Real-time High-resolution Global PWV Retrieval based on Weather Forecast Foundation Models and cross-validation with Radiosonde, GNSS, and ERA5

Junsheng Ding, Wu Chen, Junping Chen, *Member, IEEE,* Jungang Wang, Yize Zhang, Lei Bai

***Abstract*—High-quality precipitable water vapor (PWV) plays a vital role in climate change and weather prediction studies. This research introduces a novel scheme for retrieving high-resolution surface-domain PWV with real-time and forecasting capabilities with global coverage, utilizing weather forecast foundation models represented by Huawei Cloud Pangu-Weather, Google DeepMind GraphCast, and Shanghai AI Lab FengWu. The accuracy of the new scheme is cross-validated against PWVs from radiosondes, Global Navigation Satellite Systems (GNSSs), and the fifth generation ECMWF reanalysis (ERA5). Results show the new scheme achieves 3.01 mm global root mean square error (RMSE) in real-time, and the value reduce to 2.25 mm when focusing only on land areas, which is more accurate than most existing methods that rely on post-processed surface-domain data. The poor accuracy in low-latitude and mid-latitude ocean regions limits the accuracy of the new scheme and future integration of GNSS PWV data from ocean sources is expected to improve it. Overall, the proposed scheme demonstrates very satisfactory global PWV accuracy and has the potential for further improvement with the development of artificial intelligence.**

***Index Terms*—Precipitable water vapor, Surface-domain, Weather forecast foundation models, Real-time, High-resolution.**

I. Introduction

THE total atmospheric water vapor content is typically quantified as precipitable water vapor (PWV). Accurate PWV measurement and forecasting are crucial for regulating global temperature and weather patterns [1]–[3]. PWV exhibits high and rapid variability in time and space, this poses a significant challenge for acquiring high resolution and accurate PWV [4], [5].

Manuscript received XX xx, 2025; revised XX xx, 2025. Date of publication xx XX 2025; date of current version xx XX 2025. This work was supported in part by the General Research Fund of Hong Kong (Grant No.15229622), in part by the Innovation and Technology Fund of Hong Kong (Grant No.ITP/019/22LP), in part by the National Natural Science Foundation of China (No. 42474034) and in part by the DFG COCAT Project of Germany (Nr.490990195). *(Corresponding author: Junping Chen)*

Junsheng Ding and Wu Chen are with the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China (e-mail: junsheng.ding@polyu.edu.hk; wu.chen@polyu.edu.hk).

Junping Chen and Yize Zhang are with Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, China, and also with the School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: junping@shao.ac.cn; zhyize@shao.ac.cn).

Jungang Wang is the Department of Geodesy, GeoForschungsZentrum (GFZ), 14473 Potsdam, Germany, and also with the Institut für Geodäsie und Geoinformationstechnik, Technische Universität Berlin, 10623 Berlin, Germany (e-mail: jgwang@gfz-potsdam.de).

Lei Bai is with Shanghai AI Laboratory, Shanghai 200030, China (e-mail: bailei@pjlab.org.cn).

PWV acquisition techniques are broadly classified into two categories: point-domain measurements and surface-domain measurements. Point-domain measurements are predominantly ground-based, and include radiosonde (RS), Global Navigation Satellite System (GNSS), and microwave radiometer methods [6]–[8]. Radiosonde, with over a century of records, provides direct atmospheric profile measurements and serves as a benchmark for other techniques [9], [10]. GNSS, an emerging technology, is also recognized for its high-precision, all-weather observations with high temporal resolution, making it a reference for accuracy assessment [11]–[13]. These techniques yield PWV accuracies of 1–3 mm [3], [14]. Ding et al [15] used foundation models to improve the GNSS PWV accuracy to the sub-millimeter level, yet they are limited to sparse locations that with GNSS stations that are unevenly distributed, with significant gaps in oceanic, desert, and underdeveloped areas.

The second category of PWV acquisition techniques is predominantly space-based and obtained through sensors on artificial satellites or aerial platforms [16], [17]. These include aircraft-based detection and meteorological and remote sensing satellite observations [18], [19]. These methods offer weather-independent, high-resolution regional observations. However, they typically exhibit lower accuracy, ranging from 3–8 mm [3], [20], [21]. Furthermore, the spatial coverage of its observation area of single measurement is typically limited, and the temporal resolution is constrained by the revisit frequency of the satellite. Real-time data acquisition is particularly challenging, making any forecasting efforts even more complex. Special attention should be given to the fact that many existing studies rely on less than a year of experimental data, often using only a few dozen days or one to two months. The accuracy reported in these studies may be overrated or underrated and unrepresentative, as the precision of PWV measurements varies seasonally, leading to different results. In addition, surface-domain PWV estimates can be derived in real-time through empirical models. For instance, the HGPT2 model achieves accuracies of 5.0 mm, 3.5 mm, and 5.7 mm on global, inland, and oceanic scales respectively [22]. These precision levels remain consistent with the uncertainty range of surface-domain PWV measurements reported in previous studies.

The emergence of weather forecast foundation models such as Pangu-Weather and GraphCast in recent years has provided new ideas for global high-resolution PWV forecasts. Using this type of model can generate more than ten days

TABLE I
PROPERTIES OF WEATHER FORECAST MODELS.

Model (release)	Affiliation	Model Type (AI technique)	Pressure levels	Step length (hours)	Inference time
Pangu-Weather (2023-1-5) *(Bi et al., 2023)*	Huawei Cloud	Foundation model (3DEST)	13	1, 3, 6, 24	1.4 s on a Tesla-V100 GPU (112 TFLOPS, 16/32 GB)
GraphCast (2022-12-24) *(Lam et al., 2023)*	Google DeepMind	Foundation model (GNN)	37	6	60 s on a Google TPUv4 GPU (275 TFLOPS, 32 GB)
FengWu (2023-4-6) *(Chen et al., 2023)*	Shanghai AI Lab	Foundation model (Transformer)	37	6	0.6 s on a Tesla-A100 GPU (312 TFLOPS, 40/80 GB)

forecast of 0.25°×0.25° resolution of the global atmospheric state in any user's local device within a few seconds. That means, the global high-resolution real-time and even forecast surface-domain PWV can be retrieved instantaneously. In this research, we computed the 15-day global forecast from three foundation models, represented by Pangu-Weather, GraphCast, and FengWu in 2022, and then obtained the global forecast in terms of RS PWV, GNSS PWV and ERA5 PWV as references to evaluate the accuracy of the new scheme. Section II describes the data sources and retrieval methods for the different sources of PWV used in this study, and Section III evaluates the accuracy of the new scheme on the station and global grid points, respectively. The discussion is given in Section IV, and a summary is given in Section V.

II. MATERIALS AND METHODS

A. Foundation Models

Relying on Ding et al. [23], which identified that sparse pressure levels result in systematic errors in zenithal tropospheric delays (a NWM with only 13 pressure levels will result in a systematic error of 1 mm for PWV), we have chosen the GraphCast and FengWu foundation models, which support 37 pressure levels. Additionally, to demonstrate the effect of sparse pressure levels on PWV retrieval bias, the Pangu-Weather model, which supports only 13 pressure levels, is included in our experiments. Among the three models, GraphCast is based on graph neural network (GNN) technology, while Pangu-Weather is based on a 3D Earth-Specific Transformer (3DEST) and FengWu is based on the Transformer architecture. Table I shows the properties of these three weather forecast models used in this research. Due to the 6-hour step length of GraphCast and FengWu, we used 6-hour Pangu-Weather to make a fair comparison. GNSS and ERA5 PWV were resampled to 6 h resolution, and when radiosonde was used as reference, all data were resampled to 12 h resolution because radiosonde data were recorded only twice a day, i.e., for 12 h.

