 , CV R^2 , CV-RMSE, and maximum VIF values are now reported in Table S9.

Third, to examine whether the strong monthly relationships were primarily associated with the deterministic annual cycle, we repeated the complete lag-selection and model-fitting workflow after removing the calendar-month climatology from CWB and all candidate predictors. The raw-versus-deseasonalized comparison is now reported explicitly in Table S8. Deseasonalization generally reduced predictive performance, confirming that shared seasonality accounts for an important portion of the raw monthly relationships; however, appreciable anomaly-scale predictive structure remained in several region–scenario combinations. For example, CV- R^2 changes from 0.955 to 0.700 for G6-1.5K-SAI in WCA, from 0.905 to 0.529 for SSP5-8.5 in SAS, and from 0.859 to 0.545 for Geo-SAI in SAS. The corresponding deseasonalized LMG results are provided separately to show how the hierarchy of predictor importance changes once the climatological annual cycle is removed.

Fourth, the revised analysis includes substantially expanded regression diagnostics. Final models were assessed using VIF, residual-versus-fitted and Q–Q diagnostics, the Breusch–Pagan test, two-sided Durbin–Watson test, Ljung–Box test, residual autocorrelation functions, Cook's distance, and Bonferroni outlier screening. Because residual heteroscedasticity and/or serial dependence remained in some raw monthly models, coefficient inference was additionally evaluated using Newey–West heteroscedasticity- and autocorrelation-consistent (HAC) standard errors. Bonferroni-significant observations, where present, were examined through sensitivity analyses rather than being automatically removed from the primary analysis.

Finally, we revised the interpretation of the MLR and LMG results throughout the manuscript. The regression results are now described as statistical associations and relative contributions to explained variance, rather than as causal “drivers,” “controls,” or physical attribution. The raw monthly models are retained as the primary descriptive analysis of total monthly CWB variability, whereas the deseasonalized models are presented as a sensitivity analysis that distinguishes relationships associated with the climatological annual cycle from those persisting at the anomaly scale.

We thank the reviewer for this comment, which led to a substantially more rigorous and transparent MLR analysis. The new text in the method section can now be read:

2.6 Multiple linear regression (MLR) framework

MLR was used to characterize region- and scenario-specific statistical associations between monthly CWB and selected hydroclimatic and land-surface variables across all regions and experiments (Hussain et al., 2024a). MLR is widely applied in climate and environmental studies to quantify statistical relationships between hydroclimatic variables and identify the relative contribution of potential predictors (Wilks and Wilks, 2006; Hussain et al., 2025; Hussain et al., 2024b). Because precipitation and ET are defining components of CWB, they were excluded from the interpretive MLR models to avoid part–whole relationships and artificially inflated explanatory power. Their relationship with CWB was examined separately as a component-reconstruction diagnostic rather than as evidence of independent predictor importance. The remaining candidate predictors included temperature, snowfall, snowmelt, runoff, SM, and LAI. These predictors were screened using pairwise correlations, hierarchical clustering,

and variance inflation factors (VIF) to reduce predictor redundancy and multicollinearity (O'Brien, 2007). CWB was excluded from predictor-screening diagnostics, and a maximum VIF threshold of 5 was maintained throughout the analysis. Predictor sets were refined where strong redundancy prevented models from satisfying this criterion.

Temporal dependence was treated systematically rather than by manually selecting several autocorrelation-based lags. For each retained predictor X_j , temporal positions from the current month to a four-month lag were evaluated, allowing only one lag per predictor in each candidate model. This accounts for delayed hydroclimatic responses and follows established time-series modeling approaches (Vicente-Serrano et al., 2010; Shumway and Stoffer, 2006). Thus, the general model was

$$CWB_t = \beta_0 + \sum_{j=1}^p \beta_j X_{j,t-\ell_j} + \varepsilon_t, \quad \ell_j \in \{0, 1, 2, 3, 4\} \quad \text{Eq. (5)}$$

where, CWB_t is climatic water balance at month t , $X_{j,t-\ell}$ is predictor j at its candidate temporal position ℓ , β_0 is the intercept, β_j is the corresponding regression coefficient, and ε_t is the residual error.

