Error analysis in the WAM2layers moisture tracking model under data limitations

Author: Ke Yang 5953979

Supervisor: Dr. Ruud van der Ent

Dr. Arjan Droste

Date: 17/11/2024

TU Delft

Content

1 Introduction

Assessing moisture sources is crucial for understanding the hydrological cycle across various time scales. It allows for the identification of source-sink relationships and helps pinpoint the primary conveyors of moisture transport and the processes involved, ultimately leading to precipitation (Gimeno, L., et al., 2020).

The study of the origin of rain emerged in the mid-twentieth century (Van der Ent, R. J., 2014). Since then, various moisture recycling models have been developed to track the movement of water. Online simulations, such as the RCM MM5 (Knoche, H. R., & Kunstmann, H., 2013) and the WRF model (Arnault, J., et al., 2016) have been used to tag and track the moisture. Another approach involves posterior water vapour tracking models, such as the multi-layer model 3D-T (Tuinenburg, O. A., 2013) and water accounting model WAM- 2layers (Van der Ent, R. J., 2014). As the enhanced version of WAM, WAM2layers 3.1.0 is utilized as the moisture tracking model in this research, featuring an additional vertical layer. This improvement is based on earlier studies that identified a key shear layer at around the sigma level at approximately 800 hPa for standard atmospheric pressure, or about 2 kilometers above sea level (Van der Ent, R. J., et al., 2013), but varying with topography using so-called model levels, which can better capture the interactions between different atmospheric layers.

This research mainly focuses on the WAM2layers model, which has been primarily applied to studying the water cycle and moisture tracking across various domains, such as tracing moisture sources for flood-season precipitation (July–August) in the Qinghai-Tibet Plateau (Deng, J., et al.,2023), and the moisture recycling in the Colombian Andes (Bedoya-Soto, J. M., & Poveda, G., 2024).

In the process of tracking, however, many limitations will restrict the accuracy of results, especially the different configurations of input datasets from different climate models, such as divisions of levels through the atmosphere and the corrections based on Total Column Water (TCW) datasets.

Therefore, since potential errors arising from running WAM2layers with limited data have not been fully measured, the target of this research is to systematically quantify how large the errors are that will arise from running WAM2layers with varying degrees of data limitation.

2 Methods

2.1 Study domain

As shown in Figure 2.1, the study domain, delineated between 90°E to 130°E and 30°N to 35°N, spans a diverse geographical transect across China from the Qinghai-Tibet Plateau, with elevations reaching up to 7000 meters, through the Sichuan Basin, and extending to the Yellow Sea. August 2023 is selected as the research timeframe, as this period typically experiences the most active hydrological cycle across much of China (Hua, L., et al., 2017).

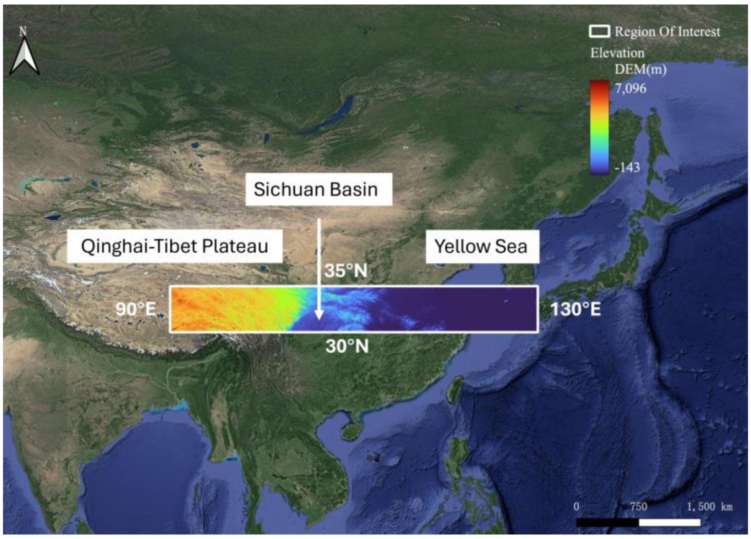


Figure 2.1 Study domain

2.2 Data

The WAM2layers moisture tracking model utilized ECMWF Reanalysis v5 (ERA5) datasets (Hersbach, H., et al., 2020). To maintain result accuracy, hourly data from August 1 to August 31 were downloaded, covering surface variables such as total precipitation, evaporation, surface pressure, 2m dew point temperature, 10m eastward and northward wind speeds, and total column water, as well as 3D variables at each level, including specific humidity and eastward and northward wind speeds.

2.3 Model configurations

2.3.1 Research questions

To measure the errors arising from the various data restrictions, surrounding the difference of Total Column Water (TCW) corrections, and the level and frequency configurations of climate model input data, five sub-questions are posed to investigate the specific impact.

1. *What are the differences between Total Column Water (TCW) and Total Column Water Vapour (TCWV)?*

Total Column Water (TCW) represents the total amount of water within a vertical column of the atmosphere in all forms, including vapour, liquid, and ice. Total Water Vapour (TCWV), however, only measures the water vapour in the atmosphere. Unlike TCWV, which only involves evaporation and condensation, TCW includes additional water content such as clouds and precipitation. In the datasets of ERA5, these two variables of the study region can be directly downloaded and compared.

2. *The impact of the correlation in WAM-2layers between Total Column Water (TCW) and Total Column Water Vapour (TCWV).*

In the model, Total Column Water Vapour (TCWV) is calculated by specific humidity across various levels. When Total Column Water (TCW) is available, it provides corrections to adjust TCWV, setting it equal to the TCW value. This step can test the differences in model results when using TCWV and TCW, analyzing the impact of the correction on model performance.

3. *What impact do the use of pressure levels and model levels have on the outcomes of the water accounting model? Does the use of model levels enhance its accuracy?*

Pressure levels are constant-pressure surfaces that are uniformly distributed over the Earth's surface, increasing at fixed intervals and unaffected by topographic features. Model levels, however, are defined based on surface topography, enabling adaptation to variations in terrain elevation. This ensures that pressure values remain meaningful even in areas with significant elevation differences, such as mountains or valleys.

4. *How do pressure level settings and time resolutions affect the accuracy of simulations in the WAM-2layers model?*

Some climate models, such as CMIP6, have fixed pressure levels and time resolutions in different scenarios. The goal is to collect and confirm the different types of pressure levels and time resolution settings and reproduce them in ERA5. These datasets will then be applied to WAM-2layers to analyze how they affect the accuracy of simulations.

2.3.2 Model settings

WAM2layers includes a configuration file (in YAML format) to configure tracking parameters, such as processing periods, tracking and tagging regions, and time frequency. The settings are adjusted according to specific research questions.
The tagging region, based on the area of rainfall events, spans from 106°E to 116°E and 30°N to 35°N, covering the period from July 13 to July 15, 2023.

As shown in Table 2.1. and Table 2.2., experiments were conducted using multiple pressure and model levels to investigate the impact of different level settings. The pressure level (Plev) configurations are based on the CMIP6 data request document from the World Climate Research Programme (CMIP6 Data Request, 2014). Specifically, the Plev7 configuration is used for simulating data with the MPI-ESM1-2-LR model.

