Author Comment (AC) for RC2

Dear Anonymous Referee #2,

Thank you very much for your positive and constructive review. We are particularly pleased that you found the paper interesting and useful as a reference for impact modelers. We appreciate the time and effort you invested in providing detailed feedback, which has helped us identify important areas for improvement in scientific precision, clarity, logical flow, and overall structure.

We have carefully considered all your comments. In particular, we acknowledge your concern that some arguments were not always scientifically precise or clearly connected to the main objectives of the paper. We have revised the manuscript accordingly to strengthen the scientific rigor and to ensure that all key arguments are clearly linked to the central aim of evaluating the effectiveness of multivariate bias correction methods in hydrological and agricultural contexts. We address each of your comments in detail below.

Major Comments

Some of the arguments are not always scientifically correct, or at times are difficult to understand. I also miss the connection to the main aim of the paper in several places when seeing those arguments.

We acknowledge that some arguments in the original manuscript were not always presented with sufficient scientific precision and that the connection to the main objectives of the paper was not always explicit. This made parts of the text difficult to follow and weakened the overall coherence of the review.

To address this, we have undertaken a thorough revision of the manuscript with the following specific improvements:

- We have carefully re-examined all arguments and revised those that were scientifically imprecise or insufficiently supported.
- We have strengthened the logical flow throughout the paper, ensuring that every major argument is clearly linked to the central aim of the study - namely, to systematically evaluate the effectiveness of multivariate bias correction methods in hydrological and agricultural impact modeling.
- We have restructured several sections (particularly in the Introduction, Results, and Discussion) to better highlight how the presented evidence contributes to answering the review's core research questions.
- We have improved the clarity of the writing by removing ambiguous statements and ensuring that claims are precisely worded and appropriately qualified.

We believe these revisions have substantially improved the scientific rigor, clarity, and focus of the manuscript. We address the specific instances you highlighted in your detailed comments below.

The whole Section 3.3 is somewhat chaotic and needs revision. I explain this further below.

We agree that Section 3.3 was overly long, somewhat repetitive, and lacked a clear logical structure. In the revised manuscript, we have substantially restructured and shortened this section. The revised version:

- Removes redundant explanations and improves the overall logical flow,
- Clearly defines key terms (such as "arbiter") to avoid ambiguity,

- Positions Section 3.3 more effectively as a concise bridge between the discussion of validation approaches and the synthesis presented in Section 4, and
- Avoids anticipating or repeating results that are properly discussed later in the paper.

I find that the text sometimes deviates from scientific writing and adopts a more popular science style. The choice of wording is occasionally exaggerated. For example, there are numerous uses of phrases such as "compelling study," etc.

We agree that some parts of the original manuscript adopted a tone that was closer to popular science writing, with occasional use of exaggerated or overly strong phrasing. This reduced the scientific neutrality and precision expected in a systematic review.

In the revised manuscript, we have adopted a consistently neutral and objective scientific tone throughout. We have systematically replaced exaggerated or informal expressions such as “compelling study,” “landmark study,” “stark contrast,” “unambiguous conclusion,” “profound and critical flaw,” and similar phrases with more precise and measured alternatives (e.g., “key study,” “notable example,” “clear contrast,” “consistent finding,” or “important limitation”).

The recommendations are somewhat synthetic and not very novel. For example, when it comes to bias non-stationarity, it is unclear how one can develop such methods. What are the clues or directions for doing so? In addition, there is now a new family of climate simulations called Single Model Initial-condition Large Ensembles (SMILEs), which allow for the assessment of time of emergence while accounting for internal variability. It would be good if the authors could cover how bias correction is affecting those and how SMILEs should be evaluated in this context. For a reference, please check https://hess.copernicus.org/articles/29/5695/2025/.

We agree that the recommendations in the current version of the manuscript remain somewhat general and would benefit from greater specificity, particularly regarding how to address bias non-stationarity in future methodological development.

