



Research Paper

Radiosonde-informed machine learning correction of ERA5 integrated water vapor: spatiotemporal generalization and external GNSS evaluation

Lingke Wang^{*} , Hansjörg Kutterer[†] 

Geodetic Institute, Karlsruhe Institute of Technology, Englerstraße 7, 76131 Karlsruhe, Germany

ARTICLE INFO

Keywords:

Integrated water vapor
ERA5
Radiosonde
Machine learning
Bias correction
GNSS meteorology

ABSTRACT

Accurate integrated water vapor (IWV) estimates are essential for atmospheric monitoring, yet reanalyses can retain systematic discrepancies relative to in situ observations. This study develops a radiosonde-informed machine learning framework to correct ERA5-derived IWV over Europe and assess spatial and temporal generalization. Radiosonde profiles from the University of Wyoming and the Integrated Global Radiosonde Archive were collocated with ERA5 and VMF3 products at 116 stations from November 2021 to November 2025. CatBoost used spatiotemporal, geographical, meteorological, atmospheric structural, and climate-zone predictors to estimate ERA5 IWV differences. Spatial generalization was assessed by strict station holdout and temporal generalization by an independent one-year test. The station holdout achieved RMSE values of 1.33 and 2.40 mm and R^2 values of 0.94 and 0.92 in the cold and warm seasons, respectively. In the temporal test, RMSE decreased to 1.11 and 2.06 mm, with R^2 values of 0.96 and 0.93 and near-zero mean biases. Relative to ERA5, RMSE was reduced by 51.5% and 40.1% in the two seasons. The model was then transferred to external GNSS station coordinates to generate radiosonde-informed, spatially collocated IWV estimates. Across 47,418 common observations from 95 valid radiosonde-GNSS pairs, the collocated product reduced RMSE from 3.15 mm for nearby radiosonde observations to 2.25 mm. These results demonstrate robust radiosonde-based correction and spatiotemporal generalization, while highlighting the value of transferring radiosonde observational constraints to locations without collocated sounding measurements.

1. Introduction

Water vapor is widely recognized as the most abundant and one of the most influential greenhouse constituents in the Earth's atmosphere (Bojinski et al., 2014; Sherwood et al., 2010). It plays a fundamental role in regulating atmospheric circulation, modulating energy transport, driving the global hydrological cycle, and influencing climate variability at both regional and global scales (Liou et al., 2001; Li et al., 2023). In addition, atmospheric water vapor is a key parameter in a variety of Earth observation applications, including atmospheric correction of optical satellite imagery and geodetic data processing, where its variability can introduce significant uncertainties if not properly accounted for (Chen et al., 2019; Dessler et al., 2008; Gimeno et al., 2021; Oikonomou et al., 2023; Wu et al., 2025). The spatial and temporal distribution of atmospheric water vapor exhibits distinct heterogeneity due to its strong dependence on dynamic meteorological processes (Ccoica-López et al., 2019; Wang et al., 2025). Integrated water vapor (IWV) represents the vertically integrated amount of water vapor within

an atmospheric column and provides a convenient metric for quantifying atmospheric moisture content (Wang et al., 2025; Xu and Liu, 2022). Given its high variability and critical role in Earth system processes, there is a growing demand for IWV products with both high accuracy and enough spatiotemporal resolution (Held and Soden, 2006; Trenberth, 2011).

IWV can be retrieved from various ground-based, satellite, and reanalysis products. Ground-based techniques include radiosonde soundings (McCarthy et al., 2009), Global Navigation Satellite System (GNSS) observations (Bevis et al., 1992; Marut et al., 2022), sun photometers, and microwave radiometers (MWR) (Fragkos et al., 2019; Wang et al., 2023). Radiosondes are widely regarded as a reference standard for water vapor validation because of their direct vertical profiling capability and well-characterized measurement accuracy (Wang and Kutterer, 2025), but their global network is spatially sparse and typically limited to twice daily observations, restricting the representation of high-frequency IWV variability (Durre et al., 2006). Sun photometers and MWRs are also limited by point-based deployment and

^{*} Corresponding author.

E-mail address: lingke.wang@kit.edu (L. Wang).

<https://doi.org/10.1016/j.atmosres.2026.109379>

Received 13 August 2026; Received in revised form 25 September 2026; Accepted 1 October 2026

Available online 6 October 2026

0169-8095/© 2026 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Open access funding enabled and organized by Projekt DEAL.

restricted coverage, while MWR measurements can be affected by rain contamination (Hocke et al., 2011; Wang et al., 2023). In contrast, GNSS-based IWV retrieval provides all-weather capability, high temporal resolution, and long-term stability (Gueroval et al., 2016), although GNSS receivers do not directly measure IWV. Instead, IWV is derived from the zenith total delay (ZTD) estimated by geodetic processing software, such as Bernese GNSS Software (Dach et al., 2015), by isolating the zenith wet delay (ZWD) and converting it to IWV using the weighted mean temperature (Bevis et al., 1992; Böhm and Schuh, 2013). This processing chain can introduce uncertainties related to processing strategies, weighted mean temperature modeling, and multipath effects (Lu et al., 2024; Ning et al., 2016). Spaceborne remote sensing further extends spatial coverage, although cloud contamination and liquid water can introduce systematic dry biases in satellite-derived IWV (Bennouna et al., 2013; Ma et al., 2023; Xia et al., 2025; Wang et al., 2024). Atmospheric reanalysis products, particularly the fifth-generation European Centre for Medium-Range Weather Forecasts (ECMWF) atmospheric reanalysis (ERA5), provide globally continuous hourly IWV fields (Senyuzi et al., 2020), but remain affected by model parameterization, data assimilation smoothing, and limited sensitivity to localized atmospheric processes (Huang et al., 2022). Additional products, such as the Vienna Mapping Functions 3 (VMF3), provide supplementary information on tropospheric water vapor structure but are not direct IWV observables (Landskron and Böhm, 2017). Consequently, the complementary strengths and limitations of these heterogeneous sources motivate multi-source data integration for improved IWV characterization.

In light of these characteristics, ERA5 has become one of the most widely used datasets for IWV estimation owing to its global coverage, high temporal continuity, and physically consistent representation of atmospheric states. Improving ERA5-derived IWV products therefore offers a cost-effective alternative to enhancing observational infrastructure through additional hardware deployment and maintenance. Although ERA5 assimilates radiosonde observations, ERA5-derived IWV still exhibits discrepancies relative to in situ measurements, indicating that further improvements in water vapor estimation are possible (Hersbach et al., 2020; Huang et al., 2022; Li et al., 2024). As direct measurements of vertical atmospheric thermodynamic profiles, radiosonde observations are widely regarded as one of the reliable references for tropospheric water vapor validation (McCarthy et al., 2009; Wang and Kutterer, 2025; Wang et al., 2025). Motivated by this, the present study combines ERA5 with radiosonde observations to characterize and predict their IWV discrepancies. By learning the long-term relationship between ERA5 estimates and radiosonde observations, the proposed method identifies reproducible discrepancy patterns for predictive correction. The corrected IWV fields retain the spatial completeness and temporal continuity of ERA5 while achieving improved consistency with radiosonde observations.

In this study, data-driven approaches, particularly machine learning methods, are employed as effective tools for integrating multi-source atmospheric observations and products. Recent advances in machine learning have substantially promoted the development of IWV modeling frameworks by exploiting nonlinear relationships across atmospheric, geodetic, and Earth observation variables. These methods have demonstrated strong potential for enhancing the spatiotemporal characterization of tropospheric water vapor. For instance, Wang et al. (2025a) used deep learning to fuse multiple satellite-derived IWV products, while Ma et al. (2023) combined ERA5 and MODIS PWV with machine learning-based spatial downscaling. Du et al. (2024) developed a stacking model for PWV retrieval from Himawari-8 observations, and Li et al. (2025) applied XGBoost for post-processing correction of Fengyun-3D PWV products. Shi et al. (2025) further combined GNSS and radio occultation observations for machine learning-based PWV vertical adjustment. More recent studies have extended these developments to GNSS tropospheric tomography and data assimilation, hybrid ERA5-GNSS spatiotemporal ZTD forecasting, and improved weighted

mean temperature modeling using radiosonde and GNSS radio-occultation observations (Tayfehrostami et al., 2026; Ma et al., 2026; Tayfehrostami and Amerian, 2026). Despite these advances, a major limitation of existing data-driven IWV approaches lies in their strong dependence on GNSS-derived products as training targets, which may introduce and propagate uncertainties associated with GNSS retrieval strategies and unstable station performance. In environments affected by strong multipath, signal reflections can further contaminate GNSS observations, introducing errors in carrier phase and code measurements that propagate into ZWD estimation and ultimately affect IWV accuracy (Lu et al., 2024). In addition, radiosonde observations have rarely been incorporated directly into model development, despite serving as a direct and physically consistent observational reference for tropospheric water vapor. As a result, the rich vertical structure information contained in radiosonde profiles remains largely underexploited.

To address these limitations, this study proposes a radiosonde-informed machine learning framework for the predictive correction of ERA5-derived IWV. The framework integrates heterogeneous atmospheric information, including ERA5 reanalysis products, radiosonde observations, VMF3-derived IWV, and spatiotemporal features such as location and time. The learning target is defined as the discrepancies between ERA5-derived IWV and radiosonde observations, enabling the model to capture nonlinear relationships among reanalysis fields, geodetic atmospheric parameters, and in-situ measurements. A Categorical Boosting (CatBoost) model based on gradient boosted decision trees was selected after comparison with ridge regression, multilayer perceptron (MLP), and XGBoost using the same training and validation datasets. The boosting-based models outperformed the linear and neural network baselines, while CatBoost provided the comparable performance. Its ordered boosting strategy and native handling of categorical variables also make it well suited to the heterogeneous atmospheric, geodetic, and spatiotemporal predictors used in this study (Prokhorenkova et al., 2017).

Once established, the radiosonde-informed correction model can be spatially transferred from the radiosonde observation stations to external GNSS station locations. Because no GNSS-derived IWV observations are used during model training, feature scaling, hyperparameter optimization, or model selection, the resulting model-corrected IWV provides a radiosonde-informed comparison product for external GNSS-based evaluation. This offers a potential advantage over conventional GNSS-radiosonde validation, in which GNSS-derived IWV is generally compared with observations from a nearby radiosonde station. Such comparisons are only approximately collocated and may be affected by horizontal separation, elevation differences, and local atmospheric heterogeneity between the two sites. By applying the learned correction relationship directly at the GNSS coordinates, the proposed framework generates a spatially collocated model-corrected IWV product at the GNSS station locations. This creates a new opportunity to evaluate GNSS-derived IWV using a radiosonde-informed product that is collocated with the GNSS observations, thereby reducing, although not completely eliminating, the spatial representativeness differences associated with nearby-station comparisons.

The main contributions of this study are twofold. First, a radiosonde-informed machine learning framework is developed to predict the discrepancy between ERA5-derived IWV and radiosonde observations and thereby generate corrected IWV products. Second, the trained model is transferred to external GNSS station locations, where the model-derived IWV is evaluated against GNSS-derived IWV and compared with nearby radiosonde observations. This allows the benefit of spatial collocation for GNSS-IWV evaluation to be assessed directly.

The remainder of this paper is organized as follows. Section 2 introduces the radiosonde, ERA5, VMF3, and GNSS datasets, together with the preprocessing and station-pairing procedures. Section 3 presents the machine learning framework, the spatial and temporal evaluation strategies, the transfer of the trained model to GNSS locations. Section 4 evaluates model performance and assesses the agreement of the spatially

collocated model-corrected IWV with GNSS-derived IWV from an external observing system. Section 5 summarizes the main findings and conclusions.

2. Study region and data

2.1. Study region

The study focuses primarily on Europe, with additional radiosonde stations in adjacent areas of North Africa and western Asia included to mitigate boundary effects and preserve the continuity of large-scale atmospheric patterns near the study-domain edges. A four-year continuous observation period was used to provide sufficient samples for representing seasonal and interannual variability in tropospheric water vapor. Europe is particularly suitable for IWV analysis because of its dense and well-maintained radiosonde network and pronounced seasonal variability, with clear contrasts between warm-season and cold-season conditions (McCarthy et al., 2009; Wang et al., 2025a; Wang and Kutterer, 2025; Wang et al., 2025). Because IWV is strongly influenced by regional temperature and precipitation regimes, the Köppen-Geiger climate classification was incorporated to characterize climatic variability. Based on Beck et al. (2018), the original 30-class Köppen-Geiger dataset was used to assign each radiosonde station to a climate class and subsequently aggregated into three dominant hydroclimatic regimes: temperate (Köppen-Geiger classes Csa, Csb, Cfa, Cfb, and Cfc), arid (BWk and BSk), and continental (Dfa, Dfb, and Dfc), comprising 46, 11, and 59 radiosonde stations, respectively. These regimes represent contrasting temperature and precipitation conditions and provide a physically consistent basis for evaluating the radiosonde-informed correction framework across different hydroclimatic environments. Fig. 1 shows the spatial distribution of the radiosonde stations and their associated climate regimes.

2.2. Data used and pre-processing

This study integrates radiosonde observations, ERA5 reanalysis products, and VMF3 tropospheric products to construct a machine learning framework for predicting IWV discrepancies. Radiosonde-derived IWV (IWV_{RS}) was used to construct the target variable, namely the IWV discrepancies between IWV_{RS} and ERA5 estimates (IWV_{ERA5}). Fig. 2 illustrates the overall data processing workflow adopted in this study.

