Review of the data descriptor:

AguaTrack-ARCO-SA: A 30-year, high-resolution atmospheric moisture tracking dataset for South America

Comments

The authors present a 30-year (1990-2019), daily, 0.25° backward moisture-tracking dataset for South America. It was produced with the Eulerian model WAM2layers v3 (Kalverla et al., 2025), forced with hourly ERA5 data over a domain spanning 90°W-25°W and 60°S-15°N.

The community value of this dataset, especially for users outside the moisture-tracking community, could be important thanks to its record length, spatial resolution and the regional focus on South America, a key region for atmospheric moisture transport studies. The effort to make a large dataset accessible (Section 4.2) is appreciable. However, in its current state, the article could address its limitations more thoroughly, and some statements may overstate its uniqueness.

I would recommend the publication of this paper after major revisions.

Please find below my comments, which may help the authors to enhance the robustness and depth of the data description and the data repository. In particular, a more robust validation and a more balanced tone would greatly improve the quality of the article and the trustworthiness of the dataset.

General comments:

(1) Appropriateness of the article to support the dataset

In its current form, the article is partially appropriate to support publication of the dataset. The article focuses more on showcasing potential applications of the dataset (subsection 4.1), with less attention to robust validation against other available moisture-tracking datasets (subsection 3.4), and the discussion of the limitations of the dataset could be expanded (subsection 4.3). In particular, the comparison is performed only against datasets forced with ERA-Interim, and the differences are attributed to the spatial resolution without isolating that effect (see detailed comments on Sections 3.4 and 4.3).

(2) Significance and uniqueness

The data set is significant and potentially very useful for its temporal coverage, spatial resolution, and region of application (South America). However, in its current form, the article may overstate its significance, uniqueness, and completeness, and could better acknowledge other examples in the literature that have released and made efforts to share large datasets. The tagging of each individual grid cell is not entirely new, and existing datasets already provide atmospheric moisture connections at 1° and 0.5° (see detailed comments on Section 1). Moreover, the spatial domain and the only backward reconstruction represent important limitations with respect to datasets with global coverage and both forward and backward reconstructions, which would benefit from being declared and quantified more clearly (see detailed comments on the Abstract and on Sections 4.1 and 4.3).

(3) Data set quality

I acknowledge the importance of the topic “Democratising Big Data” and the efforts made to address this relevant theme through the ARCO architecture. This topic could also be made more central throughout the article. However, some important aspects remain to be addressed. The repository supports analysis of the published outputs but does not allow reproducing the generation of the dataset, and several aspects needed to use the data correctly are not yet documented, such as the meaning of the time coordinate, the definition of the variables, the selection of the tagging cells, and the provenance and licence of the forcing data (see detailed comments on Sections 2 and 3 and on the dataset and repository).

(4) Publication quality

The quality of the article would benefit from a clearer declaration of the dataset's coverage limits and from a more robust validation of the output and contextualization with respect to existing literature, both for the atmospheric tracking dataset, the moisture-tracking model uncertainty discussion. This is particularly important because the authors highlight that the dataset could be useful to a broad scientific community beyond moisture-tracking modeling (lines 325-330). In addition, several statements are qualitative and not yet supported by quantitative evidence, and the tone could be revised accordingly. In general, the most novel and interesting parts of the article (Section 4.2 and the discussion in lines 205-230) should be extended and made more central and balanced with the showcase of the applications in subsection 4.1.

(5) Remarks

The dataset could be useful, especially for its daily temporal resolution and long coverage, but its trustworthiness would greatly benefit from a more robust validation, a clearer declaration and quantification of its limitations, a fully reproducible workflow and a clearer positioning with respect to existing atmospheric tracking datasets and literature on tracking methods and related applications in South America.

Detailed comments:

Abstract

Line 1: To be more precise, the authors could consider rephrasing as follows: atmospheric moisture *transport* is a key component of the hydrological cycle, linking evapotranspiration sources to downwind precipitation *sinks*.

Line 16: The regional limitations of the dataset (only land grid cells of South America tagged as receptors, sources resolved only within the tracking domain, backward direction only) could be stated in the abstract (see also Section 4.3). In line 16, the authors could close the abstract by remarking something like “[…] hydroclimatic variability across South America *at the regional/continental scale*”.