All possible combinations of the permitted temporal positions were evaluated using rolling-origin time-series cross-validation. Candidate models with maximum VIF > 5 was excluded, and the VIF-acceptable model with the minimum cross-validated root-mean-square error (CV-RMSE) was selected. Where necessary, scenario–region-specific predictor refinement was performed rather than relaxing the VIF threshold. The selected lag structures, R^2 and adjusted R^2 , CV-RMSE, and maximum VIF are reported in Table S4.

To assess the influence of the climatological annual cycle, the complete MLR procedure was repeated using calendar-month anomaly series. For each variable, the corresponding monthly climatology was removed before repeating predictor screening, lag selection, VIF filtering, cross-validation, and final model fitting. Raw monthly models were retained as the primary analysis of total CWB variability, whereas deseasonalized models were used as a sensitivity analysis of relationships that persist beyond the seasonal cycle. Raw and deseasonalized CV-R2 values are compared in Table S3, and the corresponding model specifications and performance statistics are provided in Table S4.

Model adequacy was evaluated using residual-versus-fitted plots, Q–Q diagnostics, heteroscedasticity tests, autocorrelation diagnostics, influence measures, and Bonferroni outlier screening. Bonferroni-significant observations were evaluated through sensitivity analyses rather than automatically removed from the primary models, thereby preserving the original monthly sequence used for lag construction.

Where residual heteroscedasticity or serial dependence remained, coefficient uncertainty was additionally assessed using Newey–West heteroscedasticity- and autocorrelation-consistent (HAC) standard errors.

Predictor importance was quantified using the LMG metric (Lindeman et al., 1980), which decomposes model R^2 into the average contribution of each predictor across all possible orders of entry. LMG values were expressed as percentages of total explained variance and therefore represent relative explanatory importance rather than causal effects or the direction of the relationship. The sign of each predictor–CWB association was determined separately from the corresponding regression coefficients. Raw monthly LMG results are presented in Figure 7, while deseasonalized LMG results are provided in Figure S2. Accordingly, MLR and LMG results are interpreted as statistical associations and relative contributions to explained variance rather than physical attribution or causal control of CWB.

Minor Comments

Comment 1: Seasonal decomposition methodology; Harmonic regression approach is appropriate but insufficiently documented; No discussion of phase uncertainty or robustness across regions

Response: We thank the reviewer for this important comment. We have substantially revised Section 2.4 to provide a more complete and reproducible description of the seasonal-cycle analysis. The revised text now reads:

2.4 Changes and shifts in seasonal dynamics

We evaluated changes in seasonal amplitude and peak timing for nine hydroclimatic and land-surface variables. Seasonal characteristics were evaluated for the future period (2055–2084) relative to present-day conditions (2015–2034) across regions. To preserve both ensemble and interannual variability, the analysis was performed separately for each ensemble member, year, and region, rather than deriving seasonal metrics solely from a single multi-year climatological cycle. For each member and year, area-weighted regional monthly means were first calculated using latitude-based cosine weighting. A complete 12-month annual cycle was required for inclusion in the analysis. Following the seasonal-amplitude definition of Rezaei et al. (2025a), the annual seasonal amplitude was calculated as half the difference between the maximum and minimum monthly values:

$$Amp_{y,e,r} = \frac{X_{\max,y,e,r} - X_{\min,y,e,r}}{2}, \quad \text{Eq. (3)}$$

where, $X_{\max,y,e,r}$ and $X_{\min,y,e,r}$ denote the maximum and minimum monthly values, respectively, for year y , ensemble member e , and region r . Annual peak timing was defined directly as the calendar month in which $X_{\max,y,e,r}$ occurred, so both amplitude and peak timing were first determined independently for each member-year realization before being summarized across the analysis period. In

addition to the direct peak-to-trough diagnostics, harmonic regression was used to characterize the shape and phase of the annual cycle and to evaluate the robustness of a single annual peak (North et al., 2021; Ramanathan et al., 2020). For each complete 12-month cycle, the first annual harmonic was fitted as:

$$X(m) = \beta_0 + a \cos\left[\frac{2\pi(m-1)}{12}\right] + b \sin\left[\frac{2\pi(m-1)}{12}\right] + \varepsilon_m, \quad \text{Eq. (4)}$$

where, $m = 1, \dots, 12$ denotes calendar month, β_0 is the annual mean component, a and b are the cosine and sine coefficients, respectively and ε_m is the residual. The harmonic amplitude was calculated as $\sqrt{a^2 + b^2}$, while the fitted phase provided a continuous estimate of the timing of the seasonal maximum.

