Journal of Global Positioning Systems (2026)
Vol. 22, Issue 1: 57-74
DOI:10.5081/jgps.22.1.57

Deep Learning-Driven Synergistic Integration of Satellite-Derived Water Vapor and Meteorological Radar for Enhanced Urban Precipitation Nowcasting: A Spatiotemporal Attention Architecture

Jing Sun[1,4,5], **Jiale Wang**[1,2,4,5]✉, **Chuang Shi**[1,4,5], **Yunchang Cao**[3,4], **Hong Liang**[3,4], **Yuyang Zhu**[3,4], **Wu Chen**[2,4]

1 School of Space and Earth Sciences, Beihang University, Beijing 100191, China

2 Department of Land Surveying and Geospatial Science, Hongkong Polytechnic University, Hongkong, China

3 Meteorological Observation Center, China Meteorological Administration, Beijing 100081, China

4 Key Laboratory of Satellite Navigation and Mobile Communication Fusion Technology, Ministry of Industry and Information Technology, Beijing 100191, China

5 National Key Laboratory of CNS/ATM, Beijing 100191, China

✉ the corresponding author, wang_jiale@buaa.edu.cn

Abstract: Intensifying climate variability has escalated the frequency and severity of localized extreme precipitation events, demanding innovative forecasting methodologies to bolster urban resilience and disaster mitigation. Conventional prediction systems struggle with the rapid onset and high-impact nature of short-duration intense rainfall. This investigation presents a heterogeneous data fusion architecture that synergistically combines satellite-based observations—specifically GNSS-derived Precipitable Water Vapor (PWV)—with ground-based Radar Reflectivity Composites (RRC) to advance precipitation nowcasting capabilities. GNSS-PWV serves as a sensitive indicator of atmospheric moisture accumulation dynamics, while RRC characterizes the spatial-temporal evolution of precipitation systems. We propose a Spatiotemporal Attention-Enhanced Swin U-Net (STAE-Swin) architecture that leverages Swin Transformer modules coupled with an integrated spatiotemporal attention mechanism to effectively assimilate these complementary data streams. The study domain encompasses the Beijing-Tianjin-Hebei metropolitan region, utilizing high-precision PWV retrievals from 220 GNSS Continuously Operating Reference Stations (CORS) via dual-frequency ionosphere-free Precise Point Positioning (PPP) methodology (validation correlation with ERA5 reanalysis: >0.99). Key Findings: Against independent validation datasets from the 2025 flood season, the model demonstrated 68% accuracy in detecting torrential precipitation events. Comparative analysis reveals that the integrated multi-source approach yields a 18.5% improvement in Critical Success Index (CSI) and a 21.5% enhancement in Probability of Detection (POD) relative to single-modality baseline models. Implications: The results substantiate that strategic integration of satellite-derived geophysical parameters with advanced deep learning architectures substantially elevates early-warning precision, offering a scientifically robust and operationally viable solution for regional hazard prevention and climate adaptation strategies.

Keywords: GNSS Atmospheric Remote Sensing; Satellite Meteorology; Precipitation Prediction; Artificial Intelligence; Attention Mechanisms; Extreme Weather Forecasting

1. Introduction

The intensification of the global hydrological

Editor-In-Charge: Dr. Yi Ding

cycle under anthropogenic climate forcing has substantially increased both the frequency and destructive potential of Extreme Precipitation Events (EPEs), posing unprecedented challenges to urban infrastructure and public safety. The Beijing–Tianjin–Hebei (BTH) metropolitan region, a critical economic and demographic center in East Asia, is particularly vulnerable to such hazards. The region's distinctive physiographic setting—characterized by the convergence of moisture-laden East Asian monsoon systems with the pronounced topographic relief of the Yanshan and Taihang mountain ranges—creates a dynamically unstable atmospheric environment[1]. This orographic-monsoon interaction frequently generates short-lived, high-intensity convective disturbances exhibiting rapid morphological changes, highly nonlinear evolution patterns, and pronounced spatial-temporal variability [2]. Compounding these natural vulnerabilities, accelerated urbanization across the BTH corridor has intensified the urban heat island phenomenon and localized precipitation enhancement effects, thereby increasing the unpredictability of storm trajectories and rainfall distribution [3]. The occurrence of recent catastrophic inundation events has demonstrated that high-fidelity precipitation nowcasting transcends academic meteorology, becoming instead a fundamental requirement for safeguarding urban systems, mitigating flash-flood hazards, and optimizing emergency management protocols [4]. The BTH region's complex terrain—combining mountainous, transitional, and densely urbanized zones with recurrent summer convective extremes—renders it an ideal natural laboratory for investigating short-term intense precipitation phenomena and associated societal risks.

Contemporary operational nowcasting frameworks are anchored upon two primary methodological pillars: Numerical Weather Prediction (NWP) systems and radar-based kinematic extrapolation. Although NWP models grounded in fundamental fluid dynamics principles provide authoritative guidance for extended-range forecasts, they are inherently constrained by the spin-up problem. Owing to imperfect observational initialization and incomplete parameterization of subgrid-scale processes, NWP systems frequently demonstrate limited skill in resolving convective cell genesis during the initial forecast hour, thereby creating a critical forecasting gap in immediate warning applications [5]. In contrast, radar-centric extrapolation methodologies reformulate precipitation forecasting as a computer vision task. While computationally expedient, these purely kinematic approaches fundamentally disregard atmospheric thermodynamic processes. By assuming Lagrangian persistence and tracking existing hydrometeor fields, they cannot represent storm cell genesis or decay mechanisms driven by unobserved water vapor instability and buoyancy forcing [6]. Consequently, the predictive accuracy of pure kinematic extrapolation deteriorates markedly beyond 30–45 minutes, as substantiated by comprehensive validation studies and operational performance assessments [7].

Bridging the methodological chasm between kinematic observation and thermodynamic evolution necessitates the incorporation of thermodynamic environmental precursors. Fundamentally, precipitation genesis is governed by phase transitions of atmospheric water vapor. The Global Navigation Satellite System (GNSS) has established itself as a dependable, all-weather remote sensing technique for characterizing atmospheric moisture fields [8]. Precise Precipitable Water Vapor (PWV) estimates derived from GNSS signal processing furnish critical information regarding moisture accumulation preceding precipitation onset [9]. Emerging research has substantiated that PWV fluctuations exhibit strong temporal correlation with convective triggering mechanisms and intense rainfall development [10]. Nevertheless, GNSS observations are inherently sparse in spatial distribution and represent vertically integrated atmospheric properties, whereas weather radar provides dense, high-resolution depictions of precipitation system structure [11]. This fundamental asymmetry between sparse thermodynamic measurements and spatially comprehensive radar observations constitutes a significant impediment to accurate short-term precipitation forecasting.

From an observational systems perspective, distinct monitoring platforms yield complementary geophysical information. Satellite-based remote sensing captures synoptic-scale cloud-top dynamics but predominantly reflects upper-tropospheric characteristics [12]. Meteorological radar remains the preeminent instrument for resolving precipitation morphology and convective system evolution, although its measurements are compromised by terrain masking, signal attenuation, and antenna beam blockage [13]. Conversely, GNSS meteorology enables quasi-continuous surveillance of atmospheric water vapor and facilitates detection of pre-convective environmental anomalies that are not directly discernible from radar reflectivity fields [14]. Antecedent investigations have demonstrated that PWV anomalies and their temporal gradients function as reliable diagnostic indicators for heavy precipitation occurrence [15].

The emergence of machine learning and deep neural network architectures has catalyzed a paradigm shift in precipitation nowcasting, transitioning from conventional extrapolation schemes toward data-driven deep learning frameworks [16]. Convolutional Neural Networks (CNNs) have achieved widespread adoption owing to their proficiency in spatial feature extraction, yet their restricted receptive fields constrain the representation of extended-range spatial dependencies [17]. Recurrent architectures including ConvLSTM have been deployed to capture spatiotemporal dynamics, although they frequently encounter training convergence difficulties and computational parallelization constraints [18]. Encoder–decoder structures exemplified by U-Net have demonstrated exceptional capability in hierarchical multi-scale feature learning and have become prevalent in precipitation nowcasting applications [19]. Nevertheless, standard convolution-based methodologies remain fundamentally limited to local neighborhood operations, potentially restricting their capacity to represent large-scale interactions within organized mesoscale convective systems [20].

Recent advances in attention-based neural architectures have unlocked novel possibilities for precipitation nowcasting. Transformer-based frameworks enable substantially improved modeling of global spatial dependencies and intricate spatiotemporal interactions [21]. Within the broader meteorological forecasting domain, deep learning methodologies have demonstrated considerable promise in capturing nonlinear atmospheric dynamics and complex physical processes [22]. Furthermore, contemporary breakthroughs in global-scale weather prediction leveraging deep learning underscore the efficacy of data-assimilative approaches in representing multifaceted atmospheric phenomena [23]. These developments collectively suggest that synergistic combination of hierarchical feature extraction with global receptive field modeling is indispensable for advancing nowcasting skill. However, prevailing methodologies predominantly emphasize radar-only information streams and conceptualize precipitation forecasting as a sequential prediction task, while inadequately leveraging thermodynamic constraints embedded within water vapor evolution patterns [24].

