## Response to Harald Sodemann:

### General comments:

The authors perform a sensitivity study of two methods to identify moisture origin for one selected summer month over two regions in the Tibetan plateau. From the comparison between the two methods, the authors see differences with regard to moisture contributions from Eurasia and over coastal regions, that are explored in a sensitivity study. The authors then draw conclusions about the consistency and validity of the two methods. The manuscript is overall written coherently and in a well-readable manner. However, I find the conclusions are too general given the episodic evidence presented in the manuscript itself. The authors could consider changing this paper to a shorter, research letter format. I also have some comments about the structure of the manuscript, the precision of the language, reference to code and use of literature, and the presentation and interpretation of the results. I hope my comments will help the authors to prepare an improved version of their manuscript.

**Response:** We are very grateful for your thorough review and comments, all of which provide excellent guidance on our revision and future research. Per your comments, we will thoroughly revise this manuscript in terms of language, content, and logic coherence. We hope that the revised manuscript aligns more closely with the requirements of a research paper. The main aspects to be addressed include:

**1.** Rewrite the Introduction, limiting the scope of the study to focus solely on moisture tracking over the Tibetan Plateau (TP), and avoiding the generalization of relevant statements.

**2.** Conduct a sentence-by-sentence revision to ensure accurate descriptions of the results and enhance connections between conclusions and key results.

**3.** Carry out additional analyses and sensitivity experiments, including 1) a series of sensitivity experiments for WAM-2layers (please see our Response to Main comments 6 for details) and FLEXPART-WaterSip (please see our Response to Main comments 7 for details); 2) analyses of the relationship between the simulated moisture sources and actual evaporation.

### Main comments:

**1.** In their introduction, the authors set forth a basic distinction into Eulerian and Lagrangian methods for "moisture tracking". I find this distinction too coarse with regard to the results presented in this study. The two methods that are being compared are broadly seen part of the respective categories, but there are many (other) approaches within the Lagrangian category (see for example the discussions in Keune et al., 2022), and many other within the Eulerian category, that are not compared here. For example, moisture tagging in a regional model (Yoshimura et al., 2004), or the E-P Lagrangian approach of Stohl and James (2004), and so on. The authors claim that the two methods they compare are most widely used - I think this is debatable, plus they are focussing here on the Tibetan Plateau only.

**Response:** Thanks for the comments. The motivation of this study originates from the extensive literature on precipitation moisture tracking in the Tibetan Plateau (TP) (Table 1 presents only a subset of existing efforts). However, to the best of our knowledge, no effort has been made to address the disparities or uncertainties among these TP-focused studies. This situation has led us to develop this manuscript, aspiring to encourage future researchers to critically assess the reliability of their simulation outcomes. Toward this goal, we strived to identify potential factors contributing to these disparities among models over the TP region.

As you mentioned in the comments, the descriptions of some concepts (such as those related to Eulerian and Lagrangian methods) in this manuscript are not very accurate. In the revision, we plan to thoroughly revise the Introduction section to emphasize the following three aspects:

1. We will narrow the scope of our current study to focus exclusively on moisture tracking over the TP. In this context, we will specifically highlight that the most widely used numerical moisture tracking models are WAM-2layers and FLEXPART-WaterSip. Our subsequent paragraphs in introduction will focus solely on these two representative models.

2. The aim of this manuscript is to investigate potential errors/uncertainties in existing moisture tracking research on the TP as well as to understand the underlying mechanisms that contribute to these errors/uncertainties. We will emphasize the significance of this study to inform future numerical moisture tracking over the TP.

3. We will caution readers against generalizing our comparison results from the TP to other regions in the absence of substantial evidence.

**2.** The study now only compares one month (July 2022) and two specific catchment areas of the Tibetan Plateau. It remains thus unclear if the findings here can be generalised, or are rather coincidental. Therefore, it would be adviseable to tune down the quite authoritative/concluding language and formulate more modestly, such that it be in agreement with the somewhat anecdotal evidence that is actually investigated and presented here. This concerns both the Abstract, Introduction, and Conclusions.

**Response:** Thank you. Both you and Dr. Ruud van der Ent have expressed similar concerns in our manuscript. We will thoroughly revise to address these issues:

1. We will ensure that the manuscript maintains accuracy and logical coherence throughout. Additionally, we will tune down any authoritative/concluding statements that are not fully supported by complete evidence chains.