B. Radiosonde Profile

The radiosonde profile data used in this research are obtained from the Integrated Global Radiosonde Archive Version 2 [24]. Note that we retrieved the PWV from the radiosonde pressure level records, rather than using the IGRA PWV products directly, for the IGRA data still suffered from jumps due to changes in instrumentation, etc., and the unhomogenized radiosonde data even affected most of the reanalysis data [25], [26]. We integrated data screening strategies from [5], [15], and [27], rigorously filtering the dataset to exclude non-compliant stations. It should be noted that a large percentage of the radiosonde raw data records have many missing data, such as incomplete time series records and pressure level records, which are mostly caused by weather and other reasons. Our criteria are met by fewer than 100 radiosonde stations globally after filtering, most of which are in the U.S. and a few scattered throughout Europe and Japan. The only data we have kept for this experiment comes from the U.S. stations (see Fig. 3).

C. GNSS Products

The GNSS products are obtained from the Nevada Geodetic Laboratory (NGL) [28]. Numerous studies have demonstrated that NGL tropospheric products possess high accuracy and serve as valuable data sources for earth science [3], [18], [27]. The GNSS data screening strategy is the same as that of [30]. Additionally, due to the problem of PWV in NGL products absorbing VMF model errors as mentioned in [27], the GNSS PWVs in this research are recalculated using NGL ZTD and ERA5 (see Fig. 1).

D. ERA5 Data

In this study, ERA5 pressure level data [31] and 37 layers of ERA5 single level data [32] are employed as initial inputs for the foundation models. ERA5 is delayed by 5 days, but the inference is near-instantaneous. The real-time in this research refers to forecasts of 5-day. It should be noted that all the data used in this research are from 2022 and are not involved in the models' training, testing, or validation processes. This ensures a fair and accurate assessment of model performance and avoids the risks of overfitting and data leakage.

E. PWV Retrieving and Comparison Method

The RS PWV is calculated as follows [33], [34]:

$$e = 6.112 \exp\left(\frac{17.6T_d}{T_d + 243.15}\right) \tag{1}$$

$$PWV = \frac{1}{g}\int_{p_1}^{p_2} \frac{0.622e}{p-e} dp \tag{2}$$

where T_d is dew point temperature (°C), e is water vapor pressure (hPa), p is the atmospheric pressure (hPa), g is the

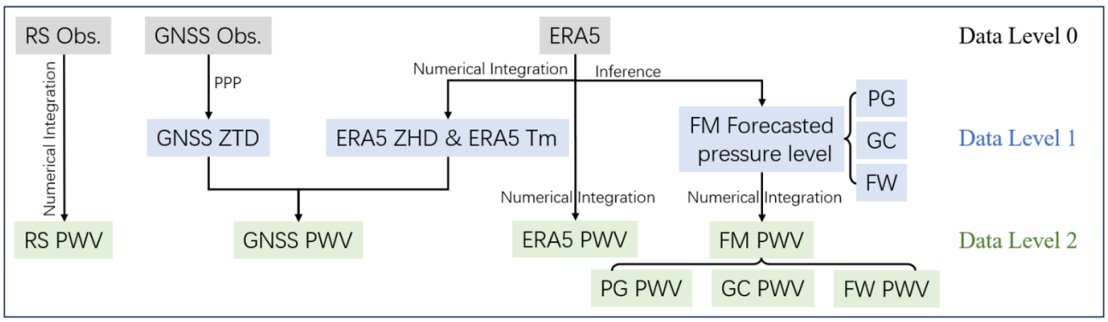


Fig. 1. Data processing sequence for PWV retrieving and comparison. RS = Radiosonde; GNSS = Global Navigation Satellite System; ECMWF = European Centre for Medium-Range Weather Forecasts; ERA5 = ECMWF Reanalysis v5; ZTD = zenith tropospheric delay; ZHD = zenith hydrostatic delay; Tm = weighted mean temperature; FM = foundation model; PG = Pangu-Weather; GC = GraphCast; FW = FengWu; PWV = precipitable water vapor.

acceleration of gravity (m/s^2), p_1 and p_2 are the atmospheric pressure at the lower and upper layers (hPa), respectively. Accurate zenith tropospheric delays (ZTD) can be calculated from GNSS observations, the GNSS derived PWV is calculated as follows:

$$ZWD = ZTD - ZHD \tag{3}$$

$$PWV = \frac{10^6}{R_w \cdot [k_2 - \varepsilon k_1 + k_3/T_m]} \cdot ZWD \tag{4}$$

Where ZHD is zenith hydrostatic delay, and $\varepsilon = M_w/M_d$, $R_w = R/M_w$ with the universal gas constant R, M_w and M_d are molar weight of water vapor and dry constituents. k_1, k_2 and k_3 are the "Rueger-best average" values, which are 77.6890 K/hPa, 71.2952 K/hPa and 375463 K2/hPa respectively [35], [36]. T_m is weighted average temperature. They can be obtained from the vertical profile of the atmospheric state through numerical integration (the method used in this research), they can also be calculated from the surface meteorological data and empirical models, but this results in a loss of accuracy [37]. Other than this, all other variables in the equation are constant coefficients. The aforementioned data processing sequences for PWV retrieving and comparisons are summarized in Fig. 1, and the different types of PWV used in this study are shown in Date Level 2 in Fig. 1.

III. Results

A. Accuracy Assessment at Stations Level - Single Site

Figure 2 shows the comparison of PWVs from different sources at the closest pair of RS-GNSS stations that satisfy the data screening strategy, with a horizontal distance of 3.83 km and height difference of 0.13 m. Figure 2a shows the time series of PWVs from these different sources, where RS PWV, GNSS PWV, and ERA5 PWV (line) are the post-processing results, and PG PWV, GC PWV, and FW PWV (dots) are the forecast results from the foundation models. The fluctuation of the ERA5 PWV is relatively small, which seems to be the result of the combination or smoothing of the RS PWV and GNSS PWV, which is consistent with the fact that the ERA5 data is the product of the assimilation of multiple observations. Among the forecast PWVs, the FW PWV is the closest to the ERA5 PWV, but lacks the temporal details of the ERA5, while the GC PWV is on the large side and the PG PWV is on the small side.

Figures 2b–2d show the PWVs obtained from the three post-processing techniques, comparing the results between the two. The RMSE between GNSS PWV and RS PWV is the largest for two reasons, the first is because the two stations are still spatially distant, and this imperfect co-located leads to part of the difference, and the other reason is that these are two completely different acquisition techniques, and they inherently have systematic biases. On the other hand, ERA5 itself comes from the data assimilation of multi-technology observations, which also includes RS and GNSS observations, so the RMSE between ERA5 and RS and RMSE between ERA5 and GNSS will be a bit smaller. Overall, there are no significant differences between the three post-processing PWVs, and the results obtained using any one of them as a reference are representative.