Regarding multi-source data integration, recent investigations have examined the synergistic coupling of GNSS-PWV and radar measurements, revealing that thermodynamic precursor signals can augment predictive performance beyond radar-exclusive systems [25]. Nonetheless, the spatial resolution mismatch between sparse GNSS networks and dense radar grids introduces non-trivial challenges in data representation and fusion architecture, potentially amplifying model uncertainty [26]. Consequently, developing a robust computational framework capable of jointly exploiting the temporal dynamics of GNSS-derived moisture signals and the spatial structure of radar echo patterns remains a fundamental challenge in operational short-term precipitation nowcasting [27].

This research endeavors to investigate whether the incorporation of GNSS-derived atmospheric moisture precursor information can enhance forecasting accuracy for short-term localized heavy precipitation, particularly during the initiation and intensification phases of convective systems across the Beijing–Tianjin–Hebei region. To address this scientific question, we propose a Spatiotemporal Attention-Enhanced Swin U-Net (STAE-Swin) architecture and establish a collaborative forecasting framework integrating GNSS PWV and radar composite reflectivity (CR) data streams.

The principal scientific and technical contributions of this work are:

1. **High-Precision Ground-Based GNSS Atmospheric Characterization:** We constructed a comprehensive meteorological dataset for the BTH region by retrieving GNSS PWV (30-minute temporal resolution, <3 mm retrieval accuracy) utilizing the dual-frequency ionosphere-free Precise Point Positioning (PPP) technique. This dataset provides a rigorous ground-based characterization of atmospheric thermodynamic precursor conditions, serving as a foundational input for climate resilience and extreme weather analysis.
2. **Intelligent Multi-Modal Data Assimilation Architecture:** We developed the STAE-Swin framework, an AI-driven system that effectively integrates the 1D temporal evolution of GNSS-PWV measurements with 2D spatial radar reflectivity patterns. Through incorporation of Swin Transformer modules and spatiotemporal attention mechanisms, the model successfully captures the nonlinear coupling dynamics between GNSS-derived water vapor precursors and precipitation system evolution.
3. **Specialized Optimization for Extreme Event Detection:** To facilitate operational early-warning systems, we implemented a composite Edge-Aware loss function that balances pixel-level prediction accuracy with geometric structural fidelity. This optimization strategy effectively mitigates the characteristic "smoothing artifact" inherent in deep learning nowcasting systems, ensuring robust detection sensitivity for high-impact torrential precipitation events and supporting reliable decision-making for urban flood risk management.

The paper is structured as follows: Section 2 describes the observational datasets, GNSS PWV retrieval methodology, and the architectural design principles of the Spatiotemporal Attention-Enhanced Swin U-Net model; Section 3 outlines the experimental configuration and validation protocols; Section 4 presents and interprets the experimental results; Section 5 synthesizes the principal findings and discusses prospective research directions.

2. Materials and Method

2.1 GNSS-Derived Precipitable Water Vapor Inversion via Dual-Frequency Ionosphere-Free Precise Point Positioning

This investigation utilized dual-frequency GNSS observations (30-second sampling interval) collected from 220 Continuously Operating Reference Stations (CORS) operated by the China Meteorological Observation Network (CMONC) distributed across the Beijing–Tianjin–Hebei region. The temporal coverage spans the May–August 2025 period, encompassing the primary convective season. The raw observational dataset comprises carrier phase and pseudorange measurements from GPS satellite constellations, supplemented by station-specific metadata including precise coordinate information and antenna specifications. Prior to PWV inversion calculations, comprehensive quality assurance procedures were implemented on the raw observations, encompassing cycle slip identification and correction utilizing the TurboEdit algorithm and systematic removal of anomalous measurements. Refined antenna phase center corrections were applied using Antenna Phase Center Offset (PCO) and Variation (PCV) parameters distributed by the International GNSS Service (IGS), thereby ensuring measurement fidelity and computational precision [28].

Rather than employing conventional differential positioning methodologies or single-equation solution approaches, this study implemented the more sophisticated and high-precision un-differenced un-combined Precise Point Positioning (PPP) technique. The GNSS observation system was formulated as a dynamic state-space estimation problem solved via Extended Kalman Filtering. For receiver station r and satellite S, the linearized dual-frequency ionosphere-free combination observation equation—designed to eliminate the dominant first-order ionospheric delay component (typically exceeding 99% of total ionospheric error)—is expressed in compact matrix notation as:

$$y_{IF} = Ax + \epsilon \tag{1}$$

where $\boldsymbol{y_{IF}}$ *denotes the observation vector comprising pseudorange* ($P_{\boldsymbol{IF}}$) and carrier phase ($L_{\boldsymbol{IF}}$) measurements after ionosphere-free linear combination, effectively attenuating signal degradation during periods of elevated ionospheric activity; A represents the design matrix containing satellite-to-receiver geometric direction cosines and tropospheric wet delay mapping function coefficients; and $\boldsymbol{\epsilon}$ characterizes measurement noise, whose stochastic properties are modeled through elevation-angle-dependent weighting functions. The state vector $\boldsymbol{x}$ to be estimated comprises:

$$x = [r_r^T, c\delta t_r, ZTD, N_{IF}^T]^T \tag{2}$$

In this formulation, $\boldsymbol{r}_r$ *represents station coordinate corrections accounting for minute-scale displacements induced by tidal loading or crustal deformation;* $c\delta t_r$ *denotes the receiver clock bias, modeled as a white noise process; ZTD (Zenith Total Delay) constitutes the primary meteorological parameter of interest, characterized as a random walk process; and* $\boldsymbol{N_{IF}}$ represents the floating-point ambiguity vector following ionosphere-free combination. To attain millimeter-level solution accuracy, IGS-provided precise satellite ephemeris and clock correction products were incorporated to eliminate satellite-side orbital and temporal errors, while the Global Mapping Function (GMF) was applied to project slant-path delays onto the zenith direction [29], thereby enabling effective utilization of low-elevation-angle observations.

Following extraction of high-precision Zenith Total Delay (ZTD) from the PPP Kalman filter, decomposition into constituent Zenith Hydrostatic Delay (ZHD) and Zenith Wet Delay (ZWD) components is required. The hydrostatic component, which exhibits high spatiotemporal stability and depends primarily on surface atmospheric pressure, is computed via the established Saastamoinen formulation [30]:

$$ZHD = \frac{0.0022768 \cdot P_s}{1 - 0.00266\cos(2\varphi) - 0.00028H} \tag{3}$$

where P_s denotes station surface pressure, φ represents station latitude, and H specifies station geodetic elevation. The wet delay component, which exhibits pronounced spatiotemporal variability and maintains intimate association with precipitation processes, is isolated through subtraction:

$$\mathrm{ZWD} = \mathrm{ZTD} - \mathrm{ZHD} \tag{4}$$

Precipitable Water Vapor (PWV) retrieval is accomplished through coupling of the Zenith Wet Delay with a thermodynamic conversion coefficient. Following the water vapor thermodynamic state equation formulated by Askne and Nordius [31], the mathematical relationship between PWV and ZWD is expressed as:

$$PWV = \frac{ZWD}{\rho_w \cdot \kappa(\overline{T_m})} \tag{5}$$

where ρ_w represents a dimensionless conversion function incorporating liquid water density, and $\kappa(\overline{T_m})$ denotes the atmospheric weighted mean temperature conversion coefficient. The rigorous mathematical formulation of this coefficient integrates empirical atmospheric refractivity relationships:

$$\kappa(\overline{T_m}) = 10^{-6} \cdot R_v \cdot \left(k_2^{'} + \frac{k_3}{\overline{\overline{T_m}}}\right) \tag{6}$$

In this expression, physical constants conform to internationally standardized values [32]: R_v represents the specific gas constant for water vapor, while $k_2^{'}$ and k_3 denote atmospheric refractivity empirical coefficients. The atmospheric weighted mean temperature $\overline{T_m}$ is estimated through application of a regionalized linear regression model calibrated using station-level surface temperature measurements, thereby optimizing inversion accuracy within the BTH geographic domain. The conversion coefficient κ exhibits adaptive sensitivity to water vapor inversion under varying thermal conditions, ensuring robust PWV retrieval performance during severe convective meteorological regimes.

It is noteworthy that GNSS-derived PWV fundamentally represents vertically integrated atmospheric moisture content and is operationalized in this investigation as a thermodynamic precursor signal (manifesting 1–2 hours antecedent to precipitation occurrence) rather than as a diagnostic tool for resolving minute-scale convective-scale moisture fluctuations.

2.2 Radar Composite Reflectivity Data Processing

In this study, radar composite reflectivity (CR) mosaic data covering the Beijing–Tianjin–Hebei region were adopted as one of the primary meteorological datasets. The temporal interval of the radar observations was 6 minutes, and the spatial resolution after unified resampling was adjusted to (0.1°×0.1°). The choice of this resolution mainly considered the distribution density of GNSS stations and the need to maintain consistency between radar grids and PWV data during multi-source feature fusion. Composite reflectivity represents the strongest radar echo within the vertical atmospheric column at each grid location, which can effectively reflect the spatial distribution and intensity variation of hydrometeors inside cloud systems.

The original radar observations are stored in polar-coordinate form and are influenced by non-meteorological interference such as ground clutter. Therefore, the raw data cannot be directly used as tensor inputs for deep learning models. To enable coordinated fusion between radar reflectivity and GNSS PWV data while satisfying the input requirements of the Swin-UNet framework, several preprocessing procedures were carried out before model training.

