General comments:

This manuscript provides valuable and much-needed insights into the sub daily variability and the diurnal cycle of precipitation across satellite and reanalysis estimates. I particularly appreciate that the authors go beyond routine intercomparisons by leveraging physical constraints to uncover the mechanisms driving observed product strengths and limitations. However, to help the manuscript reach its full potential, I would encourage the authors to refine its overall narrative structure. Currently, the study attempts to balance three distinct analytical angles simultaneously, ground validation over Northern Hemisphere land domains (CONUS, Europe, East Asia), global-scale unvalidated intercomparisons across land and ocean, and process-oriented physical diagnostics at two U.S. ground sites. Combining these different scopes creates some narrative fragmentation. In particular, the global ocean/land intercomparison adds limited new insight, given similar existing literature and strays slightly from the core focus nicely anchored by the paper's title (*"Evaluation against ground-based observations"*). Streamlining these components would reduce visual clutter and allow the true highlight of this work moving beyond descriptive bias mapping to explain the physical drivers of diurnal discrepancies to shine through much more clearly.

General comments:

- Given that the title and scope of the manuscript emphasize comparison with ground observations, I expected a direct spatial comparison map against the gauge network (Model − Obs and Satellite − Obs). Providing only spatially averaged bar plots (e.g., Fig. 3a, 4a) masks all regional performance details.

- Several figure sequences throughout the manuscript appear redundant. While they present slight variations of the diagnostic metrics, they ultimately convey the same overarching physical conclusion and seems more of supportive rather than adding new insights. This creates repetitive discussion paragraphs and makes it difficult for the reader to pinpoint the primary takeaway. I strongly advise the authors to prioritize key representative figures for the main text and relocate secondary variations to the Supplementary Materials (Specific comments are listed later).

- The mathematical decomposition presented in the manuscript involves multiple nested hierarchical tiers - moving from total precipitation variance (σ^2_{total}) to sub-daily variance (σ^2_{subD}), separating regular (σ^2_{reg}) and irregular (σ^2_{irreg}) components, and further decomposing the regular component into Fourier harmonics ($\sigma^2_{1st}, \sigma^2_{2nd}, \sigma^2_{higher}$). Because of this multi-layered hierarchy and the large number of associated statistical terms and ratios (rat_{subD}, rat_1, $\sigma^2_{1st}/\sigma^2_{total}$), the analytical framework is difficult to follow in pure text form. To improve readability and pedagogical clarity: I strongly encourage the authors to include a methodological flowchart or schematic diagram in Section 2. Providing this visual roadmap early in the paper will make it substantially easier for readers to navigate the subsequent spatial maps and diagnostic discussions in Sections 3 and 4.

- The final section of the results *'4.3 Physical mechanisms of the diurnal cycle'* is particularly difficult to follow. The authors need to revise this entire section to improve readability and structure. Currently, each paragraph feels disjointed and fluctuates back and forth without arriving at a clear conclusion. Furthermore, the datasets keep changing from one figure to the next throughout the section, which makes it even harder to follow the narrative.

- Finally, there are numerous acronyms used throughout the manuscript, which makes it challenging for the reader to keep track and forces frequent back-and-forth scrolling. I would strongly suggest defining each acronym at its first occurrence in the main text and, where possible, reducing non-standard abbreviations to improve readability and flow.

Other Comments:

Abstract:

- Please explicitly mention in the Abstract that this study is conducted at a 3-hour temporal scale. Simply referring to a "sub-daily scale" leaves the exact temporal resolution ambiguous to the reader.

- Line12: '*Global'* to '*quasi-globa*l' as it is not covering the entire global ocean. Also it would better to clarify that this study has been carried out at 60 NS.

- I would suggest removing or softening the strong claim in the Abstract stating that PMW upper-level ice scattering is the primary driver of the satellite precipitation phase lag. As noted in the main text comments later, algorithm artifacts such as PMW orbital sampling gaps and IR based morphing could be major co-contributors to these timing delays. Rephrasing this in the Abstract to avoid attributing the lag exclusively to a single physical mechanism would be more accurate.