2. We will enhance the manuscript by incorporating further analyses, discussions, and sensitivity experiments to ensure that all conclusions are substantiated (for more details please see our Response to Main comments 6 and 7).

**3.** The authors state that they use the FLEXPART-WaterSip method. I don't think this is correct, since the WaterSip code is a specific implementation of the Sodemann et al. (2008) moisture source diagnostic in C++ language which is currently not yet available publicly. The WaterSip code has first been used by Sodemann and Stohl (2009) and later my many other studies (Bonne et al., 2014; Läderach and Sodemann, 2016; Sodemann 2020 to name a few). The authors also state that all original codes are available from the official websites - this is not correct for the WaterSip method. A separate publication on this actual "WaterSip" code is in preparation by this reviewer. My impression is that the authors have written their own implementation of the algorithm of Sodemann et al. (2008), which they then use for this study. This must be stated clearly and correctly, and the authors' own code should be linked to in the Code availability section. In any case, the reference to the website at University of Bergen is no proper code reference to the WaterSip method.

**Response:** Sorry for any confusion caused by the use of WaterSip code in this manuscript. Indeed, we developed our own Python implementation of the algorithm described by Sodemann et al. (2008). In the revised manuscript, we will clearly outline all methods and code used in our manuscript, and we will make our code available in a supplementary document.

**4.** The immense literature review presented in Table 1 is never properly described and hardly used in the manuscript. I also note that a similar table has been presented already in the supplement material of Li et al. (2022), a study by the same authors that is not cited in this manuscript. I do appreciate the effort put into this table. Currently, however, there are just two sentences in the introduction that make general remarks about this table. A more systematic discussion of what was found during the literature review would be needed to justify including this table in the main manuscript. In addition, it would be useful to tie the results from this study up agains the reviewed literature in a Discussion section in the end.

**Response:** Thanks for the comments. When summarizing the literature in Table 1, we also aimed to derive insights by comparing these different studies to highlight the significance of our research. However, these studies encompass different methodologies, datasets, and even study areas (e.g., different parts of the TP region). Beyond Table 1, we did not find a suitable method to effectively compare these different studies when preparing this manuscript. We are currently exploring methods for a potential comparison and discussing potential approaches with our peers. We plan to further add comparative results to the Introduction and Discussion sections in our revised manuscript.

The main reason we did not cite Li et al. (2022) is that the literature summarized in Table 1 primarily focuses on the backward tracking of precipitation moisture over the TP, whereas Li et al. (2022) mainly focuses on forward tracking. In addition, there are numerous moisture tracking studies over the TP not yet included in Table 1. For example, extensive studies on precipitation and ice-core isotopes in the TP region often rely on moisture tracking results to support explanations for isotopic data, with HYSPLIT being the most commonly used model. These isotope-related studies have not been summarized in Table 1. We will address these aspects in our revised manuscript.

**5.** Section 2 discusses the generalities of the two selected methods. I think the broad description of these two examples as Eulerian and Lagrangian methods in general does not fit the two specific methods that are applied here. Also, how these specific methods work are described sufficiently elsewhere in the literature. Instead, the authors would need to describe more clearly how exactly the respective simulations have been set up. Specifically regarding the FLEXPART-WaterSip like method, was a domain-filling setup selected in FLEXPART? Was the calculation run in forward mode? Has convection parameterisation been used? What domain has been used? All these details are important. Furthermore, the WaterSip code is currently not available publicly, and the website pointed out in the data section only provides a manual. What code has then been used to diagnose the moisture sources from the FLEXPART particle trajectories, and where is this code accessible? How were Lagrangian moisture sources gridded? What output interval and humidity thresholds were used? These aspects are all essential aspects for reproducibility of the work, and to understand the preconditions of this comparison.