In the revised manuscript, we will substantially strengthen Section 5.3 by:

- Providing more concrete and actionable directions for tackling bias non-stationarity. We will discuss specific promising avenues, including regime-dependent bias correction, conditional approaches that explicitly account for changes in inter-variable relationships, and ensemble-aware methods that preserve both the change signal and the ensemble spread when using large initial-condition ensembles.
- Explicitly incorporating recent advances based on Single Model Initial-condition Large Ensembles (SMILEs). We will cite and discuss the recommended reference (Astagneau et al., 2025, HESS) and explain how SMILEs can be used to better evaluate bias correction performance under non-stationarity. In particular, we will highlight their value for assessing the impact of bias adjustment strategies on signal-to-noise ratios, time-of-emergence, and the preservation of internal variability in future hydrological projections.
- Clarifying how the choice of bias correction strategy interacts with the use of SMILEs and what this implies for robust climate impact assessments.

Specific Comments

L27: Please use a more fundamental reference.

The citation to Edwards (2011) appears in the opening sentence of the Introduction, supporting the statement on the scientific consensus regarding anthropogenic climate change and the resulting need for robust future climate projections to inform adaptation and mitigation. Edwards (2011) is a valuable historical overview of climate modeling, but we agree that it is not the most fundamental or direct reference for this foundational claim.

We have replaced the citation with a reference to the IPCC Sixth Assessment Report (IPCC, 2021), which provides the most authoritative, comprehensive, and widely recognized assessment of the scientific consensus on climate change. The IPCC AR6 explicitly underscores the necessity of reliable climate projections - including at the scales relevant to impact modeling in hydrology and agriculture to support adaptation and mitigation strategies.

L32: 25–1 km does not seem right.

We will correct the spatial resolution range to the accurate and commonly cited values (typically 1-25 km for impact applications).

L33: Do you mean 250?

The phrasing “often at 25–1 km spatial resolutions” in the Introduction (describing the fine-scale data requirements of hydrological and agricultural impact models) was a typographical/notation mistake. The intended meaning was the conventional range of high-resolution data needed for many impact applications typically from ~25 km down to 1 km (or finer) spatially, and daily to sub-hourly temporally.

We have corrected the text to:

“Impact applications require fine-scale data, often at spatial resolutions of 1–25 km (or finer) and daily to sub-hourly temporal frequencies.”

L35: The sentence after the hyphen is unclear.

We have revised the text for clarity and improved readability while preserving the intended technical meaning:

“It is crucial to distinguish nominal grid spacing from a model’s effective resolution — the scale at which physical features are robustly resolved. The effective resolution is typically several times coarser than the nominal grid spacing (Skamarock, 2004).”

L38: I am not sure I understand what you mean by resource allocation. The purpose of going from a GCM to an RCM is to include more processes.

We agree that the original wording was ambiguous and could be misread as referring to the GCM-to-RCM downscaling step. That was not our intention.

The sentence in question was meant to explain a key reason why GCM nominal grid spacings have not improved as rapidly as raw increases in computing power might suggest: within global modeling centers, available computational resources are often allocated to other priorities such as running larger ensembles or adding more complex Earth system components (interactive carbon cycle, dynamic vegetation, atmospheric chemistry, etc.) rather than uniformly increasing spatial resolution in every GCM simulation.

Regional Climate Models (RCMs) do indeed serve the important purpose of dynamically resolving finer-scale processes that remain parameterized in GCMs. We have revised the text to make this distinction explicit, remove any potential confusion about “resource allocation,” and better acknowledge the complementary role of RCMs.

The revised paragraph now reads (changes highlighted for your convenience):

“It is crucial to distinguish nominal grid spacing from a model’s effective resolution — the scale at which physical features are robustly resolved. The effective resolution is typically several times coarser than the nominal grid spacing (Skamarock, 2004). Furthermore, while increases in computing power have enabled important advances, GCM grid spacings have not decreased linearly with available resources. Modeling centers have frequently prioritized larger ensembles or the inclusion of additional Earth system processes (e.g., interactive carbon cycle, dynamic vegetation, and atmospheric chemistry) over uniform increases in spatial resolution across all simulations (Haarsma et al., 2016). Regional Climate Models (RCMs) are then employed to dynamically downscale GCM output, thereby explicitly resolving many of the finer-scale atmospheric, land-surface, and hydrological processes that remain parameterized in GCMs.”