Table 2.1 Pressure level settings

Plev 4	50, 200, 500, 850 hPa
Plev 7	50, 250, 500, 600, 700, 850, 925 hPa
Plev 8	10, 50, 100, 250, 500, 700, 850, 1000 hPa
Plev 14	1, 5, 10, 20, 50, 100, 150, 250, 500, 600, 700, 850, 925, 1000 hPa
Plev 20	100,200,300,400,500,550,600,650,700,750,775,800,825,850,875,900,925,950,975,1000 hPa
Model Levels	20,40,60,80,90,95,100,105,110,115,120,123,125,128,130,131,132,133,134,135,136,137 levels

Table 2.2 Experiment Configurations

	levels	Input frequency	TCW corrections
Different Pressure levels	Plev 4	1 hour	without
		1 day	without
	Plev 7	1 hour	without
	Plev 8	1 hour	without
	Plev 14	1 hour	without
	Plev 20	1 hour	without
		1 day	without
TCW corrections	Plev 20	1 hour	with
		1 day	with
Model levels		1 hour	with
		1 day	with

In the tables, "Plev" refers to "pressure levels", and the following numbers (4, 7, 8, 14, 20…) represent the quantity of levels. The default configuration for the full WAM2layers model includes TCW datasets with either 20 pressure levels or 22 model levels. To assess the impact of TCW corrections, TCW datasets were removed, allowing the direct use of total column water calculated from specific humidity across different levels. The model levels study used the default configuration. Furthermore, to explore the effect of time resolution, daily datasets were applied to the default settings (20 pressure levels and model levels with TCW), to 20 pressure levels without TCW, and to the least favorable condition (4 pressure levels without TCW correction).

3 Results

3.1 What are the differences between Total Column Water (TCW) and Total Column Water Vapour (TCWV)

To specify the difference between total column water(TCW) and total column water vapour(TCWV) in different temporal and spatial scales, comparisons are drawn both for hourly variations within a single day and daily variations throughout an entire month.

3.1.1 The hourly values of TCW and TCWV in the region.

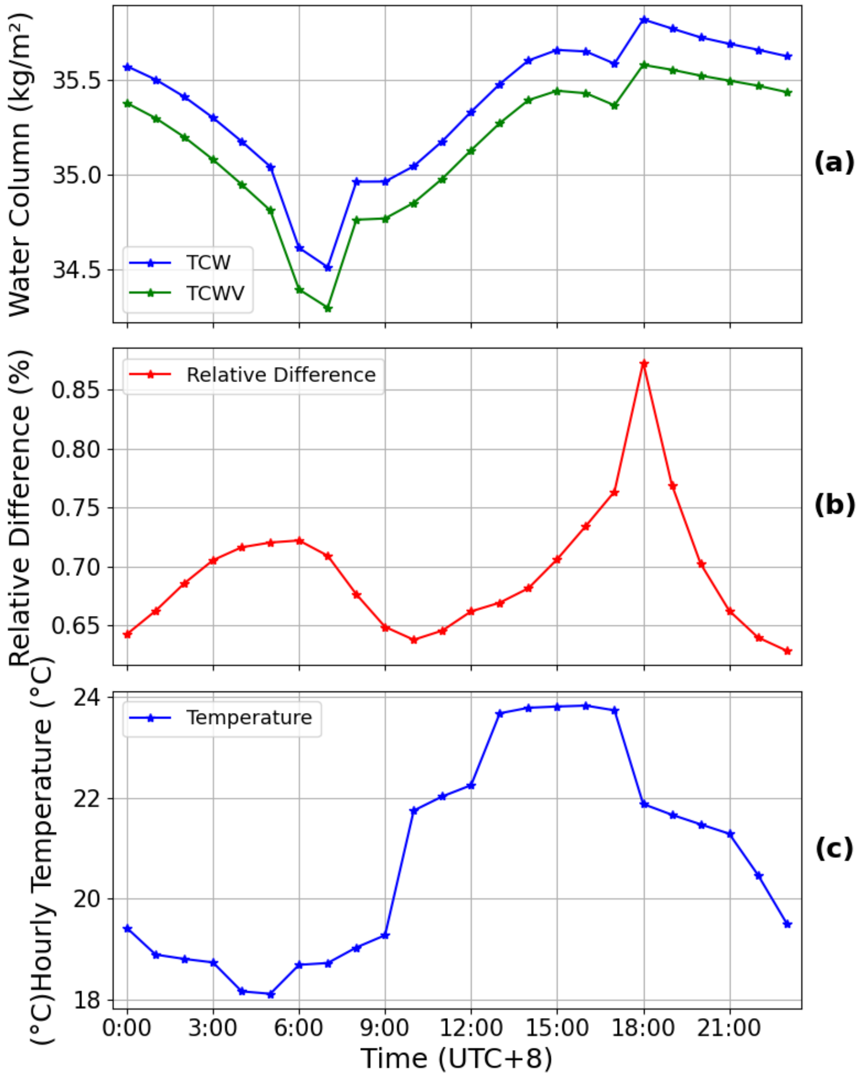


Figure 3.1.1 Comparison of hourly average values, calculated as the hourly mean across all hourly data within a month: (a) TCW and TCWV values; (b) Relative difference between TCW and TCWV, calculated as (TCW - TCWV)/TCW; (c) Hourly average temperature.

Figure 3.1.1 reveals a similar general trend, with a gap attributed to cloud ice and liquid water. The total column water (and vapour) is lowest in the early morning, around 6-7 am, when air temperature reaches the dew point, causing atmospheric water vapor near the surface to condense into dew, resulting in a peak in relative difference at this time, as shown in Figure 3.1.2. In the late afternoon, around 6 pm, as the temperature drops with sunset, the atmosphere's capacity to retain water vapor decreases, leading to condensation into liquid water throughout the atmosphere rather than just as surface dew, producing the highest relative difference at this time.