2.2.1. Radiosonde observation

Radiosonde observations were obtained from UoW and IGRA, with their spatial distribution shown in Fig. 1 (Durre et al., 2016). UoW stations are denoted by triangular markers, while IGRA stations are represented by circular markers. Radiosonde observations are typically available twice daily at most stations (0000 and 1200 UTC), while a subset of stations provides observations four times per day (0000, 0600, 1200, and 1800 UTC). The study period spans from November 2021 to November 2025. Following quality control, the retained 62 UoW and 54 IGRA records represented 116 distinct physical stations, with no physical stations shared between the two archives in the final machine learning dataset, as verified using World Meteorological Organization (WMO) station identifiers. Duplicate station-epoch records were removed prior to model development. Duplicates were identified based on physical station and launch epoch, resulting in the removal of 94 excess records and retention of 202,748 unique observations. For IGRA, profiles were required to contain at least eight moisture-valid levels, reach 300 hPa or higher in the atmosphere, and contain no pressure gap exceeding 250 hPa. Physical plausibility checks were also applied to pressure, temperature, dew point depression, vapor pressure, and derived IWV, with valid profile IWV restricted to 0.05–80 mm. Of 130,762 audited IGRA profiles, 124,708 (95.4%) passed these criteria. For UoW, the provider-derived precipitable water reported for the entire sounding was used directly. Following common practice, radiosonde profiles were referenced to their launch-station coordinates, without

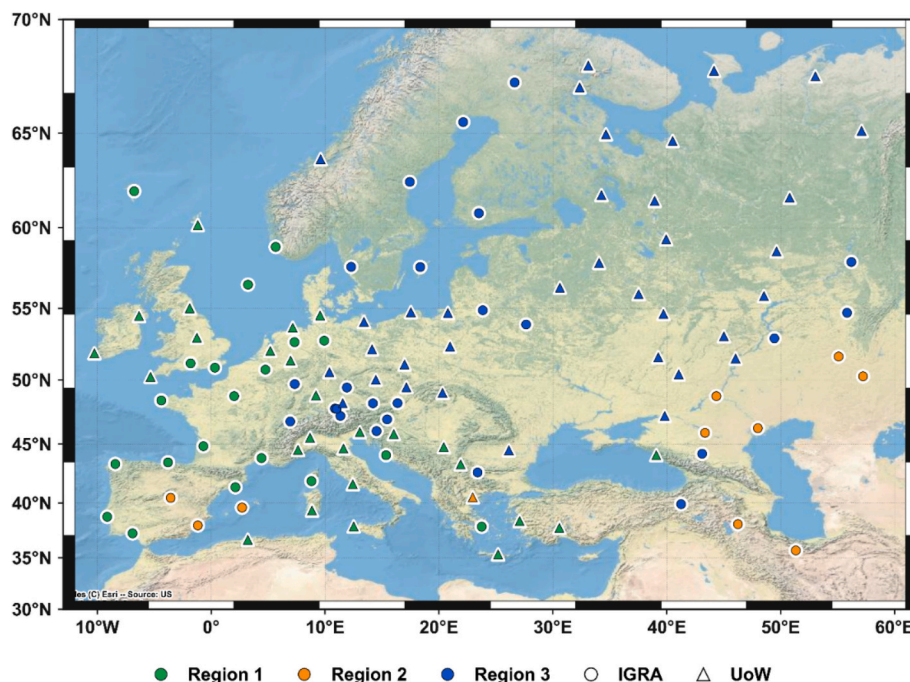


Fig. 1. Spatial distribution of the radiosonde stations obtained from the University of Wyoming (UoW) upper air sounding database and the Integrated Global Radiosonde Archive version 2 (IGRA v2). The background colors represent the three aggregated Köppen-Geiger climate regions considered in this study. Region 1 (green) represents temperate climates, mainly distributed across western Europe; Region 2 (orange) represents arid climates, primarily located in southern Europe and parts of western Asia; and Region 3 (blue) represents continental climates across central, northern, and eastern Europe. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

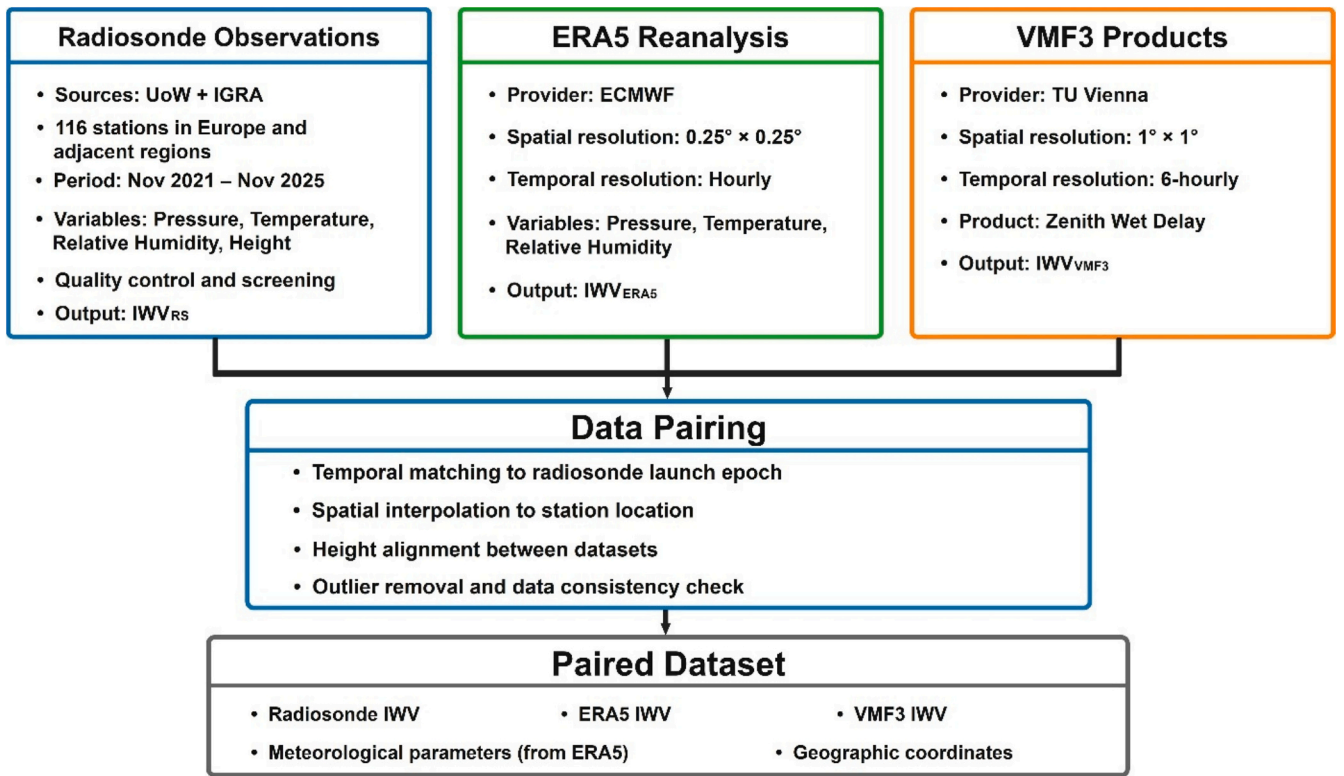


Fig. 2. Workflow of data collection, processing, and pairing adopted in this study.

explicit correction for horizontal balloon drift during ascent (Seidel et al., 2011).

For each sounding profile, atmospheric pressure, temperature, height, and relative humidity were extracted. The calculation of Radiosonde IWV from the radiosonde profiles is outlined as follows (Namaoui et al., 2021):

$$IWV = \int_{H_s}^{H_{top}} \rho_w dh = \sum_{i=1}^n \rho_{d,i} r_i \Delta h_i \quad (1)$$

where ρ_w is the water vapor density, and n is the total number of profile layers. $\rho_{d,i}$, r_i and Δh_i refer to the dry air density, the mixing ratio, and the height difference of each layer. For each layer, the parameters are expressed as:

$$\rho_d = \frac{M_d}{R} \frac{P}{T}, r = 0.622 \frac{e}{P} \quad (2)$$

with the molar mass of dry air $M_d = 0.0289644 \text{ kg mol}^{-1}$ and the universal gas constant for air $R = 8.31432 \text{ J mol}^{-1} \text{ K}^{-1}$. In addition, T is the air temperature (in K), P is the atmospheric pressure (in Pa), and e (in Pa) is the partial water vapor pressure. The detailed calculation methodology can be found in Namaoui et al. (2021).

2.2.2. ERA5 reanalysis

The ERA5 reanalysis dataset produced by ECMWF was used to provide gridded atmospheric information. ERA5 combines a global numerical weather model with a comprehensive data assimilation system to generate physically consistent atmospheric fields with high temporal and spatial resolution. ERA5 combines global observations with a numerical weather prediction model and an advanced data assimilation system to provide physically consistent reconstructions of the atmospheric state. ERA5 delivers hourly estimates of pressure, temperature, geopotential, and relative humidity on a global grid with a spatial resolution of 0.25° . ERA5 pressure-level data were obtained at 37 standard levels from 1000 to 1 hPa. For consistency with the radiosonde IWV

calculation, the lower integration boundary was adjusted to the same radiosonde station height reference, while all valid ERA5 levels above the station were retained up to 1 hPa. Based on these pressure-level fields, the ERA5-derived ZWD (ZWD_{ERA5}) and IWV_{ERA5} can be integrated as follows:

$$ZWD_{ERA5} = 10^{-6} \sum_{i=i_0}^{36} \frac{1}{2} \left[k_2 \left(\frac{e_i}{T_i} + \frac{e_{i+1}}{T_{i+1}} \right) + k_3 \left(\frac{e_i}{T_i^2} + \frac{e_{i+1}}{T_{i+1}^2} \right) \right] d_{i,i+1} \quad (3)$$

$$IWV_{ERA5} = \int \frac{q}{g} dP \approx \sum_{i=i_0}^{36} \frac{q_i}{g_i} \Delta P_i \quad (4)$$

where i_0 is the initial layer which corresponds to the height of radiosonde station. T_i are the pressure and temperature at i -th layer pressure level obtained from ERA5 directly. e_i is the water vapor pressure at i -th layer pressure level, q_i and g_i represent the mean specific humidity ($\text{kg} \cdot \text{kg}^{-1}$) and the mean local gravitational acceleration ($\text{m} \cdot \text{s}^{-2}$) at i -th atmospheric layer, respectively. ΔP_i denotes the pressure difference (Pa) between the $(i+1)$ -th and the i -th layers. To ensure consistency between ERA5 and radiosonde stations, height alignment is initially performed. In this study, ZWD_{ERA5} and IWV_{ERA5} are computed relative to the height of the radiosonde station. If the radiosonde station's elevation is lower than the surface layer height of ERA5, the surface height is used as the initial reference, and the i_0 in Eqs. (3) and (4) is determined accordingly. Conversely, if the radiosonde station's height exceeds the ERA5 surface height, the radiosonde station's height is selected as the initial reference. Due to space constraints, the detailed procedures for calculating IWV_{ERA5} are outlined in Jiang et al. (2023), with the corresponding interpolation methods based on those described by Jade and Vijayan (2008).

2.2.3. VMF3 products

To incorporate additional information describing the tropospheric water vapor state, products from VMF3 model were employed. VMF3 is a widely used empirical model provided by the Vienna University of

Technology (TU Vienna). It offers tropospheric delay and corresponding mapping function data for the Earth's surface grid, with a spatial resolution of 1° and a temporal resolution of 6 h (Landskron and Böhm, 2017). As described by Kouba (2009), estimating the ZWD at a target location using VMF3 (ZWD_{VMF3}) involves an initial linear temporal interpolation across epochs for the four surrounding grid points. This step in this research adjusts the temporal resolution to 1 h, ensuring consistency with the ERA5 products. Since VMF3 provides surface products, the elevation of the target location may differ from that of the surrounding grid points. Therefore, a vertical adjustment must be applied after time alignment. Finally, the ZWD_{VMF3} at the target location can be derived using the procedures and equations outlined in Kouba (2007), applying bilinear interpolation for spatial adjustments. Subsequently, IWV_{VMF3} can be determined using Eqs. (5) and (6).

$$IWV_{VMF3} = Q \bullet ZWD_{VMF3} \quad (5)$$

$$Q = \frac{10^6}{R_v \bullet \left[k'_2 + \frac{k_3}{T_m} \right]} \quad (6)$$

where Q is the conversion factor and the parameter $R_v=461.522 \text{ J}\cdot\text{kg}^{-1}\cdot\text{K}^{-1}$ is the specific gas constant for water vapor (Kestin et al., 1984). The parameters $k'_2=22.1 \text{ K}\cdot\text{hPa}^{-1}$ and $k_3=373,900 \text{ K}^2\cdot\text{hPa}^{-1}$ are atmospheric refractivity constants (Bevis et al., 1994). In Eq. (6), T_m is the weighted mean temperature, the methodology of obtaining T_m is outlined by Bevis et al. (1992).

2.2.4. Data pairing criteria

To ensure consistency among the radiosonde, ERA5, and VMF3 datasets, a strict data pairing strategy was utilized. First, radiosonde observations were used as the temporal reference, and ERA5 and VMF3 records were matched to the nearest radiosonde launch epoch. Spatial matching was performed using the geographic coordinates of each radiosonde station. ERA5 variables were interpolated from neighboring grid cells, whereas VMF3 parameters were obtained through spatial interpolation of the surrounding VMF3 grid nodes. Samples were retained only when all required variables were simultaneously available. To further improve data quality, extreme outliers exceeding predefined physical thresholds were excluded. The resulting paired dataset forms the basis for subsequent feature engineering, machine learning model development, and subsequent evaluation. In addition to the IWV products, vertical atmospheric profile information and surface meteorological parameters derived from ERA5 reanalysis were retained as candidate input features. Furthermore, geographical variables, including station latitude, longitude, and height, were incorporated as input features to account for spatial variability within the machine learning framework.

2.2.5. GNSS and paired radiosonde stations

To evaluate the transferability of the radiosonde-informed machine learning framework using an external GNSS dataset, GNSS stations were selected within the study region according to the joint availability of GNSS tropospheric products, auxiliary atmospheric data, and nearby radiosonde observations. Candidate GNSS stations were selected based on the availability of ZTD products from the Nevada Geodetic Laboratory (NGL) during the GNSS evaluation period from 1 December 2024 to 30 November 2025 (Blewitt et al., 2018). In addition, the stations had to be located within the study domain, have the atmospheric variables required for model inference, and remain within the elevation range represented in the model training dataset. Each radiosonde station was paired with a nearby GNSS station subject to spatial distance and elevation difference constraints. To avoid overrepresenting individual radiosonde sites through multiple nearby GNSS stations, one primary GNSS station was selected for each valid radiosonde station. The final validation network consists of 98 radiosonde-GNSS pairs involving 96

unique GNSS stations. Two GNSS stations were each paired with two radiosonde stations. As shown in Fig. 3, the selected stations provide broad spatial coverage across Europe and adjacent regions. The network is relatively dense in western and central Europe. The distribution reflects the combined requirements of NGL ZTD availability, radiosonde proximity, temporal coverage, and consistency with the model application domain rather than uniform spatial sampling.

3. Methodology

This section presents the methodology of the proposed framework for predicting and correcting discrepancies between IWV_{ERA5} and IWV_{RS} . The framework is designed to generate improved IWV products for high-accuracy atmospheric applications. The overall workflow is first presented, followed by the construction of spatiotemporal and atmospheric features, the training of the CatBoost-based prediction model, and the validation strategy adopted for performance assessment. The trained model is then transferred to external GNSS station locations to generate spatially collocated corrected IWV. Finally, these products are evaluated against GNSS-IWV and compared with nearby radiosonde observations to assess their suitability for external GNSS-based IWV evaluation.

3.1. Framework of radiosonde-informed IWV difference prediction and correction

Although radiosonde observations are assimilated into the ERA5 reanalysis system, differences in ERA5-derived IWV relative to radiosonde observations still persist. These discrepancies may arise from multiple sources, including model parameterization errors, data assimilation uncertainties, spatial representativeness differences between gridded reanalysis fields and point-based observations, and limitations in describing local atmospheric variability (Hersbach et al., 2020; Huang et al., 2022; Li et al., 2024). Previous studies have reported that such discrepancies vary across regions, seasons, and atmospheric conditions, indicating that further correction remains beneficial for applications requiring accurate large-scale atmospheric water vapor information. A data-driven framework is therefore needed to relate ERA5 and radiosonde IWV discrepancies to their spatiotemporal, meteorological, and atmospheric controls.

