Authors' Responses to Reviewer 1

General Comment:

This manuscript proposes a robust and innovative unified framework for diagnosing MJO biases in General Circulation Models (GCMs). The integration of moisture-mode theory, Linear Inverse Modeling (LIM), and the process-resolved attribution of errors (via inner-product and unit circle analysis) is a significant methodological advance. The results clearly identify systematic GCM shortcomings, particularly the slow propagation bias, and successfully attribute these errors to physical processes like horizontal moisture advection and convective-associated terms.

The paper is generally well-written and logically structured.

Major/Moderate Comments:

1. Limitations of Linear Models: While the paper mentions limitations in the data (lack of hourly data, coarse vertical resolution), The paper should more directly address the inherent limitations of *Moisture Mode Theory* and *LIM*. Is the *linear* assumption of LIM adequate for capturing the full non-linear MJO life cycle? Could the authors briefly discuss or justify why the linear approximation is sufficient for diagnostic purposes in this context?

   We thank the reviewer for the positive evaluation of our unified diagnostic framework and the constructive feedback regarding our methodological assumptions. The reviewer raises an excellent and sophisticated point concerning the inherent boundaries of linear approximations when dealing with a complex climate phenomenon.

   We completely agree that a linear assumption cannot fully capture the highly non-linear, state-dependent lifecycle of the MJO. In particular, it inherently overlooks details like MJO event diversity (e.g., Wang et al., 2019) or the region-dependent shifts in physical regimes where traditional moisture-mode theory might break down or require modifications (e.g., Mayta and Adames, 2023) (L133 – L135).

   However, for the specific objective of multi-model evaluation and systematic bias attribution, we contend that this linear approximation serves as an exceptionally powerful and necessary diagnostic baseline for the following reasons:

   1. **Capturing the Climatological Baseline:** The Linear Inverse Model (LIM) framework acts as a robust diagnostic metric for the long-term, climatological characteristics of MJO propagation and maintenance

across a wide range of models. It isolates the system's intrinsic dynamics by mapping the fundamental relationships between the MJO state and its physical tendencies (L136 – L139).
2. **Mathematical Consistency and Traceability:** The primary strength of using a linear system is its direct mathematical compatibility with the moisture budget equation. This tractability allows us to seamlessly project complex, high-dimensional physical space budgets directly into a reduced-dimensional MJO phase space. Without this linear formulation, mathematically closing the loop between a traditional performance metric (the MJO indices) and individual process-level biases (the 28 budget terms) would be exceptionally difficult (L131 – L133).
3. **A Testbed for General Circulation Models (GCMs):** Rather than attempting to serve as a comprehensive simulation of every non-linear mechanism, our framework is designed as a foundational, zeroth-order diagnostic tool. It offers a standardized and practical bridge to evaluate where state-of-the-art models systematically deviate from theoretical expectations (L139 – L141).

That being said, we recognize that our framework is not strictly confined to linear boundaries. As we noted in the text, future iterations could substitute the core linear components with advanced non-linear data-driven models (e.g., Transformers), extracting non-linear growth rates and frequencies via tangent linear modeling (L141 – L143).

To more transparently address these limitations and better justify our methodological choices, we have significantly expanded our discussion in the revised manuscript.

2. Clarification (L45): It's stated that the linear/quasi-linear nature of the theories allows for direct evaluation of metrics like growth rate and frequency. Please clarify if the framework itself diagnoses these properties from the budget terms or if it uses the traditional LIM-derived metrics and then attributes the bias.

We thank the reviewer for highlighting this ambiguity regarding how growth rates and frequencies are calculated and connected to the budget. We realize our original phrasing blurred the line between the theoretical properties of the moisture mode and the practical execution of our statistical diagnostics.

