Answer to Harald Sodemann (R2) in the Interactive comment on "Simple physics-based adjustments reconcile the results of Eulerian and Lagrangian techniques for moisture tracking" by Alfredo Crespo-Otero, Damián Insua-Costa, Emilio Hernández-García, Cristóbal López and Gonzalo Míguez-Macho

The authors present a study focused on the comparison between Eulerian and Lagrangian approaches to trace moisture and to identify the evaporation sources of precipitation. Using a regional model simulation with water tagging as a reference, they then evaluate two Lagrangian offline approaches in that framework for a set of Atmospheric River events from different regions. Two tunings are proposed to reduce a general bias towards shorter transport distances in Lagrangian methods. The study is overall interesting, presented clearly, and well written.

Thank you very much for your detailed review. We believe that the modifications you suggest have improved the manuscript and the robustness of our analysis. Please, find below the responses to your comments.

However, the fairly coarse choice of tagging regions, as well as the exclusive selection of AR cases introduces limitations that are currently not well addressed. A more careful and balanced discussion of the results and implications from this study are thus advised. I also see further issues with the proposed tuning and with regards to some parts of the literature detailed below that the authors should address when preparing a revised manuscript.

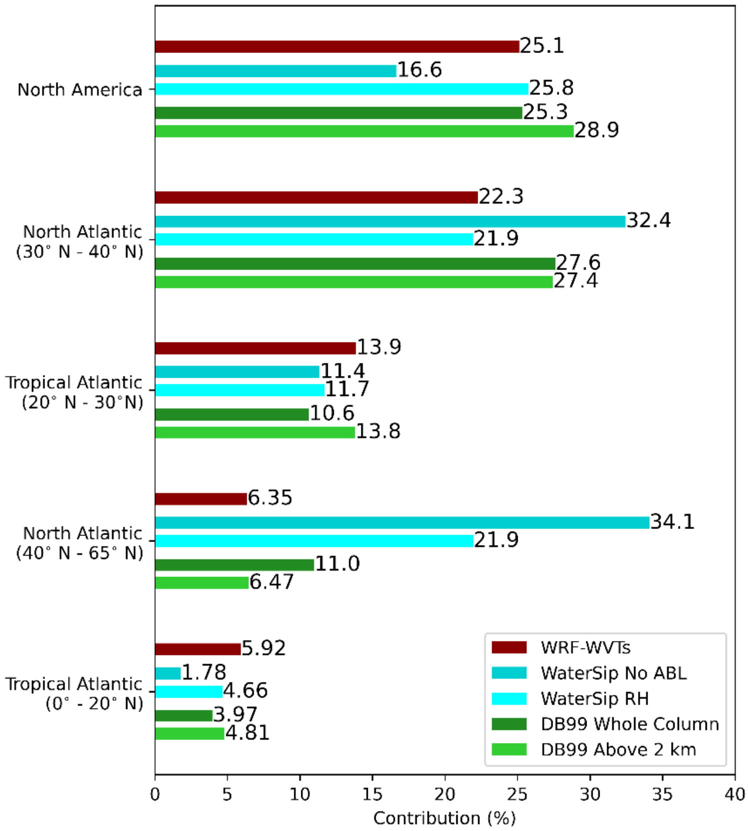
We are aware of some limitations of our study, particularly regarding the exclusive selection of AR cases. Because of this, we will soften the language in the abstract and conclusions and recognize them in the discussion. Moreover, we conducted additional experiments to address some of the issues you encountered, which will appear in the response to your major and specific comments.

Major comments:

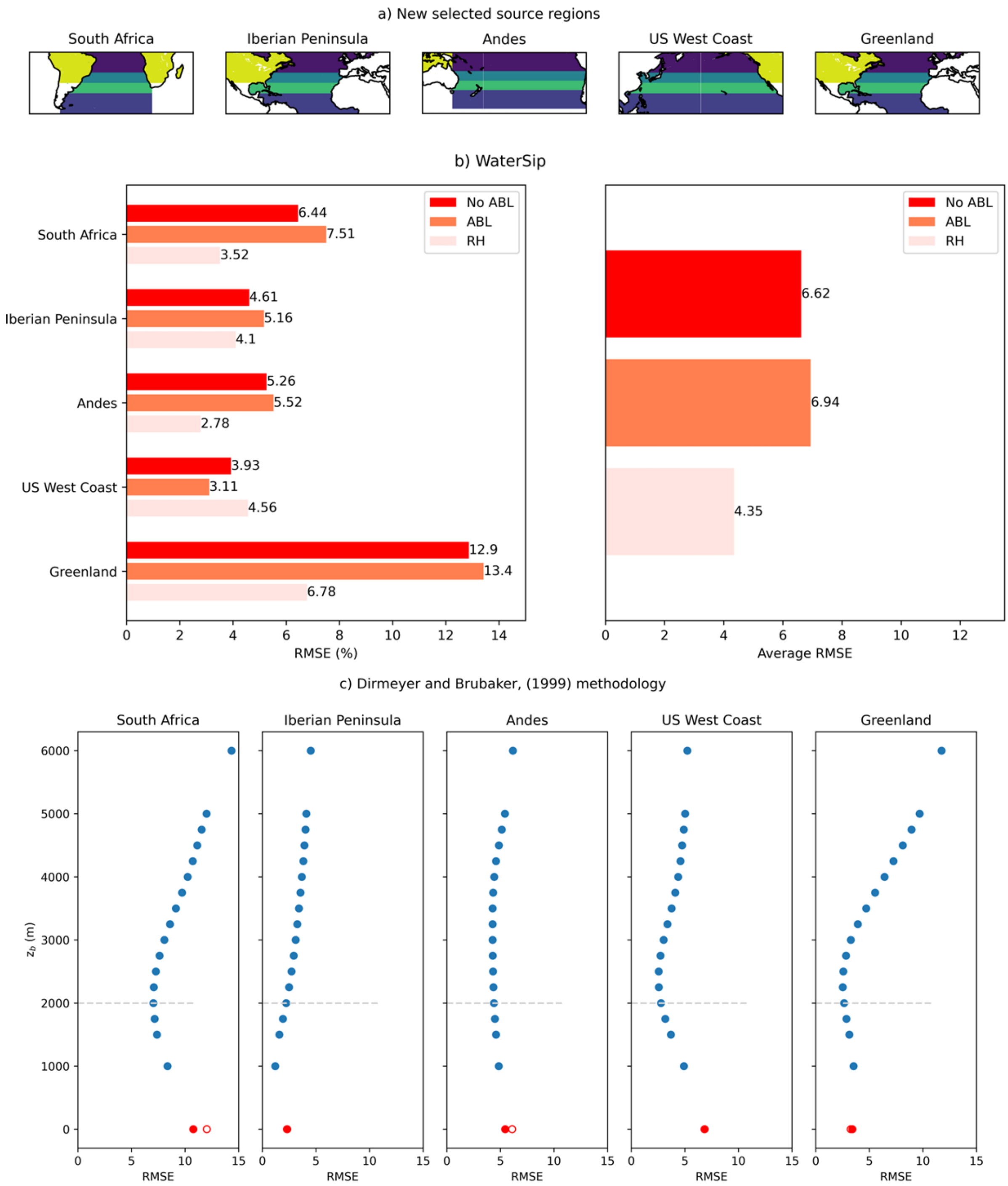
1. Coarse definition of tagging regions. The authors subdivide the hemispheric land and ocean into 9 sectors, separated along 30 N and S. This allows only for a very coarse distinction between ocean basins and continental areals and the boundaries. As the Lagrangian diagnostics are showing, the majority of sources are located at different regions within the same ocean basin. As a consequence, the RMSE computed here only picks up the outermost differences. An example for this is seen for the Greenland AR, where the structures in the North Atlantic region widely differ between the two Lagrangian approaches. The current tagging setup misses these differences entirely, and exclusively focusses on the fringe of the moisture sources. There are probably two ways to approach this deficiency: One is to increase the number of tracer subdivisions depending on every case, adding complexity to the study, but providing more sharpness in the tagging simulation (e.g. using a setup similar to Sodemann and Stohl, 2013). The other way is to openly address this deficiency in the study design, and adjust the discussion to be more nuanced, and formulate the conclusions more carefully.

We agree with the reviewer that our selection of source regions overlooks some differences that may exist between both Lagrangian approaches, this being evident in the Greenland case. This selection is based on the fact that for many applications in the field of moisture sources we are only interested in the basin as a whole and not so much in the contribution of the different sub-areas of the basin. An example is the calculation of continental recycling; in that case we are interested in the contribution of the continents as a whole. Another recurring example with atmospheric rivers is the separation between tropical and non-tropical moisture inputs, for which we only need two sources.

However, for comparison purposes, we acknowledge that this division of sources may substantially affect the results, so we have conducted an additional experiment. Specifically, for each AR case we divided the tropical and extratropical ocean where it is located in two more parts (thus four sources for each basin) and repeated the comparison with WRF-WVTs using only six source regions for each case: the four sources for the corresponding ocean, the next most important source region (continental in all cases except for the Andes) and the rest of the world. This is a slightly coarser configuration than that of Sodemann and Stohl, (2013), but it overlooks much less differences than the original one. For example, in the Greenland case this new setup reveals that the correct distribution of moisture sources - in relation to WRF-WVTs - is that of UTrack, as shown below (UTrack is now referred to as DB99). In particular, WaterSip overestimates the contribution of the northernmost part of the basin, i.e. the part closest to Greenland.



However, what we found is that the main conclusions of our study do not change with this new configuration; when the errors are averaged to obtain a single deviation per event (e.g. Fig. 5), the values obtained are very similar. Thus, this experiment validates our results, and will be included in the Supplement, where we will add the following figure with the results of Fig. 5 and 7 for this new selection of source regions. Additionally, we will briefly refer to this experiment in the main manuscript just after discussing the appropriateness of the selection of source regions.



2. Biased selection of cases. The study includes five AR cases from different parts of the world. All cases are thus potentially related to a large amount of long-range transport. While this selection in itself is no matter of concern, proposing a general tuning of the Lagrangian methods based on a selection of long-range transport cases only is problematic, as it may introduce biases during cases of more local precipitation sources (e.g. convective summertime precipitation, weaker precipitation cases). The focus on AR cases only, and the limitations following along with that, should be more clearly highlighted in the title, abstract, and conclusions.

We agree with the reviewer and therefore propose the following changes:

- New title: "Simple physics-based adjustments reconcile the results of Eulerian and Lagrangian techniques for moisture tracking in atmospheric rivers".
- Abstract. We will rewrite the last sentence of the abstract: "Although these modifications may need to be adjusted for other types of precipitation events, our results demonstrate that Lagrangian techniques are a viable and compatible alternative to Eulerian water vapor tracers, and that the main discrepancies between the different methodologies can be derived from the obviation of certain physical considerations".
- Conclusions. We will again recognize that the proposed modification to UTrack (specifically, the threshold for release height) may depend on the type of precipitation event when discussing the achieved improvements in RMSE.