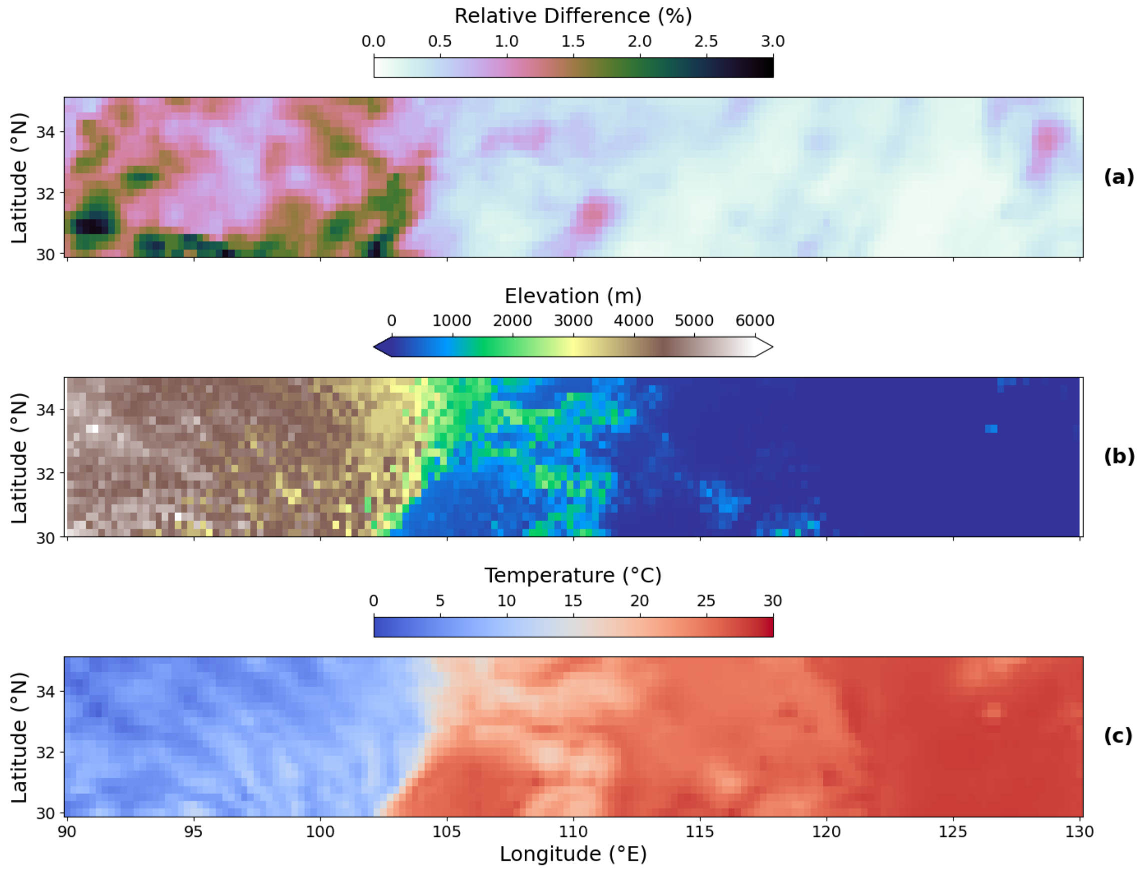


Figure 3.1.2 Spatial Relative Difference and comparison with elevation and average temperature at 18:00, calculated as the hourly mean across all hourly data within a month: (a) Spatial relative difference between TCW and TCWV, calculated as (TCW - TCWV)/TCW; (b)Spatial elevation; (c) Spatial temperature

The spatial map of the relative difference at 18:00, as shown in Figure 3.1.2, reveals peak values reaching around 3%, which closely aligns with the elevation and temperature patterns. As elevation rises, temperature drops, diminishing the atmosphere's capacity to hold water vapor. Consequently, this leads to increased condensation and a greater proportion of liquid water in the total atmospheric column.

3.1.2 The daily values of TCW and TCWV in the region.

The hourly performance of TCW and TCWV reveals the general daily pattern, however, during specific events, the values can fluctuate significantly over a concentrated few days.

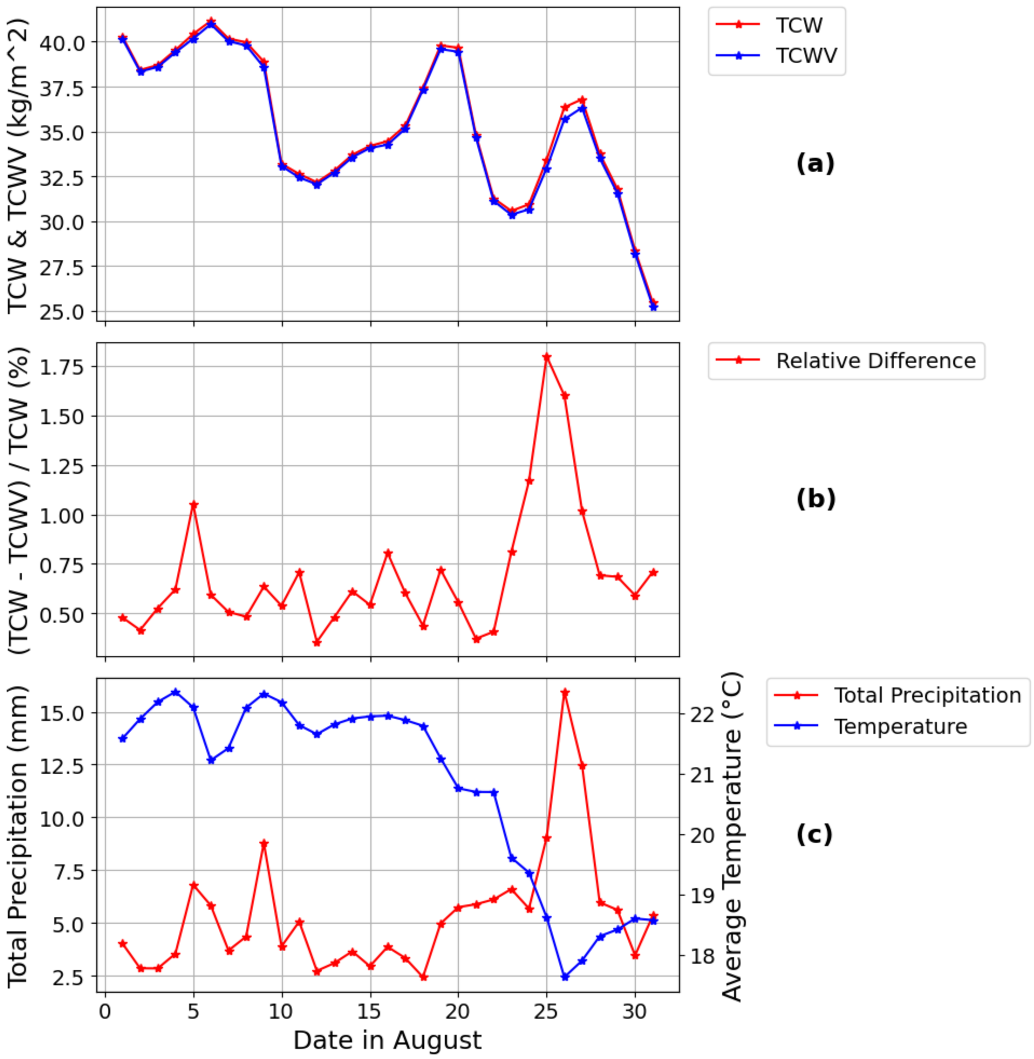


Figure 3.1.3 Comparison of daily average values about TCW and TCWV: (a) Values of daily TCW and TCWV; (b) Relative difference of daily TCW and TCWV, calculated as (TCW - TCWV)/TCW; (c) Comparison of daily total precipitation and daily average temperature

As shown in Figure 3.1.3 (a), total column water (TCW) and total column water vapor (TCWV) exhibit a similar decreasing trend throughout August, with a broader range of variation than observed within a single day. In Figure 3.1.3 (b) a notable peak in the relative difference appears on August 25–26, corresponding to a rainfall event on August 26, as depicted in Figure 3.1.3 (c). During this event, a temperature decrease led to atmospheric water accumulation, which ultimately condensed into raindrops, resulting in the highest observed difference between TCW and TCWV for the month.

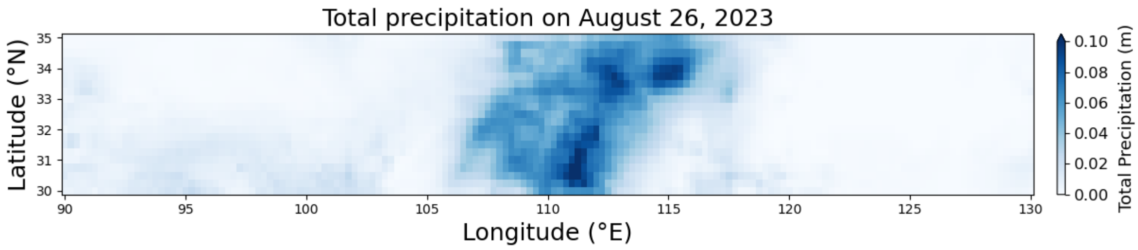


Figure 3.1.4 Spatial Map of total precipitation on August 26, 2023

Zooming in on this specific event, Figure 3.1.4 shows that precipitation was primarily concentrated in the eastern plains. To gain further insight into moisture movement leading up to the August 26 rainfall, the WAM-2layers model was employed to track moisture from August 1 to 27, with the tagging region defined between 106°E-116°E and 30°N-35°N.