We also counted the variation of the RMSE of the three forecast PWVs with forecast time at this station when RS PWV, GNSS PWV, and ERA5 PWV are used as the reference, respectively (see Fig. 2e–2g). Overall, the RMSE of the FW PWV is smaller than that of the GC PWV, which is smaller than that of the PG PWV. The accuracy of the three PWVs used for reference is not homogeneous in time and space, so there are crossovers and overlaps of the RMSEs in the plots. And the periodic pattern may due to the data assimilation window of ERA5.In addition, since there is a 5-day delay in the ERA5 data, we focus on the results for the 5-day forecasting. From the figures, we can see that the RMSE for the forecast 5 days can reach about 2.5 mm at this site.

B. Accuracy assessment at stations level - multisite

Figure 3 depicts the RMSE of PWV forecasted by the Pangu-Weather, GraphCast and FengWu models for 69 pairs of RS-GNSS stations in the U.S. and neighboring regions, respectively, with RS PWV, GNSS PWV, and ERA5 PWV serving as references. The RMSE at these station locations for a 5-day forecast, namely at 20 steps. All model outcomes demonstrate smaller values in the high-elevation region of the east-central U.S. and larger RMSE values in the Gulf of Mexico and the eastern region of the U.S. The RMSEs of the Pangu-Weather, GraphCast, and FengWu models are presented

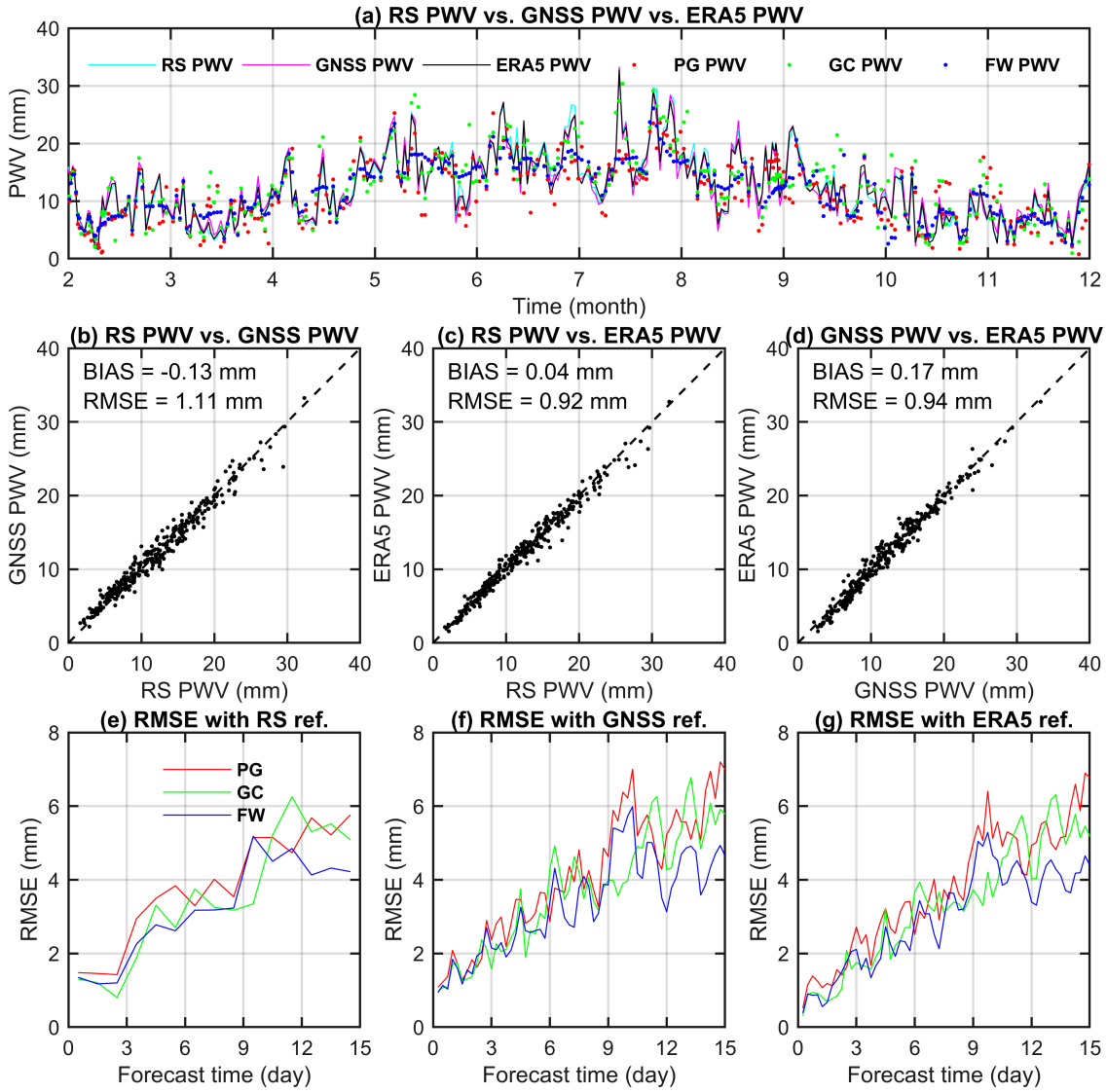


Fig. 2. PWV time series form different data sources (a) and their comparisons (b)–(g). (b)–(d) are the two-by-two comparisons between RS PWV, GNSS PWV and ERA5 PWV, respectively. (e)–(f) are the RMSEs of the PWVs of the three new schemes, PG, GC and FW, with respect to the forecast time, using these different PWVs as references, respectively.

in Figures 3a–3i. Additionally, the RS-referenced RMSE is generally greater than the GNSS-referenced results, which in turn are greater than the ERA5-referenced results. Among the three foundation models, the RMSE of Pangu-Weather is larger than that of GraphCast, and the RMSE of GraphCast is larger than that of FengWu.

Figures 3j–3l show RMSE mean value variation with forecast time for these stations. As RS observations are twice daily, only 12-hour epochs and multiples are valid when RS is the reference. Overall, FengWu outperforms GraphCast over Pangu-Weather regardless of reference (RS, GNSS, or ERA5). A half-day epoch term is evident with RS reference, likely due to different daytime and nighttime RS observation accuracies. Initial forecast accuracy (6/12 hours) is around 2 mm for RS and GNSS references and around 0.5 mm for ERA5 reference due to training on ERA5 data. RMSEs calculated with different references are around 4 mm for 5-day forecasts, and there's no significant difference after more than 5 days. This indicates the proposed method's real-time and forecasting accuracy remain consistent across data references. Given the lack of reliable in-situ measurements in many inland and oceanic regions, where existing data often lack quality or completeness, ERA5's integration of multi-source observations and physical constraints makes it a feasible and representative reference for global validation.

C. Accuracy Assessment at Global Grids Level

In this section, we show the results of the accuracy of the three models in forecasting PWV on the global grid points, Figure 4 shows the global distribution of BIAS and RMSE for forecasting 5 days, i.e., 20 steps, while Figure 6 shows the variation of BIAS and RMSE with latitude and height for forecasting PWV from 1 to 60 steps.