3. Proposed tuning to the WaterSip method. The authors propose to introduce a relative humidity threshold in the WaterSip method during the identification of precipitation/moisture loss events en route. While such a proposal seems physically plausible at first, there are some downsides as well. Importantly, a moisture loss can be due to one of two reasons, either removal of water vapour from the atmosphere due to condensation and precipitation, or due to the mixing with drier air masses. The second case will be necessarily ignored in unsaturated situations if a relative humidity threshold is introduced as proposed here. Ignoring the lowering of specific humidity due to mixing can then lead to an over-accounting of the moisture sources, i.e. a larger amount of uptakes are assigned to the specific humidity of the air parcel that are contained within. Duetsch et al. 2018 proposed a distinction between mixing events and rainout events. However, both types of situations still need to be part of the accounting method to be physically plausible.

We were not aware that the modification we are proposing had already been introduced in a previous study. Therefore, in the revised version we will acknowledge that it is an existing and used modification to the WaterSip method. This will be recognized both in the methodology and in the conclusions. To be more explicit, in the last case we will do it at the beginning of the third paragraph: "In the case of WaterSip, we assessed a modification already applied in Dütsch et al., (2018) and Cheng and Lu, (2023)". Regarding decreases in specific humidity due to dry air mixing, in our opinion, precisely what the relative humidity filter does is to prevent these decreases from being attributed to precipitation, which would be incorrect.

A more conventional tuning of the WaterSip method is to change the specific humidity thresholds and the time step. While the authors have tested different time steps, the specific humidity threshold has been set to a quite low value compared to literature (a common value is 0.2 g kg-1 6h-1). The specific humidity threshold will have a similar effect as the RH threshold, and is justified by interpolation errors in the offline approach. Can the authors report how sensitive the moisture sources are, and thus the RMSE values to a variety of changes in the specific humidity threshold?

Following the reviewer's suggestion, we have performed several experiments in which we modify the specific humidity threshold and found almost no changes. Regarding the time step, there is a clear dependence on that parameter, as explicitly shown in the Supplement. In the revised version, we will change the standard setup to be that of the literature (0.2 g kg^{-1} in a 6 h time interval) and include in the main manuscript the dependence of the average RMSE on both the specific humidity threshold (setting dq to 0.01, 0.05, 0.1, 0.2 and 0.3 g/kg) and time step (1 h, 3 h and 6 h). This will show that the optimal choice is that of the literature, although the changing dq is not very relevant. All this additional information will be included in a revised version of Figure 5, by adding three more panels reflecting the dependence of the average RMSE on both parameters (one for each of the No ABL, ABL and RH configurations).

Ultimately, I think one also has to acknowledge that offline trajectory methods do have their inherent limitations, both from the computation of trajectories, and the specifics of the moisture source diagnostic, which are sort of the price for the lower computational expense, and the more detailed spatial information on the source location. Knowing different methods' limitations may be in the end more valuable than tuning methods towards an expected or desired outcome for a specific type of cases. Maybe the authors could reflect on this perspective in their discussion and conclusions?

We agree that knowing the limitations of the tool being used is always essential in order to make a proper interpretation of the results. However, we believe that, in the long run, merely describing the limitations of tools does not make them any better, which is precisely what we are trying to do here. Finally, it is worth stressing again that the changes we are proposing have a physical basis, not a mere tuning of methods to match each other.

4. Title, abstract and conclusions appear too wide-ranging. As partly commented in the points above, the present study has limitations from the method design with respect to tracer setup and case selection, and the tuning of Lagrangian methods can lead to inconsistencies in the method. The discussion throughout the manuscript should be more nuanced and balanced by taking up these limitations. In particular the abstract is now formulated in a very definite, concluding language, which does not seem justified in the light of the limitations mentioned above. The title also suggests to a superficial reader that studies should generally apply the proposed tunings, but their general validity is questionable, or is at least not generally established. In particular the aspect of AR case selection could be included in the title. The study design with coarse tagging regions does in my view not 'reconcile' different approaches, but is rather a tuning using a particular choice of parameters. Maybe the title could be rephrased in terms of sensitivity, and mention the importance of long-range transport for the examined cases?

We have already mentioned how abstract, title and conclusions will be reformulated to reflect that our study is focused on AR-related precipitation events. Regarding the selection of source regions, we believe that the additional experiments we have conducted, together with the new figure in the Supplement and its corresponding discussion in Sect. 3.2.2 give validity to our results. Nevertheless, in the same discussion we will explicitly acknowledge that a different choice of source regions could be better for comparison purposes.

5. Use of literature. There are some citations of previous tagging studies that are missing or could be valuable to add. There are also some wrong citations (Lagrangian method cited in Eulerian context). These publications are listed in the detailed comments below.

We have updated the bibliography and citations with your comments, and this will be reflected in the revised version of the manuscript. Thank you very much for that.

Detailed comments

L. 21: What unit does the RMSE have, is this in percent, or a fraction?

The RMSE has the same units as the precipitation fractions, we are expressing them as percentages. We will correct the manuscript indicating these units where appropriate.

L. 22: "narrowly superior performance": How significant are the differences of less than 1 (%?) between both methods considering all sources of uncertainty?

We agree with the reviewer that this difference is not significant, but this statement is no longer in the new version of the manuscript.

L. 23: Maybe clarify that this is a relative improvement, since the RMSE appears to have the same units. The 50 % relative improvement could be misleading, both because the untis are the same as for the RMSE, and given the overall quite small RMSE difference. Can the overall result be presented more balanced and objective here?

In order to avoid misunderstandings, we will use the mean absolute error skill score (MAESS) to present the main results of our study in the abstract.

