Review of "Simple physics-based adjustments reconcile the results of Eulerian and Lagrangian techniques for moisture tracking in atmospheric rivers" by Crespo-Otero et al.

This is my second review of this manuscript. The authors addressed most of my comments to the previous submission in a very constructive way and adjusted the manuscript accordingly. Still, I was very surprised that the authors state they find no sensitivity to the specific humidity threshold Δq. The authors claim throughout the text that they use the method "WaterSip", which in my understanding refers to a specific computer code implementation of the Sodemann et al. (2008) method. In an attempt to understand where these differences come from, I have made a comparison of two cases with the code that was used in previous studies for cases investigated here. It turns out that the results I get from running WaterSip V3.2, based on a global run with 5 million particles using ERA-Interim (Läderach and Sodemann, 2016) look entirely different for the Greenland case (Fig. 1), and only somewhat similar for the South Africa case (Fig. 2). The differences are so large that I do not think this could be due to the WRF simulation setup, and I rather suspect that the way the Sodemann et al. (2008) algorithm has been implemented differs in some way from the original. In fact, the results in Fig. 1b suggest that only a small contribution comes from the North Atlantic domain as defined by the authors, and a larger part from the North American continent. As the findings from the Greenland case are a cornerstone in the argumentation throughout the manuscript, this is a critical situation that needs to be resolved before publication.

Please note that WaterSip refers to a specific computer code that implements the Sodemann et al. (2008) method for FLEXPART and LAGRANTO with additional diagnostics. This software is now in public discussion (Sodemann 2025). In order to avoid confusing readers with regard to what software has been used, I would like the authors to not claim that they use WaterSip in their analysis, but rather their own implementation of an algorithm that is maybe inspired or resembles the Sodemann et al. (2008) method. The major differences that I see for the Greenland case imply that something else is done in the authors' code that leads to moisture uptakes identified close to Greenland in that case, and it would be misleading to assign these differences to either WaterSip or the Sodemann et al. (2008) method in itself. The lacking sensitivity to Δq is another hint at substantial differences in the author's code to what is done in Sodemann et al. (2008).

I still find the setup of the WRF moisture source tags overly coarse to make strong conclusions. In their reply, the authors state that they made a simulation with more finely resolved source boxes, but I do not see this discussed or shown in the manuscript, except for a figure in the supplementary material that is not mentioned anywhere (maybe I missed it). If the setup stays like this in the manuscript, it is in my opinion of critical importance to openly discuss the limitations of their approach in this study. For example, the boundary for the South Africa case at 30 deg N has the source footprint just south of that boundary (Fig. 2). This makes the verification statistics very sensitive so small meridional shifts in the source footprint, while really the AR case is one dominated by zonal advection. Thus, if the aim is to verify differences in the remoteness of sources, a setup with boxes over different longitudes would be needed to make the results more robust and help to support the conclusions, rather than a setup with different latitudes as is now shown in the supplementary material. All this may be still ok for showing the principle of how such a verification can be done, but such limitations need to be stated and discussed clearly to come to a well-justified overall assessment.

I also have a number of minor and technical comments below:

L. 14 and elsewhere: Instead of "WaterSip", which denotes a particular implementation of the Sodemann et al. (2008) method, refer to your own implementation of the Sodemann et al. (2008) algorithm here and elsewhere.

L. 37: This reference is missing from the reference section:

Sodemann, H., Wernli, H. and Schwierz, C., 2009: Sources of water vapour contributing to the Elbe flood in August 2002: A tagging study in a mesoscale model, Quart. J. Royal Meteorol. Soc., 135, 205-223, doi:10.1002/qj.374.

L. 43 and elsewhere: There is occasionally an extra comma before the bracket with references.

L. 60: remove extra '.'

L. 150: What are the units of QFX?

L. 152: The term "assimilate" is used here in a non-standard way, and can be confused with data assimilation (which I don't have the impression is meant here). Please rephrase to e.g., "ingest directly from ERA5" or the like.

L. 177: This should be cited as Sodemann and Stohl (2009) as in the references.

L. 183: Here it says hourly, in the sentence in L. 187 it says 3-hourly. Which one is correct?

L. 188: Give a typical pressure height of the topmost retained level.

L. 208: add "at arrival" after relative humidity

L. 222: 0.05 g kg 6 h^{-1} is only ¼ of the typically used value for mid-latitudes of 0.2 g kg 6 h^{-1}

L. 262: "coming also": check sentence

L. 264: "basic results" -> "baseline"?

Figure 3: I don't really see a great value in using a figure instead of a table, which I would think would be clearer/more easy to read with bigger font and better contrast. Values could by highlighted by making them bold or italtic.