1) Noise Filtering and Normalization Processing: Radar echoes are often contaminated by abnormal signals generated by terrain reflections, biological targets, or other environmental interference. Referring to the quality-control approach proposed by Zhang et al. [33], a statistical threshold method was applied to eliminate invalid observations. According to the (3σ) criterion, reflectivity values lower than −30 dBZ or higher than 70 dBZ were identified as abnormal samples and replaced with 0 dBZ. This operation mainly aims to suppress noise effects during convolutional feature extraction. It should be noted that a value of 0 dBZ does not strictly indicate no precipitation; instead, it generally corresponds to extremely weak echoes.

After removing abnormal values, the reflectivity fields were further scaled using linear normalization before tensor generation. By constraining the data into a unified numerical interval, the optimization process during network training becomes more stable, and the convergence efficiency of batch learning can be improved. At the same time, normalization reduces the influence of calibration differences among radar scans at different times, allowing the model to focus more on structural evolution characteristics of precipitation systems rather than absolute intensity variations.

2) Temporal and Spatial Alignment of Multi-Source Data: GNSS PWV observations are discrete point-based sequences, whereas radar data are continuous gridded fields. As a result, the two datasets differ in both temporal frequency and spatial representation. To establish consistency between heterogeneous data sources, temporal and spatial registration procedures were performed.

For temporal matching, a nearest-time strategy was employed to synchronize radar observations with GNSS measurements. Let the radar observation time set be represented as $\mathcal{T} * radar$. For an arbitrary GNSS observation time $\mathrm{t} * \mathrm{gnss}$, the radar frame with the smallest temporal difference was selected as the matched sample [34]:

$$t^* * radar = \mathrm{ar\,g\,m\,in} * t \in \mathcal{T} * radar |t - t * gnss| \quad (7)$$

Using this method, all datasets were unified under a consistent 6-minute temporal interval.

In the spatial dimension, the original radar mosaics are commonly generated under Lambert conformal conic projection or radar polar coordinates. Since GNSS station information is expressed in the WGS84 geographic coordinate system, coordinate conversion was first performed to establish a common spatial framework. Afterwards, nearest-neighbor interpolation was adopted to project radar reflectivity values onto the target grid point (lon_i, lat_j). Through this process, each tensor element could correspond to an actual geographic location, thereby providing spatial consistency for subsequent channel stacking and feature fusion. Compared with higher-order interpolation methods, nearest-neighbor interpolation introduces fewer smoothing effects and preserves local echo characteristics more effectively.

3) Missing Data Filling and Sample Augmentation: Because of terrain obstruction, beam shielding, and radar scanning limitations, local missing regions may appear in some radar images. To maintain the continuity of CNN input features, bilinear interpolation was applied to repair incomplete grid areas. Suppose the target missing point is denoted as $P(x, y)$, and its surrounding valid grid points are $Q_{11}, Q_{12}, Q_{21}, Q_{22}$. The interpolated value can then be estimated as:

$$Z(x, y) \approx \frac{1}{(x_2 - x_1)(y_2 - y_1)} \sum_{i=1}^{2} \sum_{j=1}^{2} w_{ij} Z(Q_{ij}) \quad (8)$$

where w_{ij} represents the weighting coefficient determined according to the relative distance between the target point and neighboring reference points.

In addition, meteorological datasets usually exhibit an obvious imbalance problem because severe convective and heavy rainfall events occur much less frequently than ordinary precipitation processes. To alleviate the long-tail distribution issue during model training, an online data augmentation strategy was introduced. Specifically, random rotations together with horizontal and vertical flipping operations were applied to the input tensors. By artificially increasing the diversity of extreme precipitation samples, the model becomes capable of learning more robust spatial features and gains better adaptability when recognizing precipitation systems with different orientations, scales, and morphologies.

2.3 Design of the Spatio-Temporal Enhanced Attention U-Net Model

To improve precipitation nowcasting performance under complex atmospheric conditions, this study develops a Spatio-Temporal Enhanced Attention U-Net model based on the Swin Transformer framework, referred to as STEA-Swin. The proposed model preserves the encoder–decoder structure of the conventional U-Net architecture while replacing standard convolutional layers with Swin Transformer modules to better capture long-range spatial dependencies and temporal evolution characteristics. In addition, the Convolutional Block Attention Module (CBAM) [35] is embedded into the skip connections to strengthen critical meteorological features and suppress irrelevant background noise. A composite loss function integrating edge-aware constraints and geometric similarity is also introduced to alleviate the common problems of blurred precipitation boundaries and underestimated extreme rainfall intensity in precipitation forecasting tasks. It should be emphasized that the major contribution of this research is not limited to network optimization itself but rather lies in the physically guided fusion of GNSS-derived PWV data and radar observations. By combining thermodynamic and dynamic atmospheric information, the model aims to improve the predictability of short-term precipitation evolution.

The STEA-Swin model adopts an end-to-end encoder–decoder framework, directly establishing a mapping relationship between historical multi-source meteorological observations and future precipitation fields. The input tensor is represented as $(\mathrm{B}, \mathrm{H}, \mathrm{W}, \mathrm{T}_{\mathrm{in}} \times \mathrm{C})$, where the historical sequence contains 10-time steps with an interval of 6 minutes, corresponding to one hour of observational history. The input variables mainly include radar composite reflectivity, GNSS PWV, Digital Elevation Model (DEM) data, and temporal encoding features.

The encoder stage is composed of multiple Swin Transformer Blocks and Patch Merging layers. During the down sampling process, the spatial resolution of feature maps is gradually reduced while the number of feature channels increases correspondingly, allowing the model to extract meteorological semantic information at different spatial scales. The bottleneck layer performs deep feature abstraction at the lowest spatial resolution and captures the large-scale evolution patterns of atmospheric systems.

The decoder stage progressively restores the spatial resolution of feature maps through Patch Expanding layers combined with Swin Transformer Blocks. In this process, high-level semantic information is gradually mapped back to fine-resolution meteorological grids. The output layer adopts a dual-task prediction strategy. One branch is responsible for continuous precipitation intensity regression, while the auxiliary branch performs precipitation-category classification. By jointly optimizing regression and classification objectives, the model improves its ability to distinguish precipitation levels, particularly during heavy rainfall events.

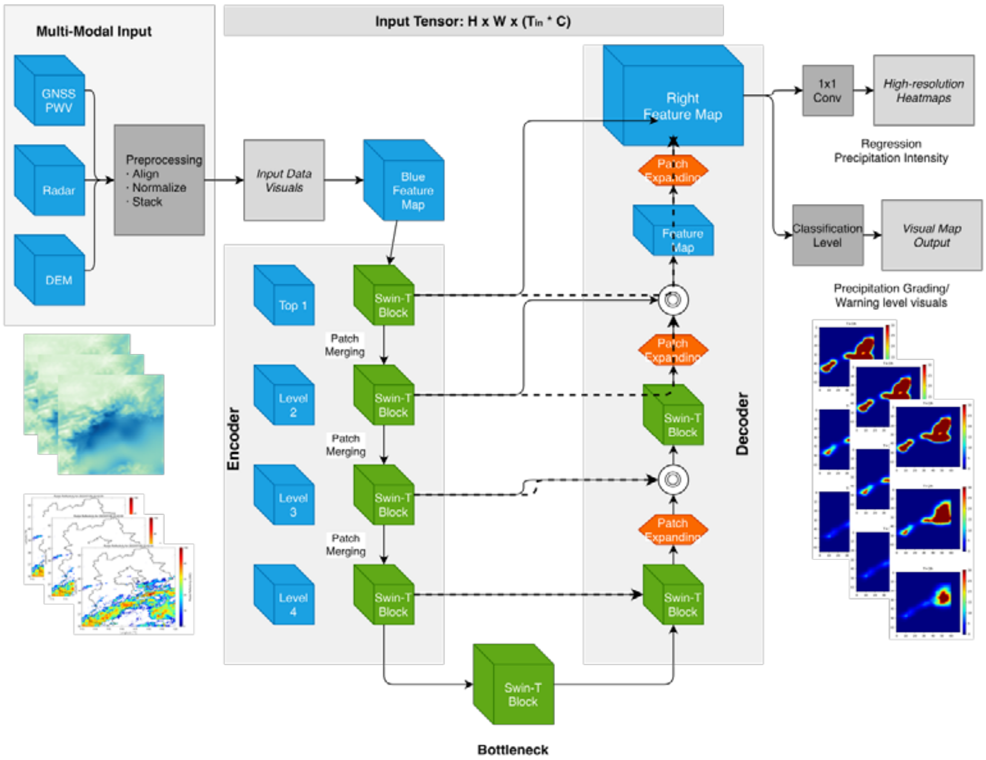


Figure 1. Architecture of the proposed Spatio-Temporal Enhanced Attention Swin U-Net (STEA-Swin) precipitation forecast model

To overcome the limited receptive field problem of traditional convolutional neural networks, the proposed model employs Swin Transformer as the fundamental feature extraction unit. The Swin Transformer mainly consists of Window-based Multi-head Self-Attention (W-MSA) and Shifted Window Multi-Head Self-Attention (SW-MSA), which achieve efficient feature interaction while maintaining relatively low computational complexity.