L39–41: This whole sentence could have a better flow.

We agree that the original passage (spanning the discussion of RCMs, GCM resource priorities, and the need to aggregate sub-hourly calculations) was overly long and had suboptimal flow, with multiple clauses connected in a way that could feel dense on first reading.

We have revised the text for improved readability, logical progression, and smoother sentence structure while retaining all scientific content. This revision also incorporates the clarification we made in response to your comment at L38 regarding resource allocation and the role of RCMs.

The revised paragraph now reads:

“Regional Climate Models (RCMs) help bridge this resolution gap by dynamically simulating many finer-scale processes. However, even as computing power has increased, GCM grid spacings have not decreased linearly. Modeling centers have often chosen to allocate additional resources to larger ensembles or to the inclusion of more complex Earth system components — such as interactive carbon cycle, dynamic vegetation, and atmospheric chemistry — rather than to higher spatial resolution in every simulation (Haarsma et al., 2016). As a result, GCMs typically aggregate their internal sub-hourly calculations (on the order of 100–300 seconds) to hourly or coarser output frequencies. This aggregation is necessary to manage the enormous data volumes generated by these models, which can range from terabytes to petabytes per simulation.”.

L50: Please avoid using assertive phrases such as "well known." Some people may not be familiar with the method.

We agree that phrases such as “well-known” can inadvertently assume prior familiarity and are best avoided, especially when introducing a specific term like “drizzle bias” to a broader readership that includes hydrologists, agricultural scientists, and impact modelers who may not work directly with climate model output.

We have revised the sentence to use more neutral and inclusive wording:

“…resulting in what is often referred to as the ‘drizzle bias,’ in which models produce rainfall too frequently but with too little intensity (Dai, 2006; Stephens et al., 2010).”

L55: Unclear sentence.

We agree that the long sentence beginning with "Because this deep technical information…" is dense and difficult to follow. The causal logic and connection to the preceding sentence about ES-DOC could be expressed more clearly.

We have revised the text for improved clarity and flow:

"Every model is unique, and developers may choose to represent different processes with alternative simplifying parameterizations. While centralized documentation initiatives such as the Earth System Documentation (ES-DOC) exist, they typically focus on high-level model configurations rather than the granular algorithmic details of individual parameterizations. As a result, detailed technical information is often scattered across technical reports and source code, making it difficult to anticipate the specific biases that may arise for a given model or region. This difficulty is further compounded by ongoing model development, as new versions are released and systematic errors must be re-evaluated."

L62–72: This whole section reads chaotic. Is bias correction really supposed to correct drift? I do not think the aim was ever to correct errors caused by internal variability. Rather, the goal of bias correction is only to correct simplification errors. The last sentence is also unclear.

We have substantially revised this section to improve structure, clarity, and conceptual accuracy. The revised text now:

Presents drift more precisely as a source of systematic discrepancies in model output (rather than implying bias correction can fully remove it).

Clearly separates the discussion of model drift from the additional challenge of inhomogeneous observations.

Rephrases the "moving target" idea in more precise and accessible language.

Maintains consistency with the paper's broader message (already strengthened in response to RC1) that bias correction adjusts statistical properties but cannot correct fundamental structural or equilibration problems in climate models.

Revised text:

"Finally, small differences in a model's initial state or parameter settings can produce long-term drifts in key variables such as ocean heat content, sea surface temperature (SST), and salinity (Sen Gupta et al., 2012, 2013). These drifts arise when the model has not fully reached equilibrium with its external forcing (Hobbs et al., 2016). Land-surface components exhibit similar systematic biases and drifts — for example in soil thermal dynamics in permafrost regions or in the representation of short- versus long-term soil moisture variability — which can affect the reliability of long-term hydrological projections (Koven et al., 2013; Xi et al., 2022; IPCC, 2013).

