Reviewer 2:

This is an excellent paper about a very interesting open-access dataset (LARA). The authors offer a state-of-the-art global Lagrangian climatology to the research community, which they calculated using FLEXPART and more than 80 years of ERA5 reanalyses. For the first time, such a comprehensive and valuable dataset can be used by researchers worldwide – and potential applications are numerous. The authors provide a selection of interesting example applications that nicely illustrate the potential of LARA. The paper is generally well written, but a few clarifications can further improve the paper. My main concern is that the first application is not at all easy to understand (the way it is presented at the moment) and this might give a wrong impression about the usability of Lagrangian diagnostics. Overall, I recommend accepting this excellent paper with minor revisions.

We thank the reviewer for the very positive evaluation of our manuscript, for the detailed feedback and constructive suggestions. We have carefully considered all comments and have made the necessary revisions to address the points that were raised.
In the following, responses are in blue, and quoted text is shown in green. Text after the little arrow '→' is newly introduced or modified manuscript text in the reaction to the reviewer's comments.

Comments (most of them are minor)

1) L17: "by time" sounds strange, maybe "as the time" works better.

We modified the sentence as follows: "a measure of continentality by time it takes air to reach land from the ocean" → "a measure of continentality ***based on the*** *time it takes for air to reach land from the ocean"*

2) L54: we typically write "airstreams" as one word

Thank you, this has been corrected.

3) L69: should read "ERA-Interim"

Thank you, we corrected this

4) Table 1: units should be "m s-1" instead of "m/s"; symbols should be in math mode, e.g., u, T, q; not sure why you use rather unusual variable names like "sp", "2t" etc.; ps, T2m would be more common à consider using more standard variable names

Thank you, we corrected the unit. The unusual variable names are the short names given of the variables of the ERA5 dataset. We clarified this by replacing the header "Short name" with "ERA5 short name".

5) L123: it would be interesting to have a bit more information about the 6 million particles: how did you decide for this number? What is the resulting mass represented by each particle? How many particles are on average in a 1x1 deg column?

A similar concern was voiced by Reviewer 1. We added a statement on why we chose 6 million particles (storage constraints, 320TB). We also added a short explanation about the domain-filling option which answers the question about the mass of each particle (= total atmospheric mass divided by 6 million). The number of particles per grid cell volume varies

greatly across the globe, due to the smaller grid cell volumes closer to the poles compared to the equator. We added the following to the text: "... fixed air mass of approximately 0.86x10^12 kg of air, ..." and "A 1°x1° column at the equator contains approximately 300 particles at any point in time."

6) L126: a bit strange "the full period … would take … to complete", maybe better "the calculations for the full period …"

We agree and took the suggestion

7) L128/129: confusing, first the overlap period is one year, then 3 months only?

This is a mistake and has been corrected (it should be 1 full year).

8) L132: FYI: the 300 s (= 5 min) time step correspond to 1/12 of the data input interval (1 h), which is the default approach used in LAGRANTO

LAGRANTO also has an iterative Petterssen correction, making the integration more correct. On the other hand, as far as we are aware, it also does not parameterise convection and turbulence, which should result in much less stochastic noise as compared to FLEXPART. For these reasons, convergence could be reached for different time step intervals. In addition, the input resolution could also be affecting the convergence point.

9) L141: the periods are 8-y long

Thank you, this has been corrected.

10) L148: I assume that BL height is directly from ERA5; what about the tropopause height? Which definition is used to calculate the tropopause height?

For clarification, the following has been added to the paragraph: "The topography is taken from ERA5, the atmospheric boundary layer height is computed with the method of Vogelezang and Holtslag (1996) based on the critical Richardson number, and the tropopause height is defined as the first stable layer to fulfill the thermal tropopause criterion (i.e. the vertical temperature gradient is smaller than 0.002 K km^{-1})."

Vogelezang, D. H. P. and Holtslag, A. A. M.: Evaluation and model impacts of alternative boundary-layer height formulations, Bound.-Layer Met., 81, 245–269, 1996.

11) L149-151: I don't understand these sentences, why are BL and TP height not simply interpolated to the horizontal position of the air parcels?

Since these values have only two spatial dimensions, to be space efficient, we decided to keep those in gridded form. We added our reasoning for doing so: "Therefore, to conserve storage space, these are saved as two-dimensional spatial gridded fields..."

