RC3 — Harald Sodemann
received and published: 27 Aug 2021

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

We are grateful for Harald Sodemann's endorsement and thoughtful comments on our work, which will help us to further improve the quality of the manuscript. We also highly appreciate his decision to forego anonymity. We reply to all comments below. <u>Underlined replies highlight planned changes in the revised version of the manuscript</u>.

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.

We fully agree with this distinction. We tried to refer to 'linear discounting' for (b) and to 'linear attribution' for (a) — if we interpret the reviewer's definitions correctly, because we are missing the word 'losses' in the discounting procedure. In our case, we referred to 'linear discounting' as the procedure, in which one 'discounts' uptakes with losses *en route*. Subsequently, we referred to 'linear attribution' as the procedure, in which one 'attributes' how much a source contributes to a sink. Please note also, that our notation differs from the one in Sodemann et al. (2008); in particular, we prefer to think of absolute contributions (Eq. 15) instead of 'fractional contributions'. Yet, the underlying concept remains identical. However, we understand from the reply above that the term 'discounting' may be used differently. Therefore, and because we also agree that this difference could be better highlighted in the manuscript, <u>we will revise parts of this section in the new version of the manuscript</u>. If the reviewer believes that those changes are insufficient, and that the terminology may still be confusing, we remain open to consider other alternatives.

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.

We thank the reviewer for highlighting this; we will better highlight the well-mixed assumption in the revised version of the manuscript.

Linear discounting and linear attribution follow the assumption that parcels are perfectly mixed; in which all sources *always* contribute with their *exact* share of specific humidity in the air parcel prior to the precipitation event. However, we believe that the random attribution also follows this same well-mixed assumption — at least on average. By construction, there may be deviations from the 'perfect' well-mixed situation, but these average out over many parcels (and long time scales).

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?

We thank the reviewer for this question. First of all, we wish to clarify: our framework builds up on (offline) Lagrangian simulations, and we 'only' evaluate the uncertainty inherent in the evaluation of these trajectories. As such, our work is limited to the accuracy of these simulations; i.e. our results depend on the accuracy of the trajectories, but also on the number of parcels that are being tracked and the time step and spatial resolution of the reanalysis data that is used to force the simulations. We will mention this dependency explicitly in the revised version of the manuscript.

Further, we want to elaborate on the reviewer's question: we agree that these simulations will always remain imperfect — inaccuracies stem from a lot of sources, such as the spatio-temporal resolution of the driving reanalysis, the reanalysis itself, the number of parcels that are being tracked, and numerical errors in the interpolation scheme, just to name a few. However, at this point, and especially due to the sparsity of measurements to validate these simulations, it remains difficult to assess how (in-)accurate these simulations are. It is true, however, that the inaccuracy can be expected to increase with increasing trajectory lengths as small errors add up. As a result, we refrain from analysing trajectories longer than 15 days; and we would not like to rely on single trajectories. For the estimation of source regions from these trajectories, however, we expect that average source regions over many trajectories and long time scales) are reliably detected as we expect some of the inaccuracies to average out.

From our perspective, there are a few ways forward to improve the accuracy of these trajectories and the resulting source region estimations. First, we believe that the time step of the driving reanalysis is critical too: we have to assume that processes such as E and P take place at the midpoint between two locations and time steps. For 6-hourly (and even 3-hourly) time steps, as employed here, this presents a large uncertainty — that could add to the numerical errors from the simulations. Other studies, such as Tuinenburg and Staal (2020) further show that the vertical structure of Lagrangian simulations can have a large influence on the recycling ratios; and the reviewer's paper (Sodemann, 2020) also shows that assumptions about vertical mixing and the employed convection scheme come along with uncertainties. Consequently, we expect that a higher spatio-temporal resolution of the driving forcing (e.g., using ERA5) and the tracking of more parcels can improve the accuracy of these (offline) trajectories, and reduce parts of the uncertainty of the source region estimation. Nevertheless, we highly encourage and support intercomparisons with other models to gauge the uncertainty inherent in such simulations.

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"?