Assuming the input feature is z^{l-1}, the computation process of the Swin Transformer Block can be expressed as:

$$\begin{aligned} \hat{z}^{l} &= \mathrm{W-MSA}\left(\mathrm{LN}(z^{l-1})\right) + z^{l-1} \\ z^{l} &= \mathrm{MLP}\left(\mathrm{LN}(\hat{z}^{l})\right) + \hat{z}^{l} \\ \hat{z}^{l+1} &= \mathrm{SW-MSA}\left(\mathrm{LN}(z^{l})\right) + z^{l} \\ z^{l+1} &= \mathrm{MLP}\left(\mathrm{LN}(\hat{z}^{l+1})\right) + \hat{z}^{l+1} \end{aligned} \tag{9}$$

where LN denotes Layer Normalization [36], which is used to improve the stability of deep network training, and MLP represents the Multi-Layer Perceptron. W-MSA restricts self-attention calculations to local non-overlapping windows. Compared with traditional global attention mechanisms, the computational complexity is reduced from $O(N^2)$ to approximately linear complexity $O(N)$, significantly improving computational efficiency. SW-MSA further establishes information exchange among neighboring windows through cyclic shifting operations, enabling the model to simultaneously capture local convective development and large-scale precipitation evolution.

In the traditional U-Net structure, skip connections usually rely on direct feature concatenation. However, this approach often introduces shallow-layer background noise into the decoder stage, including ground clutter and non-meteorological echoes. To address this issue, this study introduces the CBAM attention mechanism into the skip connections to adaptively fuse shallow texture features from the encoder and deep semantic information from the decoder. CBAM mainly contains two components: Channel Attention and Spatial Attention.

The Channel Attention module compresses spatial dimensions through global average pooling and max pooling, allowing the network to automatically identify the feature channels most relevant to precipitation forecasting, such as strong echo centers or regions with significant PWV gradients, while suppressing irrelevant interference. The Spatial Attention module performs pooling operations along the channel dimension to generate a two-dimensional attention map, guiding the model to focus on critical spatial regions such as rainband leading edges and terrain-induced uplift zones. The feature fusion process is expressed as:

$$F' = M_c(F) \otimes F, \quad F'' = M_s(F') \otimes F' \tag{10}$$

where F represents the input feature map, M_c and M_s denote the channel attention map and spatial attention map respectively, and $\otimes$ indicates element-wise multiplication. Through this mechanism, background noise can be effectively suppressed while the feature response of extreme precipitation core regions is significantly enhanced.

In precipitation forecasting tasks, the traditional Mean Squared Error (MSE) loss function often produces overly smooth prediction results, causing the loss of high-frequency details such as intense precipitation centers and sharp rainband boundaries. To address this limitation, this study designs a composite loss function named EdgeAwareLoss, which constrains the model from the perspectives of numerical accuracy, geometric consistency, and edge texture preservation.

The first component is the Huber Loss (L_{huber}) [37], which is mainly used for pixel-level precipitation intensity regression. Unlike MSE, which is highly sensitive to outliers, Huber Loss behaves as a squared error for small deviations and as a linear error for large deviations. Therefore, it provides stronger robustness when dealing with extreme precipitation samples and prevents gradient instability caused by overfitting abnormal values.

The second component is the Log-Cosh Dice Loss (L_{dice}) [38], which is used to improve the geometric matching of precipitation regions. After introducing Log-Cosh smoothing, the Dice-based loss function can be written as:

$$L_{dice} = log\left(cosh\left(1 - \frac{2\sum y_{true}\,y_{pred}}{\sum y_{true} + \sum y_{pred}}\right)\right) \quad (11)$$

This loss function has smoother gradients and is more suitable for backpropagation optimization. At the same time, it improves the Critical Success Index (CSI) of precipitation prediction by making the predicted precipitation region more consistent with the observed distribution.

The third component is the Sobel Edge Loss (L_{edge}) [39]. This term uses the Sobel operator to extract horizontal and vertical gradients from both prediction maps and ground truth maps, thereby forcing the model to learn precipitation boundary variations:

$$L_{edge} = \frac{1}{N}\sum\left|\nabla y_{true} - \nabla y_{pred}\right| \quad (12)$$

By incorporating edge constraints, the predicted precipitation field maintains sharper boundary structures and more accurately represents the contours of severe convective systems.

The overall loss function is defined as:

$$L_{total} = L_{huber} + \lambda_1 L_{dice} + \lambda_2 L_{edge} \quad (13)$$

After hyperparameter optimization, the weighting coefficients are set to $\lambda_1 = 1.0$ and $\lambda_2 = 0.5$ to balance pixel-level regression accuracy and structural similarity. In addition, the auxiliary classification branch further constrains the semantic understanding capability of the network, ensuring that precipitation intensity prediction and precipitation-category recognition maintain high accuracy simultaneously.

The overall framework of the proposed system is illustrated in Figure 2. The framework mainly consists of four components: the multi-source data preprocessing layer, the spatio-temporal feature encoding layer, the multi-scale feature decoding layer, and the dual-task forecasting output layer. The preprocessing layer performs temporal synchronization, spatial registration, and normalization for GNSS PWV, radar CR, and DEM data, followed by channel stacking to construct unified multi-modal tensors. The spatio-temporal feature encoding layer adopts Swin Transformer as the backbone network, capturing long-range spatio-temporal dependencies through shifted-window attention mechanisms while using Patch Merging for hierarchical downsampling. The multi-scale decoding layer restores spatial resolution through Patch Expanding operations and introduces CBAM attention modules within skip connections to adaptively fuse shallow texture details with deep semantic information, thereby alleviating feature inconsistency between encoder and decoder stages. Finally, the forecasting output layer employs both regression and classification heads, optimized jointly using the edge-aware composite loss function, to generate high-resolution precipitation intensity maps and graded warning results for future time periods, achieving accurate and fine-grained precipitation nowcasting.

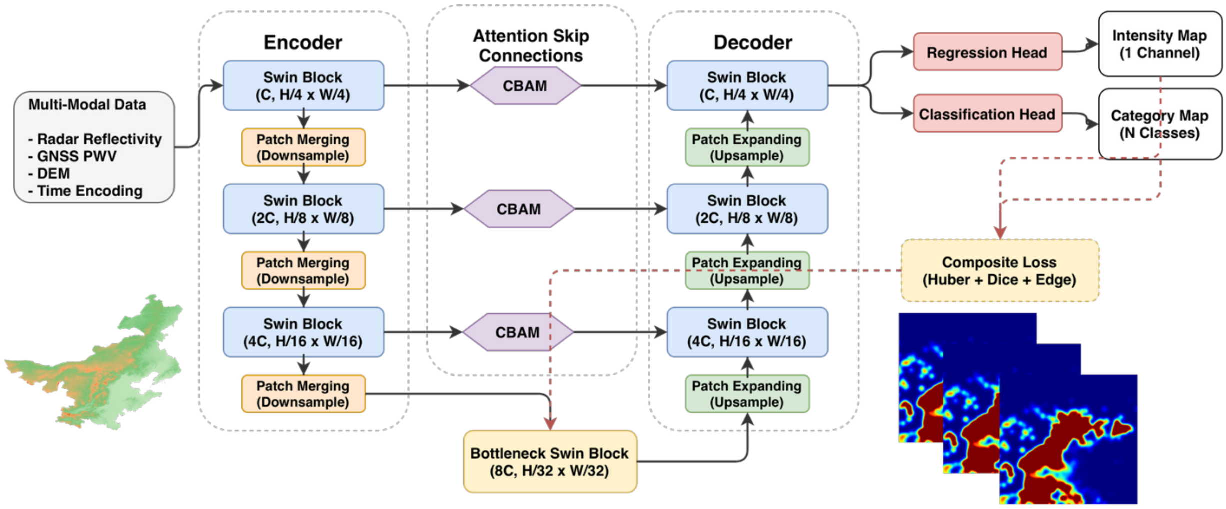


Figure 2. Overall architecture of the proposed STEA-Swin based multi-modal collaborative precipitation forecast model. The architecture includes a multi-source data input layer, Swin Transformer encoder, CBAM attention fusion module, and dual-task output heads

3. Experimental Design

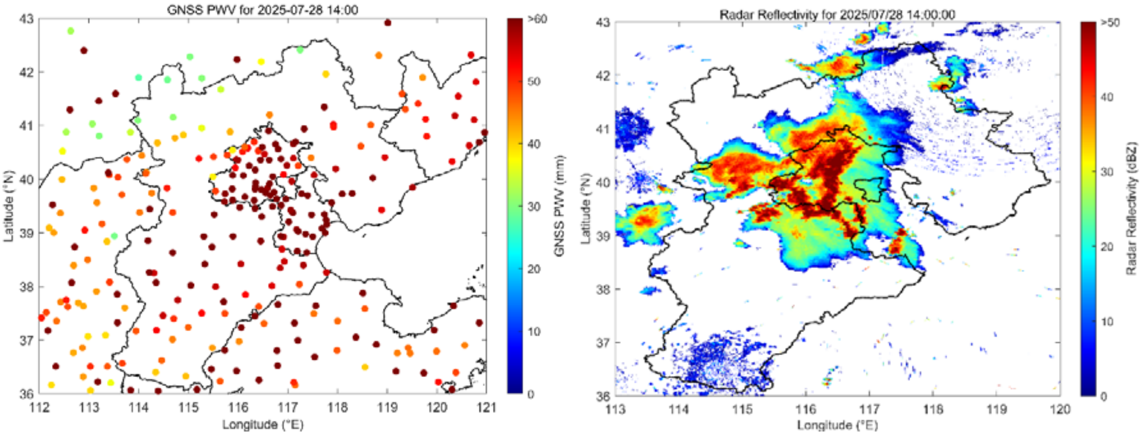
3.1 Study Area and Dataset Construction

The present study focuses on the Beijing-Tianjin-Hebei (BTH) region as the target area and develops a multi-source meteorological dataset integrating GNSS-derived atmospheric water vapor, radar reflectivity, and surface precipitation measurements. The dataset encompasses the period from May 1 to August 31, 2025, corresponding to the peak flood season in this region. For dataset partitioning, observations from May through July were designated as the training set to facilitate model feature learning, whereas August data were reserved for evaluation purposes, with 20% allocated for validation and the remaining 80% for testing. This temporal splitting approach mimics operational nowcasting conditions, where models are trained on historical observations and then deployed to forecast future events, minimizing temporal leakage and providing a realistic evaluation of the model's generalization performance.

