Review of paper

**A holistic framework to estimate the origins of atmospheric moisture and heat using a Lagrangian model**

by J. Keune et al.

submitted to *Geosci. Model Dev.*

This is a much-needed study contributing to quantitatively assess the reliability of Lagrangian source diagnostics. As pointed out by the authors, these diagnostics potentially provide very valuable insight into the atmospheric moisture and heat budgets; however, it is intrinsically difficult to quantify errors and uncertainties associated with these methods. I therefore fully support the intention of this study, and to a large degree also the used methodologies; however, in its current version the paper is difficult to read. I find the notation confusing in several (important) places and I could not understand the idea and implementation of the "random attribution method". Therefore, major revisions are required to streamline the paper, clarify concepts and notation, and make the paper in the end more reader friendly. It will then be a valuable contribution to the field.

**Major comments**

A) Line 89: "the application of these models and tools to assess diabatic heating and heat transport lags behind". I am not sure that I agree with this statement. The study by Pfahl et al. (2015) is an important one but certainly not the first one in this direction. Early applications of trajectory computations with reanalysis data in the 1980ies and 1990ies looked at latent heating in cyclones and warm conveyor belts, and how this latent heating affects the potential vorticity structure of the systems (e.g., Whitaker et al., 1988; Reed et al., 1992; Wernli and Davies, 1997; Rossa et al., 2000). These were not yet full budget studies, but I would claim that Lagrangian methods first looked at latent heating and only about 1-2 decades later also at moisture sources and transport.

References:

Reed, R. J., Stoelinga, M. T., Kuo, Y.-H., 1992: A model-aided study of the origin and evolution of the anomalously high potential vorticity in the inner region of a rapidly deepening marine cyclone. Mon. Weather Rev., 120, 893–913.

Rossa, A. M., H. Wernli, and H. C. Davies, 2000. Growth and decay of an extratropical cyclone's PV-tower. Meteorol. Atmos. Phys., 73, 139-156.

Whitaker, J. S., Uccellini, L. W., Brill, K. F., 1988: A model-based diagnostic study of the rapid development phase of the President's Day cyclone. Mon. Weather Rev., 116, 2337–2365.

B) Line 94: I think this is a slightly misleading summary of the Quinting and Reeder (1997) study. They mainly emphasized the role of adiabatic descent, and their last sentence of the abstract says "Likewise, the role of the local surface sensible heat fluxes is deemphasized.".

C) I am completely lost with understanding the "random attribution" method (section 2.3.2) for several reasons. First, I don't understand the notation "length nt": is this n times t? And then later, what is ix, nx, … an later n_min … ?? Most likely this requires a schematic where you explain also the notation. Then how can you use Delta q,random in step 1 if you calculate it only in step 3? Then I am completely lost with step 2, and I also don't understand the general motivation for doing this. Can you explain this method and the motivation for it in a much better way?

D) Fig. 3b is a key figure of this study. Since, e.g., the methods SOD08 and RH-20 vary in multiple ways (additional RH criterion, different Delta q threshold) it would be interesting to know which change had the largest effect. It would be very useful to have a more in-depth discussion of which criteria affect the results shown in Fig. 3b.

E) While I agree that this study addresses important technical aspects of moisture and heat source identification, the text is rather heavy to read, and the results are mainly presented in a statistical way, which is hiding a bit what is going on technically. To me, it would be useful to have a didactic example, starting with a single trajectory and then a set of trajectories, which helps me better understand the differences between the methods and the effects of the bias corrections etc.

F) The random attribution method has an important effect on estimating the transport time between uptake and rainout (Fig. 9c). With the random attribution method, you have much more "old uptakes" and therefore you have more long-range transport and remote sources (Fig. 10). This is very interesting and most likely an important result of this study (see also your discussion in lines 672-682). My problem is just that I didn't understand the random method (see my point C above) and that I don't find physical reasons in the paper why the random method has these effects compared to the linear attribution method. Again (see my point E above), a case study with a few trajectories might be very helpful for explaining what is going on.

G) How sensitive are your main conclusions with respect to the total number of parcels calculated with FLEXPART? I don't ask you to redo a certain analysis with more parcels (this might be too time consuming), but it might be interesting to look at the effects of reducing the number of parcels.