L. 24: I think this conclusion statement is going too far. The selection of cases and limitations in the setup does not allow this conclusion. Expressed more neutrally, the sensitivity test and tuning performed here increase the amount of long-range transport detected from the Lagrangian methods. There is not sufficient evidence presented supporting that the tuning is valid generally in all cases. Maybe instead it could be emphasized that the overall approach of using a Eulerian tagging setup to validate Lagrangian methods is promising, but needs further refinement for generally valid modifications.

We agree with the reviewer that we should soften the language, especially given our focus on AR events. Because of this we will modify the abstract and mention that the changes we are proposing may need to be adapted for other types of precipitation events. Moreover, we will emphasize that Lagrangian moisture tracking techniques are a real alternative to Eulerian water vapor tracers. All these changes were already highlighted in our response to the major comment 2.

L. 35 and elsewhere: It is customary to sort references by year of publication. Consider adding Yoshimura et al., 2004 to this list. Sodemann and Stohl (2009) is a Lagrangian study, did you mean to cite here Sodemann et al. (2009)?

We thank the reviewer for its clarification, we meant to cite here Sodemann et al., (2009). We will also add Yoshimura et al., (2004), and sort references by year of publication.

L. 36: "Lagrangian transport models": Lagrangian transport models are the general category of models that simulate airmass transport. To be more specific to the case here, consider rephrasing as "Lagrangian moisture source diagnostics".

We will rewrite “Lagrangian transport models” as “Lagrangian moisture source diagnostics” in the revised version of the manuscript.

L. 39: I do not know of an existing online implementation of a Lagrangian moisture source diagnostics. The offline/online distinction can however be made regarding the tagging and Lagrangian methods.

We only state that there exists another possibility to classify moisture source diagnostics, and this allows us to introduce the difference between offline/online methods, which is used later.

L. 40: "most academics often use": this point is debatable, there exist a range of studies that do make such comparison efforts.

We agree with the reviewer that there are different studies that do make such efforts. However, it is also true that the majority of studies use a single model.

L. 41: "results can be highly discrepant": Winschall et al., 2014 does not provide highly discrepant results, at least that is not what is said in this paper. Please rephrase to do justice to the actual state of the literature, and to better clarify the intent and actual novelty of this study. In this context, please also consider the book chapter of Sodemann and Joos (2021).

We will rephrase “results can be highly discrepant” to “results may not be in agreement”, as we understand that although Winschall et al., (2014) does not provide highly discrepant results, they do show that tropical contributions calculated in one way or another, for example, may be clearly different.

L. 46: Consider adding references to the original AR studies in this context.

We are going to add a reference to Zhu and Newell, (1998) following the reviewer’s suggestion.

L. 47: This statement does not seem to do justice to the existing literature. See Sodemann and Stohl (2013) for a tagging study focused on AR events, as well as Stohl et al. (2008) for a study with Lagrangian methods. There are also a range of studies from other regions and locations (e.g. Terpstra et al., 2021, Bonne et al., 2015). Please update this statement in light of existing studies, and clarify what this study adds to the existing literature. Please also take notice of the book chapter about AR moisture budgets (Sodemann et al. 2020). What is meant by "go beyond the identification of moisture sources to quantify them?"

We will reformulate this paragraph to include the references that the reviewer provides, in order to improve the presentation of the state of the art on moisture sources in atmospheric rivers. Since the main objective of our study is to compare two Lagrangian moisture tracking methodologies with WRF-WVTs, we will delete "go beyond the identification of moisture sources to quantify them", as our study is not focused on providing additional insights on moisture sources for precipitation in ARs.

L. 55: There are two aspect here that are a little bit mixed together. One is that the tagging simulation is also only a model representation of the actual water cycle in nature. At the grid resolution of the model (here 20 km horizontally), a large spectrum of the processes affecting the water cycle are parameterized. I assume that also a deep and shallow convection parameterisation (which one?) has been used in the Eulerian model simulation. Obviously, the model will thus not be identical with nature. However, the approach and argument of the present study is, as I understand it, that the tagging water cycle and the Lagrangian methods are internally consistent, even if the tagging results differ from nature. This is important, as the authors write, since the source information that is being sought after is not directly available from observations.

This statement, although debatable, is based on the fact that Eulerian water vapor tracers do not suffer from some of the pontentially more conflicting approximations that other methods suffer from. To give a couple of examples, WAM2Layers suffers from the “well mixed assumption”, and UTrack lacks a convection parameterization. To give another, in WaterSip, at least in the version we use, phase changes along the parcel trajectories are not taken into account. All these processes are considered in Eulerian water vapor tracers, so it is reasonable to assume that, if implemented correctly in the corresponding model, they should provide more realistic results. As for the parameterizations in the WRF simulations, we will explicitly state them in the methodology of the revised version of the manuscript, and they will also be mentioned in the response to a later comment.

L. 57: Another important limitation of the tagging approach, which also becomes apparent in this study, is the requirement to predefine moisture sources in this forward calculation approach. If more spatial detail is required, the computational overhead multiplies and can become prohibitive. In contrast, the Lagrangian backward approaches provide spatially detailed information, that can be more easily interpreted, for example in terms of the physical processes related to weather systems. This discrepancy between both approaches is important to mention here.

We thank the reviewer for his suggestion. In the revised manuscript we will explicitly mention this drawback of the Eulerian water vapor tracers, by adding the following sentence: “Additionally, the amount of information they provide is limited, as they need to predefine the moisture source to be tagged”. In addition, we will also mention that Lagrangian methods provide gridded information, which is another advantage, along with their higher computational efficiency.