In addition, the observational datasets used to evaluate and correct climate model output are frequently inhomogeneous over time, owing to changes in satellite missions, station networks, or data processing methods. Combined with spatial gaps in remote regions, this creates a moving and incomplete reference against which bias correction methods must be developed and validated (Funk et al., 2015; Merchant et al., 2019)."

We believe this version is now logically structured, conceptually accurate, and much easier to read. The changes also reinforce the paper's central theme that bias correction has important limitations when applied to certain types of model errors.

L75: Panacea?

We agree that “panacea” is somewhat informal and potentially overstated for a systematic review. We have replaced it with the more neutral and precise phrase “universal solution”:

“However, while powerful, bias correction is not a universal solution and must be applied with caution.”

L77: A citation alone does not validate a statement. Please use different phrasing.

We agree that framing a claim primarily through the lens of a single paper can give the unintended impression that the statement is validated by the paper’s popularity rather than by the strength of the underlying argument and supporting evidence.

We have rephrased the passage to present the conceptual point first and then cite Ehret et al. (2012) as a key reference that illustrates the issue, rather than letting the citation carry the statement:

“Most bias correction methods rely on the assumption of stationarity — that is, that the statistical relationship between modeled and observed climate remains constant over time. This assumption does not always hold under non-stationary conditions. Ehret et al. (2012), for example, demonstrated that while bias correction can improve agreement with historical observations, it may compromise the physical consistency of future projections.

L79: The statement that bias correction methods assume stationarity is not correct. In fact, I would argue that most modern bias correction methods account for the non-stationarity of the climate signal. Please check the bias correction papers by Mathieu Vrac, Tootoonchi, and Teutschbein.

We agree that the original wording was imprecise and potentially misleading.

The issue stems from two related but distinct concepts of “stationarity” in bias correction:

- Stationarity of the bias: Most methods assume that the statistical relationship (bias) between the climate model and observations remains constant over time. The correction derived from the historical period is applied to future periods under this assumption.
- Non-stationarity of the climate signal: Many modern bias correction methods are explicitly designed to preserve the model’s projected climate change signal (changes in mean, variability, and extremes) rather than assuming the climate itself is stationary.

The original sentence conflated these two ideas. We have revised the text to make the distinction clear and to acknowledge that several widely used methods (including Quantile Delta Mapping and variants of CDF-t) are constructed to handle non-stationarity of the climate signal while still relying on the assumption of stationary bias.

Revised text:

"Most bias correction methods rely on the fundamental assumption that the statistical bias between the climate model and observations remains stationary over time; that is, the correction algorithm derived from a historical period is assumed to be valid for future periods (Teutschbein & Seibert, 2012). Indeed, Teutschbein and Seibert (2012) identified this stationarity assumption as a major limitation and the main weak point of bias correction when applied to future climate projections. To address this, many modern methods are specifically designed to preserve the model’s projected non-stationary climate change signal. Examples include Quantile Delta Mapping (QDM) and certain implementations of the Cumulative Distribution Function transform (CDF-t) developed by Vrac and colleagues (e.g., Vrac, 2018). Recent studies, such as those by Tootoonchi et al. (2022), have further examined the performance of these modern bias correction approaches under non-stationary conditions."

L83–85: This is a separate topic in climate modeling that does not connect well with the objective of the paragraph.

We agree that the example from Boberg and Christensen (2012), while scientifically interesting, discusses how model biases can affect simulated climate feedbacks and regional warming amplification. This is somewhat tangential to the paragraph's main focus on the risks and limitations of applying statistical bias correction methods.

We have removed the Boberg and Christensen example and revised the text to stay tightly focused on the implications for bias correction:

Revised text:

"This risk is particularly acute when the magnitude of the bias approaches or exceeds the projected climate change signal itself. In such cases, statistical bias correction may inadvertently modify or even distort the climate change signal rather than simply adjusting for systematic errors."

