Summary

We thank the reviewer for providing valuable feedback on the second version of our submitted paper. We have done our best to take into account all remarks raised. However, we disagree with the first comment, which we explain in detail below. In the following we give a detailed list of all the changes made in response to the points raised by the reviewer. Thank you once more for your help in improving our paper.

Main comments

Comment 1

Both reviewers have commented that the selection of every other grid point etc. to degrade the wind field is not consistent with what the information on the model grid represents. Even if the authors find studies that have used the same approach before, the problem remains that there is a spatial sub-sampling of the discretized fluid, rather than a corser representation in terms of averaged properties. The sub-sampling at an interval violates the perception of the fluid as a continuity, which matters for interpolation. Many other members of the targeted audience of geophysical models will have the same reaction as the two reviewers, and will immediately be sceptical to your method and results.

Therefore, it seems to me a moderate but necessary adjustment to your study to average 2x2 and 4x4 grid points to obtain the coarser version of their training and test grids.

Even if this is the first study of this kind in atmospheric science, it should get the community interested into a new methodology, rather than raise scepticism. I therefore strongly recomment that such a basic aspect of the fluid no be overlooked.

Response 1:

We believe that the reviewers confuse aspects of kinematics vs. dynamics of the flow with this point. In our study, we present a method for interpolating wind fields for a kinematic trajectory model (which always uses interpolated winds), whereas the reviewers focus on the dynamical consistency of the flow. We fully agree that, if the goal was to build a neural network that reconstructs a higher-resolution flow from a coarser-resolution one, then the correct approach is as suggested by the reviewers. This would also be possible as an alternative basis for trajectory calculations.

However, to assess the quality of the interpolation, we need a ground-truth state. Now, if we average the wind fields to a coarser resolution, we would also use information from the grid points that we actually want to reconstruct by interpolation. A comparison between high-resolution and low-resolution winds is then not meaningful anymore, since any differences are a combination of **averaging and interpolating**. By contrast, we want to assess the quality of the **interpolation only**. In addition, the interpolation distance to the high-resolution ground-truth grid points (averaging 2x2 grid points, with the new grid point in the centre of these) would then only be half a grid distance, and over such a short distance, interpolation errors will always be very small. So our test, in addition to mixing the effects of averaging and interpolating, would also become less rigorous.
This is also the reason why past studies have used the same technique as we have used, in order to compare different interpolation techniques vs. a ground truth. We see no good reason for deviating from that practice.

Comment 2

The argumentation about performance is somewhat contractictory or uneven. The introduction heavily emphasizes how surprising it is that the "simple" linear regression is still being used. However, this is by no means a surprise. Rather, previous studies have shown that the cost-benefit ratio of higher-order interpolation did not justify other interpolation methods. This is actually stated in L. 59.

Towards the outlook section, you provide an estimate of 1 order of magnitude increase in computation time from the single image superresolution approach to obtain 20-50% lower AHTD. These are quite the same numbers you cite for higher-order schemes in L. 59 (1 order of magnitude, 30%).

I do not find the conclusions balanced in light of these facts. There is quite some overhead with implementing GPU-enabled model code, training, etc. If the same gains can be achieved with simple higher-order interpolation, why is it worth exploring your methods further? I am sure the authors can come up with an answer to this question, but it would be nice to see this properly stated.

Response 2:

This is the first attempt to analyze the accuracy of neural network interpolation for Lagrangian trajectory computations. We choose to compare against the standard method of linear interpolation. For future work it could be interesting to compare our method against other interpolation methods such as higher order polynomial interpolation. At this stage we are also not implementing the method in the most efficient way (we generate the higher-resolution fields offline and then read them into FLEXPART). For a true comparison we would need to implement the neural network implementation directly into FLEXPART and then compare the execution time.

Comment 3

There are still numerous hard-to-read sections in the manuscript. I make some recommendations in the minor comments below. I recommend the authors read some instructions on how to improve the clarity of scientific writing (Gopen and Swan, 1990; Schultz 2009)

Response 3:

We appreciate your remark and went through the entire manuscript again to soften some of our writing style.

Detailed comments

L. 13: we demonstrate -> we find

- ➔ We changed it.

L. 34: rephrase in light of major comment #2

- ➔ We rephrased it.

L. 45-49: This paragraph very similar to L. 24, shorten/rephrase

- ➔ We shortened the paragraph.

L. 50, 53, 56: simple/surprising: rephase in light of major comment #2

- ➔ We rephrased surprising.

L. 70: project -> study

- ➔ We changed it.

L. 71: variety -> range. Please back-up statement with a reference. Maybe it would be more correct to state that this can be the case, but there are for example differences between small-scale turbulence and horizontal turbulence.