3.2 What is the difference of result by using Total Column Water (TCW) and Total Column Water Vapour (TCWV) in the WAM2layers.

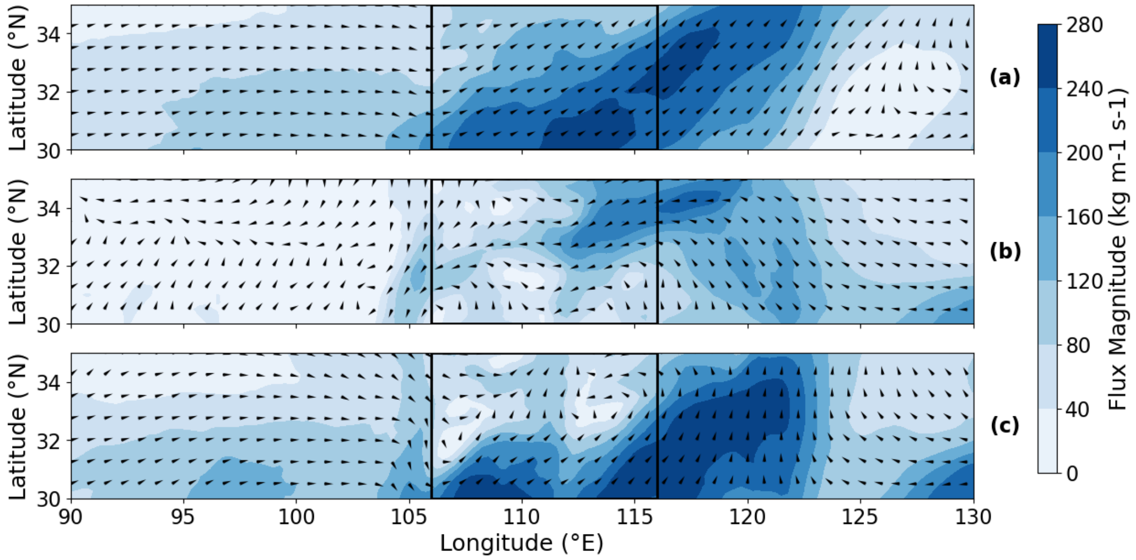


Figure 3.2.1 Magnitude and Direction of Moisture Flux: (a) Flux of upper layers; (b) Flux of lower layers; (c) Flux of the whole atmosphere (upper layers + lower layers)

Figure 3.2.1 illustrates the flux conditions from August 25 to August 27, during which tagged moisture reaches its peak and precipitation occurs. For total flux, the opposing directions of sea and land breezes generate vortices within the tagging region. Vertically, the upper layer shows a relatively uniform wind direction with high speeds, reflecting the overall distribution pattern. In contrast, the lower layer displays more variable wind directions, often opposing those in the upper layer, fostering atmospheric convection and moisture redistribution, which ultimately leads to precipitation within the tagging region.

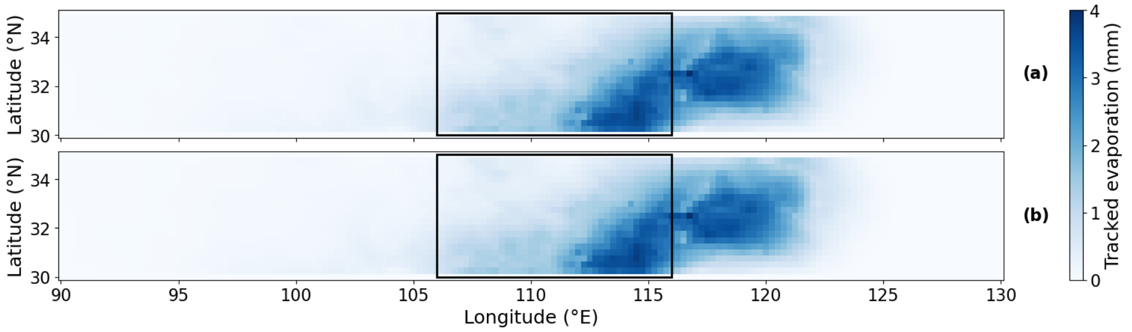


Figure 3.2.2 Spatial Map of Accumulated Tracked Moisture with and without TCW Corrections: (a) with TCW corrections; (b) without TCW corrections

The results, including Figure 3.2.2 and Table 3, show that the accumulated moisture distribution closely mirrors the pattern of wind speed magnitude. Tagged moisture is predominantly concentrated in the eastern part of the domain and the tagging region, accounting for approximately 46%, while the western part contributes around 6%, indicating that the moisture primarily originates from eastern convection.

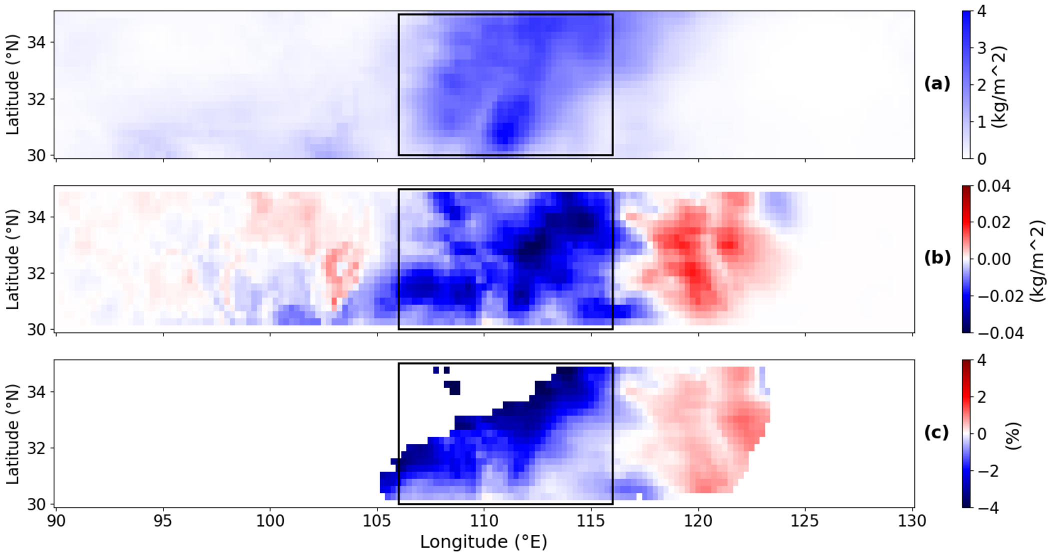


Figure 3.2.3 Spatial comparison of results with and without TCW correction:

(a) difference of TCW and TCWV values, calculated as TCW - TCWV;

(b) difference of accumulated moisture results calculated with and without TCW corrections, derived as results with TCW - results without TCW;

(c) relative difference in accumulated tracked moisture with and without TCW corrections, calculated as (with TCW - without TCW) / with TCW. To focus on significant values, only areas where results with TCW > 0.5 are selected.

The comparison of results with and without TCW corrections, as shown in Figures 3.2.3, indicates that differences are mainly concentrated in the tagging region and the eastern part of the domain. In these areas, moisture is underestimated in the tagging region but overestimated in the east, with relative differences reaching up to 4%. On August 26, the difference between TCW and TCWV is particularly evident in the tagging region, where moisture underestimation is attributed to heavy cloud cover during rainfall events. As a result, since total precipitation remains consistent across both settings, the tracked moisture around the tagging region occupies a smaller proportion.