The overall workflow of the research is illustrated in Fig. 4. The framework consists of five main stages: (1) data collection and collocation, (2) feature construction, (3) discrepancy prediction model training, (4) IWV correction, and (5) spatial transfer to external GNSS station locations and evaluation against GNSS-derived IWV. First, radiosonde observations from the UoW and IGRA databases are collocated with ERA5 and VMF3 according to the strict temporal and spatial matching strategy described in Section 2.2. For each matched sample, the IWV_{DIFF} is defined as the difference between radiosonde-derived IWV and ERA5-derived IWV:

$$IWV_{DIFF} = IWV_{RS} - IWV_{ERA5} \quad (7)$$

Subsequently, a comprehensive set of input features is constructed to characterize the factors influencing IWV_{DIFF} . These variables include temporal information, geographical characteristics, surface meteorological parameters, atmospheric vertical structural indicators, and VMF3-derived tropospheric delay products. The resulting feature set serves as the input to the machine learning model, while the corresponding IWV_{DIFF} values are used as prediction targets. The CatBoost algorithm is then employed to establish the nonlinear relationship between the input variables and the observed IWV difference. Once trained, the model can estimate IWV_{DIFF} under given conditions. The predicted correction is subsequently applied to ERA5 IWV products to obtain corrected IWV estimates:

$$IWV_{corr} = IWV_{ERA5} + \widehat{IWV}_{DIFF} \quad (8)$$

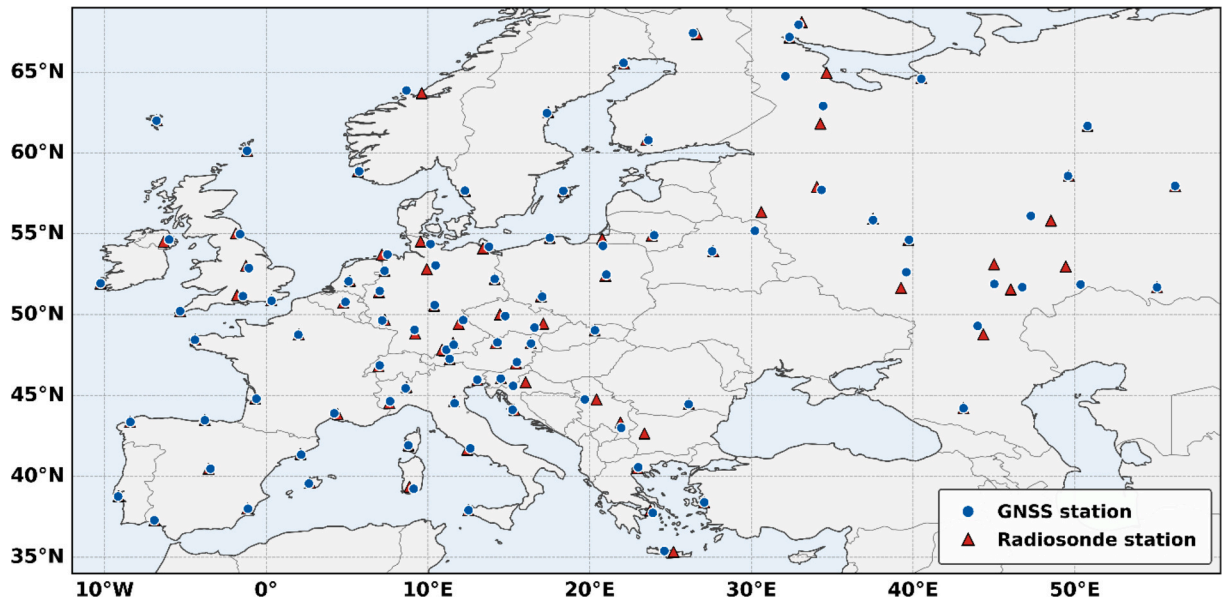


Fig. 3. Spatial distribution of the selected GNSS stations and their paired radiosonde stations used for the external GNSS-based IWV evaluation. The validation network comprises 98 radiosonde-GNSS pairs involving 96 unique GNSS stations across Europe and adjacent regions.

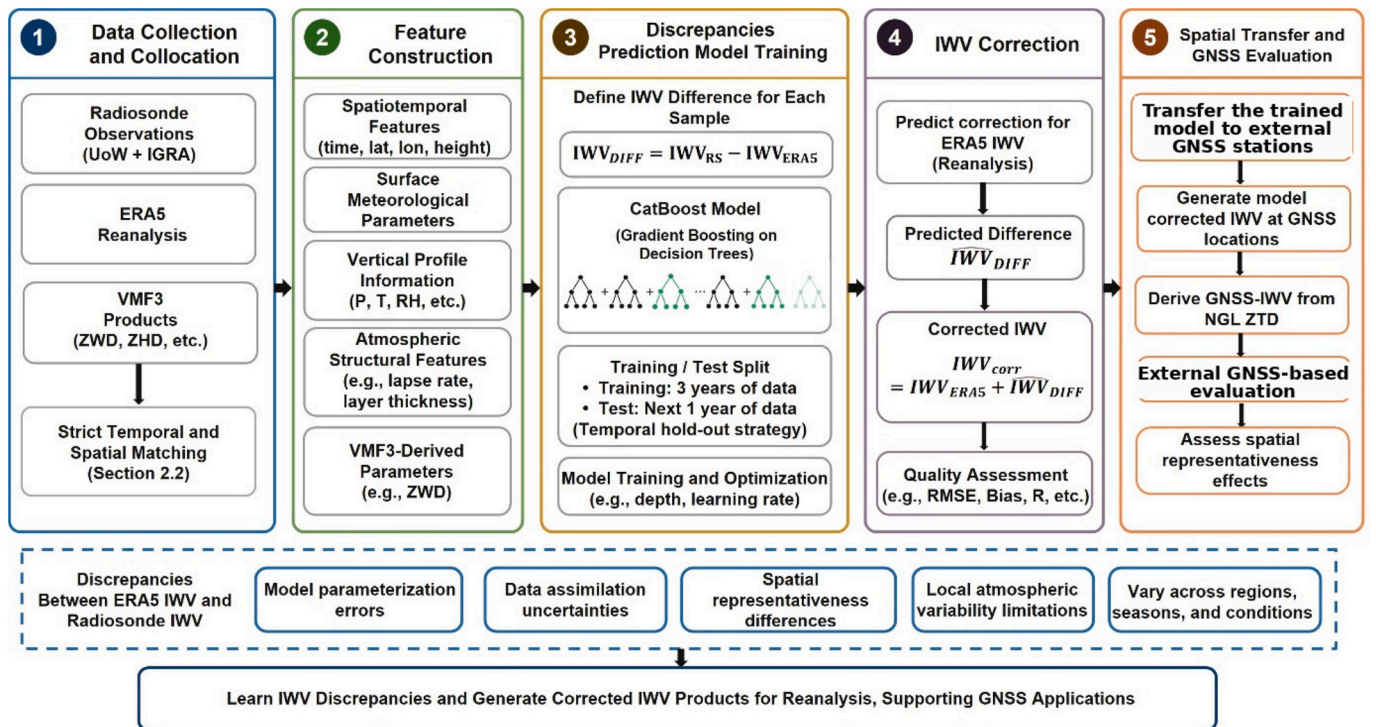


Fig. 4. The workflow of the proposed radiosonde-informed IWV difference prediction and correction framework.

where $\widehat{IWV}_{DIFF}$ is the machine learning predicted difference and IWV_{corr} denotes the corrected IWV value.

Compared with directly using reanalysis products or relying solely on sparse in-situ observations, the proposed framework aims to combine the strengths of both data sources. Radiosonde observations are utilized as a direct observational reference to characterize and learn the residual IWV differences present in ERA5 products, while the spatially and temporally continuous information provided by ERA5 is retained. Rather than replacing existing reanalysis products, the framework enhances their accuracy by correcting discrepancies through a data-driven

approach. As a result, the corrected IWV products preserve the key advantages of ERA5, including global coverage, high temporal resolution, long-term continuity, and operational accessibility, while achieving improved consistency with high-accuracy radiosonde observations. Then, the proposed framework provides an effective and scalable solution for generating more accurate IWV estimates across large spatial domains.

3.2. Feature construction

To characterize the factors influencing IWV difference, a

comprehensive set of input features was constructed from temporal information, geographical characteristics, ERA5-derived meteorological parameters, tropospheric vertical structural features, and climate zone classifications. The complete feature set is summarized in Table 1.

Tropospheric water vapor exhibits significant seasonal and diurnal variability. To preserve the cyclic characteristics of temporal information while avoiding discontinuities at calendar boundaries, the day of year (DOY) and observation hour were encoded as

$$\sin\left(\frac{2\pi(\text{DOY})}{N_y}\right), \cos\left(\frac{2\pi(\text{DOY})}{N_y}\right),$$

and

$$\sin\left(\frac{2\pi h}{24}\right), \cos\left(\frac{2\pi h}{24}\right),$$

where N_y is 365 or 366 and h is the UTC observation hour. These first-order harmonic pairs uniquely encode the annual and diurnal phases. Although semiannual and semidiurnal variability could be represented using second-order harmonics, these terms were not added because CatBoost can learn higher-order nonlinear temporal patterns from the complete phase information. In addition, geographical information, including latitude, longitude, and station height, was further incorporated to account for spatial variations in water vapor associated with topography, circulation patterns, and regional climatic conditions. Meteorological features were derived from ERA5 and include IWV, surface temperature, and surface pressure. Since radiosonde stations are often located at elevations different from the ERA5 model terrain, the corresponding meteorological variables were vertically interpolated to the station height using the ERA5 atmospheric profiles to reduce representativeness errors caused by terrain mismatches. Although ERA5 IWV is involved in the definition of the target feature, it was intentionally retained as an input variable because the magnitude and sign of the difference inherently depend on the original ERA5 estimate. Surface temperature and pressure provide complementary information describing the local thermodynamic state, further improving the model's ability to capture nonlinear relationships associated with IWV discrepancies. To further characterize the vertical structure of the atmosphere, several profile-derived variables were calculated from the ERA5 pressure level data. Specifically, the tropospheric column was divided into three layers representing the lower, middle, and upper troposphere, from which the geopotential thicknesses $Thick_{low}$, $Thick_{mid}$, and $Thick_{high}$ and temperature lapse rates $Lapse_{low}$, $Lapse_{mid}$, and $Lapse_{high}$ were derived. In addition, tropopause height was estimated from the vertical temperature structure of the ERA5 atmospheric profiles following the WMO lapse-rate criterion. These variables provide information on atmospheric stability, thermal stratification, and vertical water vapor transport processes that cannot be fully represented by surface meteorological parameters alone. Finally, a climate-zone flag derived from the Köppen-Geiger climate classification shown in Fig. 1 was introduced as a categorical feature. Based on the climatic characteristics of the study area, the selected stations were grouped into three major climate regimes: temperate, arid, and continental climates. This feature was

treated as a categorical variable within the CatBoost model, allowing the algorithm to learn climate-dependent patterns and improve its representation of regional atmospheric variability.

3.3. CatBoost-based IWV discrepancy prediction model training

To model the nonlinear relationship between the constructed input variables and the target feature, the CatBoost algorithm was employed in this study. CatBoost is a gradient boosting decision tree framework that constructs an ensemble of decision trees in a sequential manner, where each new tree is trained to minimize the residual errors of the previously generated trees. The final prediction can be expressed as

$$F(x) = \sum_{m=1}^M \gamma_m h_m(x) \quad (9)$$

where $h_m(x)$ denotes the m -th decision tree, γ_m represents the corresponding tree weight, and M is the total number of trees in the ensemble. Compared with conventional machine learning algorithms, CatBoost offers several advantages for atmospheric and geoscientific applications. For model selection, ridge regression, MLP, XGBoost, and CatBoost were evaluated using the same chronological evaluation period. First, it effectively captures complex nonlinear relationships among meteorological, geographical, and atmospheric structural variables without requiring explicit feature engineering. Second, the ordered boosting strategy adopted by CatBoost reduces prediction bias and mitigates overfitting during training. Third, CatBoost can directly process categorical variables through an efficient target encoding scheme, avoiding the need for one-hot encoding and enabling more effective utilization of categorical information. These characteristics make CatBoost particularly suitable for heterogeneous atmospheric datasets that contain both numerical and categorical features.

In this study, the input variables described in Section 3.2 were used as model inputs, while the corresponding IWV_{DIFF} values served as output targets. The climate-zone flag derived from the Köppen-Geiger classification was treated as a categorical feature in CatBoost. Model hyperparameters were optimized using Optuna (Akiba et al., 2019) with 40 trials per season and three-fold station-level GroupKFold, using validation RMSE as the objective. The random seed was 42, and the optimized parameters included iterations, tree depth, learning rate, L2 regularization, and subsampling ratio. The final cold/warm models used 2076/1407 iterations, depths of 8/6, learning rates of 0.013/0.026, L2 regularization values of 8.34/8.47, and subsampling ratios of 0.84/0.92, respectively. Three complementary evaluation settings were adopted. First, an in-sample training evaluation was performed to confirm that the model adequately learned the dominant relationships in the development data. Second, a strict station holdout experiment was conducted to assess spatial generalization. For this purpose, approximately 20% of the stations from each radiosonde data source were completely excluded from scaler fitting, hyperparameter optimization, and CatBoost training, and were used only for independent testing. All observations from each station were assigned exclusively to either the development or station-holdout partition. The original station holdout used random seed 42, with 23 of the 116 stations reserved for testing. To further assess split sensitivity, a five-fold GroupKFold evaluation was also conducted at the physical-station level using the same model configuration. Third, temporal generalization was evaluated using a chronological split of the full dataset covering November 2021 to November 2025. Observations acquired before 1 December 2024 were used for model development, whereas data from 1 December 2024 to 30 November 2025 were reserved as an independent temporal test dataset. This period was not used for feature scaling, model fitting, or parameter optimization. After completion of the strict station holdout evaluation, the final model was retrained using all eligible radiosonde stations and all observations before 1 December 2024. Then, the model was also applied to the independent temporal test period and subsequently transferred to GNSS

Table 1

Feature categories and corresponding input variables used in the CatBoost model.

Feature Category	Variables
Temporal	$\sin_{DOY}$, $\cos_{DOY}$, $\sin_{Hour}$, $\cos_{Hour}$
Spatial	Latitude, Longitude, Height
Surface	ERA5 IWV, VMF3 IWV, Surface Temperature, Surface Pressure
Meteorological	
Atmospheric	$Thick_{low}$, $Thick_{mid}$, $Thick_{high}$, $Lapse_{low}$, $Lapse_{mid}$, $Lapse_{high}$,
Structural	Tropopause Height
Climate	Climate Flag
Classification	

station locations to generate spatially collocated model-corrected IWV. This design separates model fitting, spatial generalization, and temporal generalization, while preventing information leakage between the development and independent evaluation datasets.