To clarify, the framework does not compute growth rates or frequencies directly from the individual budget terms alone. Instead, it uses a multi-step sequence to bridge the budget variables with the Linear Inverse Model (LIM):

1. **Phase Space Construction:** We first extract the dominant MJO modes from the observed moisture field using EOF analysis, establishing a reduced-dimensional state vector space.
2. **Data-Driven LIM Estimation:** We construct a data-driven LIM within this reduced-dimensional space to estimate the baseline system's angular frequency and decay/growth rates via an eigenvalue problem. These traditional LIM metrics represent the collective "intrinsic" dynamics of the system.
3. **Budget Projection and Attribution:** Finally, we project the space-time decomposed moisture budget equations directly onto this exact same MJO EOF phase space. Because the mathematical structure of the budget tendency matches the formulation of the LIM tendency, we can then cleanly attribute model biases in those growth rates and frequencies back to errors in individual physical terms (via our inner-product and unit-circle analyses).

Therefore, the growth rates and frequencies are diagnosed via the data-driven LIM , but their underlying physical errors are systematically mapped and attributed using the projected budget equations. A similar approach is also used to study the barotropic annular mode, where the barotropic zonal momentum budget is projected onto the lead EOF of barotropic zonal-mean zonal wind for analysis.

To prevent any further confusion, we have thoroughly updated and clarified this sequence of operations at the beginning of Section 2 (The Unified Diagnostic Framework, L66 – L69).

3. Methodology (L70): The data uses 6 pressure levels (1000, 850, 700, 500, 250, 100 hPa). Given the sensitivity to vertical advection (e.g., L197), can the authors comment on how much the coarse resolution is believed to affect the budget terms, especially in the boundary layer?

   We appreciate the valuable feedback from the reviewer. We believe that lack of data within the boundary layer could lead to underestimation of vertical advection terms' contribution to biases (L79 – L81). However, due to the limitation of model vertical resolution and to make the analysis consistent through datasets, we keep the analysis at the original vertical resolution.

4. Term Summation Clarity (L137, Eq. 7): The sum runs from i=1 to 27. Please clarify what the index i represents. If I am not mistaken, they are the 27 advection terms (3 terms per timescale x 3 timescales x 3 directions) plus the one apparent moisture sink term, totaling 28 processes. A brief in-line list of the terms being summed or a reference to a table listing them would improve

clarity, as the total number of terms (advection, sink, etc.) is not immediately obvious.

We appreciate the valuable feedback from the reviewer. The right-hand side of Eq. 7 (Eq. 8 in the revised manuscript) is indeed the 27 advection terms (3 directions x 3 timescales of winds x 3 timescales of moisture gradients) plus one apparent moisture sink term. To avoid confusion, we have made the clarification in the revised manuscript (L164 – L166), and added a new table (table 2 in revised manuscript).

5. Finding Detail (L197): The paper states: "The vertical advection term has significantly positive contribution to total bias in the Indian Ocean, but the result does not hold in the Maritime Continent." Given the regional importance of the MC, is there a physical reason (e.g., topography, boundary layer, mean state) for this divergence that can be speculated upon?

   We thank the reviewer for encouraging us to expand on the physical mechanisms driving this regional divergence. The lack of statistical significance for the vertical advection term over the Maritime Continent (MC) does not imply that General Circulation Models (GCMs) simulate this process accurately; rather, it reflects a massive structural disagreement among the models that washes out any clear multi-model signal.

   We hypothesize that this regional divergence stems from a fundamental shift in the thermodynamic and dynamical regimes between the open Indian Ocean and the land-sea interface of the MC. Over the complex topography of the MC, convective variability is heavily modulated by local circulations where the critical Gross Moist Stability (GMS) is relatively large. In high-GMS regimes, vertical processing and large-scale moisture/energy anomalies depend intimately on sub-grid diabatic processes and the microphysical exporting efficiency of upper-tropospheric divergence (e.g., Inoue and Back 2015, 2017, Dang and Yu 2025).