Specifically, the experimental dataset comprises: GNSS Precipitable Water Vapor (PWV) measurements obtained from 220 Continuously Operating Reference Stations (CORS) across the BTH region; radar composite reflectivity mosaics covering the same area; and rainfall data collected from 200 ground meteorological stations, which serve as reference labels for training and validating the model. As illustrated in Figure 3(a), GNSS stations are evenly distributed throughout the principal urban centers of Beijing, Tianjin, and Hebei, while also extending coverage to mountainous areas susceptible to heavy precipitation. This distribution ensures that the PWV data provide comprehensive spatial representation within the study domain. Figure 3(b) presents a radar composite reflectivity mosaic at 14:00 on July 28, 2025, corresponding to a typical heavy rainfall episode. During this event, the central and southern BTH region displayed pronounced strong reflectivity cores (>50 dBZ), closely matching the spatial extent of observed torrential rain.

To guarantee the reliability and physical consistency of multi-source inputs, a structured preprocessing workflow was applied before model training. First, a sliding window filter was used to detect and correct anomalous spikes in the GNSS PWV time series, effectively reducing observational noise while maintaining temporal continuity. Second, spatial interpolation was performed to fill missing radar data points, enhancing the completeness of radar coverage. Finally, all datasets were resampled to a uniform temporal interval of 6 minutes and a spatial resolution of 0.1° × 0.1°, producing a standardized high-resolution spatiotemporal tensor suitable for deep learning-based fusion. This preprocessing procedure ensures proper spatiotemporal alignment among heterogeneous data sources and provides consistent, high-quality inputs for model training.

The selected 0.1° spatial resolution represents a compromise between GNSS station density and radar coverage uniformity, supporting stable multimodal data fusion while acknowledging limitations in resolving fine-scale convective structures below the kilometer scale.



(a) (b)

Figure 3. Examples of multi-source input data over the study domain (28 Jul. 2025, 14:00 UTC). (a) Spatial Distribution of GNSS PWV stations. (b) Mosaic of radar composite reflectivity.

3.2 Experimental Environment and Training Strategy

All experiments were conducted on a high-performance computing platform designed for large-scale spatiotemporal deep learning tasks. The computational environment consisted of Intel Xeon Gold 6348 CPUs and NVIDIA A100 GPUs with 80 GB VRAM, accompanied by 128 GB system memory to support efficient processing of high-dimensional meteorological datasets. The software framework was implemented in Python 3.9, with PyTorch 2.0 and TensorFlow 2.10 serving as the primary deep learning libraries.

To achieve stable optimization and effective extraction of heterogeneous spatiotemporal features, the training scheme and hyperparameter settings were systematically configured. Model optimization was performed using the AdamW optimizer with an initial learning rate of 0.0001. The batch size was fixed at 16, and the network was trained for 300 epochs. In order to simultaneously improve precipitation intensity estimation and spatial localization capability, a customized edge-aware hybrid loss function was introduced. This objective function integrates three complementary components: Huber Loss for continuous precipitation regression, Log-Cosh Dice Loss for precipitation area classification, and Sobel Edge Loss for preserving fine-scale spatial boundary information. A weighted fusion strategy was adopted in which the regression component was assigned a coefficient of 1.0, while the classification-related component received a coefficient of 0.5, thereby prioritizing accurate rainfall intensity reconstruction while retaining sensitivity to precipitation morphology and spatial distribution.

The adopted temporal resolution of 6 minutes and spatial resolution of 0.1° were selected as a compromise between resolving rapidly evolving convective systems and maintaining numerical robustness during multi-source feature fusion. In addition, Digital Elevation Model (DEM) information was incorporated into the input features to help the network capture terrain-modulated precipitation processes, especially in mountainous regions where orographic effects significantly influence rainfall evolution. This terrain-assisted strategy improves the spatial realism and representativeness of precipitation forecasts.

All experimental procedures, including dataset preparation, preprocessing operations, and model training settings, were developed under a unified Materials and Methods framework to ensure methodological transparency and reproducibility in accordance with standard scientific research practices.

It should be noted that the selected spatial resolution of approximately 10 km (corresponding to a 75 × 75 grid within the study domain) was primarily constrained by the computational characteristics of transformer-based spatiotemporal networks. Although such grid sizes are relatively lightweight for conventional Numerical Weather Prediction (NWP) systems operating on CPU clusters, deep learning architectures based on self-attention mechanisms impose substantially higher GPU memory requirements. In the Swin-UNet framework adopted in this study, multi-channel and multi-temporal input sequences must be processed concurrently within limited GPU VRAM. If the spatial resolution were increased to 1 km (750 × 750 grids), the number of spatial tokens would increase dramatically, causing the computational cost and memory consumption of the self-attention operation to grow quadratically, i.e., $\mathcal{O}(\mathrm{N}^2)$, where NN denotes the number of spatial tokens. Under current hardware conditions, such high-resolution sequence modeling would lead to immediate GPU out-of-memory (OOM) errors during training. Therefore, the adopted 10 km resolution represents a practical compromise between computational feasibility and predictive performance, enabling the model to process extended spatiotemporal sequences while preserving a sufficiently large receptive field for regional precipitation evolution analysis.

4. Results and Analysis

4.1 Accuracy Evaluation of GNSS-Retrieved PWV

To ensure that the moisture inputs ingested by the STEA-Swin model possessed both physical credibility and quantitative reliability, a systematic accuracy evaluation of the GNSS-derived Precipitable Water Vapor (PWV) was first conducted. The validation interval spanned May to August 2025, a window deliberately chosen to coincide with the principal summer flood season of the study area. During this period, precipitation events occur frequently and atmospheric water vapor exhibits substantial temporal variability—conditions that rigorously challenge the retrieval scheme under meteorologically extreme regimes.

The fifth-generation atmospheric reanalysis product (ERA5) released by the European Centre for Medium-Range Weather Forecasts (ECMWF) was adopted as the reference benchmark [40]. As the operational successor to ERA-Interim, ERA5 employs an advanced 4D-Var assimilation framework that ingests heterogeneous observational sources, including satellite-based remote sensing, radiosonde profiles, and surface station measurements. Such an integration substantially enhances the dynamical consistency and quantitative fidelity of the resulting atmospheric state fields. The product delivers globally seamless coverage at a horizontal grid spacing of 0.25° × 0.25° and an hourly temporal cadence, enabling fine-scale characterization of the rapid evolution of atmospheric moisture. By virtue of its spatiotemporal continuity and the broad validation it has received in prior studies, ERA5 PWV is currently regarded by the international meteorological and geodetic communities as a credible reference for evaluating the performance of ground-based GNSS water vapor retrieval algorithms and for diagnosing regional moisture climatology [41]. To enable rigorous point-by-point comparison, ERA5 PWV time series spatially and temporally collocated with each GNSS station were obtained via bilinear interpolation [42], thereby providing an objective foundation for the subsequent error diagnostics.

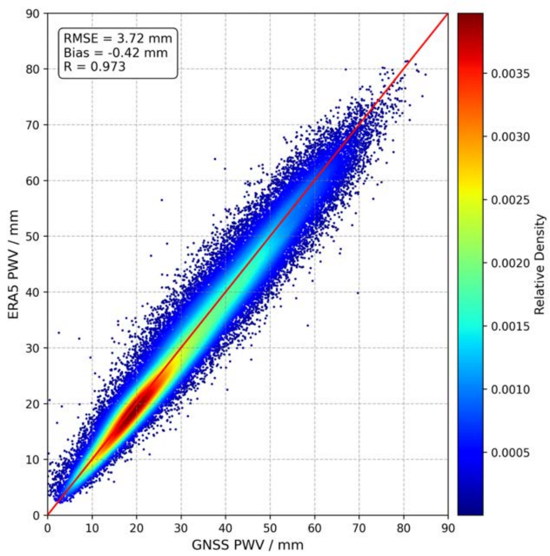


Figure 4. Scatter density plot illustrating the relationship between GNSS-derived and ERA5 PWV from May to August 2025.

The aggregate validation outcome is presented in Figure 4. Statistical diagnostics reveal that the GNSS-retrieved PWV exhibits pronounced agreement with the ERA5 reference. Quantitatively, the Pearson correlation coefficient (R) attains 0.973, signifying a strong positive linear association; the Root Mean Square Error (RMSE) is constrained to 3.72 mm, while the Mean Bias amounts to only −0.42 mm. Within the density plot, the bulk of paired samples cluster compactly along the 1:1 diagonal, and the high-density core (red region) virtually coincides with this reference line. Collectively, these features attest that the GNSS PWV retrieval is essentially free of systematic offset and exhibits neither pervasive overestimation nor underestimation. The product therefore demonstrates high precision and operational reliability, being capable of accurately quantifying the moisture loading of the regional atmosphere.

