Review of "A holistic framework to estimate the origins of atmospheric moisture and heat using a Lagrangian model" by Keune et al., submitted to GMD

Keune et al present a framework for the evaluation of Lagrangian methods for quantitative offline-diagnosis of heat and moisture from air-parcel trajectories. There exists quite a number of studies with similar yet different concepts and implementations, and the community is clearly in need of ways to enabe comparison and verification exercises. In this regard, the paper is clearly a needed and welcome contribution to the literature. In addition, the manuscript is well-written and most of the material clearly presented. I have a number of comments with respect to some of the literature and interpretation, and to the presentation of figure material, detailed below. Since my attention has mostly been on the moisture source identification, I mainly focus my comments on those aspects of the paper. I have no reason to conceal my identify, also because it will be quite evident from my comments that I am the main author of one of the methods assessed here.

Harald Sodemann

Major comments

1. The description of the accounting procedure is not entirely clear or may miss one important point. I recommend to separate two aspects more distinctly, (a) considering the fractional contributions of the uptakes (source contributions) during the uptake (i.e. how much does a source contribute to what is in the air parcel at the end of the time interval), (b) discounting all previous contributions according to their relative share of all water vapour in an air parcel. Step (a) is a fundamental change from methods without accounting, that only consider the local humidity change, rather than the fractional contribution times the arrival precipitation.
2. It should be mentioned somewhere that there is a physical/theoretical basis for the assumption that all sources contribute to precipitation en route and at the arrival point according to their share of the total water vapour in an air parcel, namely the assumption of well-mixed conditions within an airparcel within a 6-h time interval. The random accounting procedure that is presented in Sec. 2.3.2 does not have such a theoretical basis. Other than being a sensitivity test, it is unclear how reliable/meaningful the results obtained with such a random attribution approach are in terms of physical interpretation.
3. Such Lagrangian offline diagnostics as discussed here will always be imperfect approximations of how water vapour moves in a model simulation. What is your take on the question, what level of accuracy we actually can expect from such

methods?

4. The term "holistic" in the title has in my perception connotations that are not well covered by what the proposed framework actually encompasses (being valid for heat and moisture specifically, rather than "everything" as holistic could imply). How about replacing with a more limited word, such as "unified" or "generic"?
5. The paper currently seems to introduce both a verification framework, and a modified source accounting algorithm with additional parameters. A clearer statement of this dual objective, and a potentially clearer separation of both aspects in the manuscript (method/results) could be beneficial to avoid confusion with the reader about the focus and intent of the paper.

Detailed comments

L. 35: "while others trace air parcels and their properties": I first misread this to comprise also the accounting-type methods, such as S08, but then two paragraphs later understood, how you build up the story. Maybe it can be made more clear how you distinguish the different aspects, and still include S08 in the list of references in L28?

L. 35: FLEXPART and Lagranto trace airmass motion and interpolate boundary-field variables to the parcel position.

L. 73: There are a few additional references that use the S08 method, that may be relevant here, including Sodemann and Zuber, 2010; Sodemann and Stohl, 2009 (introducing the FLEXPART basis, and testing trajectory length and deltaq sensitivity for Antarctica); Winschall et al., 2013 (introducing the uptake time perspective).

L. 38: "the tracking of air parcels": add "the tracking of water vapour from air parcels" or something to that effect

L. 58: "if all parcels are homogeneously...": such a global initialisation as used in FLEXPART is just one way to initialize trajectories, one can just as well release particles from just a column or from a regular grid over a specified region

L. 65: "discounting in a linear manner": I do not find the choice of the word "linear" entirely intuitive. There could be a clearer separation between the calculation of fractional contributions and the discounting in case of precipitation in this paragraph (see main comment #1).

L. 68: Sodemann and Stohl (2009) used the dq threshold of 0.1 for polar regions, and evaluated the sensitivity to trajectory length for such regions. 15 days seemed to be a lower limit here, which may be important for the results obtained in Fig. 1 for polar regions. The ABL/no ABL distinction has been topic also in Winschall et al., 2014 and in Sodemann and Stohl (2009).

L. 80: I would express this a bit more nuanced, in that Sodemann (2020) propose to consider the lifetime distribution, and highlight that the long lifetimes that are part of the mean of the distribution are beyond reach or highly uncertain for Lagrangian diagnostics. In addition, the highly skewed lifetime distribution is probably more appropriately described by its median (as is commonly done for highly skewed distributions). See also the recently published review paper by Gimeno et al., 2021.

Figure 1: I like Figure 1 in that it clarifies the flow of the analysis. Could it be possible to add information on the different forms of uncertainty entering the diagnostic, such as from the trajectory calculations, the detection, the attribution etc., that then add to total uncertainty?

In Eq. (2), A needs to be defined, and something be said about A and m are determined in the analysis shown in the results part.

L. 182: I do not understand why a distinction is made between f_z over land and ocean, there may be a misunderstanding of the relevant passage in S08, but if f_z has been applied, an f_z = 1.5 has always been applied uniformly over land and ocean.

