Response to Comment CC1

We thank Dr. Loveridge for his interesting perspectives. We extract below key themes from his comment and respond to each in turn

1) First, I would like to push back against the language that features in the abstract and introduction stating that cloud and aerosol properties are averaged to 20 – 100 km resolution to reduce uncertainties. Averaging is not a strategy to deal with uncertainty.

   Thank you for this comment. We agree and now reword the text to focus on the practical reasons for averaging in the context of our analysis. (lines 6, 25, 116)

2) It appears to me that there is an assumed separation between the estimation of 'the rules of the game' and the knowledge that 'the rules are invariant' across a set of samples. To me, it is not clear that this is the case. When reading, I don't see a clear definition of which geophysical variables or properties we can use as evidence that 'the rules are invariant' and which we can use to determine the rules themselves (i.e., constrain processes).

   First, we would like to clarify that the 'rules of the game' are not being estimated. We are attempting to understand processes amongst geophysical variables within systems that may be experiencing meteorological changes over the course of the observation period, i.e., changing rules of the game.
   We agree that this perspective assumes a timescale separation between the variability of the meteorological conditions and the process(es), which we are studying. This assumption is implicit in our definition of Type 1 and Type 2, which are both based on the process timescale being much smaller than the meteorological timescale. While this might not always be the case, one might argue that this assumption may be sufficiently fulfilled in the successful examples of space-time exchange covered in the manuscript. We will provide further justification for this assumption below.

   We now mention the timescale and associated process scale separation in the paper Lines 206-207: *Lending success to this approach is that there is sufficient scale separation between mesoscale processes (our focus), small-scale processes such as cloud-top entrainment or local plume penetration, and longer timescale variability inversion height.*

3) For example, for the stratocumulus case, it is stated that the inversion height is horizontally homogeneous, so the 'rules are invariant', and yet variation in the inversion height appears to be integral to the intra-cell variability as well from Fig. 2.

Broadly speaking stratocumulus (Sc) are characterized by relatively flat tops and ragged bases – as opposed to cumulus with flat bases and highly variable cloud tops. The 'flat-topped Sc are a consequence of a strong inversion. In Lilly's mixed-layer model of the stratocumulus-topped boundary layer, the evolution of this inversion height has a characteristic timescale of 48 hours, while the thermodynamic properties of the boundary layer evolve on a shorter timescale of about 9 hours (Schubert et al., 1979). To the extent that 9 hours can be considered "much smaller" than 49 hours, this justifies an assumed timescale separation.

Note too that we are referring to the *domain-mean* inversion height. Of course, there are local variations in the inversion height, but they are part of another set of rules and timescales (Rayleigh-Benard convection), which are not the focus here. Thus, we select our methodology to focus on processes/timescales of interest.

Changes have been made on lines 203-204

4) Again, for the stratocumulus, do we know a priori that there are no drivers that operate at scales between the cellular scale and ~100 km? Or are we relying on observations (reanalysis?) that demonstrate a lack of variance at this range of scales?

Our analysis of the Sc cellular structure is an example of how observational evidence (cell size, cell aspect ratio, radiative cooling, circulation) can be used along with detailed modeling to build a composite characteristic cell comprising samples from many different cells collected into TWP bins. If meteorological conditions vary across the domain – e.g., changing inversion strength – one might find that various other scales emerge. Our intent has simply been to show that if meteorological conditions are reasonably constant, Sc cells manifest with remarkable self-similarity. This seems in line with assuming timescale separation as discussed above.

5) For the cold-air outbreak trajectory example, the timescales discussed only mentions the timescale of SST gradient. Could there not be meteorological changes that are significant at a timescale of ~12 hours associated with synoptic systems that cold-air outbreaks are often part of?

There are no doubt other meteorological changes that do occur, which will vary from one meteorological state to another. Having looked at a large number of MCAO cases we have seen that sometimes the meteorological gradient *does* change within ~12 h. Nevertheless, SST is a very strong controlling factor and as shown in the analysis, there are indications that one can learn about cloud processes based on our analysis approach. This is again in line with our previous arguments for assuming a timescale separation for the stratocumulus-topped boundary layer.

6) I think it would be great if the authors could be a bit more precise in how they would determine that the 'rules are invariant'.

   We pay more attention to this in the revised version while keeping the manuscript conceptual and focused on the key theme of space-time exchange, Deborah number, and ergodicity.