**Minor comments**

1) Line 18: “synergistic impacts” on what? And what is meant by “a cohesive assessment”, maybe “coherent assessment”?

2) Line 28: here reference to Sodemann et al. (2008) would be more appropriate than Sprenger and Wernli (2015).

3) Line 33 and in other places: I think references should be listed in chronological order.

4) Lines 36-74: I appreciate this nice summary of Lagrangian approaches to identify moisture sources. What may be missing is a remark that Lagrangian approaches suffer from accuracy errors of trajectory computations, which can be substantial for trajectory integrations over several days. These errors stem from limitations of the numerical schemes, and most likely more substantially from the limited temporal resolution of wind fields available for offline trajectory computation.

5) Line 88: “A myriad” seems a bit exaggerated.

6) Line 98: maybe this summary of recent Lagrangian heat wave studies should also include the one by Zschenderlein et al. (2019): Zschenderlein, P., A. H. Fink, S. Pfahl, and H. Wernli, 2019. Processes determining heat waves across different European climates. Quart. J. Roy. Meteorol. Soc., 145, 2973–2989.

7) Line 159: I think that the notation Delta_q($t_0 - t_{-1}$) is not ideal. Delta_q does not so much depend on the time difference but rather on the two times themselves. I therefore suggest that Delta_q($t_0$; $t_{-1}$) would be more appropriate, or maybe even Delta_q($t_{-1}$; $t_0$).

8) Line 174 and elsewhere: units should not be in italics.

9) Line 181: I was first confused and thought that *z* is a function of *m*, but your m is the unit of z. I don’t think that you need to mention units in the text, or you write “*z* (in m)”.

10) Line 190: either “applied” or “used”

11) Line 198: is Delta q_i the absolute change? I assume that Delta q_i is negative if mixing with free tropospheric air occurs and then the Delta q_i condition is trivially fulfilled and does not help to exclude mixing with dry tropospheric air. Please clarify.

12) Lines 213/215: I am not sure that I understand these RH criteria. Evaporation is particularly intense for dry air, and so why shouldn’t intense (ocean) evaporation not lead to a strong increase in RH? And for the heat flux H, I assume that H leads to warming and therefore to a

reduction of RH, so Delta RH should be negative, meaning that the criterion Delta RH < 10% is trivially fulfilled(?). Please clarifiy.

13) Line 235: I expect the opposite: with 6-hourly data we estimate the diurnal cycle of the ABL height poorly and therefore the ABL height criterion might be important. If we had hourly data (e.g., with ERA5) then there should be less sensitivity with respect to the design of the height criterion.

14) Line 260: “for E for P” should read “for E and P”.

15) Line 274: Strange formulation “Due to the consideration …, mass and energy are conserved …”. I think mass and energy conservation is valid independent of what is considered by the algorithm(?).

16) Line 285: “time step (t)” should read “time step *t*” (italics).

17) Line 288: I don’t see the need to introduce a new notation 1_Delta Phi … for this function. This is the Heaviside step function, which in this case could be written as H(–Delta Phi_j).

18) Lines 285-299: please clarify whether everything here is identical to SOD08, or whether you introduced some modifications.

19) Lines 300-307: I don’t understand why this explained here after the linear discounting, appears a bit out of place.

20) Line 358: I again struggle with the notation: is x here an index? If yes, why then do you write it as a superscript of S_LM? I realize that at this point of the paper I cannot really follow any further, mainly because of confusing notation. What are indices, what are coordinates, what are just subscripts/superscripts …? Does “LM” mean “Lagrangian model” or something else?

21) Line 497: typo in “heat”

22) Fig. 6: I cannot find the information how you define “local”; does this “local region” have the same size for all cities?

23) Figs. 7 and 8: I find it very difficult to see something in these many panels, except that they all look very similar. I think the smooth blue-only color bar does not help. Can you find an alternative way of visualizing the results that is more insightful for the reader and that makes the differences more apparent? Maybe by showing difference fields from a “reference setup”.

24) Line 661: why are the new criteria better to assess global warming trends?