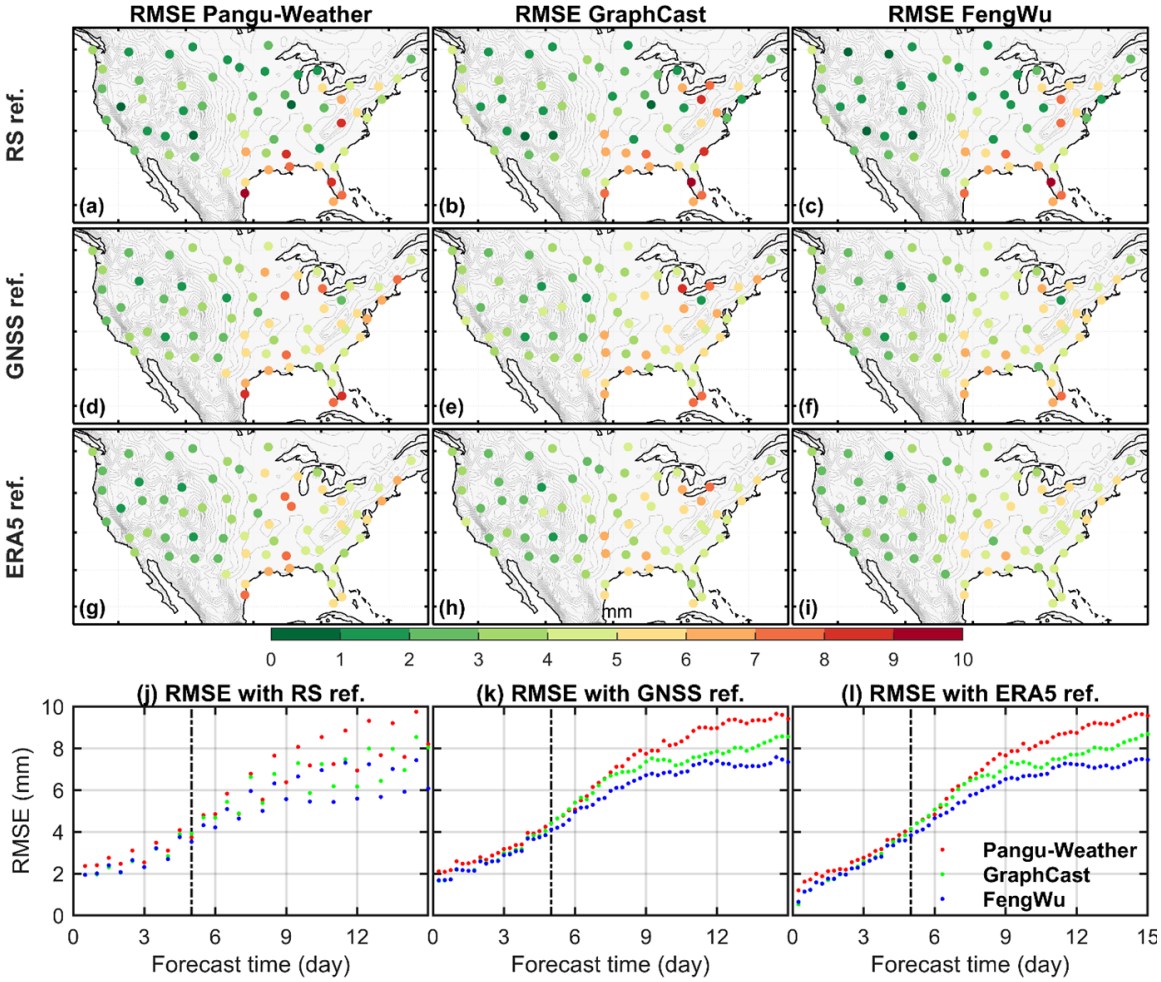


Fig. 3. The root means square errors (RMSEs) of PWV forecasts by Pangu-Weather (a, d, and g), GraphCast (b, e, and h), and FengWu (c, f, and i) in the US and surrounding regions for each station, and the variation of RMSEs with forecast time (j, k, and l), are compared to the results from RS, GNSS, and ERA5, respectively. (a)–(i) corresponds to the dotted line moments in (j)–(l), i.e., the result at the time of the 5-day forecast. Note that each point on the map represents a pair of RS-GNSS stations. RS is merely two observations per day, thereby data are only accessible when the forecast time is a multiple of 12 hours.

The left panels (4a, 4c and 4e) of Figure 4 respectively display the BIAS of Pangu-Weather, GraphCast and FengWu. Overall, on land, all the three models exhibit more positive biases with relatively smaller magnitudes. Additionally, Pangu-Weather has a considerable number of negative biases in the oceanic regions at low and middle latitudes, particularly in the Pacific and Indian Oceans. Meanwhile, GraphCast has a large negative bias in Indonesia and its surrounding areas. The positive and negative biases are approximately the same in the other regions. FengWu performs the best, with no significant large areas featuring positive or negative biases globally, and the bias is generally more uniform.

The right panels (4b, 4d and 4f) in Figure 4 respectively present the RMSEs of Pangu-Weather, GraphCast and FengWu. The results of the three models on land are consistent with one another, featuring smaller values. The values are negatively correlated with height, that is, the higher the altitude, the smaller the RMSE. Additionally, the values are significantly smaller in the high-latitude region than in the low-latitude region. The disparities in the results of these models mainly lie in the oceanic region, and the areas with higher RMSE values are also mainly located in the oceanic region. The locations where the larger values are concentrated are also basically the same. However, the coverage of these larger values varies among the three models, with Pangu-Weather being larger than GraphCast, and GraphCast being larger than FengWu. In the 5-day forecast, i.e., the real-time results, the average RMSE in the land region is only about 2.25 mm, while in the ocean region this value reaches 3.53 mm (see results in Figure 5). The ETOPO5 model is used here to classify grid points.

Figures 6a–6c depict the BIAS variation with latitude for PWV forecast results from 1 to 60 steps. Among the three models, Pangu-Weather performs the worst due to the sparsity of pressure levels, leading to negative bias (underestimation of PWV), especially in low and mid-latitude regions, and this underestimation grows with an increase in forecast steps. FengWu outperforms GraphCast as the results of FengWu show a latitude variation closer to the zero-mean value, while GraphCast has more negative bias at low latitudes. Addition-

