General Comments:

This manuscript represents an important advancement by incorporating observation-based, time-dependent forcing into the stochastic skeleton model of the Madden-Julian Oscillation (MJO). It demonstrates effectively that the model reproduces key statistical characteristics of the MJO, clearly outlining the model's capabilities and limitations, especially regarding seasonal variations and ENSO modulation. However, to strengthen the conclusions and scientific rigor of the study, additional discussions and analyses are necessary, particularly regarding the following points:

Major Comments:

1. Validity and Impact of the 3-month Moving Average

The use of a 3-month moving average for smoothing the forcing profiles significantly influences the results. The authors acknowledge that this smoothing may be responsible for the inadequate representation of seasonality. Thus, further in-depth discussions are required:

- Rationale for choosing a 3-month window: Provide clear theoretical or empirical justifications for this specific smoothing window. Stating merely that it captures "long-term trends" is insufficient. Discuss explicitly how results might differ if shorter or longer smoothing windows (e.g., 1-month or 6-month averages) were employed.

- Necessity of sensitivity experiments: Ideally, sensitivity experiments with different smoothing windows (e.g., no smoothing, 1-month average) should be conducted and presented, at least as supplementary materials. Such analysis would quantitatively reveal how significantly this smoothing affects seasonal and interannual signals.

Without this analysis, the conclusion that observation-based forcing enhances realism remains insufficiently supported.

2. Diagnosing the Lack of ENSO Modulation

A key finding is the model's inability to reproduce the observed modulation of MJO characteristics under different ENSO phases. The authors plausibly attribute this to the model's structural simplicity, such as the lack of ocean coupling and mean-state wind interactions. However, to strengthen this conclusion, it is essential to first diagnose the source of this discrepancy. Another possibility is that the 1D forcing profiles themselves do not adequately represent the ENSO signal.

I strongly recommend that the authors analyze the forcing functions (Sq and Sθ) to determine if they contain a statistically significant ENSO signal. This can be done by comparing the forcing profiles during El Niño, La Niña, and neutral periods.

- If the forcing does contain a clear ENSO signal, yet the model fails to respond, this would strongly support the authors' conclusion about the model's structural deficiencies.
- Conversely, if the forcing signal is weak or absent, it would suggest that the limitation lies within the forcing methodology itself (i.e., the projection of 2D observational data onto a 1D profile).
- This analysis would allow for a much clearer separation of causes and significantly enhance the scientific impact of the findings.

3. Statistical Evaluation and p-value Interpretation (Lines 314–320)

The current statistical interpretation confusing. Typically, a significance level of 5% (p=0.05) is standard, yet the manuscript mentions a level of 0.04 without clear justification.

The choice of a 0.04 significance level seems arbitrary and post-hoc. The authors should evaluate this result against the standard $p < 0.05$ threshold or provide a clear justification for their non-standard choice.