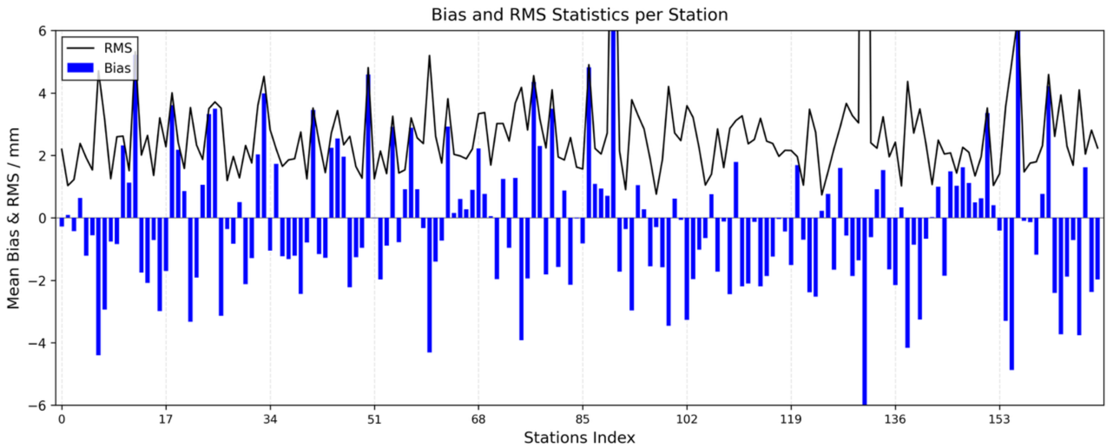


Figure 5. Station-wise histogram of GNSS PWV retrieval errors.

Figures 5 further elucidates the error structure from the perspectives of inter-station dispersion and spatial heterogeneity, respectively. As shown in Figure 5 (station-level error histogram), among the 153-plus observation sites, the Mean Bias of the majority of stations (blue bars) oscillates narrowly around zero and is predominantly contained within ±4 mm. Although the RMSE (black line) of a small subset of stations rises to 6–8 mm, the overall behavior remains stable and is sustained at approximately 3 mm.

4.2 Assessment of Precipitation Forecast Skill over the Beijing-Tianjin-Hebei Region

To rigorously and impartially benchmark the predictive capability of the STEA-Swin model across a spectrum of precipitation intensities, four standard verification metrics widely employed in operational meteorology were adopted: the Probability of Detection (POD), False Alarm Ratio (FAR), Critical Success Index (CSI), and Heidke Skill Score (HSS). The combined deployment of these indices constitutes a balanced diagnostic framework—POD characterizes the model's responsiveness to genuine precipitation occurrences, CSI synthesizes overall hit fidelity, FAR quantifies the suppression of spurious alarms, and HSS gauges the skill enhancement relative to random chance. In aggregate, this metric ensemble affords a holistic appraisal that addresses both detection sensitivity and operational reliability concurrently.

A suite of carefully designed ablation and lateral-comparison experiments was implemented, with the corresponding configurations summarized in Table 1. Two categories of baseline schemes were selected to serve as references for the proposed STEA-Swin:

1) Data-Source Ablation: To interrogate the necessity of multimodal data fusion, an architecturally identical variant ingesting radar echoes alone (denoted STEA-Swin) was constructed, thereby isolating the marginal contribution conferred by GNSS PWV assimilation to precipitation forecasting.

2） Architectural Benchmarking: To substantiate the merit of the Swin Transformer backbone, two canonical recurrent paradigms prevalent in atmospheric prediction were chosen as counterparts: a convolution-augmented LSTM (ConvLSTM) representing classical spatiotemporal sequence learning, and a sequence-to-sequence framework based on Gated Recurrent Units (GRU_Seq2Seq). This pairing enables a rigorous examination of the proposed model's competence in resolving long-range spatiotemporal dependencies and mitigating gradient vanishing during extended sequence prediction.

The heatmap presented in Figure 6 indicates that the proposed F_STEA-Swin attains the most favorable comprehensive scores across the four primary verification metrics—POD, FAR, CSI, and HSS. Specifically, under the torrential-rainfall category, the POD of the fused F_STEA-Swin reaches 0.676, marking a 21.5 % gain relative to the radar-only STEA-Swin counterpart (0.556); concurrently, the CSI ascends from 0.345 to 0.409, equivalent to an improvement of approximately 18.5 %. These gains substantiate that the synergistic coupling between GNSS-derived PWV and radar composite reflectivity (CR) effectively compensates for the moisture-evolution information that radar alone fails to register during the initiation and dissipation phases of intense convection, thereby alleviating the intensity underestimation that typically afflicts single-radar extrapolation.

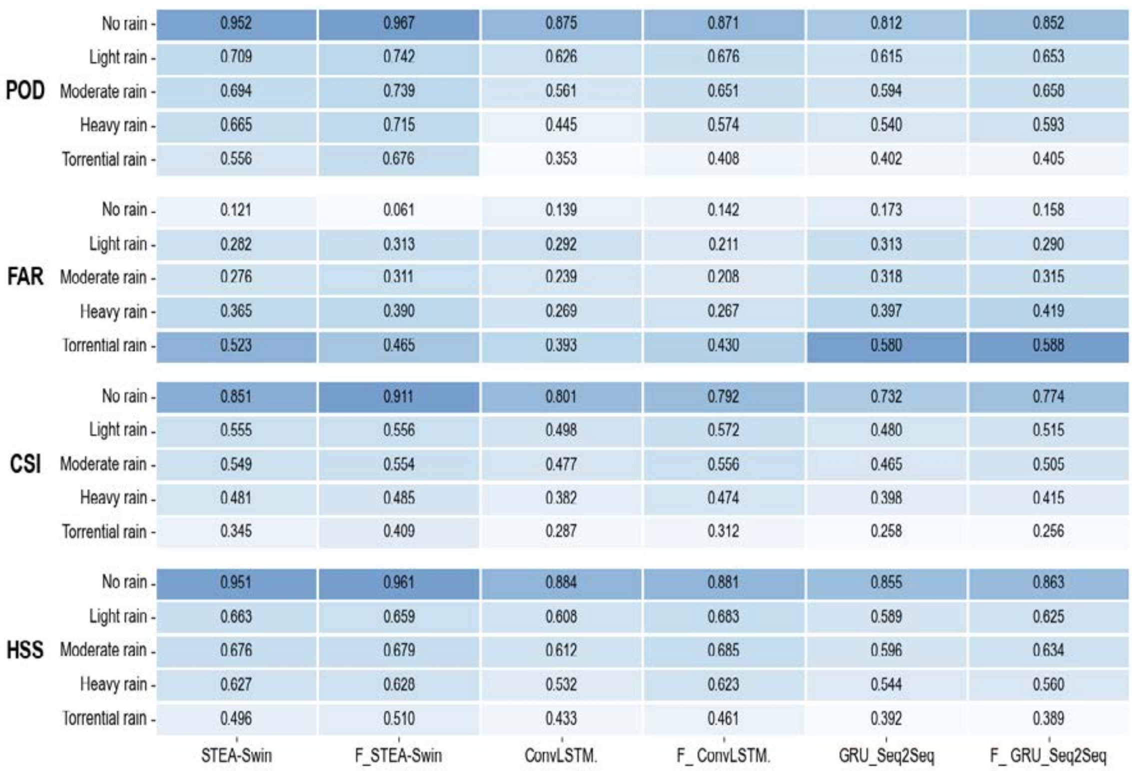
When juxtaposed across architectural families, the STEA-Swin series exhibits a consistent margin of superiority over both the ConvLSTM and GRU_Seq2Seq baselines, and this performance gap progressively widens as precipitation magnitude intensifies. In the torrential-rainfall regime, the CSI of STEA-Swin (0.409) markedly surpasses those of ConvLSTM (0.312) and GRU_Seq2Seq (0.256). This advantage is principally rooted in the hierarchical self-attention mechanism inherent to the Swin Transformer, which constructs global spatiotemporal dependencies and effectively circumvents the gradient attenuation and error accumulation that commonly afflict RNN-based architectures during prolonged forecast horizons. As a consequence, the structural fidelity of the predicted convective fields is preserved more reliably under highly variable severe-weather scenarios.

Examining the trajectory of the CSI metric across precipitation tiers, all evaluated models display a monotonic decline as rainfall intensity escalates. Nevertheless, the rate of degradation differs substantially: traditional architectures such as ConvLSTM undergo a precipitous CSI drop in the transition from heavy to torrential rainfall (0.474 → 0.312, equivalent to a 34 % reduction), whereas F_STEA-Swin sustains a markedly gentler attenuation (0.485 → 0.409, only a 15 % reduction). This contrast underscores the heightened responsiveness of the proposed model to extreme-value regimes and its superior robustness against perturbations. With regard to the trade-off between false alarms and missed events, F_STEA-Swin curtails the FAR at the torrential-rainfall threshold to 0.465, distinctly below the 0.588 recorded by F_GRU_Seq2Seq.