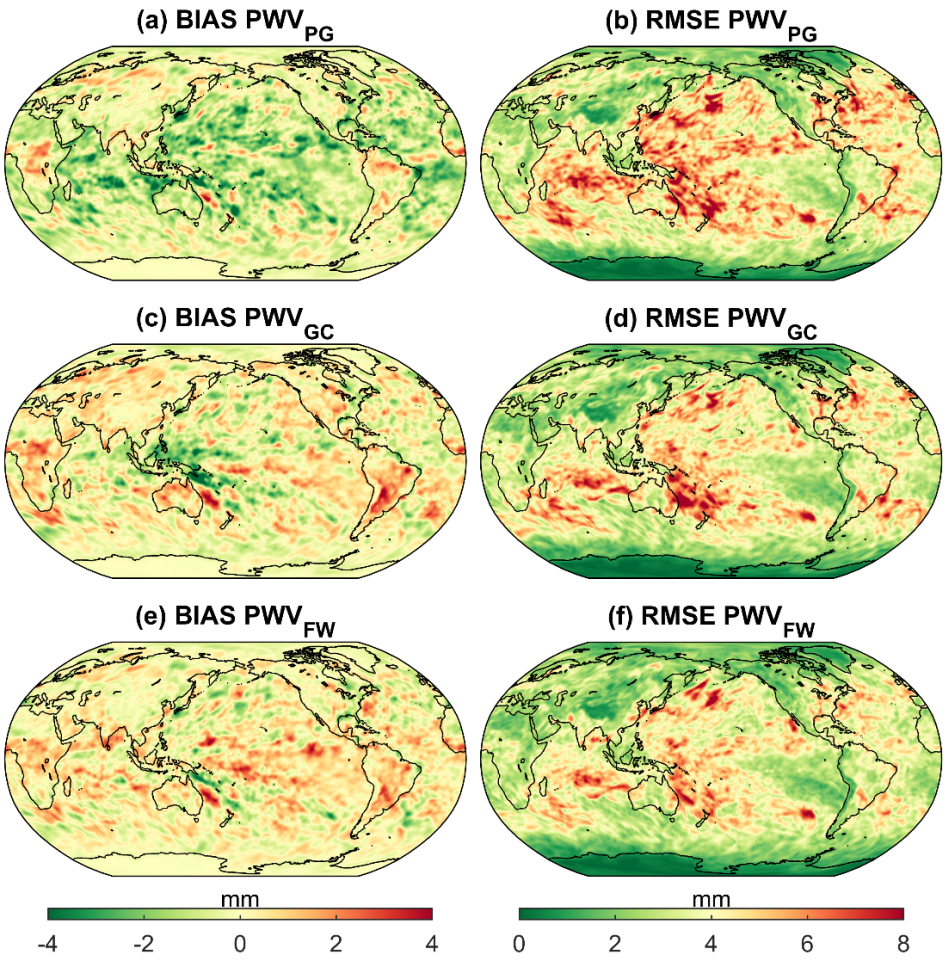


Fig. 4. Global distribution of mean bias (BIAS) and root mean square error (RMSE) of 5-day (120 hours, 20 steps inference) precipitable water vapor (PWV) forecast using the foundation models of Pangu-Weather (PG), GraphCast (GC) and FengWu (FW) with ERA5 PWV as references.

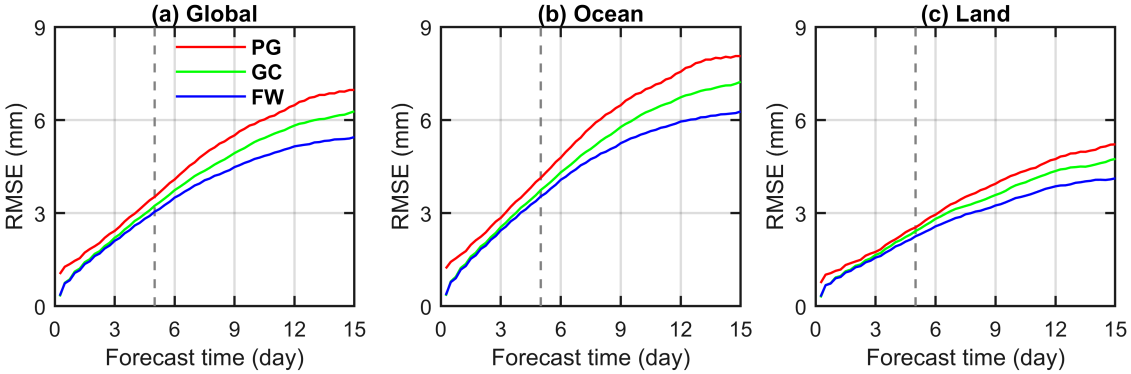


Fig. 5. The variations of root-mean-square errors (RMSEs) of PWV with forecast time of the three foundation model solutions, using ERA5 results as a reference. The left, middle, and right panels are global, ocean, and land results, respectively, with the dotted line being the forecast 5 days (i.e., real-time).

ally, the BIAS of the three models is generally positively correlated with the increase in forecast steps, i.e., moving with steps in the direction of their original bias (positive or negative).

Figures 6d–6f illustrate the RMSE variation with latitude for PWV forecasts results from 1 to 60 steps. The results from the three models exhibit a highly similar shape, all presenting an "M" pattern, meaning the RMSE is larger in middle and low latitudes and smaller in equatorial and polar regions. Additionally, the Antarctic region has a greater RMSE than the Arctic region due to the presence of Antarctica. Regarding RMSE values, FengWu is smaller than GraphCast and GraphCast is smaller than Pangu-Weather. The best-performing FengWu can achieve a value better than 4 mm if forecasts are limited

to 5 days. Moreover, there is almost no intersection among these lines, indicating that the accumulation of errors in the inference process of the foundation models is relatively homogeneous and isotropic.

Figures 6g–6i display the variation of BIAS with height for PWV forecast results from steps 1 to 60. All three models perform optimally between heights of 2.0–3.8 km, with BIAS close to zero. Above 3.8 km, the BIAS of all three models rapidly increases and after the height is over 3.8 km, FengWu and GraphCast tend to be more negatively biased, while Pangu-Weather shows no significant positive or negative bias. In the area less than 100 m, both Pangu-Weather and GraphCast have significant negative bias, followed by a jump toward zero. After the jump, in the range from 100 m to 2 km, Pangu-Weather shows negative bias, GraphCast shows positive bias, and FengWu has no significant positive or negative bias below 2 km.

Figures 6j–6l show the variation of RMSE with height for PWV forecast results from steps 1 to 60. All three models exhibit a consistent shape. At heights less than 3.8 km, the RMSE of all three models decreases with increasing height. In the region of 3.8–4.1 km, it suddenly increases. At elevations greater than 4.1 km, it becomes smaller with increasing height. These phenomena are attributed to the underestimation of forecasted PWV at high altitudes (as shown in the BIAS results in Figures 6g–6i), while also being influenced by the decreasing PWV values with increasing altitude. In terms of RMSE values, FengWu is smaller than GraphCast, which is smaller than Pangu-Weather. For forecasts limited to 5 days, the RMSEs of the three models are basically less than 4 mm at most heights.

IV. Discussion

In this study, we propose a new scheme of surface-domain PWV retrieval with high-resolution real-time and forecasting with global coverage using foundation models represented by Huawei Pangu-Weather, Google GraphCast, and Shanghai AI Lab FengWu. The accuracy of PWV forecasting in the new scheme is evaluated using the PWV obtained by radiosonde, GNSS and ERA5 as references. Site-level evaluations show that in the U.S. region, the forecasting method provides real-time PWVs with an accuracy of 4 mm, using either RS, GNSS, or ERA5 as the reference. In addition, evaluations at global grid sites show that 5-day forecasts, using ERA5 as a reference, are able to achieve RMSEs of less than 4 mm over most of the global region, and an average RMSE of 2.25 mm is achieved in the global land area, which is even better than the accuracy of most post-processed surface-domain measurements such as near-infrared (NIR) measurements. It is important to note that the new scheme is an important addition to the existing methodology and is not intended to replace the traditional surface-domain PWV measurements, which are the data source for the reanalyzed data, i.e., the initial field inputs that contribute to the proposed method.

