A multidimensional daily precipitation dataset (1998-2024) over the Third Pole with uncertainty estimates and precipitation phase information

This paper introduces a novel precipitation dataset for the Tibetan Plateau. Given that the Himalaya–Tibetan Plateau region continues to stand out in global precipitation datasets (both in satellite retrievals as well as in high-resolution model simulations) continued efforts to improve precipitation quantification by combining different data sources are highly valuable. Such datasets can serve both practical hydrological applications and as useful references for model evaluation.

However, I think there are still a few issues that should be addressed before the paper is ready for publication. In particular, the manuscript would benefit from a clearer explanation of the added value of the new dataset and from additional evidence demonstrating that it provides improvements over existing gridded precipitation datasets, as this appears to be one of the main motivations and objectives of the study. I also strongly recommend clarifying several aspects of the methodology and being more cautious in some instances where the results or novelty of the dataset may currently be overstated. Overall, I see value in the work, and I believe that addressing these points would substantially strengthen the manuscript. Please see my detailed comments below.

Major comments

- **Data availability:** First of all, the link to the dataset in the abstract is not working, which constitutes a major obstacle to properly assessing the dataset and, consequently, reviewing the publication. There is another link in the data section of the paper but I could not succeed to download any files. It says that I should use the FTP client tools but no proper documentation on how to download the data is given. Given that the dataset is only 5 GB (Is that really true? I would have expected at least a few hundred GB for 30 years of daily data), there should be ways to easily download the files with one click. Alternatively, please provide instructions on how to access the data. Since this is a dataset-based publication, I would prefer to reserve my final assessment until the dataset is made available and can be evaluated directly. In my view, the final decision regarding the publishability of the manuscript should take into account the quality, completeness, and usability of the underlying dataset.
- **Added value of the new precipitation estimates:** While there is definitely value in merging different data sources, adding user-friendly information and making the data accessible in a nice way, I think it is overstated that new information is really created here, rather than bias correction of the dry bias in precipitation in the satellite data that becomes as the new product uses gauge measurements as the target data. I acknowledge that it is hard to fully validate the correction against ordinary uncorrected station precipitation because the reference gauge data contain the same systematic bias

in wind-induced undercatch. Instead the validation can focus on whether the correction improves consistency with independent hydrological and atmospheric constraints which is partly done here in Fig. 12 by looking at the water balance. However, I think the science problem behind this validation is not discussed in sufficient detail and references to how others have addressed this issue could be important (e.g. Kochendorfer, J., Rasmussen, R., Wolff, M., Baker, B., Hall, M. E., Meyers, T., ... & Leeper, R. (2017). The quantification and correction of wind-induced precipitation measurement errors. *Hydrology and Earth System Sciences*, *21*(4), 1973-1989.) In addition, it is not explained where the SWE measurements come from and what biases they may contain. Can the SWE from ERA5 be added to show that the novel estimates are closer to the observed SWE than what ERA5 precipitation would be to the ERA5-derived SWE?

- **ERA5 uncertainty for small-scale processes:** A crucial part of this dataset is based on ERA5 but its uncertainties in both near-surface winds and precipitation are discussed in an insufficient manner. I suggest first to add some references and a discussion of what the expected uncertainty in ERA5 near-surface winds is (e.g. Minola, L., Zhang, G., Ou, T., Kukulies, J., Curio, J., Guijarro, J. A., ... & Chen, D. (2024). Climatology of near-surface wind speed from observational, reanalysis and high-resolution regional climate model data over the Tibetan Plateau. *Climate Dynamics*, *62*(2), 933-953). Second, the fact that ERA5 uses a convective parametrization is not discussed at all. A valuable addition would be 1) to quantify how much of the precipitation in ERA5 is sub-grid scale. The authors briefly talk about surface pressure as a proxy for vertical motion, but why do the authors choose to not include any variables that are more directly linked to convection, e.g. vertical velocity or any instability parameters such as CAPE and CIN?
- **Comment on hourly data:** Can you comment on what applications might need hourly data and if your framework could somehow be used to correct hourly precipitation estimates in the future?
- **Validation with independent stations:** Throughout the manuscript, the new precipitation estimates are validated against a set of independent observations which I believe belong to the red circle stations in Fig. 2. Can the authors comment on the expected deviations from those measurements compared to the rest of the station network given that the red stations seem to represent a quite different environment and terrain compared to all blue stations?
- **Comparison between datasets:** It is not entirely clear if for the validation part of the paper, the datasets are compared at the same resolution (have they been all regridded to the same grid?) or not. Similarly, Fig. 4 shows a slightly different domain covering less of the wet parts of Northern India for TPHiPr compared to the other sets. Has this been taken into account in the analyses that follow?

Detailed comments

- The machine learning method is explained quite late in the introduction and it does not immediately become clear how the authors utilize the input datasets to provide novel

information. I suggest revising the introduction and moving a simple explanation of the developed framework relatively close to the start of the paper.