The term ‘holistic’ referred to the merge of tracking heat and moisture; also to the fact that we introduce a workflow that encompasses all steps, from the unconditional detection of fluxes to the bias-correction, along with possible variations in all of them. However, this comment is in line with another reviewer's comment and we may reconsider replacing ‘holistic’, e.g., with ‘generic’ or something similar.

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.

We wish to highlight uncertainties along all steps of the evaluation workflow, which leaves us with a few more aspects that require separation. To keep the manuscript short and readable, we highlighted only the steps that introduce the largest uncertainty. However, we understand that a bit more explanation is needed to make the paper accessible to a broader audience. Therefore, and in line with comments from another reviewer, we plan to update the introduction and specify our objectives better; we will introduce additional figures that highlight the difference between the verification part and the source region estimation, and we will revise the structure of the methods section.

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?

Our introduction follows the following structure: we first introduce models (e.g., Eulerian, Lagrangian), and then explain differences between Lagrangian approaches (water vs. air parcels) — and in l. 35, we simply mention the main references to Lagrangian models that trace air parcels. Just from there onwards, we specify the methods used to evaluate the latter — which is where we see Sodemann et al. (2008). Our references in l. 28 were not intended to be complete (and we will add ‘e.g.’ to this reference list), but we do not mind adding Sodemann et al. (2008) as a main references for Lagrangian models too.

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

Indeed — FLEXPART (and Lagranto) interpolate from the large-scale air mass motion to the parcel position. This is, to our knowledge, not restricted to the atmospheric boundary layer and boundary-layer variables. Further, while this describes the technical procedure, we believe that the ‘tracing air parcels’ picture is an accurate visual description of the underlying idea.

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).

We thank the reviewer for providing further references, and we will incorporate some of them where appropriate.

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

In our understanding, the former includes the latter — which is why we refer to the ‘tracking of air parcels and their properties’ in most parts of the text.

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

That is true. However, from our experience and understanding, most (global) FLEXPART analyses of moisture sources use such a global initialization. While the framework is equally applicable to regional domains with a homogeneous initial condition (and a corresponding boundary condition), it is not applicable to point/column releases. In the latter case, not only the mass of each parcel changes with time; but the (unconditional) detection of fluxes is impossible — and further prohibits a source bias correction. We will add a sentence to the manuscript that highlights that our framework is designed for homogeneous parcel distributions.

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).

We agree on the separation issue (see our reply above). We note, however, that both steps of the linear discounting/attribution, apply some linearity assumption: (i) moisture (or heat) uptakes between a source and a receptor are ‘discounted’ by any losses en route — using the ratio of uptake to specific humidity content of the air parcel; and similarly, (ii) the contribution of a source region to the final precipitation event is also estimated using the ratio of the (discounted) uptake to specific humidity content of the air parcel, multiplied with the ‘final’ moisture loss through precipitation.

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).

We wish to emphasize that the global diagnosis as presented in Fig. 2 (we believe that the reviewer is referring to this Figure when referring to polar regions) is based on two consecutive time steps only — the results in Figs. 2–3 are thus independent of the trajectory length. Only the estimation of source region contributions is based on longer trajectories, i.e., up to 15 days (60 consecutive time steps backward from the receptor region). While we agree that trajectory lengths may play a role, our source region analysis is restricted to the cities of Denver, Beijing and Windhoek and hence does not consider polar regions. Further, we find that the linear discounting/attribution often shortens remote impacts, especially for moisture — which is in line with the rather short (average) residence times from Läderach and Sodemann (2016). Consequently, we do not necessarily see the need to examine other trajectory lengths. We will highlight, however, that the dq thresholds can be calibrated using the trajectory length.

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.

We will revise this sentence and update the reference.

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?

We will consider updating this figure but wish to remain restricted to the uncertainty inherent in the evaluation of trajectories. While we will elaborate on additional uncertainties in the trajectories/simulations in the text (see previous replies; and replies to other reviewers), we refrain from adding this uncertainty to the figures, as it would divert from our objectives and workflow. Further, and also in response to other comments, we will add figures on the detection/attribution processes to the main text and/or the supplementary material.