Eq. (7): The maximum RH change criterion is not yet obvious to me. You state that "large RH changes are typically associated with ABL growth and warming": why is that inconsistent with evaporation?

L. 260: How dependent are your verification results on the chosen thresholds?

L. 269: This reads as if Sodemann (2020) said that 15 days is a proxy for the globally averaged maximum lifetime. I may have overlooked it, but I do not find this statement in the cited paper.

Sec. 2.3.1: I believe this description would be clearer by separating into the fractional accounting of the arrival precipitation, and into the discounting due to precipitation en route (major comment #1).

Sec. 2.3.2: The reasoning in this section is hard to follow. Could you clearer lay out the idea behind the random attribution, and contrast to the idea of the well-mixed air parcel (not well-mixed atmosphere) in the "linear" (or fractional/sequential/well-mixed) accounting? Maybe an example would also help.

L. 346-365: I find the question of bias correction for evaporation quite intriguing. If studies indicate an overestimation, this would cause too local sources (due to overly large contributions at each time step). Potentially it would make sense to mention this already here? Note that such bias correction as applied here is only possible for global-scale studies, at least local studies suffer from the fact that only the share of evaporation contributing to a certain region is diagnosed.

L. 386: Reanalysis data include humidity perturbations from data assimilation (Läderach and Sodemann, 2016), which are another source of uncertainty of these diagnostics, and one of the motivations for using a (large enough) threshold value for humidity changes.

L. 435: The verification is done using your newly introduced additional thresholds. Here a clearer separation from the verification framework introduced just before will be useful.

Fig. 2: How do the results in Fig. 2 compare to the same kind of evaluation for the S08 method?

L. 449: This paragraph starts with the conclusion, before presenting the facts. Consider reverting the order of the paragraph.

L. 462: This statement seems to conflict with the statement in L. 449.

Fig. 3: I believe the regional results here are obtained with fixed dq thresholds. To what extent do the findings argue for the need to adapt the method to a specific study region?

L. 497: Typo in "heat"

Fig. 4, 7, 8 and similar: A more distinct colour bar, with a clearer separation from white will print better. Consider using less colour categories to allow reading off numbers/categories.

Fig. 4, lower row: these graphs are almost identical. Do you have an explanation why the source correction is overriding the diagnostics so strongly?

Fig. 6: This figure may be more informative as a table, maybe with the addition of numbers for the bias corrected results.

L. 532: Change to "There are ..."

L. 540: I am not used to the term "recycling" for heat, is this a well established expression?

L. 572: In what sense do you find the similarity of the source region maps reassuring?

Fig. 10: The colors are very similar and do not print well on all printers. Consider using patterns or a white/light region in the middle segment.

Sec. 4 (Discussion): This section needs a clearer distinction between the part of the study dealing with a verification framework, and with a modified accounting method. Consider combining the Discussion with the Conclusion section, which is now rather a summary of the study, similar to the abstract. You could also list the main findings again as bullet points to facilitate grasping the take-away messages for the reader.

L. 655: This seems to fit better to the conclusions than the discussions (or could re-appear in the conclusions)

L. 677: I think Sodemann (2020) does not claim that the discrepancies is entirely an issue of definition, see the comment to L. 80 above.

L. 707: Given the lack of a real theoretical basis for the random accounting, I would formulate this conclusion more carefully. There is certainly uncertainty in the accounting, but how large the uncertainty stemming from the accouting is in relation to the overestimation of evaporation is not finally answered from your study - and deserves further investigation.

Supplemental material

Sec. 4: I could imagine this section to better be placed in the main manuscript (see major comment #1).

L. 123: "Contrary to Sodemann...": It is not entirely clear what you consider to be the sources of the ABL uptakes, if not convective detrainment of BL air into the free troposphere, and on what basis you make your argument here. A more direct reference for the cited statement is Winschall et al., (2014).

References

Gimeno, L., Eiras-Barca, J., Durán-Quesada, A.M., Dominguez. F., van der Ent, R., Sodemann, H., Sánchez-Murillo, R., Nieto, R. and Kirchner, J. W.: The residence time of water vapour in the atmosphere. Nat. Rev. Earth Environ., https://doi.org/10.1038/s43017-021-00181-9, 2021.

Winschall, A., Sodemann, H.,Pfahl, S. and Wernli, H., 2014: How important is intensified evaporation for Mediterranean precipitation extremes?, J. Geophys. Res., 119: 5240–5256, doi:10.1002/2013JD021175

Sodemann, H.and Zubler, E., 2010: Seasonality and inter-annual variability of the moisture sources for Alpine precipitation during 1995-2002, Int. J. Climatol., 30: 947-961, doi:10.1002/joc.1932.

Sodemann, H., and Stohl, A., 2009: Asymmetries in the moisture origin of Antarctic precipitation, Geophys. Res. Lett., 36, L22803, doi:10.1029/2009GL040242.