The architecture proposed herein is purposefully engineered for multi-step precipitation forecasting. Rather than yielding a solitary future snapshot, the model autoregressively emits predictions at hourly intervals extending up to a maximum lead time of three hours. Accordingly, each inference pass produces three temporally distinct frames corresponding to T+1 h, T+2 h, and T+3 h. This multi-horizon output paradigm enables a comprehensive interrogation of the model's capacity to track the continuous spatiotemporal evolution and intensity modulation of convective rainfall systems within the short-range nowcasting window. Building upon this design, the forecasting performance was further dissected across three lead times (+1 h, +2 h, and +3 h) to characterize the decay of forecast skill, as illustrated in Figure 7. The results reveal that, although all models inevitably suffer skill degradation as the lead time extends—a manifestation of the intrinsic chaotic dynamics of convective systems—STEA-Swin sustains comparatively higher stability across the entire prediction window. Both the CSI and POD of the proposed model decay at appreciably slower rates than those of the baseline schemes, attesting to its superior preservation of spatiotemporal coherence over extended forecast intervals.

Table 1. Experimental Groups

Groups	**Description**
F_STEA-Swin	STEA-Swin model (fusing radar and pwv)
STEA-Swin	STEA-Swin model (using radar data only)
F_ ConvLSTM	ConvLSTM model (fusing radar and pwv)
ConvLSTM	ConvLSTM model (using radar data only)
F_ GRU_Seq2Seq	GRU_Seq2Seq model (fusing radar and pwv)
GRU_Seq2Seq	GRU_Seq2Seq model (using radar data only)



		STEA-Swin	F_STEA-Swin	ConvLSTM.	F_ ConvLSTM.	GRU_Seq2Seq	F_ GRU_Seq2Seq
POD	No rain	0.952	0.967	0.875	0.871	0.812	0.852
	Light rain	0.709	0.742	0.626	0.676	0.615	0.653
	Moderate rain	0.694	0.739	0.561	0.651	0.594	0.658
	Heavy rain	0.665	0.715	0.445	0.574	0.540	0.593
	Torrential rain	0.556	0.676	0.353	0.408	0.402	0.405
FAR	No rain	0.121	0.061	0.139	0.142	0.173	0.158
	Light rain	0.282	0.313	0.292	0.211	0.313	0.290
	Moderate rain	0.276	0.311	0.239	0.208	0.318	0.315
	Heavy rain	0.365	0.390	0.269	0.267	0.397	0.419
	Torrential rain	0.523	0.465	0.393	0.430	0.580	0.588
CSI	No rain	0.851	0.911	0.801	0.792	0.732	0.774
	Light rain	0.555	0.556	0.498	0.572	0.480	0.515
	Moderate rain	0.549	0.554	0.477	0.556	0.465	0.505
	Heavy rain	0.481	0.485	0.382	0.474	0.398	0.415
	Torrential rain	0.345	0.409	0.287	0.312	0.258	0.256
HSS	No rain	0.951	0.961	0.884	0.881	0.855	0.863
	Light rain	0.663	0.659	0.608	0.683	0.589	0.625
	Moderate rain	0.676	0.679	0.612	0.685	0.596	0.634
	Heavy rain	0.627	0.628	0.532	0.623	0.544	0.560
	Torrential rain	0.496	0.510	0.433	0.461	0.392	0.389

Figure 6. Comparative diagnostic indicators of forecast accuracy among the evaluated deep-learning models across distinct precipitation grades.

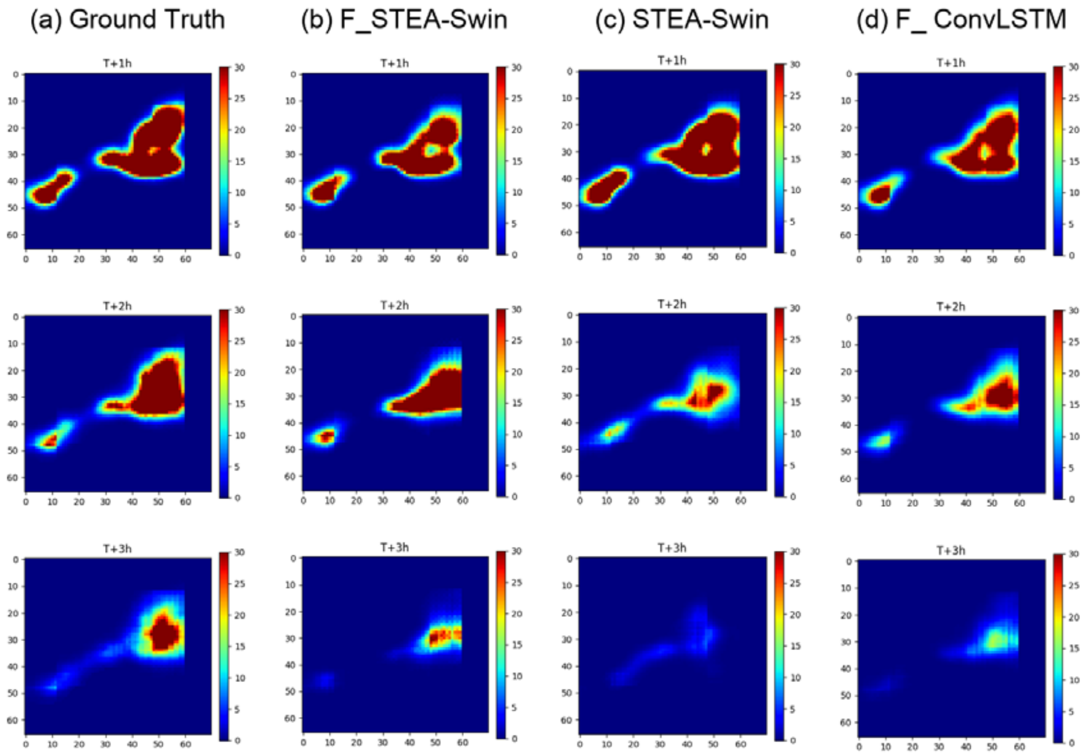


Figure 7. Side-by-side comparison of precipitation forecasts generated by (a) ground truth, (b) F_STEA-Swin, (c) STEA-Swin, and (d) F_ConvLSTM at lead times of +1 h, +2 h, and +3 h for the heavy-rainfall event initialized at 16:00 on 19 August 2025

Figure 7 visualizes the predictive behavior of F_STEA-Swin, STEA-Swin, and F_ConvLSTM relative to ground-truth observations for a severe rainfall event initialized at 16:00 on 19 August 2025, evaluated at the +1 h, +2 h, and +3 h horizons. By means of a discrete color scheme that explicitly delineates precipitation tiers, the visual evidence affirms that F_STEA-Swin delivers the most robust reconstruction. The model accurately renders the spatial morphology, sharp intensity gradients, and temporal progression of the convective ensemble; notably, the predicted heavy-rainfall cores (magenta zones, > 30 mm h^{-1}) remain strikingly consistent with the observations even at the +3 h horizon. By contrast, the radar-only STEA-Swin variant exhibits a pronounced underestimation of rainfall intensity, with the extreme cores noticeably eroded by T+2 h and T+3 h. The F_ConvLSTM, in turn, suffers from substantial structural smoothing and spatial dislocation, failing to retain the fine-scale boundaries of the convective system and producing an overly diffuse, attenuated forecast field. Collectively, these observations spotlight the marked advantage of F_STEA-Swin in reproducing the intricate spatiotemporal dynamics that characterize heavy precipitation.

In summary, by harmonizing multi-source physical descriptors with the global-receptive modeling capacity of the Transformer paradigm, the F_STEA-Swin model not only attains elevated accuracy under conventional precipitation conditions but also decisively suppresses the prevalent blurring artifacts associated with deep-learning extrapolation. The model achieves substantive progress in the accurate localization of heavy-precipitation cores and in the effective curtailment of spurious alarms, thereby substantiating its merit as an advanced next-generation nowcasting algorithm.

5. Conclusion and Future Work

Against the backdrop of intensifying extreme weather events driven by global warming, the establishment of dependable short-term forecasting techniques for heavy precipitation has become indispensable for safeguarding urban communities and underpinning timely disaster mitigation. Conventional forecasting methodologies frequently fall short of resolving the rapid evolution of convective storm systems, owing primarily to their constrained spatiotemporal granularity or their reliance on isolated observational platforms. Rather than tackling long-term climate adaptation, the present investigation centers on a seasonal-scale nowcasting paradigm that seeks to elevate forecasting precision through the synergistic exploitation of heterogeneous data sources. Specifically, by coupling Precipitable Water Vapor (PWV) retrieved from ground-based Global Navigation Satellite System (GNSS) networks with radar composite reflectivity within an artificial intelligence (AI)-driven architecture, this work substantiates the merit of jointly harnessing thermodynamic and dynamic atmospheric signatures to enhance nowcasting skill.

The principal findings of this study corroborate the effectiveness of both refined multi-GNSS processing strategies and multi-modal AI integration. Employing the dual-frequency ionosphere-free Precise Point Positioning (PPP) algorithm, we successfully derived high-resolution PWV fields from 220 Continuously Operating Reference Stations (CORS) distributed across the Beijing–Tianjin–Hebei (BTH) region. Cross-validation against ERA5 reanalysis yielded a correlation coefficient of 0.974 and a root-mean-square error (RMSE) of 3.63 mm, attesting to the capability of ground-based GNSS observations to faithfully characterize regional atmospheric water vapor dynamics. It merits emphasis that the PWV estimates utilized herein originate exclusively from ground-based GNSS networks rather than satellite-borne radiometric retrievals, thereby affording superior temporal granularity and pronounced sensitivity to localized moisture transitions.