The percentage analysis indicates that, without TCW corrections, moisture in the region shifts slightly eastward. This shift is likely due to the higher proportion of atmospheric water in liquid form rather than water vapor before precipitation occurs.

3.3 How much error is introduced in WAM2layers by using pressure levels instead of model levels?

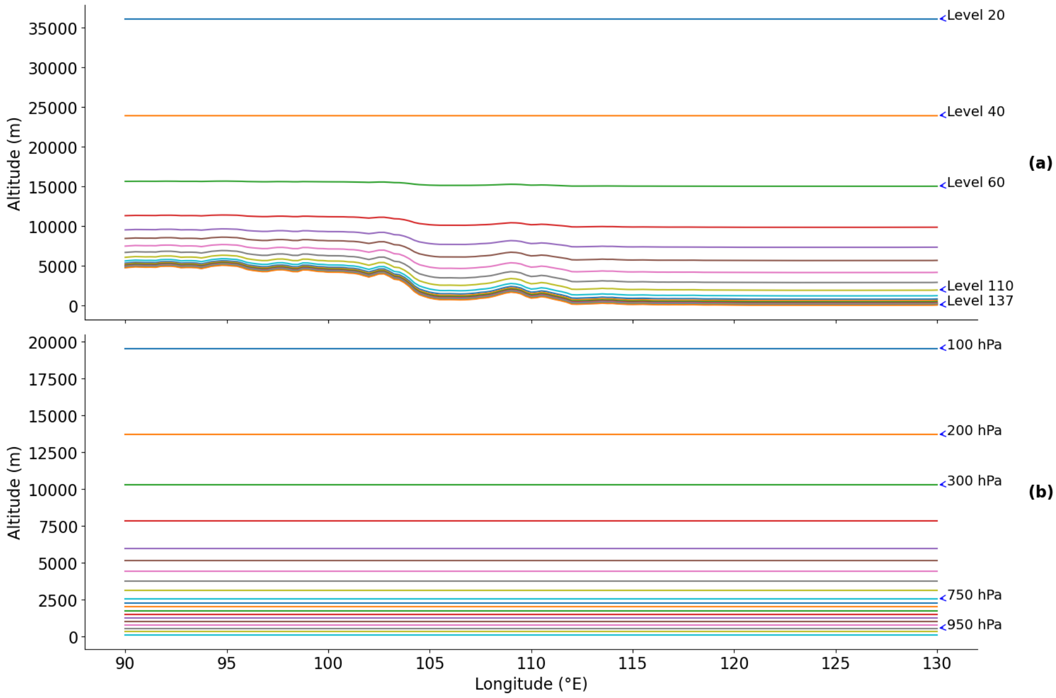


Figure 3.3.1 Comparison of model levels and pressure levels: (a) schematic representation of model level altitudes along the 32° latitude cross-section; (b) schematic representation of 20 pressure level setting's altitudes

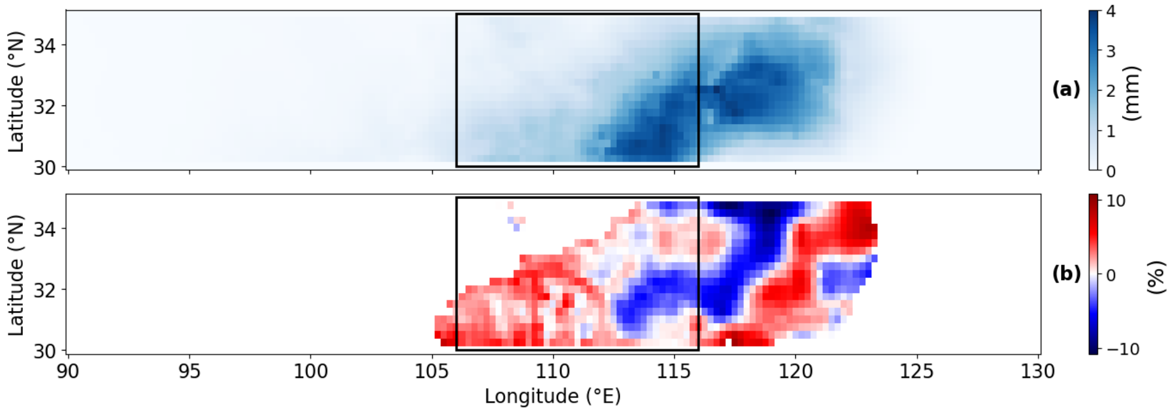


Figure 3.3.2 Model Level results and comparison: (a) accumulated moisture tracking results of model levels; (b) related difference between the results of model levels and pressure levels, calculated as (with TCW pressure levels results – model levels results) / with TCW pressure levels results. To focus on significant values, only areas where results with TCW > 0.5 are selected

As shown in Figure 3.3.1, unlike pressure levels, which are divided strictly by altitude, model levels adjust according to surface topography, enhancing efficiency. In high-elevation mountainous regions, this approach prevents values within the mountain itself, which would be meaningless, thus improving data utility.

Therefore, as shown in Figure 3.3.2, the difference in accumulated moisture results between model levels and pressure levels is primarily influenced by terrain, particularly in high-altitude areas. However, in the tagging region, where the altitude is approximately 300 meters, this difference is less pronounced. in areas where accumulated moisture is more concentrated and at higher elevations, the results from model levels are comparatively lower.

3.4 How do pressure level settings affect the accuracy of simulations in the WAM-2layers model?

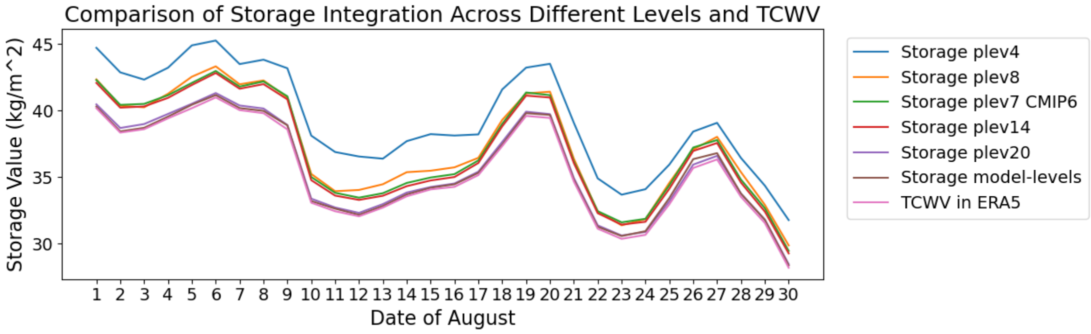


Figure 3.4.1 Comparison of Calculated Integrated Moisture Storage at Different Pressure Levels and Model Levels, Compared with the TCWV Dataset in ERA5

