Review of Erfani et al. "Building a comprehensive library…"

This work presents a novel methodology to generate representative samples of marine stratocumulus for LES simulation based on remote observations and reanalysis. The motivation, methods, and analysis are well-written and thoroughly documented. I have a few major comments related to the trajectory and PCA methods and the performance of the LES which warrant additional justification and explanation. After these minor revisions, I believe this work will make an excellent contribution to the literature.

Major Comments:

- Trajectory modeling (section 2.2): I would like to see further details about how the trajectory forward modeling is performed.
    - Which fields are assimilated and which, if any, are freely evolving?
    - Why does the uw-trajectory package utilize such a huge gridbox (2deg x 2deg) to provide input data when finer-resolution data is available from many resources?
    - I am also confused about the purpose of creating an 86h trajectory when only 48h trajectories are relevant to the study.
- PCA & dimensionality reduction: I would like to see better justification for the PCA and sampling methods used, specifically…
    - Can you describe the meaning and purpose of including the difference in each CCF between the beginning and end of the trajectory? What quality of the Sc are you trying to capture by distinguishing the mean and the difference?
    - I'm confused by the decision to remove SST and P_MSL from the list of standard CCFs, considering that PCA is already a dimensionality reduction. The covariance values cited in L260-261 to exclude these factors seem like an arbitrary cutoff, especially considering your determination that correlations above 0.14 are statistically significant (L275-279). Since the ultimate selection and display of factors is performed along PCA space, you could continue to include SST and/or P_MSL in the analysis while still ultimately selecting two principal components to represent the majority of variability.
    - Related to the point above, why did you stop at 2 principal components to analyze variability in the data? PC3 is not that much less important than PC2 (15% versus 19%) and is strongly correlated with delta-omega in a way that is not captured by the first two PCs.

    - Given that your eventual analysis samples from the original dataset rather than strictly along the axes of PC1 and PC2, the downsampled data includes axes of variability that are beyond what PC1 and PC2 alone represent, which questions the utility of the PCA altogether. How well do these downsampled trajectories span higher order PC spaces? If you were to randomly sample your trajectories along the CCF axes instead, would you obtain a substantially less representative downsampled set?
    - How did you select the two trajectories used for LES modeling? In terms of both CCFs and the two PCs, they appear more similar than not within the spread of the data.
- Mismatch in performance of the CTRL case between the two LES examples: While the authors do an excellent job discussing the impacts of varying Na within a given LES setup, I would like to see further explanation or analysis of the fact that the MERRA-2 diagnosed aerosol concentration is not a reliable choice to use for initializing the model aerosol concentration.
    - Given that this work seeks to develop methods for analyzing MCB and SCT, the ability of the model to predict cloud behavior under different aerosol concentrations is of utmost importance. Considering that the LES cannot reliably predict the real cloud behavior using a renanalysis-derived value of Na, how much would you trust the model to predict a *hypothetical* brightened or thinned cloud, where there is no observation for comparison?
    - In L895, you propose initializing with CERES Nd instead, but figures 8 and 9 both indicate (a) very poor temporal coverage; and (b) strong variability in the CERES value. How would you mitigate these issues to come up with a trustworthy value for your forward model?
    - If MERRA-2 offers a predicted value of Nd, it would be useful to show that against the CERES retrieved Nd and LES experiments. A close match between MERRA-2 and CERES Nd would indicate that the prognosed MERRA-2 Na is actually useful, but that some difference or deficiency in the aerosol-cloud parameterizations in the LES leads to the underprediction of Nd in certain cases.

Minor Comments:

- L87-88: LES resolves some subgrid-scale turbulence that is not resolved in a GCM, but it does not actually represent aerosols or cloud microphysics any differently than most GCM approaches (moment and modal models), unless you are specifically referring to using a sectional or superdroplet approach.

- L116+: It would be helpful to clarify at the start of this paragraph that you are referring to the present study or the previous Erfani 2022 work.
- Figure S1 is fuzzy / poor image quality
- L602: typo in “initiation”