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.

The area *A* is defined in l. 173. For the analysis, we introduce the area / receptor region around l. 408 ("a 3°x3° box around each city center is used as a receptor area"). However, we will consider adding a description of the parcels mass *m* for our global FLEXPART simulations with 2 million air parcels, and explicitly mentioning the area *A* for the three regions in the methods section and the captions of Figs. 6+10.

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.

We thank the reviewer for clarifying! This was indeed not clear to us. We will fix the corresponding sentences in the revised version of the manuscript.

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?

It is not 'inconsistent' with evaporation. However, one could argue that the relative humidity can change only because the ABL is growing and warming. To clarify: we wanted to question the criteria already published and provide complementary alternatives — to gauge the full uncertainty inherent in these criteria. If only Delta(q) values larger than a minimum are considered for the detection of E (as in Sodemann et al., 2008), one could also question the appropriateness: (i) because Delta(q) displays the difference between e and p, and hence Delta(q) will often be smaller than e (or E); and (ii) because also evaporation can be small. Consequently, small uptakes should be considered as well. To assess the uncertainty associated with that, we wanted to introduce a criterion that does the opposite and filters for a maximum increase.

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

The results do, of course, depend on the thresholds for verification. However, we chose small thresholds to evaluate if any parcel detects the flux — which is better than no detection at all, as no detection cannot be bias-corrected. We tested several small thresholds (e.g., 0.01 mm day^{-1}, 0.1 mm day^{-1}, 1 mm day^{-1}) and found minor differences for the global evaluation (Figs. 2-3). Larger fluxes, using larger thresholds, are instead more difficult to detect (not at least because of the e-p issue mentioned above). Such an analysis is, however, out of scope for this study. The chosen thresholds represent adequate values that highlight the differences between the results.

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.

<u>We will delete the reference from this sentence.</u>

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).

We partly agree. This differentiation is already mentioned in Section 2.31: we define the discounting first (Eq. 13), before the 'attribution' of source region contributions to a sink quantity is introduced (Eq. 14). However, <u>we understand that the term 'contributions' also for the former (e.g. in l. 286) is misleading and will revise our wording in this section.</u>

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.

We thank the reviewer for this suggestion. In response, <u>we will add a figure to the main manuscript and/or the supplementary material that illustrates the idea behind the random attribution</u>. Further, as also suggested by another reviewer, <u>we will add examples to the supplementary material, that highlight the differences between both attribution methodologies.</u>

For the reviewer's interest, we also wish to highlight that the random attribution could also be performed on the ‘discounted’ uptakes along a trajectory (that take losses *en route* into account) — and thus could be used to evaluate the fractional or ‘linearity’ assumption in the attribution step. This is, however, out of scope for our current study.

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.

We fully agree with the statement that the overestimation of fluxes could indicate an overestimation of these source regions — if one wishes to identify ocean/land sources only. We understand, however, that the conceptual idea behind some studies is not restricted to oceanic and land origins and instead evaluate large-scale convergence and divergence zones of the vertically-integrated moisture transport; thus also comprising, e.g., phase changes and above-ABL mixing as sources of moisture. Therefore, we do not wish to dive into this discussion and restrict our analysis to the detection of surface source regions — and indeed, using global-scale or regional-scale studies, which track air parcels that are homogeneously distributed over a specific domain. As noted above, we will clarify this restriction.

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.

We agree with the reviewer that the data assimilation impacts humidity perturbations and that this depicts another source of uncertainty. However, as outlined in one of our previous replies, one could also argue against a minimum threshold for the detection of evaporation from humidity changes. As we wish to evaluate the overall uncertainty arising from these thresholds, we only wish to evaluate the ‘post-processing’ uncertainty and refrain from diverting towards the uncertainty inherent in the setup of the simulations (such as the number of parcels and/or the driving forcing).

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.

We are very sorry, but we do not understand this comment. If the reviewer believes that this is important, we kindly ask him to clarify what he means.