**Response:** Thanks for all the questions. We recognize the importance of avoiding overly general descriptions of the methods used in this manuscript. In our revised manuscript, we will:

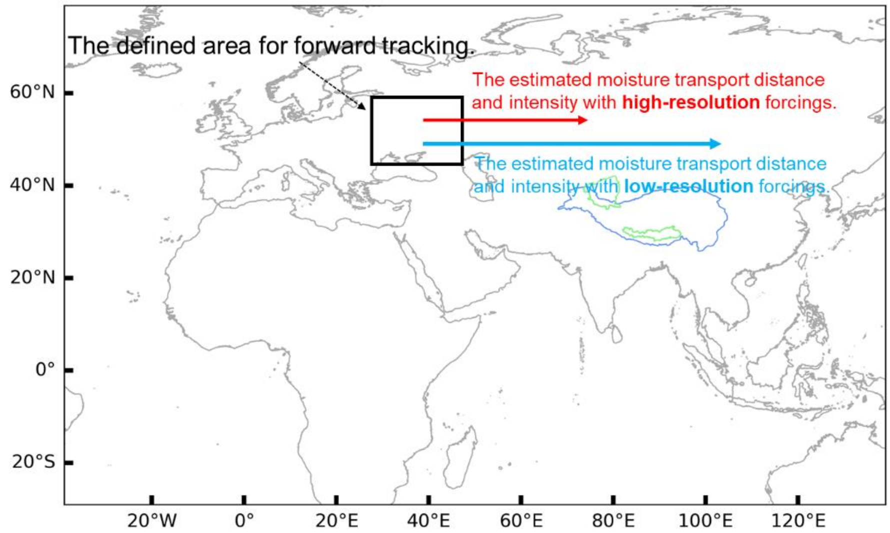
1. Provide detailed descriptions of the two selected models.
2. Clearly specify all numerical settings for these two models.
3. Share all the algorithms we developed in the Supplementary to enhance reproducibility and help readers understand the prerequisites of this comparison.

**6.** The difference in moisture source contribution from Eurasia between the two methods is quite interesting. We don't know what is the truth from the two approaches, but a gridded map of air parcel location density for trajectories arriving in the study domains could help indicate if FLEXPART (based on ERA5) does identify transport pathways from Europe. In this context, I find the sensitivity of the WAM2layer method to finer resolution quite striking. What is possibly going on that leads to such a strong senstivity to grid resolution in the results? Maybe be there is numerical diffusion at coarser resolution (see Sodemann 2020, Sec. 7)? Additional sensitivity experiments or analyses of different time snapshots could be useful.

**Response:** Thank you for the comments. To address these issues, we will make improvements in the following two aspects:

1. Expand the sensitivity experiments for the WAM-2layers to four configurations: 1°×1° at 3-hourly resolution, 1°×1° at hourly resolution, 0.25°×0.25° at 3-hourly, resolutions, and 0.25°×0.25° at hourly resolutions.

2. Design an additional experiment that involves: (1) identifying a westly source area with substantial simulation discrepancies between the two models (as depicted by the rectangular box in the Figure below), and (2) conduct forward tracking of evaporated moisture from this rectangular box using WAM-2layer with both low-resolution and high-resolution forcings. This experiment will help determine whether using lower resolution forcing overestimates the distance and intensity of moisture transport.

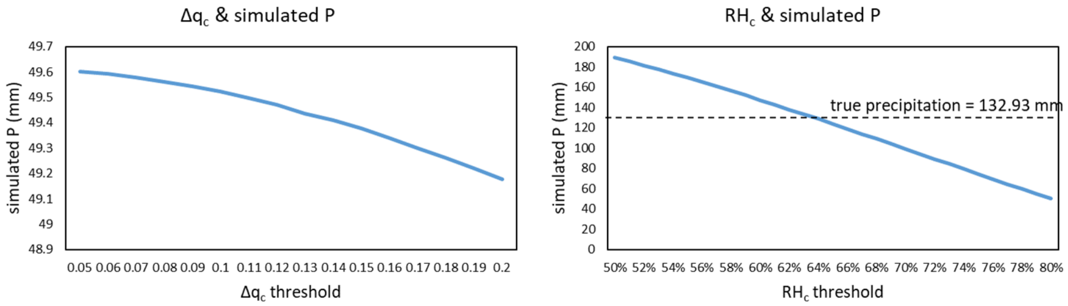



**7.** The sensitivity study in Sec. 5 is quite interesting, but does not really include the most important sensitive parameters of this approach, as discussed widely in the literature. Instead of number of particles (Fremme et al., 2023), it would be more important to test the threshold of specific humidity (dqc in Sodemann et al., 2008) as well as the relative humidity at arrival (RHc in Fremme and Sodemann, 2019). The areal source-receptor attribution method comes a bit out of the blue here. It is an entirely different method of the Lagrangian category. The difference between this method and the others should be described in the methods.