The coefficient of determination (R^2) was used only to assess how well the annual harmonic represented the seasonal cycle and was not used as the primary definition of peak timing.

Following the seasonal-cycle classification framework of (Rezaei et al., 2025a), based on Berghuijs and Woods (2016) and (Knoben et al., 2019), 12-month and 6-month harmonic models were fitted to each member-year-regional seasonal cycle to diagnose unimodal and bimodal seasonality. The model with the lower mean absolute error was considered the better representation, and the fraction of cycles favoring the 12-month harmonic was used as an indicator of unimodal-seasonality robustness. An ordinary harmonic formulation was applied because some variables may take negative values. Harmonic analysis was therefore treated as a complementary diagnostic of seasonal shape and phase rather than as the primary basis for identifying the annual peak.

Peak timing was summarized across years and ensemble members using circular statistics. The circular mean represented the characteristic peak month, while circular standard deviation quantified interannual and inter-member variability. Bootstrap resampling ($n=5000$) was used to estimate 95% confidence intervals for mean peak timing and future-minus-present shifts. Modal peak month and the fraction of peaks occurring within ± 1 month of the circular mean were used as additional measures of timing consistency. Future changes were evaluated relative to the corresponding present-day baseline, and Welch's two-sample t-test ($p < 0.05$) was applied to member-year distributions to identify significant peak-timing shifts, supported by circular spread and bootstrap confidence intervals. Peak months shown in figures were rounded to the nearest calendar month, whereas continuous circular estimates and uncertainty statistics were retained for the supplementary analysis.

Seasonal-cycle figures were based on raw monthly values without detrending or demeaning. As a sensitivity test, annual-mean-removed cycles were also examined by subtracting each year's mean from its monthly values. Because subtracting the annual mean does not alter either the month of the annual maximum or the peak-to-trough difference, the primary peak-timing and seasonal-amplitude diagnostics are unchanged by this transformation.

Comment 2: Redundancy in results presentation; Similar interpretations repeated across variables (precipitation, AW, runoff); A synthesis figure or consolidated discussion is recommended

Response: We thank the reviewer for this important suggestion. We agree that some interpretations across hydroclimatic variables were repetitive and could reduce the clarity of the main findings. In the revised manuscript, we have streamlined the Results from section 3.1-3.5, by reducing repeated descriptions of individual variables and emphasizing the dominant regional patterns and scenario-dependent responses. We have also added a more integrated synthesis in the discussion of hydroclimatic responses, highlighting the coupled behavior among precipitation, evapotranspiration, climatic water balance (CWB), runoff, and soil moisture rather than interpreting each variable independently. These revisions improve the overall readability while preserving the key scientific results and regional differences.

Comment 3: Terminology clarity; “Hydroclimatic efficiency” should be carefully defined or reduced in usage; “Restoration” language should be avoided unless physically justified

Response: We thank the reviewer for this important observation. We agree that the terminology required further clarification to avoid overinterpretation. In the revised manuscript, we have reduced the use of the term “hydroclimatic efficiency” and replaced it with more precise descriptions of the simulated hydroclimatic responses where appropriate. We have also clarified that sulfur-normalized responses represent a descriptive scaling diagnostic rather than a measure of physical efficiency or intervention performance. Furthermore, we have removed or revised statements involving “restoration” and replaced them with more appropriate terms such as “modification,” “adjustment,” or “movement toward present-day conditions” when describing simulated responses. These changes ensure that the interpretations remain consistent with the physical meaning of the model experiments.