L89: "All others" is unclear.

The phrase "independently of all others" is indeed unclear and could be misinterpreted. We have revised it to be more precise and explicit:

Revised text:

"For decades, the field of bias correction was dominated by univariate methods, most notably Quantile Mapping (QM) and its prominent variant, Quantile Delta Mapping (QDM), which correct the statistical distribution of each climate variable separately, without accounting for dependencies between variables (Räty et al., 2018; Cannon et al., 2015)."

L92: Please tone down words such as "profound" and "critical."

We agree that "profound and critical flaw" is overly strong and somewhat dramatic for a systematic review.

We have revised the sentence to use more measured and precise language:

Revised text:

"Despite its effectiveness at this task, the independent treatment of each variable constitutes a key limitation. By design, univariate methods ignore and can disrupt the physically meaningful dependence structure—or coherence—that links variables such as temperature, precipitation, and humidity."

L93: No, univariate bias correction cannot destroy the dependence structure. It simply reproduces what the climate model produces. Please check François et al. (2020) or Tootoonchi et al. (2022).

We agree that the original wording was too strong and not fully accurate.

Univariate bias correction methods adjust each variable independently. They therefore preserve whatever dependence structure exists in the raw climate model output. They do not actively "destroy" inter-variable relationships. However, because they do not account for dependencies, they generally fail to restore or improve the physically meaningful dependence structure observed in reality. In some cases (particularly with quantile mapping-type methods), they can unintentionally alter the relationships present in the model.

We have revised the text accordingly:

Revised text:

“By design, univariate methods correct each variable independently. As a result, they generally fail to preserve or restore the physically meaningful dependence structure between variables such as temperature, precipitation, and humidity (François et al., 2020; Tootoonchi et al., 2022). Because each variable’s time series is adjusted according to its own quantile without regard to the others, the physical relationships that bind them can be altered or left uncorrected.”

L95: "Without regard to the others" is unclear.

We have revised it for clarity and precision:

Revised text:

“Because each variable’s time series is adjusted according to its own quantile independently of the other variables, the physical relationships between them can be altered or left uncorrected.”

L97–100: This sentence is unclear.

We have revised the text to improve flow, break up overly long sentences, and make the logical progression clearer:

Revised text:

“This can produce physically implausible combinations, such as intense precipitation occurring together with unrealistically low relative humidity, which may lead to distorted results from impact models (François et al., 2020). Figure 1 illustrates the issue by contrasting the dependence structure preserved under multivariate correction with the scattered, incoherent output that can result from univariate correction. The theoretical basis for this critique was established by Bhowmik and Sankarasubramanian (2019), who mathematically demonstrated that univariate methods are generally unable to correct inter-variable cross-correlations consistently. Under certain conditions, they can even amplify existing correlation biases.”

L116: Why is it specifically needed in those locations?

We have revised the text to make the reasoning explicit:

Revised text:

“For instance, in alpine and high-latitude regions, snow accumulation and melt are strongly controlled by the joint occurrence of precipitation and temperature. The phase of precipitation (rain versus snow) and the timing of snowmelt depend on their co-variation, rather than on their individual (marginal) distributions alone (Meyer et al., 2019; Nury et al., 2021).”

L121–123: Examples?

We have revised the passage to include additional concrete examples relevant to hydrology and agriculture, while keeping the paragraph concise.

Revised text:

“The need for a multivariate perspective is further underscored by the growing scientific focus on compound events — high-impact events arising from the interaction of multiple climate drivers. Relevant examples include concurrent heatwaves and droughts (which can severely stress crops and water resources), heavy precipitation coinciding with saturated soils or high temperatures (increasing flood risk), and warm, wet conditions that alter snow accumulation and melt in mountainous regions. The assessment of such events fundamentally requires a multivariate approach to bias correction that can preserve the joint probabilities of their constituent variables (Bevacqua et al., 2017).”