The primary limitation to the accuracy of the new scheme is the poor performance over ocean regions at low and mid-latitudes, largely due to a lack of high-accuracy PWV observations in these areas. We anticipate that the incorporation of additional high-accurate oceanic observations, such as GNSS PWV data collected from maritime vessels, into future reanalysis datasets will enhance the accuracy of our scheme. Despite this, the global PWV accuracy of our scheme is overall satisfactory compared to current surface-domain measurements, and we hope that future AI advancements and the emergence of better foundation models will further enhance it.

V. Conclusions

This study establishes a surface-domain PWV forecasting framework powered by AI foundation models, enabling high-resolution, real-time global predictions with accuracy rivaling or exceeding traditional post-processed methods like near-infrared (NIR) over land areas. Validation against radiosonde, GNSS, and ERA5 data in the U.S. shows 4 mm RMSE for real-time forecasts, while global real-time maintains sub-4 mm RMSE across most terrestrial regions. The approach complements conventional observational networks by reducing latency and expanding spatial coverage, though oceanic predictions at low and mid-latitudes remain less reliable due to sparse reference data. Incorporating maritime GNSS PWV observations and refining reanalysis assimilation could mitigate these gaps, while advancements in AI weather modeling are likely to further improve precision. This work highlights the potential of AI-driven systems to transform PWV monitoring for operational meteorology and climate applications.

VI. Acknowledgments

This research processed data from more than 20 TB. The ERA5 data on pressure levels [31] and ERA5 data on single levels [32] are available at Copernicus Climate Change Service (C3S) Climate Data Store (CDS) (last accessed 8 Oct, 2023), which can be obtained from https://cds.climate.copernicus.eu/datasets/reanalysis-era5-pressure-levels?tab=overview and https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?tab=overview. The foundation model Pangu-Weather is available at https://github.com/198808xc/Pangu-Weather [38]. The foundation model GraphCast is available at https://github.com/google-deepmind/graphcast [39]. The foundation model FengWu is available at https://github.com/OpenEarthLab/FengWu [40]. The NGL GNSS tropospheric delay products are available at http://geodesy.unr.edu/gps_timeseries/trop/ [28]. The radiosonde data is available at https://www.ncei.noaa.gov/products/weather-balloon/integrated-global-radiosonde-archive [24]. The open-source AI models command-line management tool ai-models is available at https://github.com/ecmwf-lab/ai-models.

References

[1] Schneider, T., O'Gorman, P. A., & Levine, X. J. "Water vapor and the dynamics of climate changes," *Reviews of Geophysics*, vol. 48, 2010. doi: 10.1029/2009rg000302.

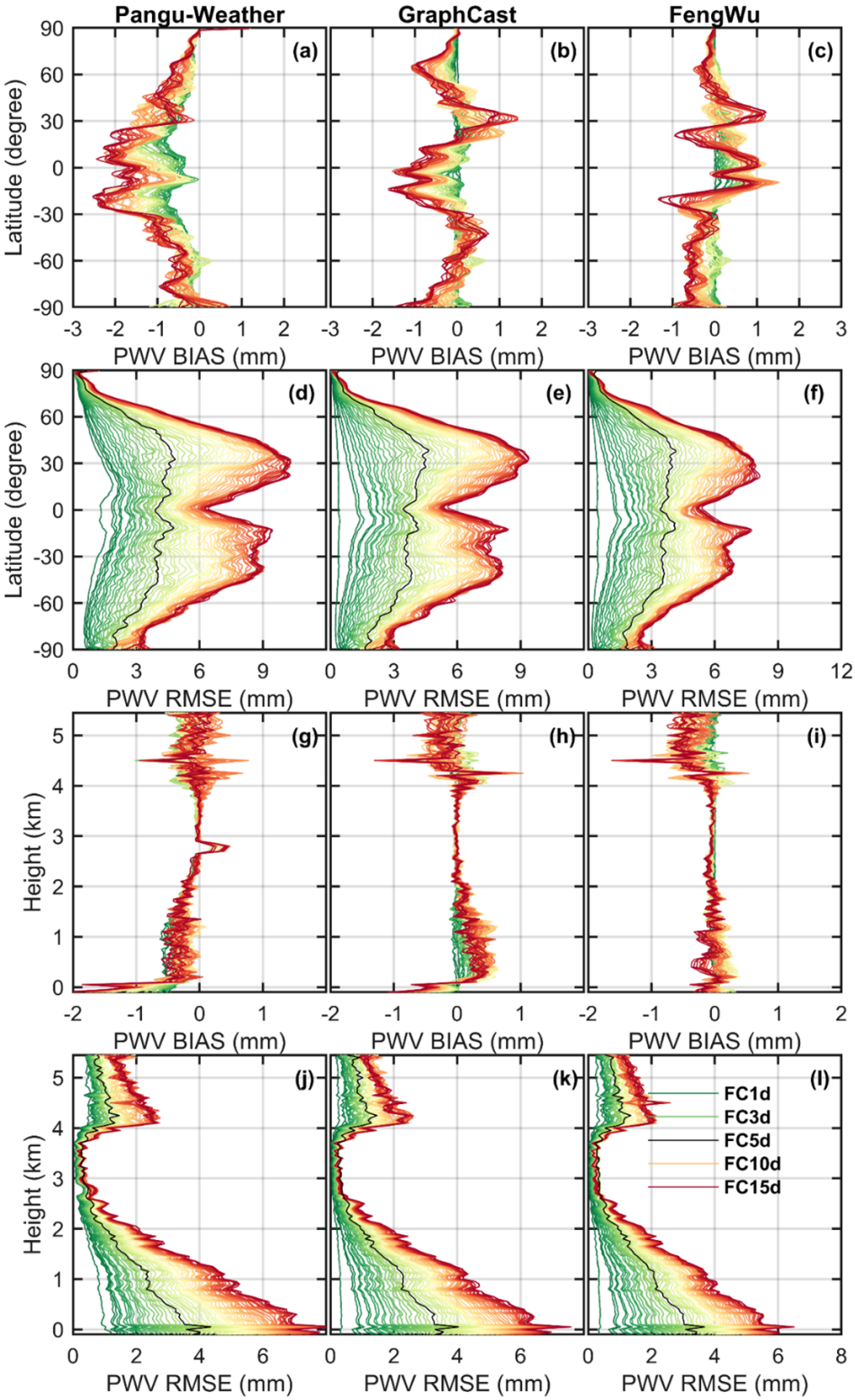


Fig. 6. Mean bias (BIAS) and root mean square error (RMSE) of forecasted precipitable water vapor (PWV) using the foundation models of Pangu-Weather (left panels), GraphCast (middle panels) and FengWu (right panels) at different latitudes (upper panels) and different heights (lower panels). Different colors in the figure represent different forecast times, ranging from 6 hours to 15 days (from green to red, a total of 60 steps, with a step length of 6 hours). The black line is the result of 5 days forecasting (the real-time results), i.e. 20 steps.

[2] Sherwood, S. C., Roca, R., Weckwerth, T. M., & Andronova, N. G. "Tropospheric water vapor, convection, and climate," *Reviews of Geophysics*, vol. 48, 2010. doi: 10.1029/2009rg000301.