- L. 45: What exactly was improved in the double machine learning approach and why was it called a double machine learning approach? This information would be useful to have here.
- L. 54: I suggest to be a bit more conservative in the language and not overemphasize the novelty of the used data formats and technologies. For example, I believe “comprehensive NetCDF4 data cube” simply refers to NetCDF4 files with multiple dimensions which has been the state-of-the-art data format for climate records for a long time. In fact, new data formats such as zarr stores that are equally practical but easier to compress have emerged. Have the authors considered storing the data as zarr files?
- Is the ERA5 regridded onto the same grid?
- Fig. 4: The colors don’t match the legend so it is hard to map those (please correct that), but is it right that iMERG and CRISP_Clim are the two curves that are very well aligned? Then what bias is really reduced here?
- L. 47: Could you please clarify what kind of uncertainty you refer to when you discuss the “uncertainty interval”? It would be helpful to add a brief discussion here what the sources of different types of uncertainties are (e.g. uncertainties in the measurements, sampling biases, model/retrieval uncertainties, etc)
- L. 49: Since much of this paper is about wind-induced undercatch, I think it would be worthwhile to explain in more detail here.
- L. 58: can can meet → can meet
- L. 78: Do you mean that you use an additional calibration based on the radar and microwave measurements of GPM or do you simply mean that IMERG includes these measurements? This should be clarified because the sentence sounds like the existing retrievals are further enhanced.
- L. 83: Please correct this statement as GSMap does not solely rely on DPR measurements, but also includes the information from multiple satellite sensors.
- Related to the comment above, it would be useful to clearly state what the difference between GSMap and IMERG is and why it is practical to combine those two datasets. It should also be clarified to what extent both datasets are built on the same information which means that the satellite observations are given more weight in the created precipitation product. How would the authors expect results to change if only one of these two datasets were included?
- L. 84: There are two Kubota references: Kubota et al., 2017 and Kubota et al., 2020 but only Kubota et al., 2020 exist in the reference list. I believe the year is wrong and the authors intention was to refer to Kubota et al., 2020 in the instances where it is mistakenly written Kubota et al., 2017.
- L. 121: Can you clarify here which model is trained with which dataset and how the different data are combined? Do you compute daily values for the satellite and ERA5 data first? Is ERA5 regridded? Are there any other preprocessing steps of the existing datasets at this point that are not made clear?
- L. 130: Can you clarify what the added value of the GTS dataset is given that both IMERG and GSMaP already contain the station data? It is stated here that “traditional

large-scale satellite calibration algorithms frequently neglect local wind-induced solid precipitation undercatch and complex subgrid orographic lifting". This is more of an explanation why it is useful to include the information of gauge measurements in satellite retrievals and not an explanation why the GTS dataset alone adds important information. Moreover, snow undercatch and potential missing precipitation associated with complex orographic lifting over small spatial scales are also inherent uncertainties of gauge measurements rather than biases that are introduced through in algorithms for satellite retrievals like IMERG? Why would the station data used for IMERG exhibit a different degree of undercatch than the additional stations used?

- L. 133 - 149: As I understand it, this is a catch ratio added to the dataset by the authors to quantify snow undercatch.
- L. 136: How often is the missing data replaced with ERA5 data? An appropriate addition to the supplementary material would be 1) a quantification of which portion of the dataset has this step based on ERA5 data and 2) how well the ERA5 data compares to the point measurements of the stations for these variables.
- L. 138: What exactly are the precipitation thresholds?
- L. 143: Replace the term "numerical instability" as this usually implies an instability in numerical modeling due to the small numerical errors growing uncontrollably during calculations. I think you mean something else in this simple equation.
- L. 150: Which figures show this independent validation?
- L. 186: I would be slightly more conservative with presenting the framework as a hybrid machine learning/physics-based approach. The machine learning algorithm is trained on physical data and you create additional variables based on thresholds but I think most people would expect a more advanced construct of combining numerical physical models with AI when we talk about "hybrid machine learning".
- L. 277: Can you clarify here what the independent data for hydrometeor phase is and how it compares to satellite-retrieved hydrometeor phase (from IMERG) as well as the hydrometeor classification in ERA5 (from the microphysics scheme)?
- L. 287: Which long-term regional climatological volume is taken as the reference value to be restored? Given the discussed uncertainties in the different datasets, do the authors think there is the possibility that the estimates in long-term regional climatological precipitation volume also suffer from biases? Or is the assumption that undercatch is the main uncertainty which is not visible in the regional climatological mean because the missing precipitation at one point will be included at another point? Please clarify here, the motivation and underlying rationale why temporal and spatial variability in precipitation is assumed to be biased but the climatological mean represents the truth.
- L. 510: perform this negative bias → exhibit this negative bias
- Fig. 6: Does the black curve labeled "Observations" here refer to all station records?