Reply to review of "The Lagrangian moisture source and transport diagnostic WaterSip V3.2", submitted to GMD by H. Sodemann

Reviewer #1

This manuscript presents a comprehensive update and documentation of the WaterSip software (version 3.2), a diagnostic tool for identifying moisture sources and transport pathways associated with precipitation and atmospheric water vapour. The tool implements the Lagrangian diagnostic framework originally introduced by Sodemann et al. (2008), and offers support for trajectory data from models such as LAGRANTO and FLEXPART.

The manuscript is technically detailed and contains extensive explanations of the algorithmic structure, parameter configuration, example case setup, and diagnostic outputs. It represents a valuable contribution to the hydrometeorological community, particularly those using Lagrangian methods for moisture tracking. However, there are several areas where the manuscript could be improved significantly, especially in the following aspects:

- *Validation of diagnostic results through comparison with observations or alternative algorithms;*
- *Technical clarity on certain assumptions and limitations;*
- *Demonstration of robustness and sensitivity through more systematic experiments;*
- *Improved structure and clarification of key terminology for broader accessibility.*

I recommend major revisions before this manuscript is accepted for publication.

Reply: Thank you for your detailed and constructive review. The specific comments are addressed below.

1. Lack of Model Validation and Performance Benchmarking

While the algorithmic principles of WaterSip are well-founded, the manuscript lacks quantitative validation of the diagnostic results. In particular:

- *No comparison with independent observational datasets (e.g., precipitation from GPM/IMERG or ERA5 reanalysis P);*
- *No benchmarking against other Lagrangian diagnostics, such as WAM-2layers, FLEXPART-WATER, or isotope-enabled models (e.g., COSMOiso);*
- *The Lagrangian precipitation estimate P~\tilde{P}P~ is claimed to have an error of 20–30%, yet this is not demonstrated empirically in the paper.*

Recommendation: Include a comparison of WaterSip-derived precipitation estimates and source regions against satellite/reanalysis precipitation and/or results from other established methods. This would help quantify accuracy and justify the use of default parameters (e.g., RHc, Δq thresholds).

Reply: The reviewer suggests to add more material that compares the Lagrangian precipitation estimate with other datasets, as well as a comparison of results from the diagnostic to other methods. There are a number of reasons why I would prefer to strongly limit this kind of analysis in the present manuscript:

- Comparison to independent observational datasets: WaterSip's primary task is not to provide precipitation, but to detect precipitation properties. While a comparison to the precipitation in the reanalysis dataset itself is useful to learn about the representativeness of the diagnostic results, a comparison to other observational or blended datasets is beyond the scope of this study. For that purpose, it would be more useful to directly use studies that compare between reanalysis data and observational/merged precipitation itself. This will be further clarified in the revised manuscript, including the addition of relevant references to comparisons between different precipitation datasets, and possibly the addition of a reanalysis precipitation map for the case study.
- Benchmarking against Lagrangian diagnostics and isotope enabled models: The manuscript shows a comparison to the results from the Stohl and James (2004) method (Fig. 2d). Including a comparison to other models either for one case study, or several case studies, is in my view far beyond of what can fit into

the scope of this manuscript - if it should be done in a way that it becomes sufficiently meaningful. Without objective 'truth' about the moisture sources at hand, it only possible to make relative comparisons. There are several papers published in the literature that attempt such comparisons, and other efforts are currently ongoing. Where possible, additional citations will be added in the revised manuscript, and the challenges of finding moisture source properties from different methods will be mentioned when discussing the case study.

- The Lagrangian precipitation estimate has been introduced by James et al. (2004) and Stohl and James (2004). The 20-30 % is a ballpark number, but results will vary depending on study region. Several studies have shown that there can be over- or underestimation, and some studies that are cited have systematically evaluated dependency on thresholds. Users will have to make their own evaluations depending on thresholds and regions that are studied. It is an important aspect throughout the manuscript that users need to be aware of what they are doing. This perspective will be underlined in the revised manuscript, and potentially backed up by a comparison with reanalysis precipitation.

2. Insufficient Sensitivity Experiments

The diagnostic depends heavily on multiple user-defined thresholds, such as:

- *Moisture uptake threshold (Δqc),*
- *Precipitation threshold (Δqp),*
- *Critical relative humidity (RHc),*
- *Trajectory length (L) and time step (Δt),*
- *Boundary-layer height scale (sh).*

While some default values are provided, the manuscript does not present any systematic sensitivity tests to justify these defaults or examine result variability.

Recommendation: Provide at least one sensitivity experiment (e.g., with RHc = 60%, 80%, 90% or Δqc = 0.1, 0.2, 0.3 g/kg/6h) using the Scandinavia case to demonstrate how output fields (e.g., source footprints, $\tilde{P}$) are affected. This will help users understand uncertainty and robustness.

Reply: It is important to note that there is a large body of literature where this method was applied from a time before code was commonly published open access, and where these thresholds have been thoroughly tested. Instead of redoing such previous work in this manuscript for a limited case study, my primary choice has been to use literature references. However, I agree that a limited sensitivity experiment will be useful to include for the Scandinavia case, support users in running their own sensitivity experiments.