3.4. Performance validation

To evaluate the performance of the proposed framework, model predictions were assessed on both the training dataset and the independent test dataset. The training results were used to examine the model's fitting capability, while the independent test results provided a rigorous assessment of its generalization performance under unseen conditions. For both datasets, the predicted IWV discrepancies were compared with the corresponding radiosonde-derived reference values. Three commonly used statistical metrics were adopted to quantify model performance, including the coefficient of determination R^2 , mean bias error (MBE), and root mean square error (RMSE). The R^2 metric evaluates the proportion of variance explained by the model, while the MBE is used to evaluate the systematic tendency of the model to overestimate or underestimate the target values. The RMSE quantifies the overall prediction accuracy and is particularly sensitive to large prediction errors. These metrics are computed as follows:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (10)$$

$$MBE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i) \quad (11)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (12)$$

Together, these metrics enable a robust assessment of both the fitting capability and predictive performance of the proposed framework.

3.5. Compatibility assessment using external GNSS observations

To assess the transferability of the radiosonde-informed correction framework, the final model was applied to external GNSS station locations not involved in model development. The purpose of this experiment is not to optimize IWV estimates against GNSS observations, but to evaluate whether the radiosonde-informed discrepancy relationship can be spatially transferred to locations where no radiosonde is physically available. The resulting estimate therefore provides a radiosonde-informed, spatially collocated IWV estimate for reducing representativeness differences in conventional radiosonde-GNSS comparisons. A total of 98 radiosonde and GNSS station pairs shown in Fig. 3 were included in the evaluation. Among 98 pairs, the pairs were classified as either Core or Extended. Core pairs were limited to a horizontal separation of 100 km and an absolute elevation difference of 300 m. Where no suitable GNSS station was available, extended thresholds of 150 km and 500 m were applied. For each radiosonde site, the eligible GNSS station with the smallest combined spatial separation was selected, resulting in 89 core and 9 extended pairs. The primary analysis used the 89 core pairs, whereas all 98 pairs were retained for sensitivity assessment. Since radiosonde IWV was not adjusted to the GNSS station height, the remaining spatial separation may introduce representativeness differences, motivating the comparison with the proposed model-corrected IWV generated directly at the GNSS coordinates. For each GNSS station, the ERA5 and VMF3 variables were extracted using the same preprocessing and interpolation procedures adopted during model development. For comparison, ERA5 and VMF3 IWV at the same GNSS coordinates and epochs were additionally retained as collocated baselines to distinguish the effect of spatial transfer from that of model correction. Horizontal distance, elevation difference, and temporal

offset were also retained for sensitivity analysis. The trained model was then used to generate the corrected IWV according to Eq. (8). GNSS-derived IWV was obtained from NGL ZTD products. ZTD was used instead of the available NGL ZWD product to maintain consistency with the adopted processing strategy. The ZHD was calculated using the Saastamoinen model with ERA5 pressure adjusted to the GNSS station height, after which ZWD was derived as the difference between ZTD and ZHD. The resulting ZWD was converted to IWV using the ERA5 weighted mean atmospheric temperature.

Compatibility was evaluated by comparing GNSS-derived IWV with model-corrected IWV at the GNSS locations, and IWV from the paired nearby radiosonde stations. GNSS and radiosonde observations were matched using the nearest available epoch within 90 min. Compatibility was then assessed by comparing GNSS-derived IWV with the model-corrected IWV at the GNSS locations and with IWV observed at the paired nearby radiosonde stations. Performance was quantified using RMSE, MBE, mean absolute error (MAE), and R^2 . The comparison was designed to determine whether spatially collocated model-corrected IWV provides better consistency with GNSS observed water vapor than conventional comparisons based on nearby radiosonde observations. Statistical uncertainty was further assessed using a station-pair clustered bootstrap with the 95 valid radiosonde-GNSS pairs as resampling units, together with pair-level Wilcoxon signed-rank and exact binomial tests.

4. Results and discussion

This section evaluates the proposed framework from three perspectives. First, the model performance is assessed using the training, station holdout, and independent temporal test datasets. Second, the spatial, seasonal, monthly, and climate zone dependent characteristics of the corrected IWV products are examined across Europe. Finally, the trained model is transferred to external GNSS station locations, where the spatially collocated model-corrected IWV is evaluated against GNSS-derived IWV and compared with observations from nearby radiosonde stations.

4.1. Model development and performance evaluation

To account for the seasonal variability of IWV, the results are presented separately for the cold-season (November–April) and warm-season (May–October), respectively. Prior to model evaluation, the statistical consistency between the training and testing datasets and the characteristics of the IWV difference are analyzed. The observed correction patterns remain stable across independent years, providing a reliable foundation for the machine learning framework.

4.1.1. Dataset characteristics and IWV difference behavior

Fig. 5(a) presents the probability density distributions of ERA5-derived IWV for the cold-season training and independent testing datasets, covering November to April. The two distributions exhibit highly similar shapes, with a dominant concentration around 8–12 mm and a gradual decline toward higher IWV values. The corresponding empirical cumulative distribution functions (CDF) shown in Fig. 5(b) also closely overlap across most of the IWV range. The testing CDF lies slightly below the training curve, indicating that the independent test period contains a marginally larger proportion of higher IWV conditions. However, the KS statistic is only 0.031, demonstrating that this shift is small and that the cold-season training and testing datasets remain statistically comparable. Fig. 5(c) and Fig. 5(d) present the corresponding results for the warm season from May to October. Compared with the cold season, the IWV distribution shifts toward higher values and spans a broader range, with most observations concentrated between approximately 15 and 30 mm and some values exceeding 50 mm. The testing CDF lies slightly above the training curve, indicating a modest shift toward lower IWV conditions during the

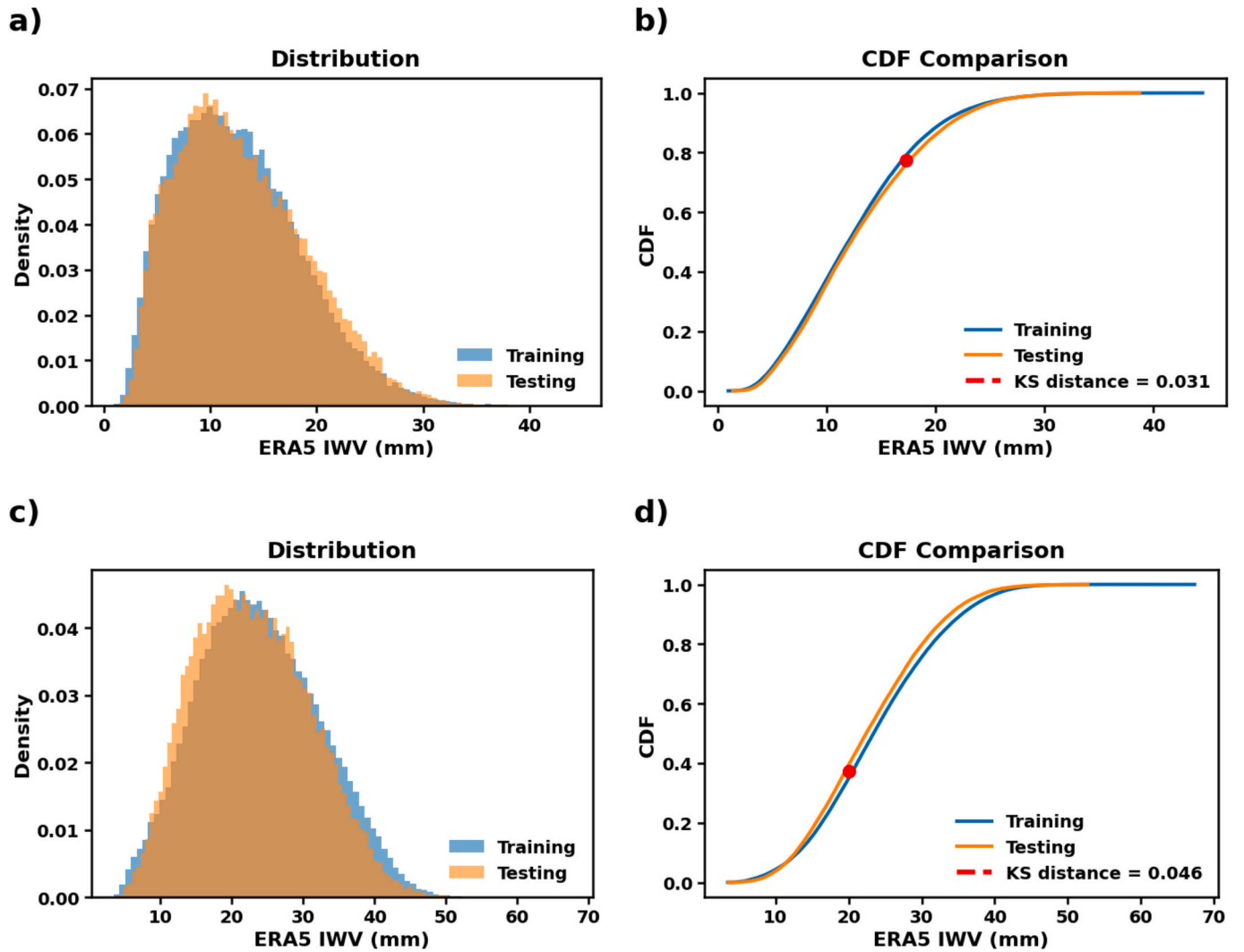


Fig. 5. Probability density distributions and empirical cumulative distribution functions of ERA5-derived IWV for the training and independent testing datasets during the cold and warm seasons: (a) cold-season probability density distribution, (b) cold-season cumulative distribution, (c) warm-season probability density distribution, and (d) warm-season cumulative distribution.

independent test period. The warm-season KS statistic of 0.046 is slightly larger than the cold-season value but remains small relative to the overall variability of the distribution. Both seasonal distributions are positively skewed because IWV is bounded at zero and most observations represent dry to moderately humid conditions, whereas high-IWV events occur less frequently and form a long upper tail. This asymmetry is further influenced by pooling observations from stations across different elevations and climate regimes. Low-IWV conditions dominate during the cold season, while larger moisture variability and occasional highly humid conditions broaden the warm-season distribution and extend its upper tail. The opposite CDF shifts observed in the two seasons are therefore interpreted as reflecting interannual variations in IWV rather than substantial changes in dataset composition. Overall, the close agreement between the training and testing distributions indicates that the chronological split preserves the principal statistical characteristics of ERA5-derived IWV in both seasons. The subsequent model evaluation can therefore be interpreted primarily as an assessment of temporal generalization.

In addition to the temporal consistency between the training and testing datasets, the consistency between the two radiosonde archives used for model development was further investigated. Since the training dataset combines observations from IGRA and UoW archive, potential source-dependent discrepancies could introduce inconsistent training

labels and consequently affect model learning. To evaluate this possibility, radiosonde observations from the two archives were collocated using identical WMO station identifiers and launch epochs. For this source-consistency analysis, ten common UoW-IGRA stations were obtained from a separate comparison dataset, yielding 17,727 collocated soundings. For these overlapping stations, only the UoW records were retained in the final machine learning dataset, and the corresponding IGRA records were excluded from model development. The IWV difference between the two archives was defined and evaluated. The comparison demonstrates excellent consistency between the two datasets. The overall median differences were only 0.003 mm during the cold-season and 0.061 mm during the warm-season. Across different IWV intervals, the median differences generally ranged from approximately -0.05 mm to 0.18 mm, while the largest statistically reliable discrepancy remained below 0.29 mm. Considering that the typical variability of IWV over Europe ranges from several to tens of millimeters, these differences are negligible. These results indicate that the IGRA and UoW archives provide highly consistent IWV observations when identical stations and launch epochs are compared. Therefore, the two datasets can be regarded as belonging to a common observational reference and can be combined directly for subsequent machine learning training without introducing significant source-dependent biases.

4.1.2. Prediction performance

The model comparison yielded cold-season RMSE values of 1.64, 4.07, 1.13, and 1.11 mm for ridge regression, MLP, XGBoost, and CatBoost, respectively; the corresponding warm-season values were 2.71, 6.51, 2.06, and 2.06 mm. CatBoost therefore provided the comparable performance across both seasons and was adopted for the subsequent analyses. To evaluate the capability of the proposed CatBoost framework in learning and predicting corrected IWV, the model was developed using ERA5 operational (ERA5 OP) and VMF3 operational (VMF3 OP) products. The performance of the model was evaluated using RMSE, R^2 , and MBE for the training dataset, the strict station holdout dataset, and the independent temporal test dataset, as summarized in Table 2. The training results were used to assess whether the models adequately captured the dominant relationships between the input features and the target feature. The independent temporal test evaluated whether these relationships remained stable during a future one-year period that was excluded from model development. In addition, a strict station holdout experiment was conducted to assess spatial transferability. Because all observations from the holdout radiosonde stations were excluded from feature scaling, hyperparameter optimization, and model training, this experiment provides a direct indication of whether the learned correction can be applied at locations not represented during model development. This spatial generalization capability is particularly important for the subsequent application of the models at external GNSS station locations.

For the cold season, the model-corrected IWV achieved an RMSE of 1.07 mm and an R^2 of 0.96 for the training dataset, with an MBE close to zero. Similar performance was retained for the independent temporal test, with an RMSE of 1.11 mm, an R^2 of 0.96, and an MBE of 0.02 mm. The station holdout evaluation yielded a moderately higher RMSE of 1.33 mm and a lower R^2 of 0.94, while the MBE remained small at -0.14 mm. Although the performance decreased relative to the temporal test, the corrected IWV still maintained strong agreement with radiosonde observations at completely unseen stations. This result indicates that the model captures spatially transferable relationships rather than station-specific characteristics. The larger warm-season errors likely reflect the broader IWV range and stronger moisture variability, which increase the complexity of the ERA5-radiosonde discrepancy. The training dataset produced an RMSE of 1.96 mm, an R^2 of 0.95, and an MBE close to zero. For the independent temporal test, the RMSE increased slightly to 2.06 mm, while the R^2 remained high at 0.93 and the MBE was only 0.03 mm. The station holdout test resulted in an RMSE of 2.40 mm, an R^2 of 0.92, and an MBE of -0.15 mm. The slightly negative MBE for the station-holdout dataset indicates that the predicted correction was, on average, marginally smaller than the correction required to match the radiosonde observations. In addition, the five-fold station-level evaluation yielded pooled RMSE values of 1.55 and 2.70 mm and R^2 values of 0.92 and 0.90 for the cold and warm seasons, respectively. Although slightly higher than the single-holdout results, these values support spatial generalization across different station partitions. Nevertheless, the magnitude of this bias remained small relative to the corresponding RMSE, indicating that the spatial generalization of the model was not substantially affected.