Fig. 4: What are P, T, and H?

We have revised the caption to make the variable definitions clearer and more prominent:

Revised Figure 4 caption:

“Figure 4: Conceptual diagram illustrating the operational structures of the three main classes of Multivariate Bias Correction (MBC) methods: (1) Marginal/Dependence, (2) All-in-One, and (3) Successive Conditional. The diagram uses three example variables: P (Precipitation), T (Temperature), and H (Humidity). This classification follows the typology proposed by François et al. (2020) and Vrac (2018).”

Table 3: The MRec method was not developed by François et al. (2020). Please find and cite the primary reference.

We have corrected Table 3 to cite the primary foundational reference for the Matrix Recorrelation (MRec) method, which was developed by Bárdossy and Pegram (2012). While later literature such as François et al. (2020) provided a valuable intercomparison of the method and utilized the "MRec" acronym, Bárdossy and Pegram (2012) are the original developers of the underlying matrix transformation framework. The manuscript and reference list have been updated accordingly.

Table 3: What about examples of Successive Conditional methods?

We have revised Table 3 to include representative examples of the Successive Conditional class. This class includes methods that correct variables sequentially, where the correction of each subsequent variable is conditioned on the values of the previously corrected variable(s). This is typically achieved through conditional distributions or pair-copula constructions.

We have updated Table 3 to include examples of the Successive Conditional class to ensure all three operational structures are represented:

- Vine Copulas / Pair-Copula Constructions (conditional): Variables are corrected sequentially using pair-copula constructions. For example, precipitation may be corrected first, after which other variables (such as temperature) are adjusted conditionally on the already-corrected precipitation. This approach is represented in our corpus by Bevacqua et al. (2017), who applied vine copulas (PCCs) for compound event analysis, as well as in several studies by Tootoonchi et al. (2022, 2023).
- Conditional Copula approaches: These methods apply conditional copulas or distributions in a step-wise manner, with the correction of each variable depending on the previously corrected variables.

L261: Examples of papers doing so?

We have revised the text to include specific examples of studies that evaluated multivariate bias correction methods by driving crop models with bias-corrected climate inputs and comparing simulated outputs (e.g., yield, biomass) against a reference.

Revised text:

“In agriculture, a similar strategy is employed. Crop models such as DSSAT and APSIM are driven by the bias-corrected climate inputs, and simulated outputs such as crop yield, biomass, or Leaf Area Index (LAI) are compared to a reference run. Examples include Galmarini et al. (2024), who evaluated multiple multivariate bias correction methods (MBCn, R2D2, ISIMIP3BASD) using an ensemble of 12 crop models within the AgMIP-Wheat project intercomparison, and Jin et al. (2022), who assessed the impact of bias-corrected climate data on wheat yield forecasts using the APSIM-Wheat model.”

The whole Section 3.3 is chaotic. This preface should be shortened and made more concise. In many ways, it repeats what is discussed in the following pages. What does "arbiter" mean in this context?

We agree with this assessment. As noted in our response to the major comments regarding clarity, structure, and conciseness, we have substantially revised Section 3.3. The opening has been shortened, repetition with later parts of the section has been reduced, and the term “arbiter” has been replaced with clearer language (“objective evaluator”).

L293–299: What is this whole section based on? We have not yet read the synthesis report, so this information is hidden from the reader.

We thank the reviewer for pointing this out. We agree that phrasing these statements as definitive conclusions before presenting the synthesis results was premature and confusing. Our intent in this section was to establish the underlying theoretical hypothesis driving our review—specifically, the idea found in recent literature (e.g., Galmarini et al., 2024) that the structural formulation of an impact model theoretically dictates its sensitivity to multivariate inputs.

To correct this, we have revised the paragraph to clearly frame these statements as hypotheses and theoretical expectations rather than established facts. We believe this revision effectively sets the stage for the synthesis report without preempting the results.