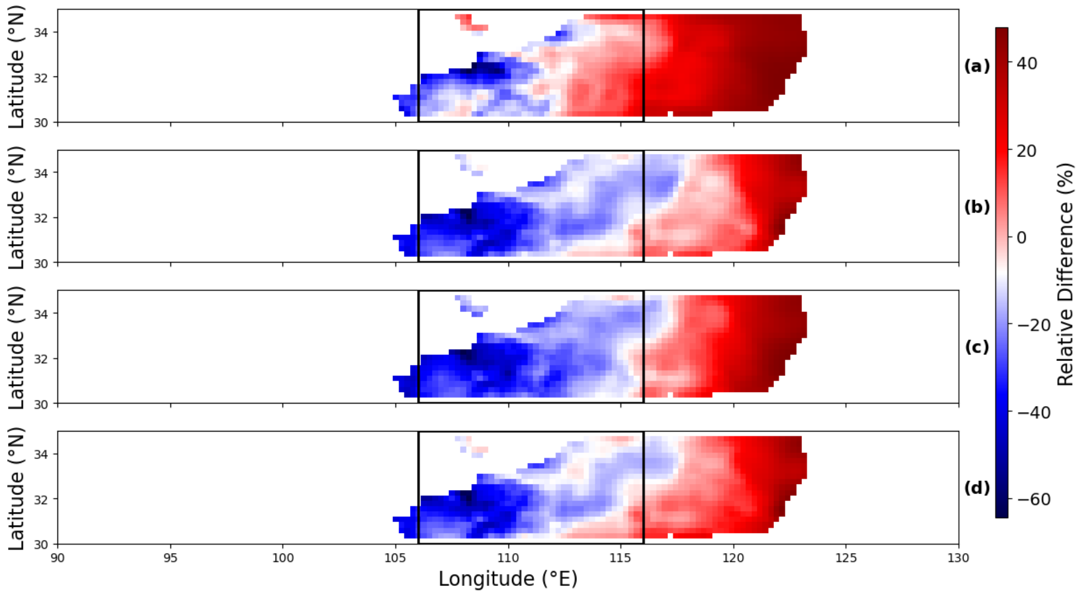


Figure 3.4.2 Spatial map of relative difference in accumulated tracked moisture between different pressure levels, compared with default settings (plev20 without TCW): (a) plev 4; (b) plev 7; (c) plev 8; (d) plev 14. To highlight significant values, only areas where default settings result > 0.5 are selected.

As shown in Figure 3.4.1 and 3.4.2, the pressure level settings can significantly affect the preprocessing of moisture storage, as demonstrated in the comparison figure. With more pressure levels, the results approach the true values in the ERA5 dataset. However, when comparing proportions across different settings, the results lack continuity; for instance, the results with 8 pressure levels perform worse than those with 7 or 14 levels. This discrepancy may stem from the distribution of pressure level settings. In the Plev8 configuration, where pressure levels are set to [10, 50, 100, 250, 500, 700, 850, 1000] hPa, levels such as 10, 50, and 100 hPa are too distant from the surface to be relevant, while 1000 hPa is so close that it may exceed the surface pressure. Consequently, the number of useful pressure levels in Plev8 is actually lower than in the 7-level configuration, resulting in a greater bias. Additionally, the west part of the domain occupies a larger proportion when fewer pressure levels are used, likely because variable differences become less distinct with a reduced number of pressure levels.

3.5 How do time resolution settings affect the accuracy of simulations in the WAM-2layers model?

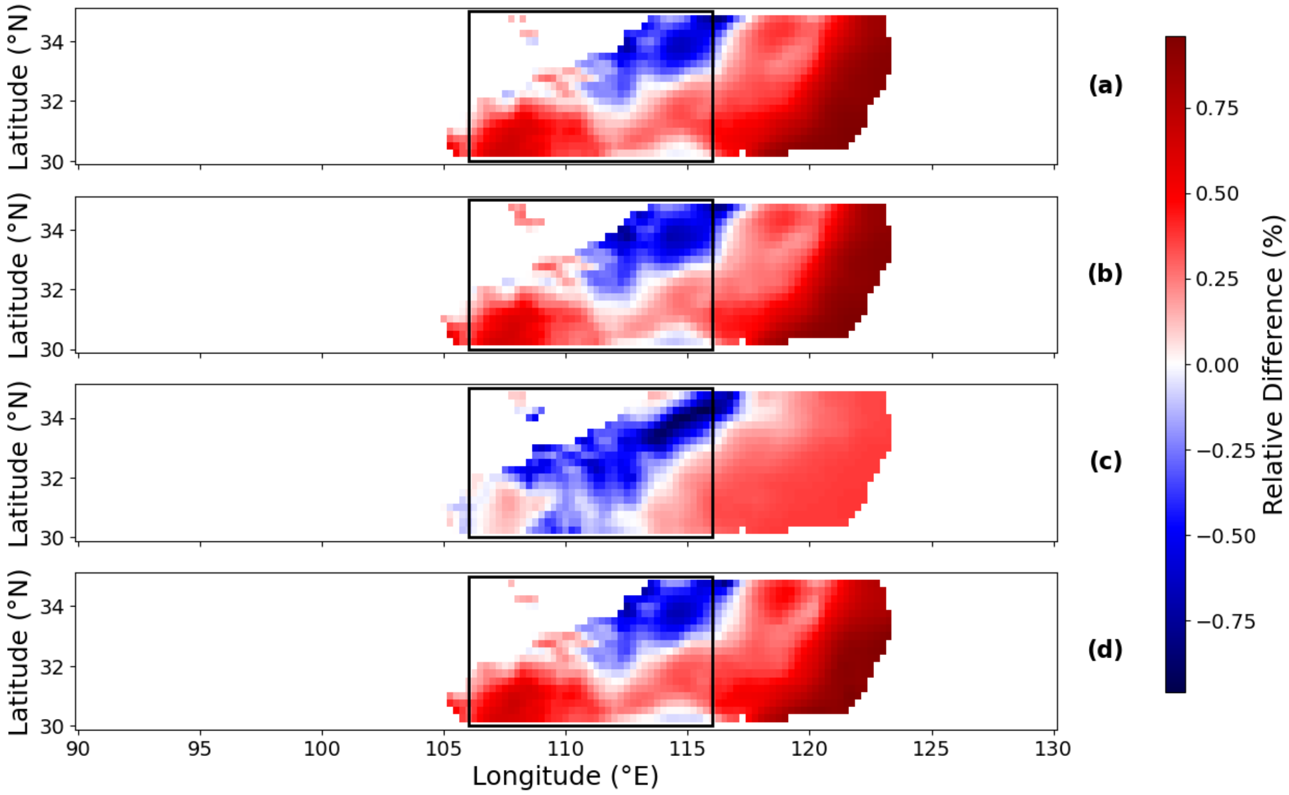


Figure 3.4.2 Spatial relative difference in accumulated tracked moisture between 1-day and 1-hour input time resolutions, calculated as (1-hour - 1-day) / 1-hour: (a) plev 20 results with TCW corrections; (b) plev 20 results without TCW corrections; (c) plev 4 without TCW corrections; (d) results with model levels

Compared to a 1-hour input frequency, the 1-day input frequency yields lower accuracy, though the relative difference pattern remains similar, with moisture underestimated in the tagging region and overestimated in the east. In the worst-case scenario, using 4 pressure levels, no TCW corrections, and a 1-day input frequency, the percentage of boundary transport and accumulation is higher in the west and lower in the east, while the tagging region maintains a comparable percentage to other settings. With 4 pressure levels and a 1-hour input frequency, the limited data volume results in rapid, high moisture influxes that exceed the air's saturation capacity, introducing errors. However, with a 1-day frequency, these peaks are averaged over time, significantly reducing this error.