Section 1

Line 31: Tuinenburg et al. (2020) is not an appropriate reference here, since this line refers to applications of WAM2layers.

Line 61: Instead of Tuinenburg and Staal (2020), which refers to the UTrack model, it would be more appropriate here to cite Tuinenburg et al. (2020), which refers to the UTrack dataset.

Lines 59-63: the authors state that existing datasets frequently provide data at monthly timescales or at spatial resolutions exceeding 1.0°. However, there are existing atmospheric moisture tracking datasets with 1° and 0.5° spatial resolution (e.g., Tuinenburg et al., 2020; De Petrillo et al., 2025; Zhang et al., 2025). I would suggest revising this statement. Further, along these lines, the authors cite Zan et al., 2024, and Berger et al., 2025, which use the datasets rather than presenting the datasets themselves, and the sentence, as written, may be confusing.

Lines 70-73: the tagging of each individual grid cell ("finest tagging area") is not entirely new, since it was already adopted by existing datasets (e.g., Tuinenburg et al., 2020). The novelty of the dataset lies rather in the combination of 0.25° output, daily temporal resolution, 30-year coverage, and cloud connection through Zarr, which could be stated more clearly. In light of this, the authors could consider toning down the language of some statements throughout the manuscript, for example:

- "unparalleled spatial flexibility" (line 142)
- "finest possible tagging area" (line 146)
- "completely removes the computational barrier" (line 291)
- "we have proven the computational feasibility" (line 322)
- "an unprecedented leap" (line 331)

Section 2

Line 100 (Section 2.2): I would suggest including here a passage on the limitations of ERA5 for moisture tracking applications, such as the condensation values affecting the tracking in the nearest regions to the poles and how they are treated in the tracking, ERA5 performance over South America, and the non-closure of the atmospheric water balance. Some of these aspects are briefly mentioned in Section 4.3 (lines 295-299), but it would be helpful to introduce them, supported with references. Please also consider stating here the source of the ERA5 data (see the comment on the dataset and repository).

Section 2.3 (lines 129-136): the daily resolution is one of the main strengths of the dataset, but the meaning of the time coordinate, the choice of releasing one dataset per year, the definition of the variables, and the selection of the tagging cells could benefit from a clearer explanation. Section 2.3 suggests that precipitation is tagged in monthly windows with a 30-day spin-up ("to map the evaporative sources of precipitation occurring in January 2019, the backward tracking process continuously ran backward through December 2018"). Could the authors clarify:

- whether the time coordinate of e_track refers to the date of evaporation or the date of precipitation;
- where evaporation from December of year Y-1 that feeds precipitation in January of year Y is stored, given the annual Zarr stores.

This is relevant for users interested in daily analyses: if daily e_track is the evaporation date for precipitation tagged over a whole month, the daily event attribution shown in Section 4.1.2 (Fig. 4) may not be derived directly, and its interpretation may need to be revised. Relatedly, the claim that more than 95% of tagged precipitation is allocated within 30 days (lines 133-136) is

currently supported only by a literature citation; it would be helpful to show it with the diagnostics of the dataset building.

Section 2.4: this section seems to mix the scientific benefit of tracking individual receptor cells with the merits of a particular storage layout. Tracking each land cell separately preserves the flexibility to construct receptor regions afterwards; however, it does not require a tagging_mask dimension.

The same information could be represented with regular destination latitude/longitude dimensions, i.e., (time, source_latitude, source_longitude, destination_latitude, destination_longitude), with missing values at untracked destinations. Therefore, statements such as "users are no longer restricted by predefined boundaries" and the possibility to "isolate individual department-scale agricultural townships, or aggregate thousands of receptors" (lines 146-148) would apply equally to a classic 5D array (see also the comment on the compact tagging_mask dimension in the dataset and repository section).

The land-mask approach is not discussed. It would be useful to document how the tagging cells were selected, including the land-mask source and threshold, and the treatment of coastal cells.

Lines 144-145: Could the authors clarify how they obtain the ~230 TB dimension and where the compression to 8 TB comes from? Please note that the dimension order given at line 144 ([time, source_latitude, source_longitude, tagging_mask]) differs from Table 2 (time, tagging_mask, latitude, longitude).