12) L160: I don't understand "a month's worth of files normalised per variable"

The NetCDF files are organised as daily files containing all variables, while the Zarr files are organised per month and per variable, so we divided the time it took to open the NetCDF files by the number of variables to make it comparable to opening the Zarr file. This is not very relevant and we see it is confusing, so we rephrased it as follows: "...for example, opening files is ~60% faster."

13) L172: difficult to understand what these Spearman coefficients refer to (particle positions or tracer concentrations?)

The coefficient refers to tracer concentrations, which is now added to the sentence: "...with Spearman coefficients of tracer concentrations between 0.56-0.68"

14) L210: in line with my comment 5: please provide more information about the density of parcels typically available in a box of, e.g., 100 km x 100 km x 100 hPa

We added the following statement: "For example, at a single point in time, a 1°x1° column at the equator contains approximately 300 particles, of which ~150 are below 5 km in altitude and only 10 are above 20 km."

15) L220: "or are fully explored on themselves" sounds strange, I suggest something like ", nor are they investigated in full detail here"

Thank you, we have replaced it following your suggestion.

16) L227: "periods that use different assimilation data" – maybe this aspect could be discussed briefly in Sect. 2.1. The ERA5 dataset is as consistent as it can be, but there are still issues with changes in the datasets available for assimulation.

Thank you for the suggestion. We added the following in section 2.1: "ERA5 uses a consistent model framework over the whole reanalysis period. However, the observation system has changed tremendously over this period, and thus the assimilation of observation may lead to spurious variability and trends in certain variables. Such artifacts may extend to the LARA data set, such that care should be taken when applying trend analyses."

17) L238: "The left panel of figure 2" à please use panel lables and write "Figure 2a …"

Thank you, this has now been corrected.

18) L241: "within the lower parts of the lower atmosphere" is not very clear, can you give a pressure or height range?

By simplifying the figure and text, we removed the paragraph containing this sentence.

19) L243: "in panel (a2)" should maybe read "in Fig. 2a2" (however quite unusual), or you better change panel labels to (a, b, … j)

The panel labels have been changed to a-h.

20) L243-266: honestly, I am a bit lost with the text and the figure … all rather complicated. Do you need so many panels to make the main point? Do you need the correlations with AMO, PDO, …? If yes, then the reader requires a more careful explanation of what is shown and why and how to interpret the results. It would be a pity if the first application of LARA was so complicated that readers get the impression that Lagrangian investigations are hard to understand.

We agree that Figure 2 was too complicated. The purpose of the figure is to highlight the methods of tracing particles and diagnosing trends in circulation patterns. We feel this example shows one of LARA's core strengths. To explain this better, we decided to remove panels, lines, text and added two maps of the particle selection at t=0 and t=20 days for a step-by-step explanation of the methods.

21) L285: 500 hPa should read 600 hPa

It should be 500 hPa.

22) L290: I don't understand the criterion "air mass within a WCB per square metre to be larger than 1 kg", is this a criterion to guarantee a certain density of particles that fulfill the WCB criterion of ascent?

Indeed, it reduces noise from the convection parameterisation.

23) L293: maybe worth noting that your WCB climatology uses different units than, e.g., Madonna et al. – if I interpret your unit correctly, then it refers to a vertical mass flux in WCBs (kg / m2/ s). The comparison with WCB frequencies (as in Madonna et al.) is therefore qualitative (which is fine, just maybe helping the readers to get along with the different units and values)

We added the following sentence: "Different units were used in earlier studies and we therefore only do a qualitative comparison."

24) L321: I very much liked this application; it is easy to understand and clearly demonstrates something that could not been obtained by Eulerian analyses

Thank you!

25) L344: this could be better mentioned already in Sect. 3.1

We agree and have added text in Section 2.1 (see also response to comment 16).

26) L388: it is a very interesting results that "… ATCEc values for specific humidity increase throughout the ERA5 period" – do you have a hypothesis why this is the case?

This is indeed puzzling. We do not have a certain answer for this, but hypothesise that specific humidity is a value that is more 'untouched' by assimilation in the past than in the present.

27) L390: "Similar results to above" sounds strange, maybe "to the ones discussed above".

Thank you, we adjusted the text as suggested.