Table 3 Boundary Transport and Accumulated Moisture Percentage in Different Portions

	Boundary Transport North	Boundary Transport South	Boundary Transport East	Boundary Transport West	Boundary Transport Internal	Result Region	Result West	Result East
plev4	18.55%	75.00%	0.0001%	2.72%	3.72%	49.23%	33.00%	17.77%
plev7	15.32%	84.27%	0.0001%	0.40%	-	49.70%	10.97%	39.33%
plev8	13.84%	85.52%	0.0001%	0.64%	-	56.30%	16.17%	27.53%
plev14	15.33%	84.21%	0.0001%	0.46%	-	49.44%	11.41%	39.15%
plev20 without TCW	14.88%	85.07%	0.0002%	0.06%	-	46.11%	6.36%	47.53%
plev20 with TCW	14.85%	85.10%	0.0002%	0.05%	-	47.87%	6.36%	45.77%
Model level with TCW	15.00%	84.97%	0.0002%	0.03%	-	46.21%	5.13%	48.66%
plev4 - 1d without TCW	20.23%	76.53%	0.0000%	3.21%	0.02%	57.99%	28.19%	13.82%
Plev20 – 1d without TCW	17.19%	82.75%	0.0000%	0.06%	-	56.80%	8.14%	35.06%
Plev20 – 1d with TCW	17.35%	82.59%	0.0000%	0.06%	-	57.05%	8.06%	34.89%
Model level – 1d With TCW	17.64%	82.31%	0.0001%	0.05%	-	56.61%	7.02%	36.37%

4 Discussion

In this study, since the study domain and tagging region were initially defined, several results reveal limitations in these area selections. The wind direction analysis appears overly restricted, suggesting a need to expand the domain to capture comprehensive sea-land air flow patterns. Additionally, the tagging region is disproportionately large relative to the domain. Ideally, it should be defined as a small unit, allowing for a focused analysis of interactions between this region and the surrounding domain. Moreover, considering the results from model levels that do not present differences within the domain, selecting precipitation events in higher-altitude regions will better reflect typical issues. In such cases, data from pressure levels may become unusable, while model level settings remain above the surface, providing a richer dataset.

Based on the existing results, several suggestions about the selection of datasets can be provided. Through all these settings, the pressure levels number and distribution make the biggest difference through the results, either the calculations of water column vapour, or the results of tracking accumulated evaporation. To better capture useful information about the weather, at least 6-7 levels should be selected, and they should be more concentrated in 600 - 900 hPa, which is near the surface, but less than the surface pressure, around 1000 hPa. Model levels are better choices if the target region has high elevations or uneven terrain, for they can avoid meaningless values under the surface. For intense, short-duration rainfall, if the focus is on longer-term tracking rather than that brief period of rain, the data frequency larger than one hour, such as one day, may be suitable to prevent overflow in simulations and reduce errors. For the additional Total Column Water datasets, while they may provide more accurate corrections, the impact on results is minimal. Therefore, many climate models that cannot provide this data can still achieve relatively accurate results by calculating specific humidity.

5 Conclusion

Despite limitations in selecting the study region and precipitation events, this study examines how data constraints influence the accuracy of the WAM-2layers moisture tracking model, with focuses on total column water (TCW) corrections, pressure level configurations, and time resolutions.

Qualitatively, while TCWV calculated from specific humidity is reliable, including TCW data can further improve accuracy. For regions with significant terrain variation, it is recommended to use model levels. Using 7 or more pressure levels, with a substantial proportion in the 600–900 hPa range, yields more accurate results. For concentrated rainfall, when the levels are insufficient, using a one-day time resolution rather than one hour helps better minimize errors.

Specifically, the differences between TCW and TCWV, influenced by terrain, become more pronounced during precipitation events, reaching up to 1.8%. Corrections to TCW introduce an approximate 4% change in accumulated moisture, leading to reduced values in the tagging region and increased values to the east.

The number and distribution of pressure levels significantly impact results. When levels are sufficient and well-distributed, results with 7 and 14 levels show reasonable accuracy, with an error margin of about ±20% compared to the default 20 levels. However, with fewer near-surface levels, such as 4 or 8, errors increase significantly, reaching 50%–100%.

In this study, model levels show a relative difference of 5%, effectively filtering out irrelevant data below the surface. However, since the tagging region is at a lower elevation, result differences may be more pronounced in areas with substantial altitude variation, such as high plateaus.

The impact of time frequency is minimal, with a 1-day interval showing only about a 0.5% error compared to 1-hour data, peaking at 0.8%. However, in boundary transport calculations with fewer levels, using 1-hour data for intense rainfall can introduce a 3% moisture error within the region, making a 1-day frequency more suitable in such cases.

Based on these patterns, different climate models and regional characteristics can guide the model configurations, enabling WAM-2layers to adapt to a wider range of moisture cycle studies with enhanced accuracy and robustness.

Data and code availability

The data and code underpinning the findings of this study are deemed valuable to the broader WAM2layers user community. To support transparency and facilitate further research, the following resources is accessible in the 4TU Research Data Repository:

- ERA5 Data: The ERA5 data used in this study.
- YAML Configuration Files: Configuration files that detail the setup and parameters used for simulations.
- Preprocessed Data: Data that has undergone preliminary processing steps necessary for subsequent analysis.
- Tracked Data: Output from the tracking algorithms applied to the preprocessed data.
- Notebooks: Jupyter notebooks containing the code used for generating the visual representations of the data.

These resources are available for access at the following link: https://doi.org/10.4121/8491b1ea-9964-4db6-9c64-93382945eb24

Access to these materials will be governed by the data sharing policies outlined by the 4TU Research Data Repository, ensuring compliance with relevant data protection regulations.

Bibliography

[1] Gimeno, L., Vázquez, M., Eiras-Barca, J., Sorí, R., Stojanovic, M., Algarra, I., ... & Dominguez, F. (2020). Recent progress on the sources of continental precipitation as revealed by moisture transport analysis. *Earth-Science Reviews*, *201*, 103070.

[2] Van der Ent, R. J. (2014). A new view on the hydrological cycle over continents.

[3] Knoche, H. R., & Kunstmann, H. (2013). Tracking atmospheric water pathways by direct evaporation tagging: A case study for West Africa. *Journal of Geophysical Research: Atmospheres*, *118*(22), 12-345.

[4] Arnault, J., Knoche, R., Wei, J., & Kunstmann, H. (2016). Evaporation tagging and atmospheric water budget analysis with WRF: A regional precipitation recycling study for West Africa. *Water Resources Research*, *52*(3), 1544-1567.

[5] Tuinenburg, O. A. (2013). *Atmospheric effects of irrigation in monsoon climate: the Indian subcontinent*. Wageningen University and Research.

[6] Van der Ent, R. J., Tuinenburg, O. A., Knoche, H. R., Kunstmann, H., & Savenije, H. H. G. (2013). Should we use a simple or complex model for moisture recycling and atmospheric moisture tracking?. *Hydrology and Earth System Sciences*, *17*(12), 4869-4884.

[7] Deng, J., Quan, Q., & Zhou, G. Analysis and Research of Precipitation in Qinghai-Tibet Plateau Based on WAM-2layers Calculation Model.