[3] Jiang, N., Wu, Y., Li, S., Xu, Y., Wang, Y., & Xu, T. "First PWV retrieval using MERSI-LL onboard FY-3E and cross validation with co-platform occultation and ground GNSS," *Geophysical Research Letters*, vol. 51, no. 8, p. e2024GL108681, 2024. doi: 10.1029/2024GL108681.

[4] Adams, D. K., Fernandes, R. M. S., Holub, K. L., Gutman, S. I., Barbosa,

H. M. J., Machado, L. A. T., & Tanaka, L. M. S. "The Amazon dense GNSS meteorological network a new approach for examining water vapor and deep convection interactions in the tropics," *Bulletin of the American Meteorological Society*, vol. 96, no. 12, pp. 2151–2165, 2015. doi: 10.1175/bams-d-13-00171.1.

[5] Zheng, Y., Lu, C., Wu, Z., Liao, J., Zhang, Y., & Wang, Q. "Machine Learning-Based Model for Real-Time GNSS Precipitable Water Vapor Sensing," *Geophysical Research Letters*, vol. 49, no. 3, p. e2021GL096408, 2022. doi: 10.1029/2021GL096408.

[6] Li, Z., Muller, J. P., & Cross, P. "Comparison of precipitable water vapor derived from radiosonde, GPS, and Moderate-Resolution Imaging Spectroradiometer measurements," *Journal of Geophysical Research: Atmospheres*, vol. 108, no. D20, 2003. doi: 10.1029/2003JD003372.

[7] Liu, J., Sun, Z., Liang, H., Xu, X., & Wu, P. "Precipitable water vapor on the Tibetan Plateau estimated by GPS, water vapor radiometer, radiosonde, and numerical weather prediction analysis and its impact on the radiation budget," *Journal of Geophysical Research: Atmospheres*, vol. 110, no. D17, 2005. doi: 10.1029/2004JD005715.

[8] Zhang, Y., Cai, C., Chen, B., & Dai, W. "Consistency evaluation of precipitable water vapor derived from ERA5, ERA-Interim, GNSS, and radiosondes over China," *Radio Science*, vol. 54, no. 7, pp. 561–571, 2019. doi: 10.1029/2018RS006789.

[9] Zhao, T., Dai, A., & Wang, J. "Trends in tropospheric humidity from 1970 to 2008 over China from a homogenized radiosonde dataset," *Journal of Climate*, vol. 25, no. 13, pp. 4549–4567, 2012. doi: 10.1175/JCLI-D-11-00557.1.

[10] Zhou, C., Wang, J., Dai, A., & Thorne, P. W. "A new approach to homogenize global subdaily radiosonde temperature data from 1958 to 2018," *Journal of Climate*, vol. 34, no. 3, pp. 1163–1183, 2021. doi: 10.1175/JCLI-D-20-0352.1.

[11] Pérez-Ramírez, D., Whiteman, D. N., Smirnov, A., Lyamani, H., Holben, B. N., Pinker, R., ..., & Alados-Arboledas, L. "Evaluation of AERONET precipitable water vapor versus microwave radiometry, GPS, and radiosondes at ARM sites," *Journal of Geophysical Research: Atmospheres*, vol. 119, no. 15, pp. 9596–9613, 2014. doi: 10.1002/2014JD021730.

[12] Bai, J., Lou, Y., Zhang, W., Zhou, Y., Zhang, Z., & Shi, C. "Assessment and calibration of MODIS precipitable water vapor products based on GPS network over China," *Atmospheric Research*, vol. 254, p. 105504, 2021. doi: 10.1016/j.atmosres.2021.105504.

[13] Tan, J., Chen, B., Wang, W., Yu, W., & Dai, W. "Evaluating precipitable water vapor products from Fengyun-4A meteorological satellite using radiosonde, GNSS, and ERA5 Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2022. doi: 10.1109/TGRS.2022.3146018.

[14] Wang, J., Wu, Z., Semmling, M., Zus, F., Gerland, S., Ramatschi, M., ..., & Schuh, H. "Retrieving precipitable water vapor from shipborne multi-GNSS observations," *Geophysical Research Letters*, vol. 46, no. 9, pp. 5000–5008, 2019. doi: 10.1029/2019GL082136.

[15] Ding, J., Chen, W., Chen, J., Wang, J., Zhang, Y., Weng, D., ..., & Bai, L. "AI Foundation Models Facilitate Real-time Global GNSS Precipitable Water Vapor Retrieval with Sub-millimeter Accuracy," *Authorea Preprints*, 2024. doi: 10.22541/essoar.172081357.72805131/v1.

[16] Jedlovec, G. J. "Precipitable water estimation from high-resolution split window radiance measurements," *Journal of Applied Meteorology*, pp. 863–877, 1990. Retrieved from https://www.jstor.org/stable/26185675.

[17] Vaquero-Martínez, J., Antón, M., de Galisteo, J. P. O., Cachorro, V. E., Álvarez-Zapatero, P., Román, R., ... "Inter-comparison of integrated water vapor from satellite instruments using reference GPS data at the Iberian Peninsula," *Remote Sensing of Environment*, vol. 204, pp. 729–740, 2018. doi: 10.1016/j.rse.2017.09.028.

[18] Lee, Y., Han, D., Ahn, M.-H., Im, J., & Lee, S. J. "Retrieval of total precipitable water from Himawari-8 AHI data: A comparison of random forest, extreme gradient boosting, and deep neural network," *Remote Sensing*, vol. 11, no. 15, p. 1741, 2019. doi: 10.3390/rs11151741.

[19] Xu, J., & Liu, Z. "Enhanced all-weather precipitable water vapor retrieval from MODIS near-infrared bands using machine learning," *International Journal of Applied Earth Observation and Geoinformation*, vol. 114, p. 103050, 2022. doi: 10.1016/j.jag.2022.103050.

[20] Du, J., Kimball, J. S., & Jones, L. A. "Satellite microwave retrieval of total precipitable water vapor and surface air temperature over land from AMSR2," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2520–2531, 2014. doi: 10.1109/TGRS.2014.2361344.

[21] He, J., & Liu, Z. "Comparison of satellite-derived precipitable water vapor through near-infrared remote sensing channels," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 12, pp. 10252–10262, 2019. doi: 10.1109/TGRS.2019.2932847.

[22] Mateus, P., Mendes, V. B. & Plecha, S. M. "HGPT2: An ERA5-Based Global Model to Estimate Relative Humidity," *Remote Sensing*, vol. 13, no. 11, Art. no. 2179, 2021. doi: 10.3390/rs13112179.

[23] Ding, J., Mi, X., Chen, W., Chen, J., Wang, J., Zhang, Y., ..., & Tang, W. "Forecasting of Tropospheric Delay using AI Foundation Models in support of Microwave Remote Sensing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, p. 5803019, 2024. doi: 10.1109/TGRS.2024.3488727.

[24] Durre, I., Yin, X., Vose, R. S., Applequist, S., & Arnfield, J. "Enhancing the data coverage in the Integrated Global Radiosonde Archive," *Journal of Atmospheric and Oceanic Technology*, vol. 35, no. 9, pp. 1753–1770, 2018. doi: 10.1175/JTECH-D-17-0223.1.