   Consequently, the vertical advection and the apparent moisture sink ($-Q_2/L$) in this region are exceptionally sensitive to how models parameterize the interaction between prevailing large-scale flows and complex island topography. Because state-of-the-art cumulus parameterizations handle these localized structural features very differently, the resulting inter-model spread is highly diverse, ultimately masking no systematic multi-model consensus for this term.

   We agree that exploring how specific parameterization choices modulate this local moisture balance is a highly compelling avenue for future research. We have revised the discussion in the manuscript to incorporate this physical speculation (L231 – L240).

6. Applicability Beyond MJO (L252): In the conclusion, the authors state the framework is a "transferable pathway for diagnosing the fundamental dynamics of diverse climate phenomena beyond the MJO." It would strengthen the claim to briefly elaborate on this. Could the authors provide an example of which *other* climate phenomena this framework might be applied to (e.g., equatorial waves, monsoonal intraseasonal variability) and why?

   We appreciate the valuable advice from the reviewer. While this represents the first application of LIM to the model diagnosis of tropical intraseasonal variability, analogous concepts have been successfully applied to other climate phenomena. Notable examples include the Bjerknes-Jin index in ENSO research (e.g., Heede and Fedorov, 2023) and the zonal momentum budget used in annular mode studies (e.g., Chen et al., 2025). By projecting the governing equations onto a reduced-dimensional space and extracting the key structural features, this analytical framework effectively isolates the specific physical processes responsible for the observed dynamical behavior. We have added the discussion in the manuscript (L339 – L342).

Minor Comment:

1. Abstract (L1): "An unified diagnostic framework..." should be "A unified diagnostic framework".
   We appreciate the advice from the reviewer. The modification has been made in the revised manuscript.

2. Definition/Context (L40): The paper relies heavily on the Weak Temperature Gradient (WTG) approximations. A very brief sentence on why the WTG approximation is justified for the MJO's large scale and slow evolution would be helpful for non-specialist readers.

   We appreciate the advice from the reviewer. We have spelled out the assumption made during the simplification of governing equations in the moisture-mode theory (L48 - L53).

3. L168: The sentence "For WH region (Fig. 5a), most of the GCMs have bias vectors pointing upward, which indicates that the propagation speeds are generally slower in GCMs within this region." Remove the final two words for conciseness: "...slower in GCMs."

   We appreciate the advice from the reviewer. The modification has been made in the revised manuscript.

4. L197: The phrase "...than of meridional." Should be "...than that of meridional."

   We appreciate the advice from the reviewer. The modification has been made in the revised manuscript.

References:

- Inoue, K., & Back, L. E. (2015). Gross Moist Stability Assessment during TOGA COARE: Various Interpretations of Gross Moist Stability. Journal of the Atmospheric Sciences, 72(11), 4148-4166.
- Inoue, K., & Back, L. E. (2017). Gross Moist Stability Analysis: Assessment of Satellite-Based Products in the GMS Plane. Journal of the Atmospheric Sciences, 74(6), 1819-1837.
- Wang, B., Chen, G., & Liu, F. (2019). Diversity of the Madden-Julian oscillation. Science Advances, 5(7), eaax0220.
- Heede, U. K., & Fedorov, A. V. (2023). Towards understanding the robust strengthening of ENSO and more frequent extreme El Niño events in CMIP6 global warming simulations. Climate Dynamics, 61(5), 3047-3060.
- Mayta, V. C., & Adames Corraliza, Á. F. (2023). Is the Madden‐Julian oscillation a moisture mode?. Geophysical Research Letters, 50(15), e2023GL103002.
- Chen, T. C., Goosse, H., Davrinche, C., Libera, S., Roberts, C., Aengenheyster, M., ... & von Storch, J. S. (2025). Southern Annular Mode persistence and westerly jet: a reassessment using high-resolution global models. Weather and Climate Dynamics, 6(4), 1179-1193.
- Dang, D. P., & Yu, J. Y. (2025). A process-oriented GMS plane for Madden–Julian oscillation analysis. Climate Dynamics, 63(9), 350.