Fig. 6 further illustrates the agreement between radiosonde observed IWV and model-corrected IWV during the independent temporal test. In

the cold season, shown in Fig. 6(a), the corrected IWV values are tightly distributed around the 1:1 line, and the fitted slope of 0.956 indicates only a slight compression of the IWV variability. For the warm season in Fig. 6(b), the scatter is broader and the fitted slope decreases to 0.933, indicating a stronger tendency to underestimate the magnitude of high IWV conditions. This behavior is consistent with the larger warm-season RMSE and suggests that extreme moisture conditions remain more difficult to reproduce than moderate conditions. However, the MBE values remain close to zero in both seasons, indicating that the models do not introduce a substantial overall systematic offset.

4.2. Spatial, seasonal, and climate-zone-dependent performance

This section further examines the spatial, seasonal, and climate-zone-dependent variations in model performance to characterize the conditions under which the proposed correction framework achieves its highest accuracy and robustness.

4.2.1. Spatial performance across Europe

To further assess the geographical robustness of the proposed framework, model performance was evaluated separately at each radiosonde station using the independent temporal test dataset. Fig. 7 shows the spatial distribution of station-wise RMSE values for the cold and warm seasons, while Fig. 8 summarizes their statistical distributions. For the cold season, station-wise errors were generally low and spatially homogeneous across most of the study region. As shown in Fig. 7(a), the majority of stations exhibited RMSE values close to or below 1.5 mm, with only a small number of isolated stations showing clearly large errors. The boxplot in Fig. 8 confirms this relatively compact distribution, with a median RMSE of 0.96 mm. The warm season exhibited a wider spatial spread of errors and a clearly higher overall RMSE level. Fig. 7(b) shows that large RMSE values occurred more frequently across southern Europe, the Mediterranean region, parts of eastern Europe, and several northern maritime locations. The corresponding boxplot in Fig. 8 has a median RMSE of 1.69 mm and a substantially larger interquartile range than that of the cold season. This broader distribution reflects the greater spatial variability of warm-season water vapor, during which convection can produce more localized and nonlinear IWV discrepancies.

Several stations were identified as outliers using the upper whisker criterion of the boxplot distribution, as listed in Table 3. Four cold-season outliers were detected and their RMSE values ranged from 2.00 to 2.99 mm. During the warm-season, three outliers were identified. The recurrence of SPM00008430, UUOO, and FOM00006011 in both seasons suggests that their higher errors are not solely caused by seasonal atmospheric variability, but may also be related to persistent local representativeness differences or station-specific discrepancy characteristics. The largest warm-season error occurred at SPM00008430 in southeastern Spain, where the RMSE reached 6.93 mm and the correlation decreased to 0.26. This station is located in a region strongly influenced by Mediterranean moisture transport, land-sea contrasts, and episodic convective activity. Such processes can generate sharp spatial and temporal water vapor gradients that are difficult to represent using large scale reanalysis predictors. UUOO in western Russia also showed consistently high RMSE values in both seasons, together with positive MBE values of 1.63 mm in the cold season and 2.30 mm in the warm season. The comparatively low correlations of 0.57 and 0.48 suggest that the local discrepancy behavior at this site differs from the dominant relationships learned from the broader training network. FOM00006011, located in the North Atlantic, exhibited elevated RMSE values of 2.69 mm and 4.50 mm in the cold and warm seasons, respectively. Its warm season correlation was particularly low at 0.06, indicating limited reproduction of temporal variability during that period. This may reflect rapidly changing maritime atmospheric conditions, frequent frontal passages, and strong moisture advection, all of which can reduce the representativeness of large-scale model features at

Table 2

Performance of the proposed model-corrected IWV products for the training, station holdout, and independent temporal test datasets.

Season	Dataset	RMSE (mm)	R^2	MBE (mm)
Cold	Training	1.07	0.96	0.00
Cold	Station holdout	1.33	0.94	-0.14
Cold	Independent temporal test	1.11	0.96	0.02
Warm	Training	1.96	0.95	0.00
Warm	Station holdout	2.40	0.92	-0.15
Warm	Independent temporal test	2.06	0.93	0.03

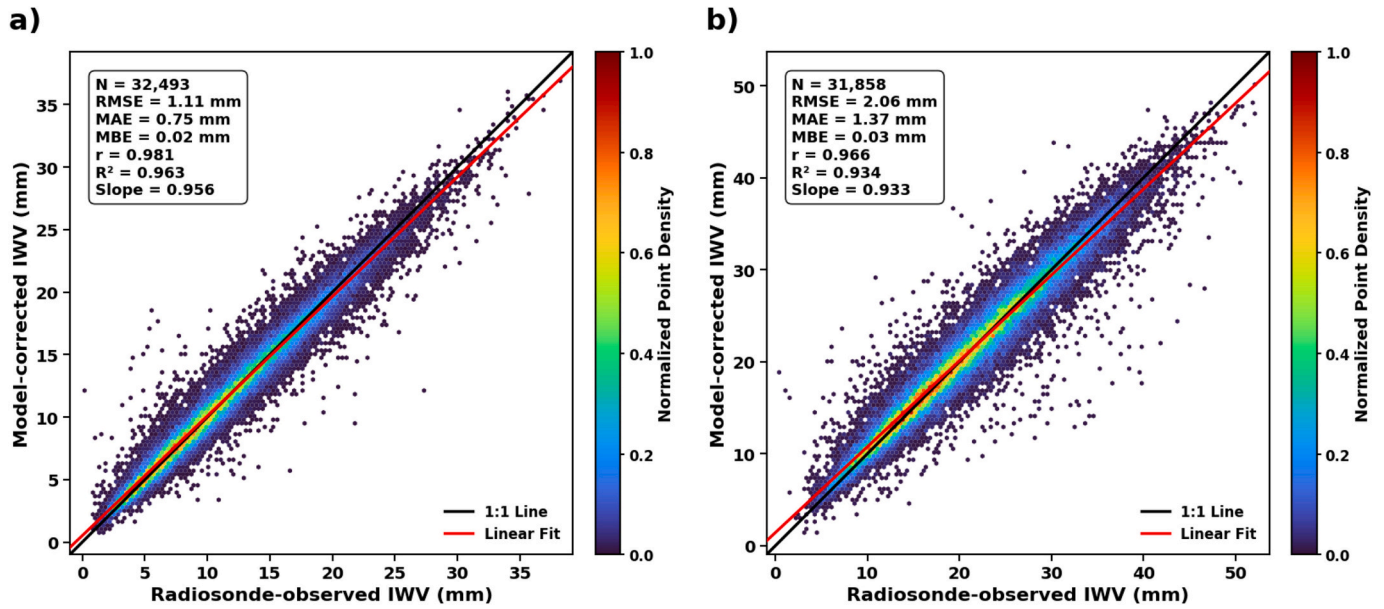


Fig. 6. Density scatterplots between radiosonde observed I WV and model-corrected I WV for the independent temporal test dataset: (a) cold season and (b) warm season. The black line denotes the 1:1 relationship, and the red line represents the linear regression fit. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

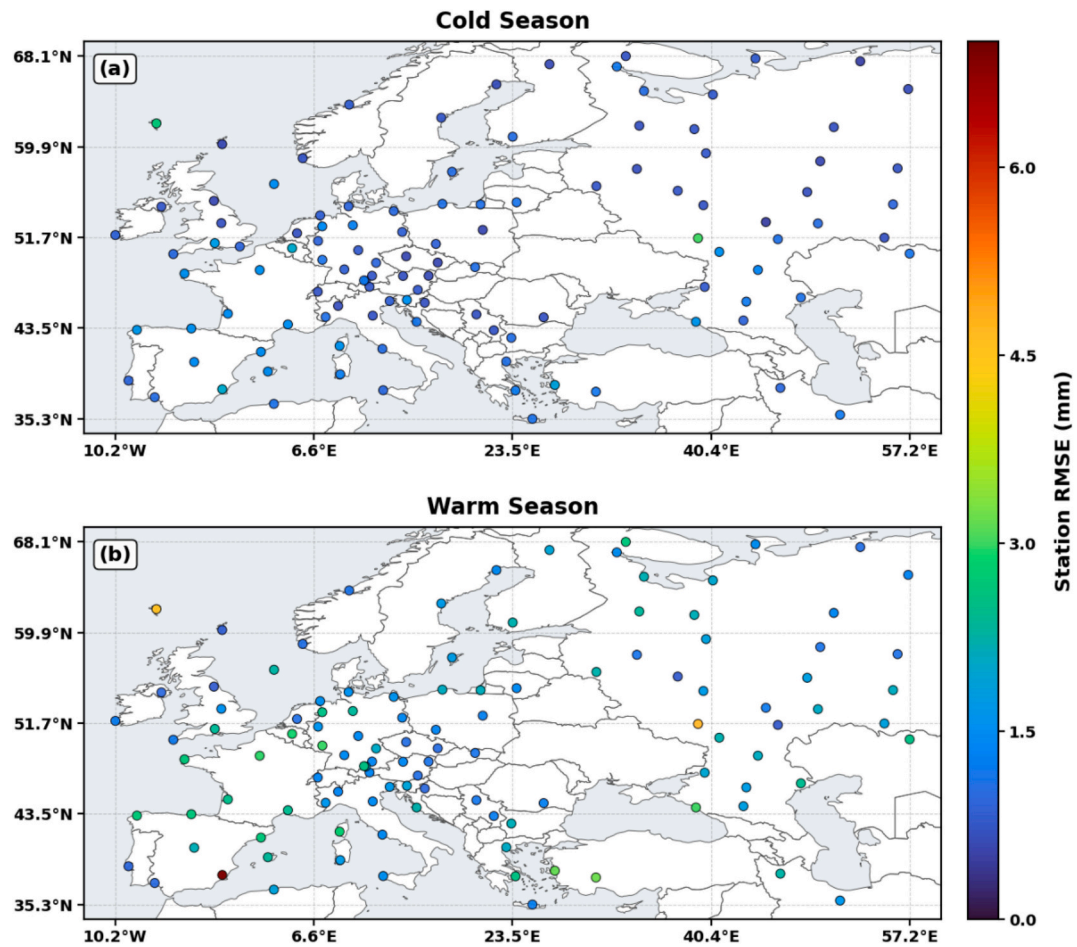


Fig. 7. Spatial distribution of station-wise RMSE values for the independent test dataset: (a) cold-season model and (b) warm-season model.

an isolated island station. By contrast, BEM00006458 and SPM00008430 showed relatively moderate cold season RMSE values

close to 2.0 mm, although SPM00008430 became the most evident outlier during the warm-season.

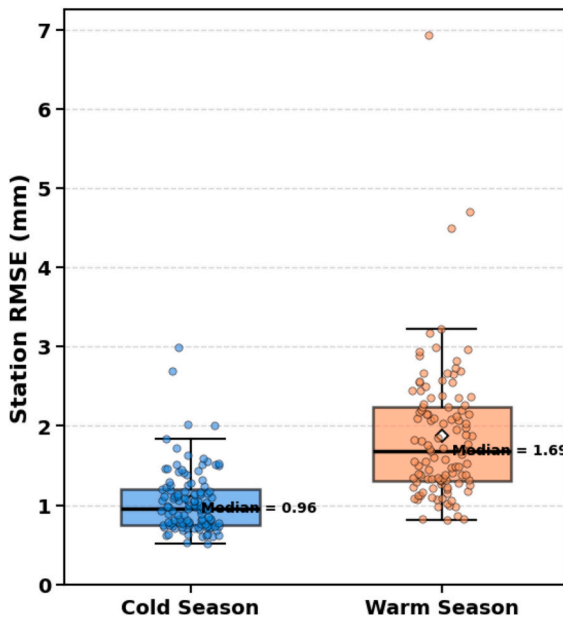


Fig. 8. Distribution of station-wise RMSE values for cold- and warm-season models.

Despite these localized high error stations, the spatial maps do not indicate a systematic degradation across any major geographical sector. Most stations retained low to moderate RMSE values, particularly during the cold season, and the overall station distributions remained compact relative to the full IWV range. The results therefore demonstrate that the radiosonde-informed correction framework shows good spatial transferability within the conditions represented by the training dataset, while also showing that transferability is reduced at locations affected by strong local moisture variability or persistent site-specific discrepancy patterns.

4.2.2. Monthly performance

To assess temporal stability, model performance was evaluated separately for each month using the independent temporal test dataset (Fig. 9). RMSE remained below approximately 1.5 mm during most cold-season months, with the best performance observed in February at 0.98 mm, but increased progressively from late spring and reached its maximum of approximately 2.74 mm in August. Monthly R^2 showed the opposite pattern, decreasing from about 0.97 in February to 0.87 in August before recovering in autumn. MBE remained close to zero during most months, whereas August exhibited a stronger negative bias of approximately -0.59 mm, indicating a tendency to underestimate IWV. Station level analysis showed that GMM00010618 and FRM00007145 contributed substantially to this degradation, with RMSE values of 4.58 and 4.30 mm and MBE values of -4.01 and -3.26 mm, respectively. The poorer performance coincided with anomalous weather conditions in August 2025. Germany experienced a strong heatwave in mid-August, with widespread maximum temperatures of $32\text{--}35^\circ\text{C}$ and prolonged dry conditions across central regions, followed by renewed rainfall toward

the end of the month (Deutscher Wetterdienst, 2025). France was affected by an exceptional 11-day heatwave from 8 to 18 August, including severe convective passages and strong regional contrasts in temperature and precipitation (Météo France, 2025). These anomalous conditions coincided with the observed performance degradation and may have contributed to increased atmospheric variability and localized water vapor gradients, although their individual effects cannot be isolated from the present analysis.

4.2.3. Climate-zone performance

Station-wise results were further aggregated according to the three climate regions defined in Section 2 to evaluate the influence of large-scale climatic conditions on model performance. Fig. 10 summarizes the RMSE, R^2 , and MBE for each climate zone during the cold and warm seasons. Climate Zone 3, representing continental climates, achieved the lowest RMSE in both seasons, with values of 0.98 mm in the cold-season and 1.83 mm in the warm-season. Climate Zone 1, representing temperate climates, showed intermediate RMSE values of 1.19 and 2.02 mm, whereas Climate Zone 2, characterized by arid and semi-arid conditions, exhibited the largest errors, increasing from 1.32 mm in the cold-season to 3.12 mm in the warm-season. The larger warm-season error in Climate Zone 2 suggests that the learned correction relationships were less transferable to this region, where IWV conditions may shift rapidly between persistently dry states and episodic moist events. This reduced transferability may reflect the stronger contrast between dry background conditions and episodic moist events, which results in a wider range of IWV states represented in this region. In contrast to RMSE, R^2 did not follow the same ranking. Climate Zone 2 produced the highest R^2 values, reaching 0.958 and 0.932 in the cold and warm seasons, respectively, while Climate Zone 1 showed the lowest values of 0.944 and 0.846. MBE remained small in all climate zones, ranging from -0.04 to 0.12 mm in the cold season and from 0.01 to 0.09 mm in the warm season, indicating no substantial climate-dependent systematic bias. Overall, absolute prediction accuracy was highest in continental climates and lowest under warm arid and semi-arid conditions, while the differing R^2 patterns highlight the importance of considering both error magnitude and explained variance when evaluating regional model performance.