L. 285 and onward: The terms "good" and "bad" or "better" and "worse" could be more neutrally be rephased by using expressions such as "more consistend with reference" or "less consistent", or "larger deviation", "smaller deviation". There is not really a way of telling objectively here at what value a difference can be considered good or bad.

L. 320: Such fluctuations are always a remaining part of the uncertainties from offline trajectory calculations. This is why the Δq threshold has been introduced. Given that there is virtually no impact from the threshold on the results, as in Fig. 5 indicates that there could be an error in the way the method has been implemented.

Figure 5: The low contrast in some of the tables makes the numbers hard to read.

L. 342: It was not easy to follow the narrative here, please rephrase.

L. 353: important for what?

L. 358 and L. 362: significantly: rephrase if no statistical test has been performed

L. 369: Move to figure caption

L. 374: What means "the moisture field is less intense"?

Harald Sodemann

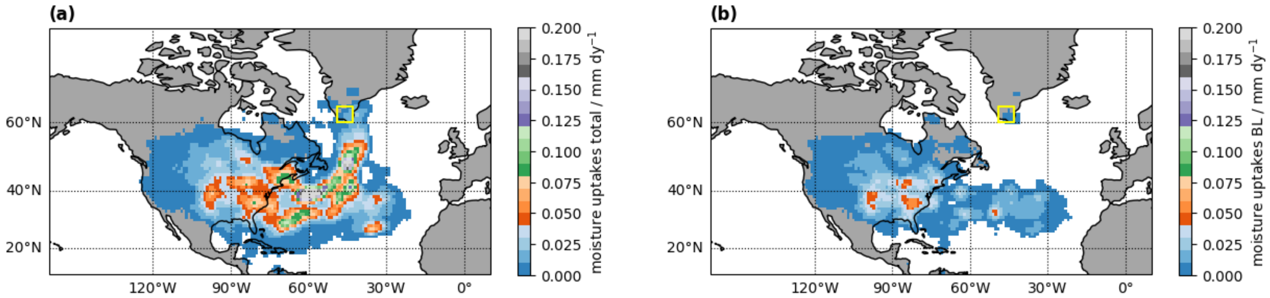

Figure 1: Moisture sources from WaterSip V3.2 extracted from the dataset of Läderach and Sodemann (2016) for the AR arriving in Greenland during a 48 h period before . (a) Combined boundary-layer and free-troposphere moisture uptakes (mm day^{-1}). (b) Boundary-layer moisture uptakes (mm day^{-1}). The WaterSip extraction was set up with a Δq_c = 0.2 g kg^{-1} 6 h^{-1} and a trajectory length of 20 days.

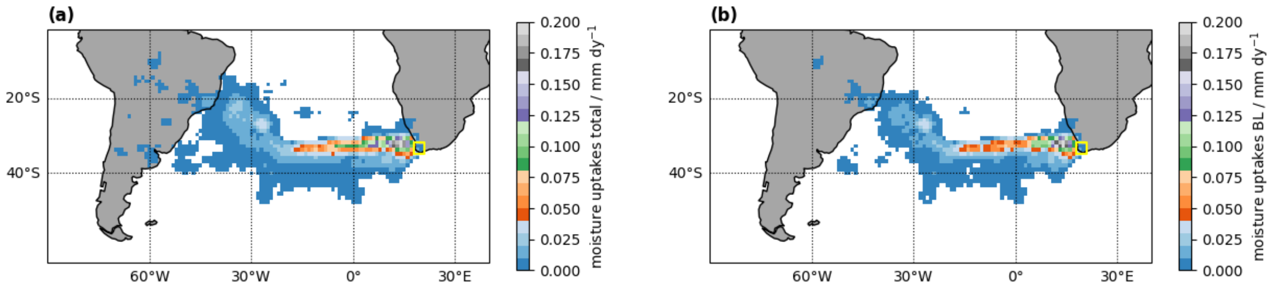

Figure 2: Moisture sources from WaterSip V3.2 extracted from the dataset of Läderach and Sodemann (2016) for the AR arriving in South Africa during . (a) Combined boundary-layer and free-troposphere moisture uptakes (mm day^{-1}). (b) Boundary-layer moisture uptakes (mm day^{-1}). The WaterSip extraction was set up with a Δq_c = 0.2 g kg^{-1} 6 h^{-1} and a trajectory length of 20 days.

References

Sodemann, H.: The Lagrangian moisture source and transport diagnostic WaterSip V3.2, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2025-574, 2025.

Läderach, A. and Sodemann, H., 2016: A revised picture of the atmospheric moisture residence time, Geophys. Res. Lett., 43, doi:10.1002/2015GL067449.