Introduction

- Line 109-130: One of the primary strengths and novel contributions of this study is that it goes beyond simply documenting diurnal variability across estimates to investigate the underlying physical mechanisms driving these patterns. However, this key novelty is currently underemphasized due to how the introduction is structured. While Lines 109–120 effectively lay out the existing literature gaps, the subsequent paragraph (Lines 123–130) jumps directly into procedural steps rather than framing overarching research goals. To better highlight the paper's main novelty and give readers a clear roadmap, I suggest first explicitly stating 2–3 central research objectives that directly address the identified research gaps, followed by a brief outline of the analytical steps taken to achieve them.

Data:

- What is the main rationale behind selecting those particular datasets. Is there any selection criteria of filters the authors used of only these particular datasets passes or find eligible. If so, please explain briefly that criteria. If no such criteria have been employed briefly mention why these particular datasets have been selected.

- Considering TRMM has long been deprecated and superseded by GPM, what was the motivation behind including this legacy dataset in the intercomparison? Please add a brief justification in the Data section explaining why TMPA was selected alongside IMERG.

- Moreover, why was the TRMM Realtime version selected here instead of the gauge-corrected one, whereas gauge-corrected versions were selected for the rest of the satellite products? If the idea was to verify the difference between near-real-time and gauge-corrected satellite estimates, wouldn't selecting both products from the same satellite framework make more sense, such as comparing IMERG-F with IMERG-E or IMERG-L.

- Also, haven't IMERG products been available back to 1998? Even though the authors mention that the record length won't impact the results, this statement seems contradictory to Line 380, where they state: '*We therefore utilize the maximum available period for each dataset to maximize statistical robustness*'.

- Also, why did the authors limit the study domain to 60°N–60°S? Is this due to data availability issues, concerns regarding dataset quality beyond 60°N/S, or because diurnal cycle and sub-daily variations are less pronounced beyond 60°N/S, making this region the primary focus of interest?

- Line 153: I expected Table.1 to be placed immediately underneath the text, as it was cited earlier in the text then of Fig.2.

- Although the case study location and metadata are described in the text (Data section), they are missing from the figures and tables. Because the case study represents a major focus of the results, it would be much clearer for the reader if its boundaries were explicitly marked in Figure 1 (e.g., with a bounding box or inset map) and its key specifications included in a summary table rather than remaining text-only.

- Line 196-209: The sensitivity analysis evaluating the 'variance-then-average' method is thorough and appreciated. However, to improve the conciseness and readability of the main manuscript, I recommend moving Lines 196–209 and Figure 2 to the Supplementary Information, leaving just a brief sentence and reference in Section 2.

Results:

- The section titles *'4.2 Diurnal cycle patterns and underlying mechanisms'* and *'4.3 Physical mechanisms of the diurnal cycle'* are repetitive and confusing. I suggest renaming them to clearly distinguish between their focus areas to make the structure more reader friendly.

- Use of Sub-subsections: To improve readability and navigation, I suggest introducing sub-subsections in Section 4. Currently, the narrative flows across large, uninterrupted paragraphs, making it difficult to discern where the analysis shifts. For instance, creating distinct subsections for the Irregular Sub-daily Variance (V_{irreg}, Fig. 3) and the Regular Sub-daily Variance (V_{reg}, Fig. 4) would provide a clearer structural hierarchy and help readers follow the results much more easily.

- Lines 474–478: Please consider removing the opening sentences of Section 4. Restating literature background and research gaps here is redundant with the introduction and creates unnecessary distraction at the start of the Results section. Removing these sentences will allow Section 4 to launch directly into the observational findings and keep the narrative concise.