L. 59: Maybe mention here that the Lagrangian methods, being offline diagnostics, require a range of assumptions and parameter choices to which these methods are sensitive. Your comparison framework allows to assess what biases exist with the different diagnostics, and how those are related to parameters and assumpations in the Lagrangian methods.

Once more we thank the reviewer for this suggestion. We will explicitly state here that Lagrangian methods “are sensitive to a range of hypotheses and parameter choices, which significantly increases their uncertainty”. These hypotheses and parameters are discussed further in the methodology, and the uncertainty they cause becomes apparent when introducing our physics-based modifications.

L. 61: "fully validated": I assume this relates to the internal consistency of the tagging approach. Validation can be misunderstood as a comparison to observable quantities. Please clarify/rephrase.

We are going to rephrase “fully validated” to “internally consistent”.

L. 70: This is not correct, Sodemann et al. (2008) used trajectories from the LAGRANTO model (Sprenger and Wernli, 2015).

We thank the reviewer for his clarification. We will include this corrected information in the revised version of the manuscript.

L. 73: "limited to highlighting ... large discrepancies": Please rephrase to do more justice to what is presented in the cited studies. For example, Winschall et al. (2014) specifically investigated the basis of the boundary layer vs. free troposphere distinction in the WaterSip method.

What we mean here is that in these studies the authors did not propose any modifications to reconcile the results of the methods used. In any case, we agree that this sentence is misleading, as it implies that these studies only compared the tools, when in fact they looked at more, as in the case of Winschall et al., (2014). This sentence will therefore be rephrased in the manuscript.

L. 77: "two of the most widely used" -> "two widely used"

We will correct this typo in the revised version of the manuscript.

L. 84: "vast majority ... force": there is no evidence supporting this statement. I don't think it is necessary to make this statement, adding reanalysis data is useful because unlike forecast data, it includes analysis increments from data assimilation, see Fremme et al., 2023.

We will delete this sentence because we have not tested it, but from our experience we believe that in most studies on moisture sources with FLEXPART, the model is forced with reanalysis data. A more thorough literature review would serve to verify this.

L. 93: It is certainly positive with different AR cases, but these cases are all long-range transport events. Can you add some clear justification for this focus in the introduction? Some of the writing makes the impression that you seek general validity, while the focus on AR events only seems in contradiction to this.

On the one hand, we focus on AR cases because we plan to compute moisture sources from a climatological perspective using one of the Lagrangian moisture tracking methodologies assessed here in the future, which is mentioned in the introduction. On the other hand, we agree that considering other types of precipitation events would give more validation to our study. However, we are aware of the existence of a model intercomparison effort in the moisture tracking community, where different types of cases are analyzed, so we decided to focus our study exclusively on ARs in order to interfere as little as possible.

Figure 1: Please add panel labels, and mention all figure panels in the caption. It would be a large advantage to have common color bars for the left and right column panels each. For precipitation amount, it is quite common to use a categorized color bar to that end. This would avoid the saturation of the color scale that now seems to occur.

We will add panel labels and a reference to the geographic region in which precipitation is tracked, and use a categorized color bar as suggested.

Table 1: Could this table include information about the total rainfall amount of these events in the model and maybe observations? Is it correct that the two last events have the same date and time, but different regions?

We thank the reviewer for his suggestion. We will include two more columns in Table 1 with the precipitation in mm from the WRF simulations and the reanalysis ERA5. Regarding the last question, there was a typo, the correct initial date and time for the Greenland case is 2012-07-09 12.

L. 119: What has been used in terms of deep and shallow convection, microphysics schemes in the WRF simulation?

In the WRF simulations, the main parameterizations used were the Yonsei University (YSU) for the boundary layer, the WRF single-moment-6-class (WSM6) for microphysics, and the Kain-Fritsch for convection. In the revised version, we will move this information from the supplement to the main manuscript, specifically at the end of the first paragraph of Sect. 2.2.

L. 121: Which fields have been nudged, only winds or also specific humidity? How does the nudging affect the tagging? The authors emphasize the importance of the nudging, but actually I think for the study objective it does not make a difference if the results resemble the actual events closely or not.

Apart from the winds, temperature and geopotential height are nudged, and this will be mentioned in the revised version of the manuscript. As to whether nudging is important or not, it is crucial when comparing with FLEXPART forced with ERA5 data.

L. 126: What is meant by this statement, and how does it relate to this citation? Consider maybe citing Gimeno et al., 2021 here.

We meant to cite van der Ent and Tuinenburg, (2017), instead of van der Ent and Tuienburg, (2013). In the first case they explicitly show the positively skewed probability density functions for the residence time of atmospheric water vapor. This justifies the long duration of our simulations, 30 days. This typo will be corrected in the revised version of the manuscript.

L. 133: If QFX is assimilated from ERA5, this can introduce an inconsistency into WRF due to differences in resolution. What is the reason for this choice? How different are the results when using the WRF-internal evaporation flux?

Before assimilating QFX from ERA5 we have interpolated the reanalysis moisture flux to the WRF domain. This was something that came up when we started to compare the results of WRF-WVTs with those of both Lagrangian methodologies directly using simulations from FLEXPART forced with ERA5, but in the end we have found that it does not make a big difference.

L. 138: Is a time interpolation used here?

No, in our case we are not using a time interpolation, as we use hourly evaporation data from the ERA5 reanalysis.

L. 149: It may be useful to have some basic information in the main manuscript, such as the chosen parameterisation schemes, and the fact that simulations are hemispheric (?)

We agree with the reviewer that moving the information about the chosen parameterizations from the supplement could be useful, and we will do it in the revised version of the manuscript. Specifically, the chosen parameterizations will be mentioned at the end of the first paragraph of Sect. 2.2. Regarding the domains, simulations are not hemispheric, as Fig. 2 shows. The domain in the Northern Hemisphere, for example, covers latitudes from 0º to 65ºN.