Line 147: What do the authors mean by "department-scale agricultural townships" (also "agricultural township" at line 290)? Please clarify.

Remarks for section 2.4: claims about efficient access would benefit from measurements for representative queries. The missing production workflow and the undocumented mask construction currently limit reproducibility. Statements such as "unparalleled spatial flexibility" and "finest possible tagging area" may overstate what is demonstrated.

Section 3

Section 3.1, Table 2: please define more precisely the variables reported in Table 2. In particular, tagged_precip is only described as "Precipitation falling on the tagged receptor": please clarify whether it corresponds to the ERA5 total precipitation of the receptor cell, and to which area e_track and tagged_precip are normalised (source cell or receptor cell). Similarly, please clarify what gains and losses contain, since in Section 4.3 the boundary inflow is reported as gains, while in Table 2 gains are described only as "Moisture gained during the tracking".

Section 3.2 (lines 172-175): could the authors clarify the mass-conservation metric, and in particular what the 0.002% refers to? It would be useful to report separately the numerical error and the fraction of precipitation attributed to boundary inflow (see also the map suggested for Section 4.1).

Section 3.4: the comparison of the dataset with the existing literature could be strengthened. Specifically, they perform the comparison with output from a previous version of the same tracking method (WAM2layers v2) provided by Link et al. (2020) forced with ERA-Interim, and with the Quasi-Isentropic Back-Trajectory (3DQIBT) method by Dirmeyer et al. (2009), also

forced with ERA-Interim (as provided by Link et al., 2020). The comparison covers the average 2001-2018, using country-level TES and CIES values (Table 3) and the La Plata precipitationshed (Fig. 2). I would recommend validating the output of the AguaTrack-ARCO-SA with other more up-to-date available datasets forced with ERA5, such as:

- Tuinenburg, O. A., Theeuwen, J. J. E., and Staal, A.: High-resolution global atmospheric moisture connections from evaporation to precipitation, Earth Syst. Sci. Data, 12, 3177-3188, https://doi.org/10.5194/essd-12-3177-2020, 2020.
- De Petrillo, E., Monaco, L., Tuninetti, M. et al. Cell-scale atmospheric moisture flows dataset reconciled with ERA5 reanalysis. Sci Data 12, 629 (2025). https://doi.org/10.1038/s41597-025-04964-3

and/or with ensemble datasets, such as:

- Zhang, Y., Zhong, D., Cai, R. et al. A multi-model based dataset of global atmospheric moisture source-sink relationships and atmospheric basins. Sci Data 12, 1842 (2025). https://doi.org/10.1038/s41597-025-06123-0

Lines 205-227: the authors attribute the differences with respect to WAM2layers v2 and 3DQIBT to the higher spatial resolution of their dataset, concluding that it provides a "more constrained and physically realistic representation of continental water balances" (lines 215-216). However, the compared datasets differ at the same time in model version (WAM2layers v2 vs v3), forcing reanalysis (ERA-Interim vs ERA5), spatial and temporal resolution, and tracking domain (global vs regional). Therefore, the effect of the resolution alone cannot be isolated, and it is difficult to consider lower source fractions as more realistic without an independent reference. I would recommend supporting this statement and the comparison with other datasets with a dedicated sensitivity test, for example, running WAM2layers v3 with ERA5 coarsened to 1.5° for a subset of years, similarly to the approach adopted by Li et al. (2024), which the authors cite.

Lines 210-212: the cited reference (Li et al., 2024) compares Eulerian and Lagrangian moisture tracking models over the Tibetan Plateau and does not discuss land-sea mask or coastal blurring effects. Moreover, its sensitivity experiment on the forcing resolution shows that a finer resolution increases the contributions from nearby sources, which does not clearly support the interpretation provided here. I would suggest providing a more appropriate reference or demonstrating this effect directly (see the previous comment).

Section 4

Section 4.1: for receptors close to the domain boundaries, a relevant fraction of precipitation could originate from moisture entering the domain, which is not attributed to any source location (see the comment on Section 4.3). Some conclusions of the showcase applications could depend on this. It would also be helpful to clarify how continental recycling is defined (also with respect to the TES).