"This potential divergence in added value raises an important theoretical premise for our review: the structural complexity of an impact model likely dictates its sensitivity to the inter-variable coherence that multivariate bias correction (MBC) methods aim to preserve (e.g., Galmarini et al., 2024). For instance, it is hypothesized that simpler conceptual or empirical models—such as a statistical crop-yield model based on aggregated monthly inputs, or a hydrological model employing Horton’s infiltration formulation (Horton, 1940)—would be structurally insensitive to these fine-scale multivariate dependencies. In contrast, more complex process-based models—such as DSSAT (Jones et al., 2003) or those solving the Richardson-Richards equation (Richardson, 1922; Richards, 1931)—are expected to be highly sensitive to the daily inter-variable correlations that univariate methods might inadvertently distort. Therefore, a guiding hypothesis of this systematic review is that the choice of impact model is not a neutral decision; its fitness for purpose is a crucial factor in evaluating the true added value of bias correction methods."

L305: What is meant by "stark contrast"?

We agree that the phrase “stark contrast” is overly strong and somewhat dramatic. We have revised the text to use more measured language while still accurately reflecting the findings.

Revised text:

“This divergence in outcomes is summarized in Figure 5. In hydrology, results were mixed and appeared to depend on the type of application. For studies focused on process-specific modelling (such as snow processes or compound events), a majority of studies found clear added value from multivariate bias correction methods. However, for studies assessing general streamflow simulation, the benefits were more variable: approximately 28% reported clear superiority of MBC, while 50% found nuanced benefits and 22% reported limited or no added value over simpler univariate methods. In contrast, all studies focused on agricultural applications reported clear benefits from multivariate methods.”

L311: I think the hydrology part should come before the agricultural one. This has been the order throughout the paper.

We agree that maintaining the established "Hydrology then Agriculture" sequence improves the logical flow and readability of the manuscript. We have restructured this paragraph accordingly. Furthermore, based on feedback from another reviewer, we have refined this section to highlight that the divergence in MBC effectiveness is driven primarily by the structural complexity and multi-variable dependencies of the specific impact models used in these domains.

L316: What is meant by an "unambiguous conclusion"?

We agree with the reviewer that the word "unambiguous" is overly definitive and potentially too strong. Our intention was to highlight the strong, systematic trend reported by the authors across the wide variety of models and locations tested. To ensure the tone is scientifically appropriate, we have toned down this phrasing and replaced "unambiguous conclusion" with "consistent conclusion."

Revised text in manuscript:

"This comprehensive study involved 12 different AgMIP-Wheat crop models, 8 climate models, and 14 different bias adjustment methods (5 multivariate and 9 univariate) across 21 locations. The core conclusion was consistent: the error in crop model outputs (such as yield, biomass, and Leaf Area Index) is 'systematically lower' when using multivariate bias-adjusted climate data compared to data from univariate methods (Galmarini et al., 2024)."

L342: Please tone down "compelling."

We agree that the word "compelling" was overly subjective in this context. To maintain a strictly objective scientific tone, we have toned down this phrasing by changing it to "notable."

Revised text in manuscript: "The study by Tootoonchi et al. (2023) is particularly notable, as it found that across 16 signatures and 55 catchments, univariate methods generally outperformed their multivariate counterparts."

L370: What is meant by a "landmark study"?

We agree with the reviewer that the term "landmark" is subjective and overly editorial in this context. Our intention was simply to highlight that this specific paper provided a particularly clear and explicit demonstration of the compounding effects of multivariate bias correction on snowpack dynamics. To maintain a rigorously objective scientific tone, we have removed the subjective phrasing.

Revised text in manuscript: "An illustrative study by Meyer et al. (2019) in alpine catchments demonstrated this clearly. They showed that the simultaneous correction of P and T with the MBCn method..."

L414: What is a "canonical example"?

We apologize for the use of unnecessary jargon. By "canonical," we simply meant the classic or most standard example of this phenomenon in the literature. To make the text clearer and more accessible to all readers, we have removed this term and replaced it with "a prime example."