4.3. Application to corrected IWV products

Following the evaluation of the predictive capability of the proposed framework, the predicted IWV bias was applied to the original ERA5 IWV to generate spatially continuous bias-corrected IWV fields at a spatial resolution of $0.25^\circ \times 0.25^\circ$, consistent with the ERA5 grid used in this study. The corrected product was evaluated against held-out radiosonde observations from the independent temporal test dataset, while VMF3-derived IWV was included as an established benchmark. Fig. 11 compares the spatial distributions of ERA5 IWV, VMF3 IWV, model-corrected IWV, and radiosonde IWV at four representative epochs over Europe. The selected epochs include two cold-season cases, 28 January and 1 November 2025, and two warm-season cases, 1 May and 1 June 2025, thereby representing substantially different atmospheric moisture conditions. Within each epoch, the four panels share the same color scale to facilitate direct comparison among the products. The four

Table 3
Outlier stations of the proposed seasonal discrepancy prediction models.

Season	Name	Latitude ($^\circ$)	Longitude ($^\circ$)	RMSE (mm)	MBE (mm)	Correlation
Cold	UUOO	51.6500	39.2500	2.99	1.63	0.57
Cold	FOM00006011	62.0231	-6.7647	2.69	0.26	0.62
Cold	BEM00006458	50.7456	4.7633	2.02	-0.04	0.85
Cold	SPM00008430	38.0019	-1.1708	2.00	0.70	0.69
Warm	SPM00008430	38.0019	-1.1708	6.93	-1.09	0.26
Warm	UUOO	51.6500	39.2500	4.70	2.30	0.48
Warm	FOM00006011	62.0231	-6.7647	4.50	1.50	0.06

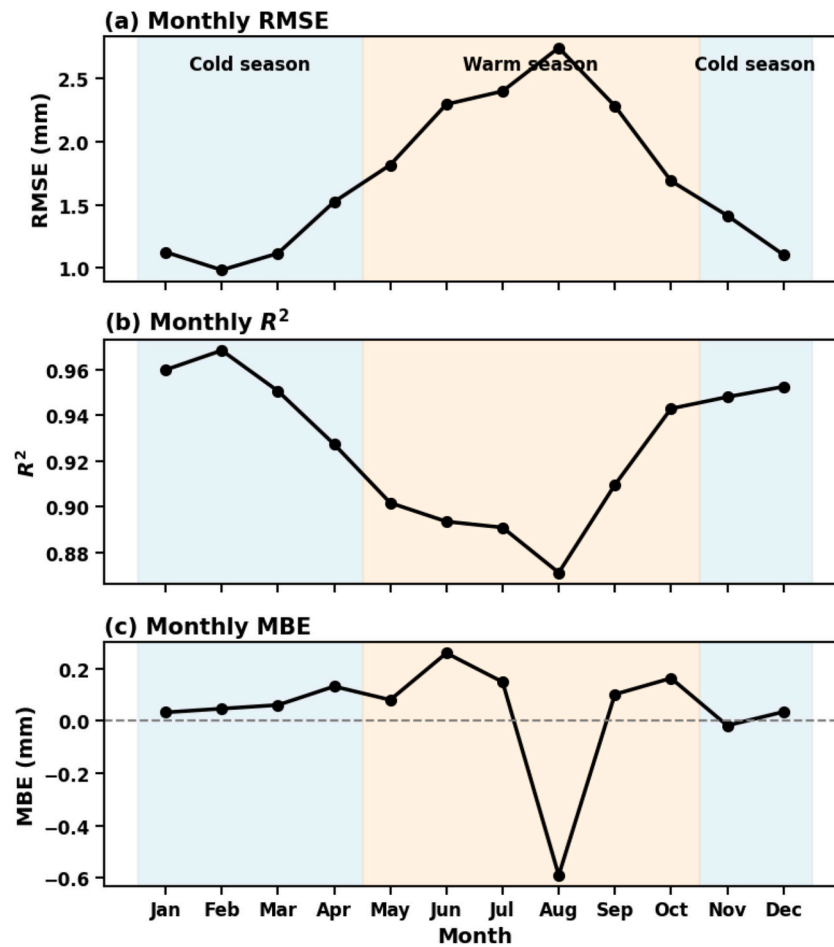


Fig. 9. Monthly variations in (a) RMSE, (b) coefficient of determination (R^2), and (c) MBE for the independent test dataset.

cases illustrate the larger seasonal and regional variability of water vapor across Europe. ERA5, VMF3, and the corrected product reproduce broadly consistent large-scale moisture structures, while differences in IWV magnitude and local spatial variability remain evident. At many radiosonde locations, the model-corrected fields show improved consistency with the observed IWV values while preserving the large-scale spatial patterns represented by ERA5. The effect of the correction is particularly apparent in regions characterized by significant moisture gradients, including central Europe, the Mediterranean region, and parts of eastern Europe. Unfilled areas in the model-corrected IWV fields correspond to grid cells where one or more predictor variables fall outside the ranges represented in the training data and do not indicate missing ERA5 data or gridding failures. Among these predictors, elevation was the dominant limiting factor because the training stations did not cover the highest-elevation regions of Europe. Because CatBoost is tree-based, predictions beyond the training feature ranges are not physically constrained extrapolations. Therefore, grid cells containing out-of-range predictor values were masked and are displayed as white areas in Fig. 11. This treatment reduces unsupported extrapolation and improves the reliability of the corrected IWV fields.

The corresponding quantitative performance is summarized in Table 4. Relative to the original ERA5 and VMF3 products, the proposed framework substantially improved IWV accuracy. During the cold-season, the RMSE decreased from 2.29 mm for ERA5 and 1.88 mm for VMF3 to 1.11 mm after correction, corresponding to reductions of 51.5% and 41.0%, respectively. The corrected product also achieved the highest R^2 , reaching 0.96. During the warm season, the corrected product yielded an RMSE of 2.06 mm, representing reductions of 40.1% relative to ERA5 and 33.3% relative to VMF3. Its MBE was reduced to

0.03 mm, while R^2 increased to 0.93. These results indicate that a substantial component of the discrepancies between ERA5 and radiosonde IWV is systematic and predictable within the applicability domain represented by the training data. The near-zero MBE values further demonstrate that the proposed framework effectively reduces systematic discrepancies while preserving the underlying spatial and seasonal variability of atmospheric water vapor. Overall, the corrected IWV product consistently outperformed ERA5 and VMF3 across both seasonal regimes, demonstrating the capability of the radiosonde-informed machine learning framework to generate more accurate IWV fields without introducing unsupported extrapolation beyond the training domain.

4.4. Spatial transfer and evaluation using external GNSS observations

To further assess the spatial transferability of the proposed framework, an additional evaluation was conducted using external GNSS stations. The NGL GNSS-derived ZTD observations were not used in model development, although the derived GNSS-IWV relies on ERA5-derived pressure and weighted mean temperature in the ZTD-to-IWV conversion. The analysis is therefore interpreted as a GNSS-based evaluation independent of model development, rather than as validation against a fully ERA5-independent reference. Its purpose is to assess whether the radiosonde-informed relationship can be transferred to locations without collocated soundings, rather than whether the resulting IWV outperforms existing products in absolute agreement with GNSS. A total of 98 radiosonde-GNSS station pairs were initially constructed, of which 95 pairs contained valid collocated observations after quality control and were retained for the final analysis. The evaluation

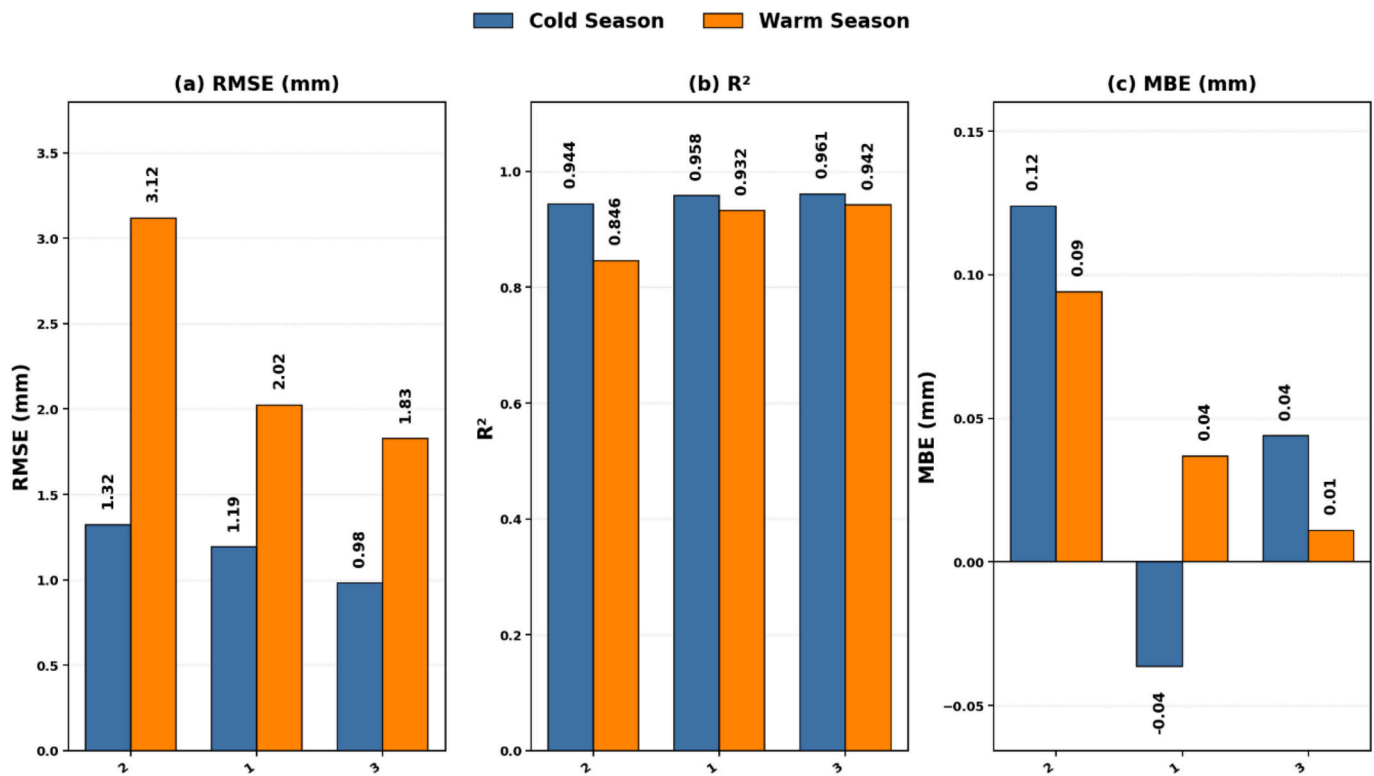


Fig. 10. Climate-zone-dependent performance of the proposed discrepancy prediction framework. Comparison of (a) RMSE, (b) R^2 , and (c) MBE for the three climate zones during the cold and warm seasons.

comprised 47,418 temporally matched observations. At identical GNSS coordinates and epochs, ERA5 and VMF3 showed closer agreement with GNSS-derived IWV than the radiosonde-informed product, with RMSE values of 1.33, 2.09, and 2.25 mm for ERA5, VMF3, and the radiosonde-informed product, respectively. However, the objective of this analysis is not to optimize agreement with GNSS, but to generate a radiosonde-informed virtual IWV observation directly at the GNSS station location. For each pair, this spatially collocated estimate and the physically observed IWV from the nearby radiosonde station were independently compared with GNSS-derived IWV. This design allows the benefit of transferring radiosonde information to the GNSS location to be evaluated while reducing the representativeness mismatch associated with a nearby but non-collocated radiosonde.

In addition to RMSE, MBE, and R^2 , MAE was included in the GNSS assessment to characterize the typical absolute difference without cancellation between positive and negative errors. This metric is particularly useful for evaluating whether spatially collocated radiosonde-informed IWV reduces the representativeness mismatch associated with nearby, non-collocated radiosonde observations. The overall results are summarized in Table 5. Relative to GNSS-derived IWV, the model-corrected product achieved an RMSE of 2.25 mm and an MAE of 1.86 mm, compared with 3.15 and 2.20 mm, respectively, for the nearby radiosonde observations, corresponding to reductions of 28.5% and 15.5%. The R^2 also increased from 0.857 to 0.927. These improvements primarily reflect the benefit of transferring radiosonde-informed information to the GNSS location and reducing spatial representativeness mismatch.

The reduction in MAE demonstrates that the model-corrected IWV was generally closer to the GNSS-derived IWV at individual epochs. The MAE decreased from 1.96 to 1.78 mm during the cold season and from 2.43 to 1.94 mm during the warm season, corresponding to reductions of 9.2% and 20.2%, respectively. The larger warm-season improvement is consistent with the stronger spatial variability of water vapor under moist conditions, for which the separation between the radiosonde and

GNSS stations introduces larger representativeness differences. By comparison, the overall MBE values remained similar, at -1.73 mm for the model-corrected product and -1.72 mm for the nearby radiosonde IWV. These negative values indicate that both products were systematically lower than the GNSS-derived IWV by approximately 1.7 mm. Thus, although the proposed framework substantially reduced the average absolute discrepancy and overall error variability, it did not remove the systematic offset between the radiosonde-informed and GNSS-derived IWV domains. The similar offsets for the model-corrected and radiosonde IWV suggest that this bias is more likely associated with systematic differences between the radiosonde and GNSS observing and processing chains, including GNSS ZTD processing and the pressure and T_m dependent conversion to IWV, rather than with the machine learning correction itself. The similar MBE values indicate that the observed improvements do not result from direct calibration toward GNSS observations, because no GNSS observations were used during model development or subsequent bias adjustment. The reductions in RMSE and MAE and the improvements in correlation and R^2 therefore reflect improved agreement with the external GNSS observations.