[25] Zhang, W., Lou, Y., Huang, J., Zheng, F., Cao, Y., Liang, H., ..., & Liu, J. "Multiscale variations of precipitable water over China based on 1999–2015 ground-based GPS observations and evaluations of reanalysis products," *Journal of Climate*, vol. 31, no. 3, pp. 945–962, 2018. doi: 10.1175/JCLI-D-17-0419.1.

[26] Durre, I., Korzeniewski, B., & National Center for Atmospheric Research Staff (Eds.). "The Climate Data Guide: Integrated Global Radiosonde Archive (IGRA)," 2023. Retrieved from https://climatedataguide.ucar.edu/climate-data/integrated-global-radiosonde-archive-igra.

[27] Yuan, P., Blewitt, G., Kreemer, C., Hammond, W. C., Argus, D., Yin, X., ..., & Kutterer, H. "An enhanced integrated water vapour dataset from more than 10,000 global ground-based GPS stations in 2020," *Earth System Science Data*, vol. 15, pp. 723–743, 2023. doi: 10.5194/essd-15-723-2023.

[28] Blewitt, G., Hammond, W., & Kreemer, C. "Harnessing the GPS data explosion for interdisciplinary science," *Eos*, vol. 99, 2018. doi: 10.1029/2018EO104623.

[29] Ding, J., & Chen, J. "Assessment of empirical troposphere model GPT3 based on NGL's global troposphere products," *Sensors*, vol. 20, no. 13, p. 3631, 2020. doi: 10.3390/s20133631.

[30] Ding, J., Chen, J., Wang, J., & Zhang, Y. "Characteristic differences in tropospheric delay between Nevada Geodetic Laboratory products and NWM ray-tracing," *GPS Solutions*, vol. 27, no. 1, p. 47, 2023. doi: 10.1007/s10291-022-01385-2.

[31] Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., ..., & Thépaut, J. N. "ERA5 hourly data on pressure levels from 1979 to present," *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*, vol. 10, no. 10.24381, 2018. doi: 10.24381/cds.bd0915c6.

[32] Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., ..., & Thépaut, J. N. "ERA5 hourly data on single levels from 1979 to present," *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*, vol. 10, no. 10.24381, 2018. doi: 10.24381/cds.adbb2d47.

[33] Bolton, D. "The computation of equivalent potential temperature," *Monthly Weather Review*, vol. 108, no. 7, pp. 1046–1053, 1980. doi: 10.1175/1520-0493(1980)108¡1046:TCOEPT¿2.0.CO;2.

[34] Wong, M. S., Jin, X., Liu, Z., Nichol, J., & Chan, P. W. "Multi-sensors study of precipitable water vapour over mainland China," *International Journal of Climatology*, no. 10, pp. 3146–3159, 2015. doi: 10.1002/joc.4199.

[35] Rueger, J. M. "Refractive index formulae for radio waves," in *Proc. 22nd Int. CongrIntegr, Techn. Corrections Achieve Accurate Eng.*, Washington, DC, USA, 2002, p. 13.

[36] Rueger, J. M. "Refractive indices of light, infrared and radio waves in the atmosphere," School Surveying Spatial Inf. Syst., Univ. New South Wales, Sydney, NSW, Australia, Tech. Rep. UNISURV S-68, 2002.

[37] Ding, J., Chen, W., Chen, J., Wang, J., Zhang, Y., Bai, L., ..., & Weng, D. "Spatiotemporal inhomogeneity of accuracy degradation in AI weather forecast foundation models: A GNSS perspective," *International Journal of Applied Earth Observation and Geoinformation*, vol. 139, p. 104473, 2025. doi: 10.1016/j.jag.2025.104473.

[38] Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., & Tian, Q. "Accurate medium-range global weather forecasting with 3D neural networks," *Nature*, vol. 619, no. 7970, pp. 533–538, 2023. doi: 10.1038/s41586-023-06185-3.

[39] Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., ..., & Battaglia, P. "Learning skillful medium-range global weather forecasting," *Science*, p. eadi2336, 2023. doi: 10.1126/science.adi2336.

[40] Chen, K., Han, T., Gong, J., Bai, L., Ling, F., Luo, J. J., ..., & Ouyang, W. "FengWu: Pushing the Skillful Global Medium-range Weather Forecast beyond 10 Days Lead," *arXiv preprint arXiv:2304.02948*, 2023. doi: 10.48550/arXiv.2304.02948.

Junsheng Ding received the bachelor's degree from Chang'an University, Xi'an, China, in 2018, and the Ph.D. degree from Shanghai Astronomical Observatory, Chinese Academy of Sciences, Beijing, China, and the School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing, in 2023.

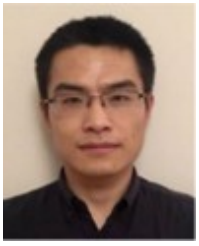
He is currently a Post-Doctoral Fellow with the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong. His research interests include Global Navigation Satellite System (GNSS) meteorology and artificial intelligence (AI) for geodesy (AI4G).

Wu Chen received the Ph.D. degree from Newcastle University, Newcastle upon Tyne, U.K., in 1992.

He is currently a Chair Professor with the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong. He has been actively working on Global Navigation Satellite System (GNSS)-related research for more than 30 years. His main research interests include GNSS positioning quality evaluation, system integrity, various GNSS applications, seamless positioning, and simultaneous localization and mapping (SLAM).

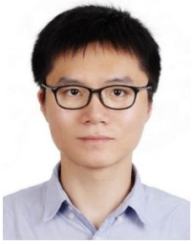
Junping Chen (Member, IEEE) received the Ph.D. degree in satellite geodesy from Tongji University, Shanghai, China, in 2007.

He is a Professor and the Head of the GNSS Data Analysis Group, Shanghai Astronomical Observatory (SHAO), Chinese Academy of Sciences, Shanghai. Since 2011, he has been supported by the "one hundred talents" programs of Chinese Academy of Sciences. His research interests include multi-Global Navigation Satellite System (GNSS) data analysis and GNSS augmentation systems.



Jungang Wang received the Ph.D. degree from the Technische Universität Berlin, Berlin, Germany, in 2021.

He is a Research Scientist at the Section of Space Geodetic Techniques, GeoForschungsZentrum (GFZ), Potsdam, Germany. His research interests are atmospheric effects in space geodesy, Global Navigation Satellite Systems, very long baseline interferometry, satellite laser ranging, and multitechnique integrated processing.



Yize Zhang received the Ph.D. degree from Tongji University, Shanghai, China, in 2017.

After his Ph.D. degree, he did his post-doctoral research at Tokyo University of Marine Science and Technology (TUMSAT), Tokyo, Japan. He is currently an Associate Professor at Shanghai Astronomical Observatory (SHAO), Chinese Academy of Sciences, Shanghai. His research mainly focuses on multi-Global Navigation Satellite System (GNSS) precise positioning and GNSS bias analysis.

Lei Bai was a Post-Doctoral Research Fellow at The University of Sydney, Sydney, NSW, Australia. He is currently a Research Scientist with Shanghai AI Laboratory, Shanghai, China. He is leading the OpenEarthLab. His research interests include machine learning, spatial–temporal learning, and their applications.

Dr. Bai was a recipient of the 2020 Google Ph.D. Fellowship, the 2020 UNSW Engineering Excellence Award, and the 2021 Dean's Award for Outstanding Ph.D. Theses.