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

The average verification statistics are quite similar (cf. for example Fig. S2, which shows the same maps but for ALL-ABL). We designed the presentation part of the manuscript in such a way that Fig. 2 shows the spatial patterns, and Fig. 3 highlights the differences for all applied detection criteria. We do not intend to include all figures for all criteria in the manuscript, as it would be overwhelming. We are happy to share all our results with the reviewer though.

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

We believe that this sentence is a good summary of the findings and makes the text easier to comprehend, and it aligns with the findings from the previous section. It further only refers to the validation of precipitation, whereas the following paragraphs highlight the verification of E and H.

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

Each paragraph in section 3.1.2 describes the validation of one flux only. I.e. the first paragraph discusses the validation of precipitation (l. 449–452, Fig. 3a), the second paragraph discusses the validation of evaporation (l. 453–462, Fig. 3b), and the last paragraph discusses the validation of the sensible heat flux (l. 470–478, Fig. 3c). Thus we do not see how these statements conflict.

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?

The reviewer is right that the results in Fig. 3 are obtained with a fixed dq threshold for SOD08 and FAS19 — but also fixed RH thresholds for the RH criteria. It is also true that this figure could be used to argue that a (regional) calibration of these thresholds is needed. We wish to emphasize that none of the presented thresholds is calibrated though. Instead, our intent is to assess the overall uncertainty from, at base, different types of criteria. We will make sure that this is emphasized in the revised version of the manuscript.

L. 497: Typo in "heat"

We thank the reviewer for noticing. The typo will be fixed.

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.

We thank the reviewer for this suggestion; however, a broader classification does not emphasize the differences between the approaches — which was our intent here (which could also be done plotting differences, as mentioned by another reviewer, but we do not want to use just one criterion as 'the' reference). We have published the data along with the manuscript, which allows any reader to analyze differences on their own. However, we will consider adjusting the color scale in the revised version.

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

Yes. The source correction has a strong impact on the estimated source region contributions for two reasons. First, the trajectories that are evaluated are the same. The criteria differ only in the identification of source locations along that trajectory and the corresponding attribution. Second, the applied criteria are emblematic of a systematic overestimation (e.g., SCH20 and ALL-ABL) or a systematic underestimation (e.g., SCH19) of the sensible heat flux. Thus, on average, the same source locations are identified — and the source bias-correction does exactly what it is supposed to do: it removes the systematic over- and underestimations of the applied approaches.

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

We believe that Figures are easier to comprehend than tables (in this case, this would be a 12x3 table) — and wish to highlight that all numbers are explicitly mentioned in the text. We thus wish to refrain from replacing this figure with a table; hopefully the reviewer is fine with it.

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

We believe that the sentence is correct and does not require changes.

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

We thank the reviewer for mentioning this; as this is indeed not well described in the current version of the manuscript — and as we do not believe that this is a well established expression (yet). While we understand that ‘recycling of heat’ is not as intuitive as the recycling of moisture, energy can be recycled. <u>We will either introduce a definition for this or revise the wording of this sentence.</u>

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

While the ‘random attribution’ does follow physical limits (e.g., through a maximum uptake set by the minimum specific humidity content of the air parcel), it contains two random factors: the source location and the magnitude of its contribution is sampled in an iterative procedure until the sink quantity is fully attributed to a set of source locations. Thus, we believe it is reassuring if the identified source regions are similar in shape and magnitude, adds credibility to the approach. <u>We will remove ‘which is reassuring’ from this sentence as it does not add much.</u>

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.

We appreciate the comment — and it is true that this color scale may not print well on all printers; but it is at least colorblind safe. <u>We will think about alternative color scales for the revised version</u>, but would prefer to keep the current colors.

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.

The discussion is already subdivided into two parts related to the comments from the reviewer: the first paragraph discusses the verification results; i.e., the different criteria and their accuracy and reliability for detecting surface fluxes and precipitation and how we expect the detection of fluxes to change with, e.g., other forcing data sets, or a changing number of parcels. This is, however, discussed in the context of source region estimation —

which is the overarching motivation for this study and the verification exercise. The remaining paragraphs then discuss the source region estimation and additional uncertainties related to that. These parts do not only deal with the impact of the attribution methodology (linear discounting and attribution in comparison to random attribution), and put these into the context of the residence time, but also with the impacts of bias correction and the potential to further develop the framework.