Figure 2: The source regions are very large in comparison to the scale of the moisture sources revealed by the Lagrangian diagnostics. A separation into e.g. 10 degree latitude bands or latitude-longitude boxes could allow for a much more detailed comparison and evaluation of Lagrangian models in the Eulerian framework.

This question has already been answered in response to the reviewer's major comment 1.

L. 166: "FLEXPART assimilates hourly data": FLEXPART does not perform data assimilation, please rephrase. It is not clear what is said from this sentence, the previous section described WRF, not FLEXPART. How exactly is FLEXPART run with WRF? Maybe some of the details from the supplement could be moved to the main text. In particular, it is important to describe how particles were initiated and released, and if any convection parameterisation was active in FLEXPART.

We are going to change "FLEXPART assimilates hourly data" by "FLEXPART ingests hourly data". FLEXPART is run with WRF by using FLEXPART-WRF, which is able to ingest input data from the output of the WRF model. Finally, we agree with the reviewer that moving some details from the supplements could be useful. Specifically, in the revised version of the manuscript we will explicitly state how parcels are released in the FLEXPART simulations: “parcels are released using the domain filling option over the black boxes in Fig. 1 such that they are vertically distributed following the density profile”.

L. 174: "it starts by assuming": This sentence and the following sound a bit strange. What you describe seems to be the basic idea of Lagrangian analysis, which is not particular to WaterSip. It would be useful to cite Stohl and James (2004) in this context, or shorten the section altogether, because all of this has already been said elsewhere.

We are going to rewrite this paragraph to retain only the most essential information, i.e., the following sentence: “the atmospheric column over the region where precipitation occurs is filled with air parcels, and that their trajectories contain information about their location and specific humidity at 6-hourly intervals for the previous days”.

L. 181 to 207: This section repeats a lot of information that is found in the original publication, and is not necessarily more easy to follow. I recommend limiting this to the most essential parts of the method which are modified here.

These paragraphs will be rewritten in the revised version of the manuscript. Specifically, we will briefly explain how WaterSip works, focusing on how the spatial distribution of moisture sources is obtained, and the remainder of that paragraph will be shortened. For example, the explanation of the discounting algorithm is omitted, as it can be found both in the original study of Sodemann et al., (2008) and in the Supplement. Once the methodology is introduced, we will present some modifications of WaterSip previously used in the literature, and which correspond with the different configurations that we test in our study. Here we acknowledge Dütsch et al., (2018) and Cheng and Lu, (2023) for introducing the modification that we later demonstrate reduce biases in WaterSip, “A less common modification is to filter the specific humidity decreases, such that previous contributions are only discounted if a specific humidity decrease occurs and the relative humidity of the parcel is higher than 80 % (Dütsch et al., 2018; Cheng and Lu, 2023)”. Finally, we also point out here the dependence of WaterSip on the specific humidity threshold and time step.

L. 188: The threshold value has been repeatedly shown to be a key sensitivity parameter (e.g., Sodemann and Stohl, 2009; Fremme and Sodemann, 2019). In addition, this value is on the very low end, that has been previously recommended for Arctic studies. How sensitive are your baseline results to this choice? To be in line with literature, I recommend a delta q of 0.2 for a 6h time interval.

As we mentioned in our response to the major comment 3, in the revised version of the manuscript Fig. 5 will include a sensitivity experiment in which we change both the threshold value and the time step. This demonstrates that our results are not very sensitive to the threshold value, but they are to the time step. Additionally, for the basic results we will use the recommended setup of 0.2 g/kg for a 6 h interval, instead of 0.05 g/kg for a 3 h interval as before.

Regarding section S3.1 referenced here, I wonder about what the role of this mathematical description is for the manuscript. There seem to be some arguments about correspondences between the two Lagrangian methods mathematically, but conceptually the two are quite different (e.g., well-mixed properties of the atmosphere). Section S3.1 could benefit from a closer connection to published literature to clarify its purpose. Does this section describe what has been published before, but mathematically in a common framework?

The original idea of section S3.1 was to unify the mathematical framework of the methodologies to show that both use a linear discounting, so that they are equivalent from a mathematical (and computational) perspective. However, it is true that they are conceptually different. Still, since we will present less details on the WaterSip methodology in the revised version of the manuscript, we believe it makes sense to keep this section of the Supplement as is.

L. 209: The authors refer to the UTrack method as the Dirmeyer and Brubaker (1999) implementation they use. However, as noted in L. 218, UTrack computes its own trajectories. Is it then not more correct to refer to the second model as the Dirmeyer and Brubaker (1999) method? What really distinguishes the approach used here from UTrack and Dirmeyer and Brubaker (1999), respectively?

We agree with the reviewer that it is more correct to refer to the second model as the Dirmeyer and Brubaker, (1999) method (we will use the abbreviation DB99). We are going to update the manuscript with this modification. The approach used here is the same as UTrack or Dirmeyer and Brubaker, (1999), but using FLEXPART trajectories instead of computing them. To do that we also need to consider how parcels are vertically released, as in the case of the Dirmeyer and Brubaker, (1999) methodology or UTrack they follow the humidity profile. Since in our case parcels are vertically distributed following the density profile, we had to weight the moisture sources of each parcel using their specific humidity. These important remarks will be now included in the methodology:

"However, in our case we use the FLEXPART and FLEXPART-WRF trajectories at hourly resolution and implement only the diagnostic tool to compute the moisture sources for precipitation. Thus, since in our simulations parcels are vertically released following the density profile, we weight the contribution of each parcel by its specific humidity to match the DB99 methodology."