3. Ambiguity in Treatment of Mixing vs. Precipitation

The distinction between moisture losses due to precipitation vs. dry mixing is briefly described but remains ambiguous in practical terms:

- *How are "mixing events" defined and treated in the accounting algorithm?*
- *Are they excluded from precipitation source attribution entirely?*
- *How does this impact attribution over dry regions or under sub-saturated conditions?*

Recommendation: Include a dedicated subsection clarifying how dry mixing events are separated and whether/how they influence the fractional contribution calculation. Provide a sample output or visualization that isolates these cases.

Reply: Identified mixing events are shown in Fig. 4c, the current manuscript text incorrectly refers to Fig. 2c. In the revised manuscript, the text describing mixing events will be extended and clarified.

4. Limited Scope of Case Study

The case study over Scandinavia is informative but lacks depth and generality:

- *It only covers a short period (10–20 Aug 2022) with one configuration;*
- *There is no validation of the Lagrangian precipitation estimates against ERA5 or in-situ observations;*
- *The transport features are discussed qualitatively without statistical summaries (e.g., source region contributions by %).*

Recommendation:

- *Add a second case study (e.g., a winter event or tropical cyclone) to demonstrate versatility;*
- *Include plots/tables showing the percent contribution of major source regions (e.g., local vs. oceanic);*
- *Overlay gridded WaterSip P~\tilde{P}P~ with observational data (e.g., E-OBS or GPCC).*

Reply: The function of the Scandinavia case study is to help users of WaterSip getting started with a concrete example. A second case study would not add to that purpose, but potentially confuse and tire readers due to the increased manuscript length. There is a large body of literature available with different geographical foci and for different time periods (Table 1), and in the revision this role of the case study will be further clarified.

A comparison between the Lagrangian precipitation estimate and ERA5 precipitation will be useful to add to illustrate the differences between both quantities. It will also be clarified again in the text that the purpose of the WaterSip diagnostic is not to estimate precipitation, but to identify precipitation properties. I consider therefore a comparison to observational datasets will be beyond the scope of this GMD paper.

Finally, a brief summary table summarising the major source regions will be added in the revised manuscript.

5. No Performance or Computational Cost Analysis

Given the tool is designed for high-volume Lagrangian data, its computational performance, memory usage, and scalability are essential for practical adoption:

Recommendation: Add a short section or table reporting:

- *Typical runtime and memory usage for the example case;*
- *Speedup with OpenMP threads;*
- *Bottlenecks or limitations for large-scale usage.*

Reply: A brief paragraph describing computational performance and memory usage for the test case will be added in the revised manuscript.

Minor Comments & Suggestions

1. *Clarify terminology early (Section 1):*
   - *Define “uptake”, “accounting”, “residual moisture”, and “arrival grid” explicitly.*
   - *Consider a graphical workflow diagram.*
2. *Equations (6)–(9):*
   - *Include variable definitions in-line with the equations, especially for readers not familiar with the 2008 method.*
3. *Section 2.5: Too long and fragmented. Suggest splitting into:*
   - *"Core algorithm parameters" (Δq, RHc, sh),*
   - *"Grid and output configuration",*
   - *"Optional diagnostics".*
4. *Figures:*
   - *Add scale bars and legends (e.g., units in mm/day);*

- *Some figures lack clarity (e.g., Fig. 2d – difficult to read e-p shading);*
- *Add observational overlay for better interpretation.*

5. *Code availability: Ensure a DOI or stable link is provided. Consider creating a GitHub/Zenodo archive.*
6. *Language & Style:*
   - *Mostly clear, but some long and nested sentences in Section 2–3 could be simplified.*
   - *Example: "Air parcels will only be retained for analysis…" → split into clearer bullet rules.*

Reply: These minor and technical comments will be implemented as suggested.

Conclusion

The manuscript presents a valuable and much-needed technical documentation of WaterSip V3.2 and the Lagrangian moisture source diagnostic algorithm. However, to be suitable for publication in a journal such as GMD or HESS Discussions, the following critical issues must be addressed:

- *Quantitative validation of results,*
- *Sensitivity and uncertainty analysis,*
- *Clear treatment of physical assumptions (e.g., mixing vs. precipitation),*
- *Extended and comparative case studies.*

Reply: all comments have been addressed in the above.

References:

James, P., Stohl, A., Spichtinger, N., Eckhardt, S., and Forster, C.: Climatological aspects of the extreme European rainfall of August 2002 and a trajectory method for estimating the associated evaporative source regions, Nat. Hazards Earth Syst. Sci., 4, 733–746, https://doi.org/10.5194/nhess-4-733-2004, 2004.

Stohl, A., and P. James, 2004: A Lagrangian Analysis of the Atmospheric Branch of the Global Water Cycle. Part I: Method Description, Validation, and Demonstration for the August 2002 Flooding in Central Europe. J. Hydrometeor., 5, 656–678, https://doi.org/10.1175/1525-7541(2004)005<0656:ALAOTA>2.0.CO;2.