Fig. 12 presents the pair-specific comparison between the model-corrected IWV and nearby radiosonde IWV relative to GNSS-derived IWV. Among the 95 station pairs containing at least 30 common observations, the model-corrected product achieved a lower RMSE at 72 pairs (75.8%), a lower MAE at 64 pairs (67.4%), and a higher R^2 at 72 pairs (75.8%). The RMSE and MAE improvements were 0.90 mm (95% CI: 0.51–1.31 mm) and 0.34 mm (95% CI: 0.15–0.58 mm), respectively, with both pair-level Wilcoxon and exact binomial tests yielding $p < 0.001$. As summarized in Fig. 12(d), improvements were obtained for all three metrics at most station pairs. The corresponding improvement frequencies for the Core pairs were 74.4%, 65.1%, and 74.4% for RMSE, MAE, and R^2 , respectively. The advantage was even more pronounced for the Extended pairs, for which 8 of the 9 pairs (88.9%) showed improvements in each metric. At the aggregate level, the RMSE for the Extended group decreased from 4.228 mm for the nearby radiosonde

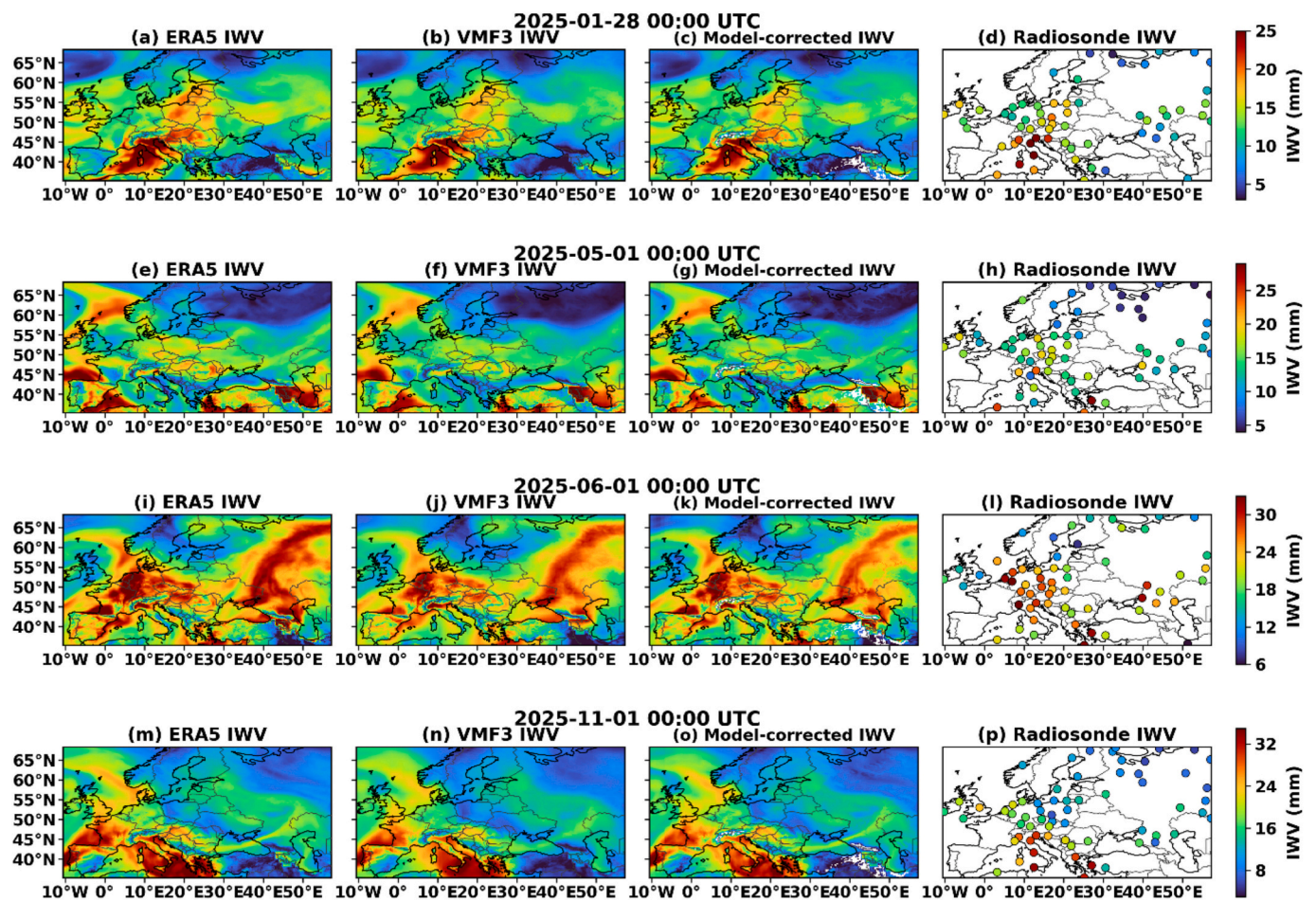


Fig. 11. Spatial distributions of ERA5 IWV, VMF3 IWV, model-corrected IWV, and collocated radiosonde IWV observations over Europe at four representative epochs: 28 January, 1 May, 1 June, and 1 November 2025 at 00:00 UTC. The four products within each row share the same color scale to facilitate direct comparison.

Table 4
Comparison of ERA5, VMF3, and model-corrected IWV against radiosonde observations for the independent temporal test dataset.

Season	Dataset	RMSE (mm)	R ²	MBE (mm)
Cold	ERA5	2.29	0.84	1.51
Cold	VMF3	1.88	0.89	0.57
Cold	Model-corrected	1.11	0.96	0.02
Warm	ERA5	3.44	0.82	1.98
Warm	VMF3	3.09	0.85	0.68
Warm	Model-corrected	2.06	0.93	0.03

observations to 2.347 mm for the model-corrected product, corresponding to a reduction of 44.5%. The MAE was reduced from 3.125 to

1.921 mm, while R² increased from 0.728 to 0.916. These results demonstrate that spatially collocated model inference becomes increasingly beneficial when the nearby radiosonde station is less representative of atmospheric conditions at the GNSS location. Sensitivity analysis further showed that the RMSE reduction increased from 26.6% for separations below 50 km to 34.2% for 50–100 km and 46.3% for 100–150 km. No consistent dependence on elevation difference was identified, whereas temporal mismatch had little influence because 99.45% of the observations were matched within 30 min, for which the RMSE reduction remained 28.5%. Spatial collocation therefore reduces the representativeness mismatch associated with site separation, although residual differences may remain because of local moisture variability, terrain and elevation differences, and small temporal offsets between the observing systems.

Table 5
Comparison of spatially collocated radiosonde-informed IWV and nearby radiosonde IWV relative to GNSS-derived IWV. RMSE reduction is calculated relative to the nearby radiosonde IWV.

	Product	N	RMSE (mm)	MAE (mm)	MBE (mm)	R ²	RMSE reduction (%)
Overall	Model-corrected IWV	47418	2.25	1.86	−1.73	0.927	28.6
	Nearby radiosonde IWV	47418	3.15	2.20	−1.72	0.857	
Cold-season	Model-corrected IWV	23764	2.03	1.78	−1.73	0.877	21.6
	Nearby radiosonde IWV	23764	2.60	1.96	−1.76	0.800	
Warm-season	Model-corrected IWV	23654	2.44	1.94	−1.74	0.899	32.4
	Nearby radiosonde IWV	23654	3.61	2.43	−1.67	0.779	
Core pairs	Model-corrected IWV	44956	2.24	1.85	−1.73	0.927	27.1
	Nearby radiosonde IWV	44956	3.07	2.15	−1.70	0.863	
Extended pairs	Model-corrected IWV	2462	2.35	1.92	−1.75	0.916	44.5
	Nearby radiosonde IWV	2462	4.23	3.13	−2.06	0.728	

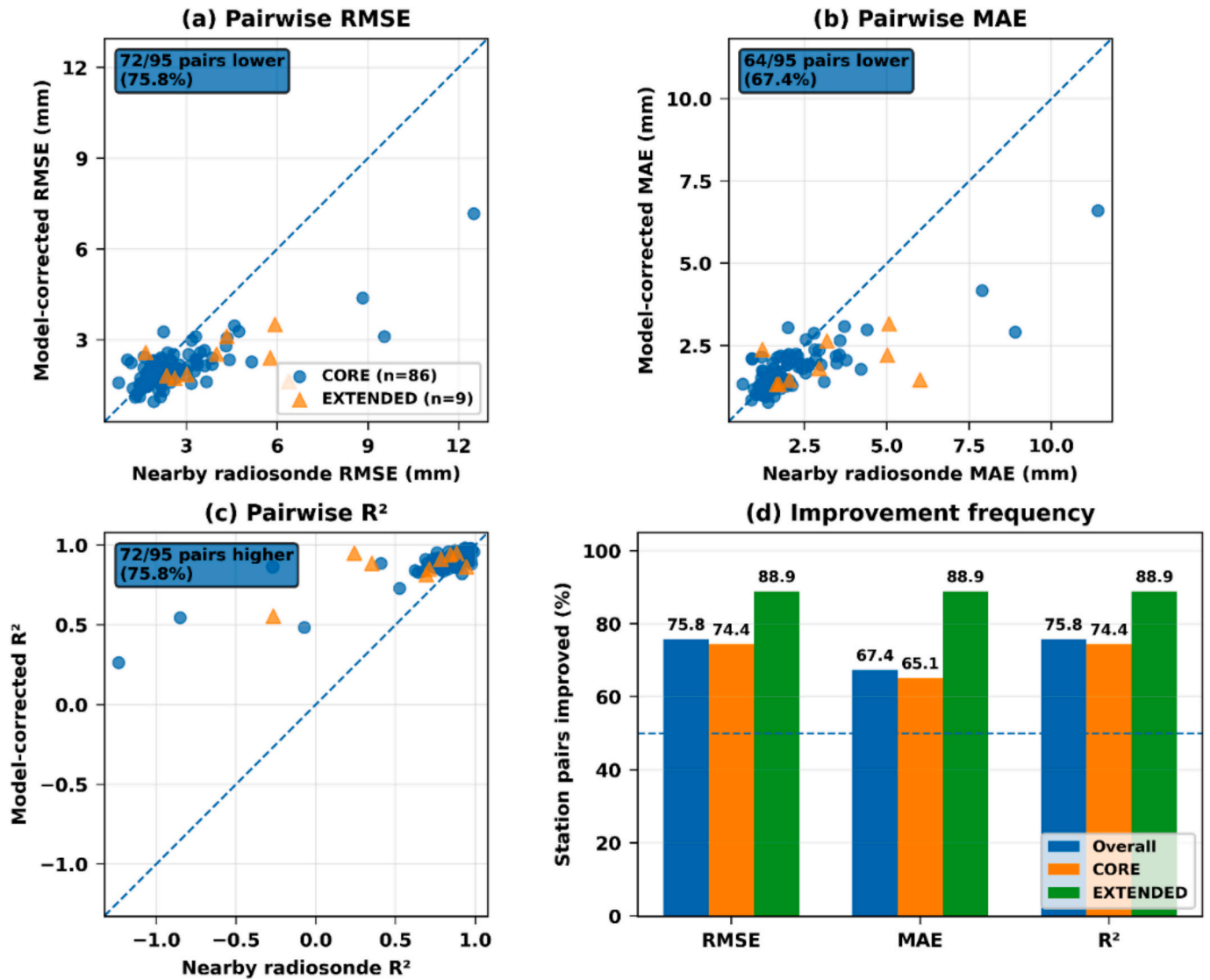


Fig. 12. Pairwise comparison of model-corrected IWV and nearby radiosonde IWV relative to GNSS-derived IWV. Panels show station-pair-specific comparisons of (a) RMSE, (b) MAE, and (c) R^2 , together with (d) the percentage of station pairs for which the model-corrected product achieved improved performance. Circles and triangles denote CORE and EXTENDED radiosonde-GNSS pairs, respectively. The dashed lines in panels (a)–(c) indicate the 1:1 relationship. Values below the line in panels (a) and (b), and above the line in panel (c), indicate better performance of the model-corrected IWV. Only station pairs containing at least 30 common observations were included.

Consequently, the proposed framework provides a radiosonde-informed, spatially collocated IWV comparison product at external GNSS locations. By transferring radiosonde-constrained correction relationships to the GNSS coordinates, it reduces representativeness differences associated with nearby, non-collocated sounding observations. The resulting product therefore supports external evaluation and inter-comparison of GNSS-derived IWV where collocated radiosonde observations are unavailable.

5. Conclusion

This study developed a radiosonde-informed CatBoost framework to correct ERA5-derived IWV and generate spatially continuous, model-corrected IWV products. Radiosonde observations from the UoW and IGRA were combined with ERA5, VMF3, and atmospheric structural features at 116 stations across Europe and adjacent regions from November 2021 to November 2025. Separate evaluations were conducted for the cold and warm seasons to account for seasonal differences in atmospheric water vapor variability.

The proposed framework demonstrated strong spatial and temporal generalization. In the station holdout experiment, the corrected IWV achieved RMSE values of 1.33 and 2.40 mm and R^2 values of 0.94 and 0.92 for the cold and warm seasons, respectively. For the independent temporal test, the corresponding RMSE values were 1.11 and 2.06 mm, with R^2 values of 0.96 and 0.93 and MBE values close to zero. Compared with ERA5, RMSE was reduced by 51.5% in the cold season and 40.1% in the warm season. Reductions of 41.0% and 33.3%, respectively, were also obtained relative to VMF3. The corrected fields preserved the large-scale spatial structure of ERA5 while improving agreement with radiosonde observations. Model application was restricted to elevations represented in the training dataset to avoid unsupported extrapolation.

The trained model was further transferred to external GNSS locations not involved in model development to generate spatially collocated, radiosonde-informed IWV estimates. Relative to nearby radiosonde observations, the RMSE decreased from 3.15 to 2.25 mm, demonstrating the value of transferring radiosonde information directly to GNSS locations and reducing representativeness differences associated with non-collocated soundings. This benefit increased with horizontal

separation, while no consistent dependence on elevation difference was identified. The resulting product therefore provides a collocated comparison reference for GNSS-derived IWV where radiosonde observations are unavailable. Because the GNSS-IWV conversion uses ERA5-derived pressure and weighted mean temperature, the evaluation is independent of model development but not fully independent of ERA5 auxiliary information.

The present applicability is therefore limited to the geographical, elevation, and atmospheric conditions represented by the training data, with greater uncertainty under highly localized or extreme moisture conditions. Future work should extend the training coverage globally and further evaluate model performance under such conditions.

CRediT authorship contribution statement

Lingke Wang: Writing – original draft, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Hansjörg Kutterer:** Writing – review & editing, Supervision, Project administration, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was financed by the Institute of Geodesy (GIK) of Karlsruhe Institute for Technology. We sincerely thank the reviewers for their valuable feedback and suggestions.

Data availability

The radiosonde, ERA5, VMF3, and GNSS datasets used in this study are available from their respective public data providers. The pre-processing scripts and final model configuration files are publicly available on Zenodo at <https://doi.org/10.5281/zenodo.22936890>.