L. 252: It has been common to initialize the domain at model time zero with all water vapour currently in the domain to achieve 100% accounting. Has this been tested here?

No, we have not tested that. This is evaluated in Insua-Costa and Míguez-Macho, (2018), where they show the internal consistency of WRF-WVTs.

L. 255: This is not correct. The Dirmeyer and Brubaker (1999) method stops accounting evaporation when 100% have been reached. The WaterSip method does not generally reach 100% (see Sodemann et al., 2008).

We agree with the reviewer that the WaterSip method does not generally reach 100 %. We meant that WaterSip typically reaches 100 % when the simulation time is 30 days and all uptakes are considered, but this is something that we have not shown. Because of this, we are going to delete that sentence from the manuscript.

L. 256: What is meant by "the bias will also be calculated after adjusting for these precipitation fractions"? This scaling should be explained in the methods section. Why is a scaling necessary at all? Is it not more correct to compare the actual identified fractions? What about comparing amounts rather than fractions?

We agree that scaling is not strictly necessary. In the revised version of the manuscript, the results will correspond to the comparison of the actual precipitation fractions without scaling, with the exception of the “ABL” and “RH” configuration, as in these cases the attributed precipitation is typically much lower or much higher than 100 % if we do not scale the precipitation fractions. On the other hand, we discarded comparing absolute quantities instead of fractions, as we can then be unconcerned whether in WaterSip the diagnosed precipitation, which comes from the specific humidity decreases at the time and location of the rainfall event, matches the WRF precipitation.

L. 266: I think the reference to Winschall et al. 2014 is not justified in such a general statement as done here. Winschall et al. 2014 did a sensitivity test of different tagging approaches, and their conclusion was: "The results of the Lagrangian diagnostics are similar to the Eulerian results, with the fraction of remote versus local moisture sources lying in between the two realisations of the tagging technique."

We agree that Winschall et al., (2014) did a sensitivity test of different tagging approaches, instead of assessing the Lagrangian method. Because of this, we will omit that reference in this part of the manuscript.

L. 268: Is the RMSE expressed as a fraction as in Eq. (6) or in percent?

We have corrected the manuscript expressing the RMSE as percentage, as the RMSE should have the same units as the precipitation fractions.

L. 274: It is interesting to note that the biases of the UTrack method are different. Why is that the case? The US West Coast case for example, UTrack has a lower performance.

This is because in the US West Coast case the moisture sources are highly dependent on altitude. This can be seen from Fig. 7, where we can check that the RMSE for the Dirmeyer and Brubaker, (1999) methodology changes a lot with the threshold for release height of parcels.

Figure 4: It is not possible the read the numbers printed in white on a light colour background.

We have updated this figure (and also Figure 9), so that all numbers can be easily read.

L. 279: I do not see a value of 29.6 in Fig. 3, nor of 14.88 for the Tropical Atlantic in Fig. 4. Is this example part of the supplement information? How does the scaling impact the results here?

We agree with the reviewer that the 29.6 value could not be easily deduced from Fig. 4., as the biases shown in Fig. 4 were computed scaling both the precipitation fractions in Fig. 3 and the precipitation fractions calculated with the Lagrangian methodologies.

In the revised version of the manuscript we will not scale the results (with the exception of the ABL and RH configurations, where the attributed precipitation is far from 100 %) so the precipitation fractions in the Lagrangian diagnostic tools can be easily deduced from Fig. 3 and 4.

L. 291: This statement applies to both Lagrangian methods. Before proceeding to tune the methods, it would be useful to quantify the overall bias of the Lagrangian vs. Eulerian methods, maybe at the end of Sec. 3.1, potentially as a function of distance from the arrival location. It may also be worthwhile to comment on the overall consistency of the results from the 3 approaches here. It would also be interesting to know more about the sensitivity here already regarding the specific setup you chose. How different are the errors/biases for a time interval of 6h, and when increasing the specific humidity threshold to 0.2 g kg-1 6h (or more)?

We agree with the reviewer that the statement applies to both Lagrangian methods, so we will mention WaterSip and the Dirmeyer and Brubaker, (1999) methodology in that sentence. Regarding the overall bias as function of distance from the arrival location, we understand that this information is somehow implicit in the estimated moisture sources, which generally gives more weight to closer regions. Finally, the dependence on the time interval and specific humidity threshold will be discussed in the revised version of the manuscript and included in an updated version of Fig. 5.

L. 295: What is presented here is exactly the argument for introducting a specific humidity threshold in WaterSip. So this needs not be formulated as a (new) hypothesis, it is part of the known uncertainty of the WaterSip diagnostic.

We will rewrite this sentence to explicitly mention that our hypothesis is that fluctuations penalize remote contributions. We know that the introduction of a specific humidity threshold tries to cope with noise in this diagnostic tool, but to our knowledge this specific implication of fluctuations has not yet been formulated. Thus, we will change "we conjecture that non-physical humidity fluctuations" to "we explore the hypothesis that non-physical humidity fluctuations" at the beginning of Sect. 3.2.1, and explain better why non-physical negative changes in specific humidity penalize earlier contributions at the end of the same paragraph, by adding "as the error caused by a single fluctuation affects all previous contributions, so the early moisture uptakes will be affected by many more non-physical changes".

L. 315: This distinction and modification have already been proposed by Dütsch et al., 2018 (Their Sec. 3.2). However, it is important to note that mixing with dry air can also lead to a specific humidity decrease. By only allowing for precipitation events to decrease specific humidity, a bias is intoduced into the method. This can also result in an over-accounting of sources (more than 100% of moisture accounted for).