I would recommend providing a map of the fraction of tagged precipitation not attributed to any source within the domain (i.e., 1 - Σe_track/tagged_precip) for each receptor, to diagnose the

effect of the regional bounds of the tracking domain, and clarifying which denominator is used to compute the ratios reported in the manuscript (tagged precipitation or the sum of the sources within the domain).

Figure 5: the legend overlaps with the plotted data; please consider adjusting it.

Section 4.2: in my opinion, this is a very relevant part of the article, together with the discussion in lines 205-230. Both parts could be extended (the latter should also be better supported; see the comments on Section 3.4) and made more central and balanced with the showcase of the applications occurring throughout subsection 4.1.

Section 4.3: this section would benefit from more references and details on the topic. Users outside the atmospheric tracking community should be aware of the intrinsic limitations of atmospheric tracking output and the differences between tracking methods, and of the existing literature on this topic. Currently, the only reference is to the WAM2layers v3 article (Kalverla et al., 2025), while some relevant references are already cited in other parts of the manuscript (e.g., Tuinenburg and Staal, 2020; Li et al., 2024) but are not used in this discussion. Section 4.3 identifies three relevant limitations (ERA5 quality, the well-mixed-layer assumption, and the domain truncation), but none of them is quantified. I would recommend expanding this section to discuss:

- the spread between methods, including Eulerian vs Lagrangian approaches;
- the sensitivity to tracking choices;
- the ERA5 biases in precipitation and evaporation in South America (see also the comment on line 100).

Authors may consider including references that acknowledge, discuss, and treat the uncertainty of atmospheric tracking methods, forcing data and final output.

Section 4.3 (lines 308-316): the spatial domain and the backward-only reconstruction are important points to discuss and frame in relation to datasets with global coverage and both forward and backward reconstructions. Regarding the spatial domain, only land grid cells of South America are tagged as precipitation receptors, so precipitation falling outside the boundaries of the selected domain (90°W-25°W, 60°S-15°N) is not tracked. Moisture sources are resolved only within the tracking domain, which includes the adjacent portions of the Pacific and Atlantic Oceans, while moisture coming from farther away (the wider oceans and the land outside the domain) is not assigned to any source location but is recorded as inflow across the domain boundary (gains variable). The authors mention this in Section 4.3; however, these regional limitations could limit some applications of the study and could be displayed more clearly in the article, including the abstract and the usage notes.

In relation to the backward perspective, the dataset provides tracked moisture values (mm/day) only in a backward direction (origin of precipitation). Since each land cell is tagged separately, the fate of evaporation over South American land could, in principle, be derived, but it would be incomplete because precipitation over the ocean and outside the domain is not tracked. Therefore, the dataset does not allow retrieval of complete atmospheric sheds in a forward direction (fate of evaporation). It would be helpful to explain this clearly, so that users do not treat the dataset as providing complete evaporation sheds.

Dataset and repository

The repository supports analysis of the published outputs but does not provide a workflow to reproduce dataset generation. In this perspective, the full code used to generate the dataset should be provided. This repository, or an additional repository, should contain the code to:

- acquire and pre-process the ERA5 data;
- run the model, including the WAM2layers v3 configuration files (e.g., the 10-min time step and the 30-day spin-up);
- generate the tagging mask coordinates and receptor indices;
- save the output in Zarr, including the chunking and compression settings.

The repository should be sufficient to reproduce the generation of the dataset. Please also state the software versions and archive the workflow with a DOI.

Although ERA5 is cited in the manuscript, its provenance and licence are not clearly documented on the Hugging Face or repository documentation. The citation of Carver and Merose (2023) suggests that the forcing data were obtained from Google's ARCO-ERA5 rather than from the Copernicus Climate Data Store ; please include a statement of the source, version, and access date, and include the relevant ERA5 references, license link, and required acknowledgments.

Could the authors explain why the daily data are distributed as separate annual Zarr stores? A single Zarr store with a continuous time coordinate would simplify multi-year analyses. This also connects to the handling of the tracking across year boundaries (see the comment on Section 2.3).

Since Zarr supports compression and omission of fill-value-only chunks, could the authors explain the choice of the compact tagging_mask dimension rather than regular destination latitude/longitude dimensions with missing values outside the tagging region, and clarify its storage, memory, or performance benefits, ideally with measurements for representative queries? (see also the comment on Section 2.4)