- ➔ We softened our statement: For the training and evaluation we consider that meteorological fields are characterized by self-similarity over a range of spatio-temporal scales. This means that the structure of the field from one resolution to a higher one is similar.

L. 97: most impressive: state objectively, e.g. results with smallest AHTD

- ➔ We changed it to: the EDSR network giving the lowest interpolation error and having the shorter training time.

L. 97: greatest ease of training -> most straightforward training. It is not clear what this means in practice.

- ➔ We changed it to: the EDSR network giving the lowest interpolation error and having the shorter training time.

L. 97: exclusively -> only

- ➔ We changed it.

L. 102: ...levels are counted... -> level indexes increase upward, contrary to ECMWF

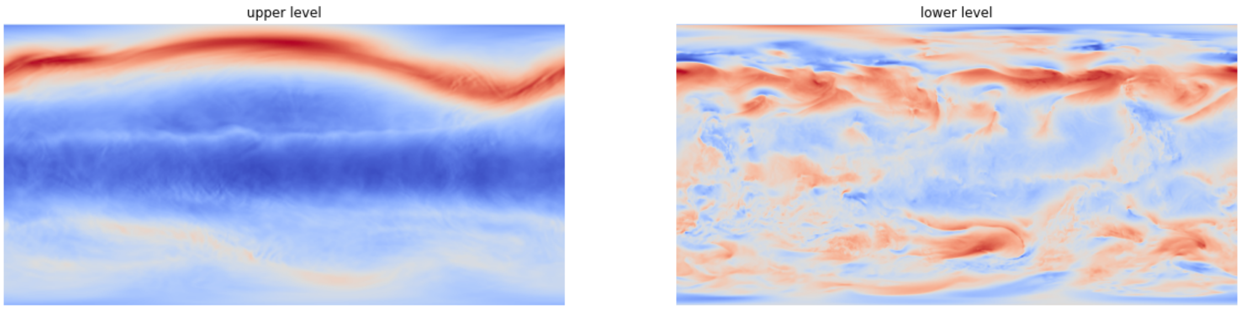
- ➔ We changed it.

L. 106: this choice of method is not consistent with the concept of a continuous fluid, which matters for interpolation, see major comment #1.

- ➔ See answer to comment #1.

L. 120: how is the structure different, and how has this been quantified?

- ➔ When looking at the windfields we saw that the lower fields have more small scale structures compared to the higher levels (see below).



L. 176: compute trajectories: this should be part of the methods, rather than the results. Maybe does not need to be mentioned here.

- ➔ We removed it.

L. 177: we demonstrate -> we compare the accuracy

- ➔ We changed it

L. 180: we demonstrate -> we investigate

- ➔ We changed it.

L. 182: we demonstrate -> we proceed with

- ➔ We changed it.

L. 184: this sentence needs to be expanded to a full description of what is seen in Fig. 2. Deciphering the meaning of this figure can not be left to the reader.

- ➔ We added the explanation: Here, each neural network is used to interpolate each resolution, starting with the resolution the model is trained on.

L. 188: we consider -> we first consider

- ➔ We changed it.

L. 211-219: It was not possible for the reviewer to comprehend what is described here, a figure or table?

- ➔ We summarize the data from the table. (?)

Figure 2: the caption needs to be rephrased to describe panel contents. Methodological statements need to be moved to the main text.

- ➔ In the main text we now also give the methodological statement.

Figure 4: Methodological statements need to be moved to the main text.

- ➔ (?)

L. 229: this is indeed the case -> restate what is "this"

- ➔ We reformulated the sentence to: Fig. 6 shows that trajectories that are advanced using neural network interpolated wind fields closer to trajectories that are advanced using the original "ground-truth" wind fields compared to linear interpolated wind fields.

L. 243: we have also checked -> how has this been done
L. 245: slightly better: how has this been quantified?

- ➔ We changed the paragraph to:We have also checked how well the quasi-conserved meteorological property of potential vorticity is conserved along the trajectories by computing absolute and relative transport conservation errors along trajectories in the stratosphere. We selected particles that were not affected by convection or boundary layer turbulence by selecting trajectories within the stratosphere that never traveled through space where the relative humidity exceeded 90\%. A full explanation of the method we used can be found in [stohl1998].The absolute and relative transport conservation errors of potential vorticity showed insignificant differences between the different trajectory data sets.

L. 251: restate what "this" refers to

- ➔ We reformulated the sentence to: This way neural network interpolation could make semi-Lagrangian advection schemes much more accurate.

L. 285: see the papers -> see the studies

- ➔ We changed it.