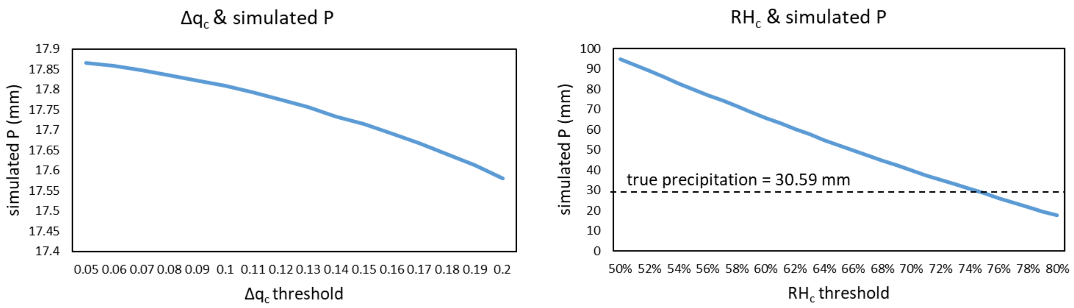
**Response:** Thanks for the suggestions.

1. Prior to preparing this manuscript, we have conducted several sensitivity experiments on precipitation in the two target areas (YB and UTB). We used two benchmark thresholds provide by Sodemann et al. (2008): 0.2 g/kg for $\Delta q_c$ and 80% for $RH_c$. The following figures show our test results:

Over the YB:

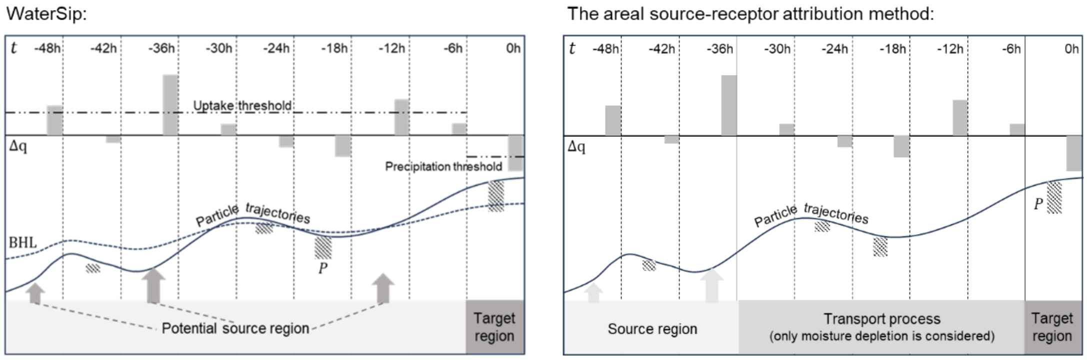



Over the UTB:



We found that simulated precipitation in both regions is more sensitive to $RH_c$, with the optimal $RH_c$ threshold being about 63% for YB and 74% for UTB. In response to the studies you mentioned above, we will carry out three sensitivity tests on $RH_c$ thresholds (60%, 70%, and 80%) for precipitation simulation and three sensitivity tests on $\Delta q_c$ thresholds (0.1, 0.2, and 0.3 g/kg) for evaporation simulations in our revision.

2. In our revised supplementary, we will include a comparison of the basic framework diagrams for WaterSip and the "areal source-receptor attribution method" (see below). We hope this comparison will help readers better understand the differences between these two methods.




**8.** I am puzzled that the authors do not discuss nor cite their own study in NHESS about the spatial distribution of moisture sources for the Tibetan Plateau using the WAM2layer model (Li et al., 2022). In the supplementary material of that paper, they show a map with Eurasian moisture sources, just as discussed here from the two methods. What could possibly be the reason that you do not discuss this previous work done with the WAM2layers method? Is this not a golden opportunity to balance or rectify any conclusions drawn in Li et al. (2022) in the light of new evidence? I also note that Li et al. (2022) contains a table similar to Table 1 presented here. A discussion of the relation between this work and your own previous work is definitely required.