From our perspective, a clear separation of this section to the 'Summary and Conclusions' section is needed. The discussion really focuses on issues that have not been dealt with in this manuscript, it provides an outlook on further applications and developments, and it remains — to a large extent — speculative. Thus, we wish to keep these sections separate if possible.
We thank the reviewer for the suggestion on the bullet-point list for the 'Summary and Conclusions' though and aim to adapt it in the revised version of the manuscript.

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

We will add a similar statement to 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.

We will revise this sentence so that it does not appear as if Sodemann (2020) claimed it.

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.

As indicated above, there is a physical basis for the random attribution. We do not claim that the comparison of both attribution methodologies spans the uncertainty inherent in the source region/contribution estimation. But as long as alternative and validating measures are lacking, we believe that the random attribution is a valid alternative to assess some uncertainty. Nevertheless, it is true that further investigation, and observations, are needed to unravel the 'true' sources of uncertainty.

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

Maybe we misunderstood comment #1, but we do believe that these are two different issues. From our perspective, the differentiation between discounting and attribution is already mentioned in the main manuscript (but will be better highlighted; see our reply above). The issue described in section 4 of the supplementary material, is a different one. Here, we mention that not all precipitation can be attributed to the identified source regions. While this is, of course, caused by and related to the linear discounting and attribution – it remains a different issue. We decided to move this part to the supplementary material, as (i) the bias correction fixes this issue; and because of that, (ii) we do not show any differences related to this discrepancy in our results. We have done some analysis in this regard and would be happy to share it with the reviewer; but we believe that a thorough investigation is beyond the scope of the manuscript.

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).

For our reply here, we assume that the reviewer means free troposphere uptakes (i.e., *above* ABL uptakes).

Our general intention is to identify the surface source regions of moisture only. Therefore, we wish to identify air parcels that are *directly* influenced by surface evaporation, which we assume to be the most dominant source of moisture in the ABL. For a sufficiently high temporal resolution, changes in specific humidity of ABL air should reflect this evaporation (disregarding phase changes for simplicity), even if some of the evaporation ends up detrained through convection. With the 6-hourly time steps from ERA-Interim, however, it proves difficult to disregard all parcels that strictly reside within the ABL. Hence, with the approach presented in the manuscript, we identify air parcels that reside within the maximum ABL of two time steps during at least one time step. As such, we might actually also sample some of the detraining air parcels — if the RH criteria are applied, however, we still disregard parcels that have already (strongly) mixed with tropospheric air and thus exhibit a very different relative humidity. Similarly, we also account for a few entraining air parcels that, e.g., gain moisture through mixing with ABL air masses. However, we wish to exclude air parcels that simply gain moisture through mixing; e.g. an air parcel that just passes over the ABL but gains moisture through mixing and/or convective detrainment. From our perspective, these above ABL moisture sources represent *indirect* moisture sources — whose surface source might be farther away than the place where the mixing occurs. Thus, we assume that these are not representative of ABL processes anymore and that further (secondary) tracking of these air masses would be required to identify the corresponding surface source.

We thank the reviewer for the alternative reference, which we will add to the manuscript.

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.

References

Läderach, A., and Sodemann, H. (2016). A revised picture of the atmospheric moisture residence time. Geophysical Research Letters, 43(2), 924-933.

Sodemann, H. (2020). Beyond turnover time: constraining the lifetime distribution of water vapor from simple and complex approaches. Journal of the Atmospheric Sciences, 77(2), 413-433.

Sodemann, H., Schwierz, C., and Wernli, H. (2008). Interannual variability of Greenland winter precipitation sources: Lagrangian moisture diagnostic and North Atlantic Oscillation influence. Journal of Geophysical Research: Atmospheres, 113(D3).

Tuinenburg, O. A. and Staal, A. (2020). Tracking the global flows of atmospheric moisture and associated uncertainties. Hydrological Earth System Sciences, 24, 2419–2435.