This has already been answered above (see reviewer's major comment 3).

Figure 8: Comparing the UTrack results with the corresponding results from WaterSip in Fig. 7, it is very interesting to note how different the spatial maps are from the two methods. UTrack basically shows almost no sources at all in the vicinity of Greenland. While we don't know which one of the results is more correct, this difference is not picked up by the comparison to water vapour tagging in the present setup. This fact points to the current tracer setup being not sufficiently sharp (or detailed enough) to resolve and quantify such differences.

This has already been answered above (see reviewer's major comment 1).

References

Bonne, J.-L., Steen-Larsen, H. C., Risi, C., Werner, M., Sodemann, H., Lacour, J.-L., Fettweis, X., Cesana, G., Delmotte, M., Cattani, O., Vallelonga, P., Kjær, H. A., Clerbaux, C., Sveinbjörnsdóttir, A. E., and Masson-Delmotte, V., 2015: The summer 2012 Greenland heat wave: In situ and remote sensing observations of water vapor isotopic composition during an atmospheric river event, J. Geophys. Res., 120, 2970-2989, doi:10.1002/2014JD022602.

Cheng, T. F. and Lu, M.: Global Lagrangian Tracking of Continental Precipitation Recycling, Footprints, and Cascades, J. Clim., 36, 1923–1941, https://doi.org/10.1175/JCLI-D-22-0185.1, 2023.

van der Ent, R. J., Tuinenburg, O. A., Knoche, H.-R., Kunstmann, H., and Savenije, H. H. G.: Should we use a simple or complex model for moisture recycling and atmospheric moisture tracking?, Hydrol. Earth Syst. Sci., 17, 4869–4884, https://doi.org/10.5194/hess-17-4869-2013, 2013.

van Der Ent, R. J. and Tuinenburg, O. A.: The residence time of water in the atmosphere revisited, Hydrol. Earth Syst. Sci., 21, 779–790, https://doi.org/10.5194/hess-21-779-2017, 2017.

Dirmeyer, P. A. and Brubaker, K. L.: Contrasting evaporative moisture sources during the drought of 1988 and the flood of 1993, J. Geophys. Res. Atmospheres, 104, 19383–19397, https://doi.org/10.1029/1999JD900222, 1999.

Dütsch, M., Pfahl, S., Meyer, M., and Wernli, H.: Lagrangian process attribution of isotopic variations in near-surface water vapour in a 30-year regional climate simulation over Europe, Atmos. Chem. Phys., 18, 1653–1669, https://doi.org/10.5194/acp-18-1653-2018, 2018

Fremme, A. and Sodemann, H., 2019: The role of land and ocean evaporation on the variability of precipitation in the Yangtze River valley, HESS, 23, 2525–2540.

Fremme, A., Hezel, P. J., Seland, Ø., and Sodemann, H.: Model-simulated hydroclimate in the East Asian summer monsoon region during past and future climate: a pilot study with a moisture source perspective, Weather Clim. Dynam., 4, 449–470, https://doi.org/10.5194/wcd-4-449-2023, 2023.

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.

Sodemann, H., Schwierz, C., and Wernli, H.: Interannual variability of Greenland winter precipitation sources: Lagrangian moisture diagnostic and North Atlantic Oscillation influence, J. Geophys. Res. Atmospheres, 113, D03107, https://doi.org/10.1029/2007JD008503, 2008.

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

Sodemann, H. and Stohl, A.: Moisture Origin and Meridional Transport in Atmospheric Rivers and Their Association with Multiple Cyclones, Mon. Weather Rev., 141, 2850–2868, https://doi.org/10.1175/MWR-D-12-00256.1, 2013.

Sodemann, H. and Joos, H., 2021: Numerical methods to identify model uncertainty in: Ólafsson, H. and Bao, J.-W. (Eds), Uncertainties in Numerical Weather Prediction, Elsevier, 309-329, doi: 10.1016/B978-0-12-815491-5.00012-4, 2020.

Sprenger, M. and Wernli, H.: The LAGRANTO Lagrangian analysis tool – version 2.0, Geosci. Model Dev., 8, 2569–2586, https://doi.org/10.5194/gmd-8-2569-2015, 2015.

Stohl, A., Forster, C. and Sodemann, H., 2008: Remote sources of water vapor forming precipitation on the Norwegian west coast at 60°N - a tale of hurricanes and an atmospheric river, J. Geophys. Res., 113, D05102, doi:10.1029/2007JD009006.

Terpstra, A., Gorodetskaya, I. V., Sodemann, H.: Linking sub-tropical evaporation and extreme precipitation over East Antarctica:an atmospheric river case study, J. Geophys. Res., 126, https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2020JD033617, 2021

Yoshimura, K., Oki, T., Ohte, N., and Kanae, S.: Colored Moisture Analysis Estimates of Variations in 1998 Asian Monsoon Water Sources, 気象集誌 第2輯, 82, 1315–1329, https://doi.org/10.2151/jmsj.2004.1315, 2004.

Winschall, A., Pfahl, S., Sodemann, H., and Wernli, H.: Comparison of Eulerian and Lagrangian moisture source diagnostics – the flood event in eastern Europe in May 2010, Atmospheric Chem. Phys., 14, 6605–6619, https://doi.org/10.5194/acp-14-6605-2014, 2014.

Zhu, Y. and Newell, R. E.: A Proposed Algorithm for Moisture Fluxes from Atmospheric Rivers, Mon. Weather Rev., 126, 725–735, https://doi.org/10.1175/1520-0493(1998)126<0725:APAFMF>2.0.CO;2, 1998.