**Response:** Thanks for noticing our earlier work (Li et al., 2022) published in HESS. In the supplementary material of Li et al. (2022), we tracked moisture sources of the entire TP using WAM-2layers driven by ERA-I, MERRA2, and JRA55. In comparison, this manuscript focuses on using ERA5 to drive two moisture tracking models in one monsoon dominated region and one westerlies dominated region of the TP. The different driving datasets and study areas present challenges for comparing these results. Nevertheless, your comments have prompted us to carefully examine the connections between the present study and Li et al. (2022). We will include additional analyses and discussions on these different studies and integrate into both the Introduction and Discussion sections.

## Detailed comments:

**1.** Figure 2: The gridding of the FLEXPART-WaterSip results in Fig. 2 looks more spotty than the WAM2layers - I would argue that either a larger grid spacing or larger gridding radius of the identified sources should be used, or the number of particles increased to mute these distracting artifacts. Maybe just show the same resolution as used in Fig. 3 where the same grid was used for both models?

**Response:** Yes, as your pointed out, we used an output resolution of 1°×1° in Figs. 2a and 2b (corresponding to the original forcing dataset used by WAM-2layers), but a resolution of 0.25°×0.25° in Figs. 2c and 2d (FLEXPART-WaterSip). Unlike the grid-based WAM-2layers, FLEXPART-WaterSip outputs the particle-scale data. In our initial submission, we considered interpolating the particle-scale data to a 0.25°×0.25° resolution for more detailed visualization. However, this seems to have caused confusion for readers. Following your comments, we will standardize all outputs to 1°×1° resolution in our revised manuscript (including Figs. 2, S5, S9, and S10).

**2.** Figure 6: I find panels a and b hard to interpret objectively, as there are subjective/conceptual arrows superimposed on the panels. Are these two panels adding new information compared to the trajectory examples shown in panels c-f?

**Response:** Thanks for pointing this out. Panels a and b were meant to show the spatial distribution of particles, but the conceptual red arrows do not add any new information to the explanations provided elsewhere in this paper. We will remove these arrows and rewrite the description of Figs. 6a and 6b in the revision.

**3.** Figure 7: Why do you show 300hPa vertical velocity in panel b? Maybe it would be more useful to add a figure that shows the average/median vertical air motion as a view of trajectory (pressure) altitude vs time arriving at the two selected regions. These vertical pathways seem to be quite different.

**Response:** Thanks for the comment. We chose ~300 hPa (~9000 m) as an illustration of the vertical air motion over the TP region (cf. ~700 hPa for the entire domain), which is a bit arbitrary. Per your comments and trajectory results shown in Figs. 6c and 6d, we will adjust the vertical velocity height to 500 hPa (~5500 m) for moisture transport of the westly region (as well as over the TP) and 850 hPa (~1500 m) for moisture transport of the monsoon region in the revision.

**4.** Figure 10: These two examples from a set of 5 million trajectories can hardly be considered representative. What is really the value of discussing exactly these two examples? It does not become entirely clear to me what to take away from these examples, and I think i is not justified to draw as general conclusions about the weaknesses of the Lagrangian diagnostics (L. 399 onward) as the authors do on this basis alone. Also, I got confused by the time axis at first, it should be made clear where the arrival point is. Winschall et al. (2014) have discussed with similar examples before that (deep) convection can contribute to moistening at upper levels that is not captured by motion of individual trajectories. Is this the case here as well? Do you use a convection parameterisation in FLEXPART? Are these locations over land or ocean? It would also be helpful to indicate the specific humidity threshold adopted in this study, and maybe include specific humidity and relative humidity in addition.

**Response:** Thanks for your suggestion. We did use a convection parameterization scheme in FLEXPART. In the revision, we will mark out the arrival point, the range of land/ocean, and move Fig. 10 to the supplementary material as only a special case. To more thoroughly investigate the characteristics and discrepancies between the two models, we will expand our manuscript with additional analyses. First, we will add a set of sensitivity experiments for WAM-2layers (please see our Response to Main comments 6 for details) and FLEXPART-WaterSip (please see our Response to Main comments 7 for details). Second, we will compare the moisture tracking results with actual evaporation and atmospheric vertical motion over the source regions. Particularly, in the WaterSip, the regional moisture uptake derived from the $\Delta q_c$ could be directly compared with actual evaporation.