- Line 478: The statement *'Total variance of precipitation at observational sites during boreal summer is* 0.9 $mm^2\ h^{-2}$*...'* is ambiguous. It is unclear whether this value represents an average across all validation sites in the Northern Hemisphere land domain or a specific sub-region. Furthermore, the text does not cite where this specific value appears in the Figures. Please explicitly state the spatial domain being referenced and cite the exact Figure panel/sub-panels (e.g., Fig. 3a/b) at the start of the paragraph. This need for precise Figure citations applies across the entire Results section whenever numerical values are introduced.

- Figure 3: Please enhance the legibility of Figures 3a and 3b. The numerical values overlaid inside the bar plots are currently very small and low-contrast, making them difficult to read. Increasing the font size or placing the numerical labels above the bars (or in an adjacent summary table) would significantly improve visual clarity.

- In Figure 3c, d, the authors present spatial distributions of rat_{subD} for the Multi-Satellite (MS) and Multi-Model (MM) ensembles, using spatial correlation coefficients (r) against ground

observations as a summary metric. For example, while CMORPH appears closest to ground observations in the global bar plot (0.0091 vs. 0.0079 mm^2 h^{-2}), it is impossible to tell their regional distribution. To substantiate the manuscript's core claims regarding ground truth, the authors should include a direct spatial difference maps against gauge observations over land domains with sufficient gauge coverage (e.g., CONUS, Europe, East Asia) rather than relying solely on global mean aggregations. Explicitly mapping the spatial differences would immediately highlight geographical hot spots where reanalysis convective parameterizations succeed or fail (e.g., complex topography vs. flat plains vs. monsoon regions), providing much greater physical utility than global correlation values alone.

- Could the authors present a brief regional consistency check (e.g., selecting representative gauge clusters across CONUS, Europe, and China) to examine the variance spread among 3 to 4 nearby stations? Demonstrating this local gauge-to-gauge variability will establish a baseline for observational uncertainty, helping readers better contextualize the performance differences seen in model and satellite estimates.

- Line:530: "*higher spatial correlation (Δr=0.09)*" is this a typo?

- 660-664: The authors attribute the observed phase lag in satellite estimates primarily to the physical delay between upper-level ice scattering (detected by PMW sensors) and surface rainfall. How can authors be sure this is the primary mechanism driving the lag? Couldn't PMW sampling gaps and the reliance on morphing would have also be playing a major role here, or at least contributing significantly to the observed delay? It would be helpful if the authors could clarify how the impact of algorithm morphing and sampling density was accounted for.

- Lines 671**:** What is the primary physical mechanism behind this behaviour? Specifically, how are the satellite products able to capture the timing so well over CONUS, and what accounts for the underestimation of phase? Please add some physical discussion here to help explain these contrasting behaviours.

Figure.7

- Coul the author clarify what percentage of global/regional land grid cells meet the $rat_1 \geq 50\%$ threshold for each dataset.

- Please consider showing unconditioned phase/amplitude domain statistics (or lowering the threshold) in the Supplementary Information. This would help confirm whether the reported peak phase leads (12–14 LST in reanalyses) are robust across all continental grid cells or if they are sensitive to the chosen threshold

Figure. 8

- It appears CMORPH is being used as the representative product for the satellite ensemble in the subsequent intercomparisons. Please add a brief explanatory line prior to this transition explicitly stating that CMORPH serves as the satellite benchmark, along with a short justification for choosing it over other evaluated products.

- Line 704: Are the author referring to Figure 7 or 8?

- In the caption of Figure 8, the authors refer to *'backward and forward masking in Fig. 6'*. This terminology is somewhat ambiguous. Rephrase text references to use standard descriptive terms

such as *'low-precipitation/amplitude mask (\ hatching)'* and *'harmonic dominance mask (/ hatching)'* to avoid ambiguity.

- It would be insightful to see a brief comparison (at least as a supplements) of how well the reanalysis estimates agree among themselves when separating convective from non-convective precipitation. Does the relative contribution of convective vs. stratiform rain show strong inter-product consistency, or do different parameterization schemes yield significant discrepancies?