7) Is this separability real or do we impose (assume?) a scale-break between the resolution of global reanalysis/climate model and the domain size of Large Eddy Simulations that is just an artifact of computational limitations? This is a critical assumption that also appears to underly the authors' arguments, so it would be great to get their opinion on it.

   Global models are good at representing large scale phenomena that LES with their limited domains cannot capture. Likewise, LES captures small-scale physical processes that global models cannot. Modelers always deal with the issue of scale-filtering whereas clearly the atmosphere does no such thing. This doesn't mean, however, that models of a certain filter scale are not useful for a selected problem. The examples presented here include a mix of complementary modeling and observational studies and demonstrate that in ideal cases – particularly Type 1 where meteorological conditions are 'invariant'– much can be learned from snapshots, and that assuming scale separation can be a useful assumption.

8) Am I correct in understanding that ergodicity implies that we can interpret the droplet effective radius profile in cumulus or the cellular structure in closed-celled stratocumulus using a parcel model, rather than requiring a whole LES?

   No, not directly because the r_e profile depends on the dynamics of the cloud system and details of entrainment. A parcel model typically does not represent entrainment, and when it does it is heavily parameterized. It could however be argued that ergodicity implies that it not fundamentally impossible – for given meteorological conditions and potentially very complex sets of parameters – to parameterize cloud evolution as simulated by an LES through a parcel model.

9) The notion that processes can be extracted from Type 1 snapshots suggests to me that we might get more value from observing systems that provide high detail and accuracy in

select conditions (at the expense of sparse sampling) rather than those that sample everything but with little detail or precision. Does this align with the authors' understanding? If so, it might be worth making a recommendation along those lines.

There are multiple aspects to this question including what geophysical variables/processes are being targeted for which scientific question and what observational systems are being considered. Even though certainly of interest, we feel that this question is beyond the scope of the current paper.

10) Perhaps I am misunderstanding, but I have some concerns about the Type 2 cases, where it is stated that they may be useful after careful stratification by meteorology. If drivers such as aerosol and 'meteorology' are correlated across snapshots, then stratification (or other statistical models and their counterfactuals) will produce biased estimates of how clouds respond to an aerosol driver under 'constant meteorology' (and vice versa).

This is true when one only considers a specific meteorological regime. However, even if meteorology and aerosol are correlated, one can obtain causal understanding under those specific meteorological conditions. One can then collect many such covarying conditions. Finally, one can scale up by taking into account of the frequency of occurrence of all these conditions. In this case we argue that the scaled-up understanding is not biased (e.g., https://doi.org/10.5194/acp-22-861-2022).

11) I think it would be helpful for the authors to be a bit more precise about the conditions required for Type 2 cases to be helpful for process understanding in terms of controlling for the variation of slow processes across snapshots.

We have attempted to do so for the examples given here but now add more discussion. For example, we now put more focus on the role of dynamics in the Stephens and Haynes Type 2 retrieval and its connections to the Z-COD composites of Suzuki et al. (2010).

Lines 262-266: A related topic is the use of space-based radar and spectrometer retrievals of Z and COD, respectively, to interpret the relative importance of condensation growth (higher COD but almost no change in Z) and collision-coalescence growth (higher Z but little to no change in COD) \citep{Suzuki10}. Based on the arguments above, when applied to single storm systems one expects such data to be of Type 1, but when compositing over many storms with different dynamics the analysis is expected to be of Type 2.

12) As an opinion piece, our intent is to provide food for thought with a number of specific examples. A comprehensive discussion lies beyond our scope.

13) This statement that events within the SZA < 65 are optimal and therefore valuable for studying cloud processes is not consistent with the available evidence.

Thank you for clarifying this important point. SZA<65 was recommended by Grosvenor et al. 2018 to filter out highly uncertain Nd retrievals but this doesn't mean that after this screening the retrievals are free of uncertainty. We have made the appropriate changes on lines 425-426: *This approach would have to take into account uncertainties in retrievals, particularly at high solar zenith angles \citep[e.g.,][]{Grosvenor18}.*

14) I suggest that the authors simply follow the spirit of their closing statement and avoid distracting from their main point by discussing details of measurement performance. The main point of this paragraph, that measurements with wide field of view and high temporal frequency will be useful, has the same caveat as all measurements (sufficient accuracy) that are discussed in the article. I don't think the authors should stress over justifying this particular type of measurement.

Given that temporal evolution is inherent to process, we consider it important to emphasize temporally evolving (geostationary) data in trying to address process. Measurement inaccuracies from the suite of instruments involved are certainly an issue.