References

- Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., 2019. Optuna. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. <https://doi.org/10.1145/3292500.3330701>.
- Beck, H.E., Zimmermann, N.E., McVicar, T.R., Vergopolan, N., Berg, A., Wood, E.F., 2018. Present and Future Köppen-Geiger Climate Classification Maps at 1-Km Resolution. <https://doi.org/10.6084/M9.FIGSHARE.6396959>.
- Bennouna, Y.S., Torres, B., Cachorro, V.E., Ortiz de Galisteo, J.P., Toledano, C., 2013. The evaluation of the integrated water vapour annual cycle over the Iberian Peninsula from EOS-MODIS against different ground-based techniques. *Quart. J. Royal Meteorol. Soc.* 139, 1935–1956. <https://doi.org/10.1002/qj.2080>.
- Bevis, M., Businger, S., Herring, T.A., Rocken, C., Anthes, R.A., Ware, R.H., 1992. GPS meteorology: remote sensing of atmospheric water vapor using the global positioning system. *J. Geophys. Res. Atmos.* 97, 15787–15801. <https://doi.org/10.1029/92JD01517>.
- Bevis, M., Businger, S., Chiswell, S., Herring, T.A., Anthes, R.A., Rocken, C., Ware, R.H., 1994. GPS meteorology: mapping zenith wet delays onto precipitable water. *J. Appl. Meteorol.* 33, 379–386. [https://doi.org/10.1175/1520-0450\(1994\)033<0379:gmmzwd>2.0.co;2](https://doi.org/10.1175/1520-0450(1994)033<0379:gmmzwd>2.0.co;2).
- Blewitt, G., Hammond, W.C., Kreemer, C., 2018. Harnessing the GPS data explosion for interdisciplinary science. *Eos* 99. <https://doi.org/10.1029/2018EO104623>.
- Böhm, J., Schuh, H. (Eds.), 2013. Atmospheric Effects in Space Geodesy, Springer Atmospheric Sciences. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-36932-2>.
- Bojinski, S., Verstraete, M., Peterson, T.C., Richter, C., Simmons, A., Zemp, M., 2014. The concept of essential climate variables in support of climate research, applications, and policy. *Bull. Am. Meteorol. Soc.* 95 (9), 1431–1443.
- Ccoica-López, K.L., Pasapera-Gonzales, J.J., Jimenez, J.C., 2019. Spatio-temporal variability of the precipitable water vapor over Peru through MODIS and ERA-Interim time series. *Atmosphere* 10, 192. <https://doi.org/10.3390/atmos10040192>.
- Chen, Z., Xu, J., Zhang, H., 2019. Evaluation of HJ-1A/B CCD surface reflectance products using the VNIR and MODIS-based atmospheric correction approaches. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 12, 437–449. <https://doi.org/10.1109/JSTARS.2018.2875263>.
- Dach, R., Lutz, S., Walser, P., Fridez, P., 2015. Bernese GNSS Software Version 5.2. User Manual. Astronomical Institute, University of Bern, Bern Open Publishing. <https://doi.org/10.7892/BORIS.72297>.
- Dessler, A.E., Zhang, Z., Yang, P., 2008. Water vapor climate feedback inferred from climate fluctuations, 2003–2008. *Geophys. Res. Lett.* 35. <https://doi.org/10.1029/2008gl035333>.
- Deutscher Wetterdienst, 2025. Deutschlandwetter im August 2025: Warmer August 2025 mit viel Sonnenschein und Teils Erheblicher Trockenheit. Deutscher Wetterdienst.
- Du, Z., Yao, Y., Zhang, B., Zhao, Q., 2024. Precipitable water vapor estimation from Himawari-8/AHI observations using a stacking machine learning model. *Atmos. Res.* 301, 107281. <https://doi.org/10.1016/j.atmosres.2024.107281>.
- Durre, I., Vose, R.S., Wuertz, D.B., 2006. Overview of the integrated global radiosonde archive. *J. Clim.* 19, 53–68. <https://doi.org/10.1175/jcli3594.1>.
- Durre, I., Yin, X., Vose, R.S., Applequist, S., Arnfield, J., Korzeniewski, B., Hundermark, B., 2016. Integrated Global Radiosonde Archive (IGRA), Version 2. <https://doi.org/10.7289/V5X63K0Q>.
- Fragkos, K., Antonescu, B., Giles, D.M., Ene, D., Boldeanu, M., Efstathiou, G.A., Belegante, L., Nicolae, D., 2019. Assessment of the total precipitable water from a sun photometer, microwave radiometer and radiosondes at a continental site in southeastern Europe. *Atmospheric Meas. Tech.* 12, 1979–1997. <https://doi.org/10.5194/amt-12-1979-2019>.
- Gimeno, L., Eiras-Barca, J., Durán-Quesada, A.M., Dominguez, F., van der Ent, R., Sodemann, H., Sánchez-Murillo, R., Nieto, R., Kirchner, J.W., 2021. The residence time of water vapour in the atmosphere. *Nat. Rev. Earth Environ.* 2, 558–569. <https://doi.org/10.1038/s43017-021-00181-9>.
- Guerova, G., Jones, J., Douša, J., Dick, G., de Haan, S., Pottiaux, E., Bock, O., Pacione, R., Elgered, G., Vedel, H., Bender, M., 2016. Review of the state of the art and future prospects of the ground-based GNSS meteorology in Europe. *Atmos. Meas. Tech.* 9, 5385–5406. <https://doi.org/10.5194/amt-9-5385-2016>.
- Held, I.M., Soden, B.J., 2006. Robust responses of the hydrological cycle to global warming. *J. Clim.* 19, 5686–5699. <https://doi.org/10.1175/jcli3990.1>.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J., 2020. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* 146, 1999–2049. <https://doi.org/10.1002/qj.3803>.
- Hocke, K., Kämpfer, N., Gerber, C., Mätzler, C., 2011. A complete long-term series of integrated water vapour from ground-based microwave radiometers. *Int. J. Remote Sens.* 32, 751–765. <https://doi.org/10.1080/01431161.2010.517792>.
- Huang, L., Wang, X., Xiong, S., Li, J., Liu, L., Mo, Z., Fu, B., He, H., 2022. High-precision GNSS PWV retrieval using dense GNSS sites and in-situ meteorological observations for the evaluation of MERRA-2 and ERA5 reanalysis products over China. *Atmos. Res.* 276, 106247. <https://doi.org/10.1016/j.atmosres.2022.106247>.
- Jade, S., Vijayan, M.S.M., 2008. GPS-based atmospheric precipitable water vapor estimation using meteorological parameters interpolated from NCEP global reanalysis data. *J. Geophys. Res.* 113. <https://doi.org/10.1029/2007jd008758>.
- Jiang, C., Gao, X., Wang, S., Zhu, H., Xu, A., An, Q., Zhu, M., Liu, G., 2023. Comparison of ZTD derived from CARRA, ERA5 and ERA5-Land over the Greenland based on GNSS. *Adv. Space Res.* 72, 4692–4706. <https://doi.org/10.1016/j.asr.2023.09.002>.
- Kestin, J., Sengers, J.V., Kamgar-Parsi, B., Sengers, J.M.H.L., 1984. Thermophysical properties of fluid H₂O. *J. Phys. Chem. Ref. Data* 13, 175–183. <https://doi.org/10.1063/1.555707>.
- Kouba, J., 2007. Implementation and testing of the gridded Vienna mapping function 1 (VMF1). *J. Geod.* 82, 193–205. <https://doi.org/10.1007/s00190-007-0170-0>.
- Kouba, J., 2009. Testing of global pressure/temperature (GPT) model and global mapping function (GMF) in GPS analyses. *J. Geodyn.* 83, 199–208. <https://doi.org/10.1007/s00190-008-0229-6>.
- Landskron, D., Böhm, J., 2017. VMF3/GPT3: refined discrete and empirical troposphere mapping functions. *J. Geod.* 92, 349–360. <https://doi.org/10.1007/s00190-017-1066-2>.
- Li, H., Choy, S., Wang, X., Liang, H., Zhang, K., 2023. Monitoring the migration of water vapor using ground-based GNSS tropospheric products. *IEEE Geosci. Remote Sensing Lett.* 20, 1–5. <https://doi.org/10.1109/lgrs.2023.3282362>.
- Li, H., Zhu, G., Huang, L., Mo, Z., Kang, Q., 2024. An improved method for developing the precipitable water vapor vertical correction global grid model. *Atmos. Res.* 311, 107664. <https://doi.org/10.1016/j.atmosres.2024.107664>.
- Li, M., Yang, L., Ji, W., Bilal, M., Pei, X., Zheng, X., Fan, Y., Lu, X., Cheng, X., Du, W., 2025. Improving Fengyun-3D satellite atmospheric precipitable water vapor products through machine learning-based post-processing correction. *Atmos. Res.* 322, 108133. <https://doi.org/10.1016/j.atmosres.2025.108133>.
- Liou, Y.-A., Teng, Y.-T., Van Hove, T., Liljegren, J.C., 2001. Comparison of precipitable water observations in the near tropics by GPS, microwave radiometer, and radiosondes. *J. Appl. Meteorol.* 40, 5–15. [https://doi.org/10.1175/1520-0450\(2001\)040<0005:copwoi>2.0.co;2](https://doi.org/10.1175/1520-0450(2001)040<0005:copwoi>2.0.co;2).
- Lu, R., Zhang, M., Yuan, P., Li, Z., Chen, W., Cai, M., Chen, Y., Dong, D., Jiang, W., 2024. Enhancing multi-GNSS positioning performances in harsh environments via a refined joint troposphere-multipath hemispherical map. *GPS Solutions* 29. <https://doi.org/10.1007/s10291-024-01763-y>.
- Ma, Y., Liu, T., Yu, Z., Jiang, C., Xu, G., Lu, Z., 2023. All-weather precipitable water vapor map reconstruction using data fusion and machine learning-based spatial

- downscaling. *Atmos. Res.* 296, 107068. <https://doi.org/10.1016/j.atmosres.2023.107068>.
- Ma, X., Huang, H., Zhang, Q., Zhang, B., Zhao, Q., Lin, X., Huang, Y., Zhao, Y., Yao, Y., 2026. A hybrid deep learning and empirical model for short-term spatio-temporal ZTD forecasting for China. *J. Geod.* 100, 27. <https://doi.org/10.1007/s00190-026-02050-z>.
- Marut, G., Hadas, T., Kaplon, J., Trzcina, E., Rohm, W., 2022. Monitoring the water vapor content at high spatio-temporal resolution using a network of low-cost multi-GNSS receivers. *IEEE Trans. Geosci. Remote Sensing* 60, 1–14.
- McCarthy, M.P., Thorne, P.W., Titchner, H.A., 2009. An analysis of tropospheric humidity trends from radiosondes. *J. Clim.* 22, 5820–5838. <https://doi.org/10.1175/2009JCLI2879.1>.
- Météo France, 2025. Vague de chaleur du 8 au 18 août 2025: Retour sur un épisode exceptionnel. *Météo France*.
- Namaoui, H., Kahlouche, S., Belbachir, A.H., 2021. Evaluation of MODIS water vapour products over ALGERIA using radiosonde data. *Anu. Inst. Geosci.* 44. https://doi.org/10.11137/1982-3908.2021_44.40110.
- Ning, T., Wang, J., Elgered, G., Dick, G., Wickert, J., Bradke, M., Sommer, M., Querel, R., Smale, D., 2016. The uncertainty of the atmospheric integrated water vapour estimated from GNSS observations. *Atmos. Meas. Tech.* 9, 79–92. <https://doi.org/10.5194/amt-9-79-2016>.
- Oikonomou, C., Karolos, I.-A., Bitharis, S., Pikridas, C., Haralambous, H., 2023. Performance of a low-cost dual-frequency GNSS receiver for near real-time water vapor estimation. In: 2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW).
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2017. CatBoost: unbiased boosting with categorical features. *arXiv*. doi:10.48550/ARXIV.1706.09516.
- Seidel, D.J., Sun, B., Pettey, M., Reale, A., 2011. Global radiosonde balloon drift statistics. *J. Geophys. Res.* 116. <https://doi.org/10.1029/2010jd014891>.
- Sherwood, S.C., Roca, R., Weckwerth, T.M., Andronova, N.G., 2010. Tropospheric water vapor, convection, and climate. *Rev. Geophys.* 48. <https://doi.org/10.1029/2009RG000301>.
- Shi, J., Li, M., Steiner, A.K., Li, W., Zhang, M., Fan, Y., Gao, W., Zhang, K., 2025. Stacking machine learning model for precipitable water vapor vertical adjustment using GNSS networks and radio occultation data. *Atmos. Res.* 325, 108212. <https://doi.org/10.1016/j.atmosres.2025.108212>.
- Ssenyunzi, R.C., Oruru, B., D'ujanga, F.M., Realini, E., Barindelli, S., Tagliaferro, G., von Engeln, A., van de Giesen, N., 2020. Performance of ERA5 data in retrieving precipitable water vapour over East African tropical region. *Adv. Space Res.* 65, 1877–1893. <https://doi.org/10.1016/j.asr.2020.02.003>.
- Tayfehrostami, A., Amerian, Y., 2026. A new model for estimating atmospheric weighted mean temperature from radiosonde and multi-mission GNSS radio occultation data. *Geodesy Geodyn.* 17, 197–210. <https://doi.org/10.1016/j.geog.2025.09.009>.
- Tayfehrostami, A., Amerian, Y., Izanlou, S., Azadi, M., 2026. TOMO4D operator for assimilation of GNSS tomography-derived water vapor fields into the WRFDA 4DVAR system to improve regional rainfall forecasting. *J. Adv. Model. Earth Syst.* 18, e2025MS005610. <https://doi.org/10.1029/2025MS005610>.
- Trenberth, K., 2011. Changes in precipitation with climate change. *Clim. Res.* 47, 123–138. <https://doi.org/10.3354/cr00953>.
- Wang, L., Kutterer, H., 2025. An advanced regional integrated water vapor estimation model utilizing least squares support vector machine for the upper Rhine Graben Region. *GPS Solutions* 29. <https://doi.org/10.1007/s10291-024-01796-3>.
- Wang, W., Murk, A., Sauvageat, E., Fan, W., Dätwyler, C., Hervo, M., Haefele, A., Hocke, K., 2023. An indoor microwave radiometer for measurement of tropospheric water. *IEEE Trans. Geosci. Remote Sensing* 61, 1–13.
- Wang, Y., Yang, F., Li, P., Gong, X., Liu, M., Xu, T., Lin, X., Wang, Y., 2024. An optimal calibration method for MODIS precipitable water vapor using GNSS observations. *Atmos. Res.* 309, 107591. <https://doi.org/10.1016/j.atmosres.2024.107591>.
- Wang, L., Wang, D., Awange, J., Kutterer, H., 2025. High-resolution integrated water vapor estimation using the Gaussian mixed long short-term memory network: a satellite-based intercomparison and data fusion. *IEEE Trans. Geosci. Remote Sensing* 63, 1–20. <https://doi.org/10.1109/tgrs.2025.3638337>.
- Wang, D., Wang, L., Kutterer, H., 2025a. An advanced tropospheric delay model based on Gaussian mixed long short-term memory network. *IEEE Trans. Geosci. Remote Sensing* 63, 1–14. <https://doi.org/10.1109/tgrs.2025.3549124>.
- Wu, J., Ma, H., Wu, W., Cheng, D., 2025. Monitoring water vapor transport in near real-time with low-cost GNSS receiver network. *Sci. Rep.* 15. <https://doi.org/10.1038/s41598-025-10603-z>.
- Xia, P., Chen, B., Huang, N., Xie, X., Zhang, Q., 2025. A high-precision real-time PWV grid model for the China region and its preliminary performance in WRF assimilation. *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.* 18, 3433–3447. <https://doi.org/10.1109/JSTARS.2025.3525770>.
- Xu, J., Liu, Z., 2022. Enhanced all-weather precipitable water vapor retrieval from MODIS near-infrared bands using machine learning. *Int. J. Appl. Earth Obs. Geoinf.* 114, 103050. <https://doi.org/10.1016/j.jag.2022.103050>.