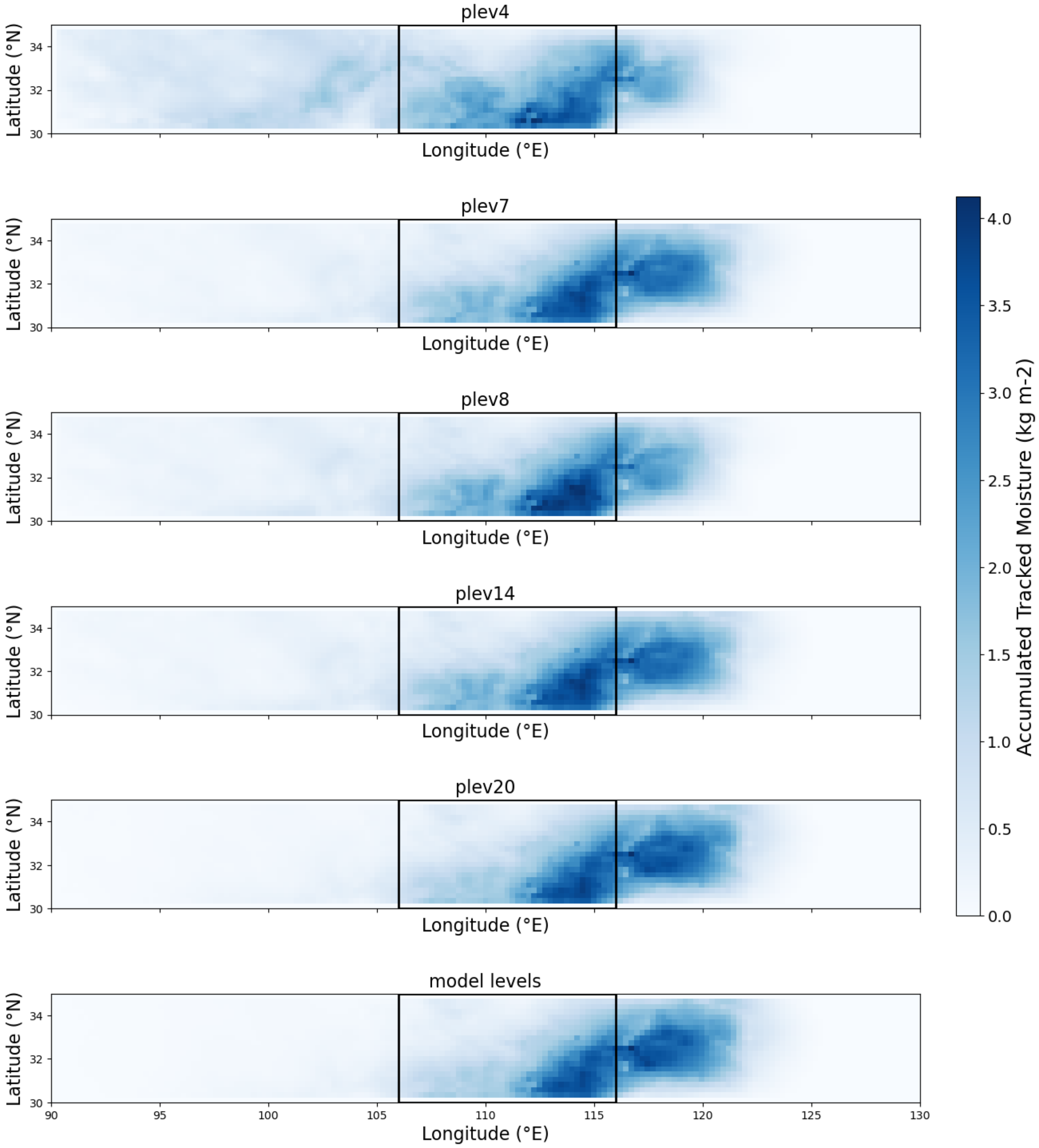
[8] Bedoya-Soto, J. M., & Poveda, G. (2024). Moisture recycling in the Colombian Andes. *Water Resources Research, 60*(1), e2022WR033601.

[9] Hua, L., Zhong, L., & Ke, Z. (2017). Characteristics of the precipitation recycling ratio and its relationship with regional precipitation in China. *Theoretical and Applied Climatology*, *127,* 513-531.

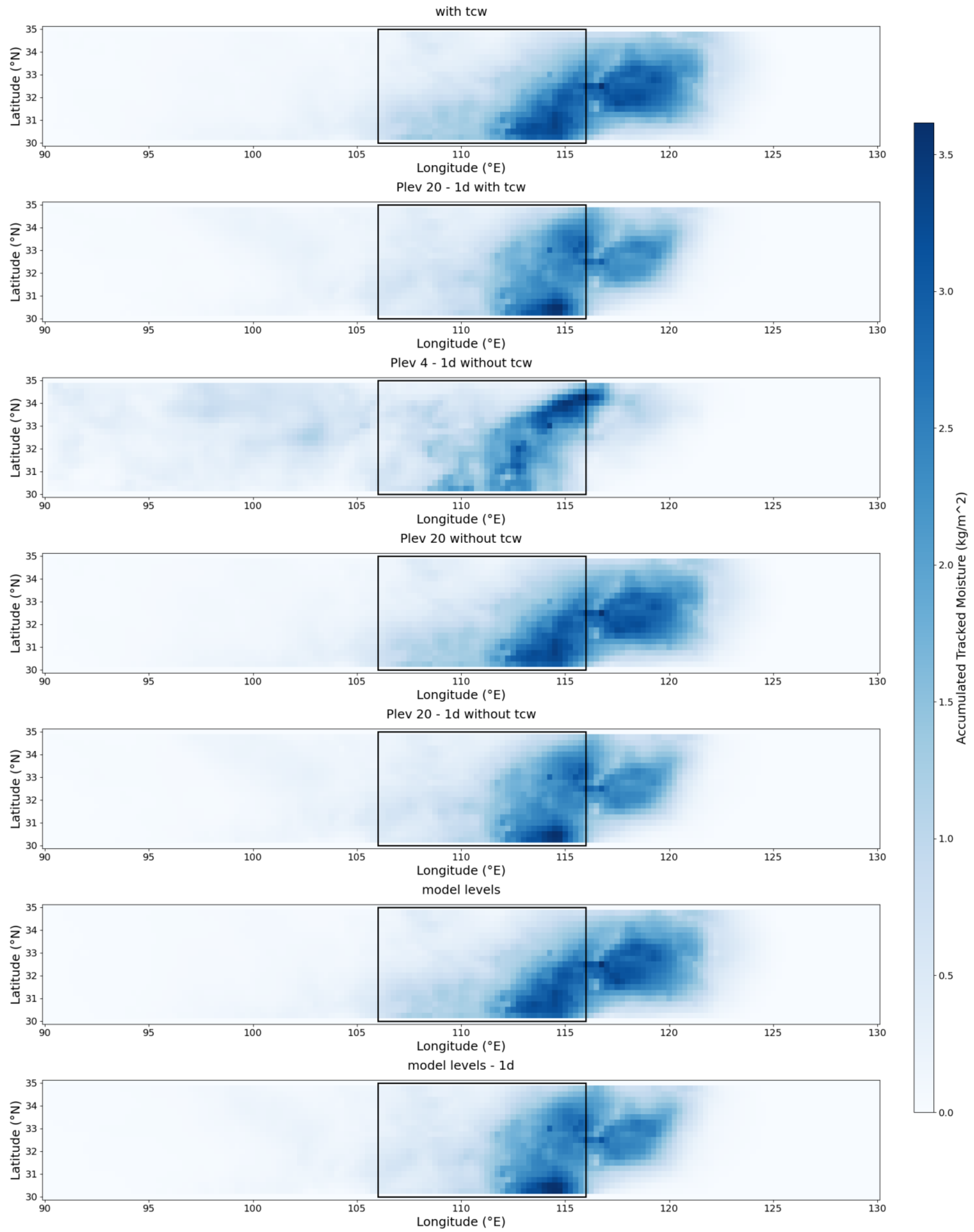
[10] Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... & Thépaut, J. N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999-2049.

[11] https://cmip6dr.github.io/Data_Request_Home/Documents/CMIP6_pressure_levels.pdf

Appendix A: Additional Figures



A1. Comparison of accumulated tracked moisture in different level settings



A2. Comparison of accumulated tracked moisture in different temporal resolutions