Revised text in manuscript: "...The partitioning of precipitation into rain or snow at the 0°C threshold is a prime example. In this case, the model's sensitivity..."

L415–418: The whole paragraph is unclear, and there are no references.

We agree that the phrasing was overly abstract and that making such a synthesis statement requires proper grounding in the literature. Our goal was to describe how different models process climate data—where some models smooth out errors over time/space, while others react sharply to daily multivariate inconsistencies.

To clarify this, we have rewritten the paragraph to remove the confusing jargon and explicitly unpack the "filter" concept. Furthermore, we have added relevant citations (including Maraun et al., 2017; François et al., 2020; Meyer et al., 2019; and Galmarini et al., 2024) to support this theoretical framework and ground our synthesis in established literature.

Revised text in manuscript:

"Ultimately, the impact model acts as a mathematical filter for climate data (e.g., Maraun et al., 2017; François et al., 2020). When a model functions primarily as a broad spatiotemporal integrator—such as accumulating run-off over a large basin to simulate general streamflow—it tends to smooth out day-to-day physical inconsistencies between variables, reducing the apparent added value of MBC. Conversely, when a model operates as a highly sensitive daily process simulator governed by sharp physical thresholds (such as the 0°C snow-rain boundary or complex crop phenology stages), preserving the exact daily coherence between multivariate inputs becomes essential (Meyer et al., 2019; Galmarini et al., 2024). This framework provides a coherent physical explanation for both the mixed results of MBC in general streamflow studies and its critical importance for snow-dominated or agricultural processes. The core components of this framework are illustrated in Figure 6."

Additional References Cited in this Repsonse:

Astagneau, P. C., Wood, R. R., Vrac, M., Kotlarski, S., Vaittinada Ayar, P., François, B., & Brunner, M. I. (2025). Impact of bias adjustment strategy on ensemble projections of hydrological extremes. *Hydrology and Earth System Sciences*, 29(20), 5695–5718. https://doi.org/10.5194/hess-29-5695-2025

Bárdossy, A., & Pegram, G. (2012). Multiscale spatial recorrelation of RCM precipitation to produce unbiased climate change scenarios over large areas and small. *Water Resources Research*, 48(9). https://doi.org/10.1029/2011WR011524

Bevacqua, E., Maraun, D., Hobæk Haff, I., Widmann, M., & Vrac, M. (2017). Multivariate statistical modelling of compound events via pair-copula constructions: Analysis of floods in Ravenna (Italy). *Hydrology and Earth System Sciences*, 21(6), 2701–2723. https://doi.org/10.5194/hess-21-2701-2017

François, B., Vrac, M., Cannon, A. J., Robin, Y., & Allard, D. (2020). Multivariate bias corrections of climate simulations: which benefits for which losses?. *Earth System Dynamics*, 11, 537-562. https://doi.org/10.5194/esd-11-537-2020

Galmarini, S., Solazzo, E., Ferrise, R., et al. (2024). Assessing the impact on crop modelling of multi- and uni-variate climate model bias adjustments. *Agricultural Systems*, 213, 103846. https://doi.org/10.1016/j.agsy.2023.103846

IPCC. (2021). Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., et al. (eds.)]. Cambridge University Press.

Jin, H., Li, M., Hopwood, Z., Hochman, Z., & Bakar, K. S. (2022). Improving early-season wheat yield forecasts driven by probabilistic seasonal climate forecasts. *Agricultural and Forest Meteorology*, 317, 108832. https://doi.org/10.1016/j.agrformet.2022.108832

Maraun, D., et al. (2017). Towards process-informed bias correction of climate change simulations. *Nature Climate Change*, 7(11), 764-773. https://doi.org/10.1038/nclimate3418

Teutschbein, C., & Seibert, J. (2012). Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. *Journal of Hydrology*, 456–457, 12-29. https://doi.org/10.1016/j.jhydrol.2012.05.052