Figures 6, 7, and 8: Figures 6, 7, and 8 all focus primarily on the diurnal phase and amplitude (with Figure 8 adding the precipitation type perspective). Because these figures largely repeat the same core message with slight variations, the authors might consider moving some of them (or their redundant panels) to the Supplementary Material to keep the main text concise and focused.

Physical mechanisms of the diurnal cycle:

- **(**Figure.9): Domain-averaging thermodynamic variables (CAPE, CIN, MFC, E) across the entire 0°--60°N region can conflate contrasting convective mechanisms (such as oceanic convection, afternoon monsoon storms, and nocturnal LLJ systems). It might be more informative for the authors to replace or complement the hemispheric mean in Figure.9 with regional diagnostic breakdowns across a few representative climate zones. This would help highlight how specific parameterizations perform under different regional regimes.

- Line 767–768: The authors state that *'CIN remains strongly negative during nighttime hours and begins to erode from 12 LST onward, reaching its minimum around 15 LST.'* Looking at the corresponding panel (middle colorbar), there do not appear to be any positive values shown at all. Could the authors please clarify the sign convention being used here for CIN (e.g., whether they are referring to absolute magnitudes or signed values) and ensure the text and colorbar scale are aligned?

- Another issue with the 0°N–60°N average figures and results is that the values appear quite low (e.g., CAPE reaching around 350 J/kg , which makes it hard to believe they represent real convective environments. Are the authors implying that we should interpret these values strictly in relative terms rather than absolute physical magnitudes?

- Furthermore, a broad 0°N–60°N spatial average is highly generic and overly simplistic. Such large-scale domain averages should serve primarily as general reference background rather than the basis for making strong claims about reanalysis performance or convective physics. Please address this dilution effect or qualify these broad claims.

- Additionally, what explains the large differences in CAPE across the reanalysis products? Is this driven by differences in convective parameterizations, vertical resolution, or parcel definitions (e.g., Surface-Based vs. Mixed-Layer)? Please add a brief explanation in the text.

- Also, why do the CAPE values appear so low here? Could the authors clarify how CAPE was computed and provide a brief note on the typical CAPE range expected for this particular site during convective conditions?

- Line 783: Could the authors clarify what is meant here by 'large-scale dynamical forcing'? Please be specific by providing concrete physical examples (e.g., synoptic-scale vertical motion, frontal lifting, or thermodynamic barriers like CIN) so the reader understands the exact mechanism being referenced.

- Lines 800–825: It is again very hard to follow what the authors are trying to convey here. By the end of this paragraph, there is no clear synthesis or takeaway; it reads more like a simple list of sequential events rather than a meaningful physical interpretation. Please either rewrite this section to clearly articulate the main conclusion or remove it if it does not add substantial value to the manuscript.

- Line:813: *The text states, 'CAPE reaches its diurnal minimum (approximately 200 J kg^{-1}) at 06 LST',* but looking at Figure 10, the minimum appears to occur at 18 LST instead. Could the authors please check this line and correct the timing if this is a typo?

- Lines 815–819: Does this imply that Convective Inhibition (CIN) plays no significant role here? Since the CIN temporal patterns and values look virtually identical throughout the diurnal cycle, please clarify its relevance to the triggering mechanism or address why CIN remains flat while the diurnal cycle evolves.

- Line 835 (Figure 11): It is quite confusing why another figure is introduced for the exact same location (the ARM SGP site) right after Figure 10. If Figure 11 focuses on the same site, why aren't these panels combined into a single figure, or why isn't one moved to the Supplementary Materials? Please clarify the distinct physical insight provided by Figure 11 or consolidate it with Figure 10.

Conclusion:

Streamlining the Conclusion Section (Lines 918–961): The section title should be simplified from *'Summary and Conclusion'* to simply 'Conclusion'. Furthermore, lines 918 to 961 currently read more like an extended discussion section rather than a concise synthesis. I recommend condensing this text to focus strictly on the key physical findings and actionable take-home messages. Removing detailed narrative re-evaluations from this section will make the core conclusions far more distinct and impactful for the reader.