To fully exploit these multi-source observations, we introduced the STEA-Swin architecture, purpose-designed to reconcile the dimensional heterogeneity between one-dimensional water vapor time series and two-dimensional radar reflectivity fields. By embedding Swin Transformer modules within a dual-attention scheme and supervising the network through a composite Edge-Aware loss function, the proposed model proved adept at capturing the intricate nonlinear interplay between large-scale moisture accumulation and microphysical hydrometeor evolution. Throughout the 2025 flood season, the integrated framework demonstrated marked superiority over single-source baselines: it attained a 68% hit rate for torrential rainfall episodes and elevated the Critical Success Index (CSI) by approximately 18.5% relative to radar-only counterparts, while concurrently curtailing the incidence of false alarms. Notably, these performance gains were consistently observed across forecast horizons spanning +1 h to +3 h, indicating that the proposed model preserves comparatively stable predictive competence as lead time increases.

Collectively, these outcomes underscore that embedding physically interpretable atmospheric variables—exemplified by GNSS-retrieved PWV—into deep learning frameworks can simultaneously sharpen detection sensitivity and refine the spatial structural fidelity of precipitation forecasts. Rather than positing a fully operational or universally robust solution, our findings provide compelling evidence that multi-source data assimilation constitutes a promising avenue for advancing short-term heavy precipitation nowcasting at the regional scale.

Notwithstanding these promising results, several caveats warrant acknowledgment. First, the present analysis is confined to a single flood season (May–August 2025), which inherently restricts the appraisal of inter-annual variability and long-term transferability of the model. Second, the adopted grid spacing of 0.1° may prove insufficiently fine to delineate sub-kilometer convective features, an issue rendered particularly acute within densely built urban landscapes. Third, residual uncertainties arising from topographically induced biases in PWV interpolation, together with artifacts associated with radar vertical-profile phenomena (notably the bright-band effect), merit dedicated future investigation.

Subsequent research will be oriented toward redressing the aforementioned shortcomings and broadening the operational scope of the proposed framework. We intend to enlarge the dataset to encompass multi-year archives, enabling rigorous appraisal of model robustness across heterogeneous synoptic regimes and rare extreme events. Furthermore, additional observational streams—such as geostationary satellite brightness temperatures and atmospheric vertical profiling—will be assimilated to construct a richer three-dimensional depiction of the pre-convective environment. To bolster physical fidelity, prospective work will explore the embedding of physically grounded constraints (e.g., atmospheric conservation laws) within the learning pipeline, alongside adaptive modeling strategies tailored to capture the multi-scale character of precipitation processes. Finally, the framework will be deployed and recalibrated across additional regions historically vulnerable to extreme rainfall, thereby facilitating location-specific optimization and a more comprehensive assessment of its cross-regional generalizability.

References

[1] Li, H.; Wang, X.; Wu, S.; Zhang, K.; Chen, X.; Qiu, C.; Zhang, S.; Zhang, J.; Xie, M.; Li, L. Development of an improved model for prediction of short-term heavy precipitation based on GNSS-derived PWV. *Remote Sensing* 2020, *12*, 4101.

[2] Li, H.; Wang, X.; Wu, S.; Zhang, K.; Chen, X.; Zhang, J.; Qiu, C.; Zhang, S.; Li, L. An improved model for detecting heavy precipitation using GNSS-derived zenith total delay measurements. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 2021, *14*, 5392-5405.

[3] Song, X.; Mo, Y.; Xuan, Y.; Wang, Q.J.; Wu, W.; Zhang, J.; Zou, X. Impacts of urbanization on precipitation patterns in the greater Beijing–Tianjin–Hebei metropolitan region in northern China. *Environmental Research Letters* 2021, *16*, 014042.

[4] Xiong, T.; Wang, W.; He, J.; Su, R.; Wang, H.; Hu, J. Spatiotemporal Feature Fusion Transformer for Precipitation Nowcasting via Feature Crossing. *Remote Sens.* 2024, *16*, 2685.

[5] Ravuri, S.; Lenc, K.; Willson, M.; Kangin, D.; Lam, R.; Mirowski, P.; Fitzsimons, M.; Athanassiadou, M.; Kashem, S.; Madge, S. Skilful precipitation nowcasting using deep generative models of radar. *Nature* 2021, *597*, 672-677.

[6] Ayzel, G.; Scheffer, T.; Heistermann, M. RainNet v1. 0: a convolutional neural network for radar-based precipitation nowcasting. *Geoscientific Model Development* 2020, *13*, 2631-2644.

[7] Espeholt, L.; Agrawal, S.; Sønderby, C.; Kumar, M.; Heek, J.; Bromberg, C.; Gazen, C.; Carver, R.; Andrychowicz, M.; Hickey, J. Deep learning for twelve hour precipitation forecasts. *Nature communications* 2022, *13*, 5145.

[8] Zhao, Q.; Yang, P.; Yao, W.; Yao, Y. Adaptive AOD forecast model based on GNSS-derived PWV and meteorological parameters. *IEEE Transactions on Geoscience and Remote Sensing* 2021, *60*, 1-10.

[9] Li, L.; Zhang, K.; Wu, S.; Li, H.; Wang, X.; Hu, A.; Li, W.; Fu, E.; Zhang, M.; Shen, Z. An improved method for rainfall forecast based on GNSS-PWV. *Remote Sensing* 2022, *14*, 4280.

[10] Li, H.; Choy, S.; Zaminpardaz, S.; Carter, B.; Sun, C.; Purwar, S.; Liang, H.; Li, L.; Wang, X. Investigating the inter-relationships among multiple atmospheric variables and their responses to precipitation. *Atmosphere* 2023, *14*, 571.

[11] Nie, T.; Deng, K.; Shao, C.; Zhao, C.; Ren, K.; Song, J. Self-attention UNet model for radar based precipitation nowcasting. In Proceedings of the 2021 IEEE sixth international conference on data science in cyberspace (DSC), 2021; pp. 493-499.

[12] Liu, Z.; Hu, H.; Lin, Y.; Yao, Z.; Xie, Z.; Wei, Y.; Ning, J.; Cao, Y.; Zhang, Z.; Dong, L. Swin transformer v2: Scaling up capacity and resolution. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022; pp. 12009-12019.

[13] Fan, L.; Zhou, Y.; Liu, H.; Li, Y.; Cao, D. Combining Swin transformer with UNet for remote sensing image semantic segmentation. *IEEE Transactions on Geoscience and Remote Sensing* 2023, *61*, 1-11.

[14] Gao, Z.; Shi, X.; Wang, H.; Zhu, Y.; Wang, Y.B.; Li, M.; Yeung, D.-Y. Earthformer: Exploring space-time transformers for earth system forecasting. *Advances in Neural Information Processing Systems* 2022, *35*, 25390-25403.

[15] Min, M.; Li, J.; Wang, F.; Liu, Z.; Menzel, W.P. Retrieval of cloud top properties from advanced geostationary satellite imager measurements based on machine learning algorithms. *Remote Sensing of Environment* 2020, *239*, 111616.

[16] Chen, Y.; Wang, G.; Menenti, M. Short-Term Rainfall Forecasting Using GNSS-Derived PWV and ZTD Variations: Case Studies of Four Hong

Kong Rainstorm Events. *Advances in Space Research* 2025.
[17] He, Q.; Shen, Z.; Wan, M.; Li, L. Precipitable water vapor converted from GNSS-ZTD and ERA5 datasets for the monitoring of tropical cyclones. *IEEE Access* 2020, *8*, 87275-87290.
[18] Ziv, S.Z.; Reuveni, Y. Flash floods prediction using precipitable water vapor derived from GPS tropospheric path delays over the eastern mediterranean. *IEEE Transactions on Geoscience and Remote Sensing* 2022, *60*, 1-17.
[19] Jones, J.; Guerova, G.; Douša, J.; Dick, G.; De Haan, S.; Pottiaux, E.; Bock, O.; Pacione, R.; Van Malderen, R. Advanced GNSS tropospheric products for monitoring severe weather events and climate. *COST action ES1206 final action dissemination report* 2020, 563.
[20] Li, H.; Wang, X.; Choy, S.; Jiang, C.; Wu, S.; Zhang, J.; Qiu, C.; Zhou, K.; Li, L.; Fu, E. Detecting heavy rainfall using anomaly-based percentile thresholds of predictors derived from GNSS-PWV. *Atmospheric Research* 2022, *265*, 105912.
[21] Li, H.; Choy, S.; Wang, X.; Liang, H.; Purwar, S.; Zhang, K. Investigating the optimal spatial resolution for assimilating GNSS PWV into an NWP system to improve the accuracy of humidity field. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 2023, *16*, 6876-6888.
[22] Wang, Y.; Wu, H.; Zhang, J.; Gao, Z.; Wang, J.; Yu, P.S.; Long, M. Predrnn: A recurrent neural network for spatiotemporal predictive learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2022, *45*, 2208-2225.
[23] Gao, Z.; Tan, C.; Wu, L.; Li, S.Z. Simvp: Simpler yet better video prediction. In Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022; pp. 3170-3180.
[24] Bi, K.; Xie, L.; Zhang, H.; Chen, X.; Gu, X.; Tian, Q. Accurate medium-range global weather forecasting with 3D neural networks. *Nature* 2023, *619*, 533-538.
[25] Lam, R.; Sanchez-Gonzalez, A.; Willson, M.; Wirnsberger, P.; Fortunato, M.; Alet, F.; Ravuri, S.; Ewalds, T.; Eaton-Rosen, Z.; Hu, W. Learning skillful medium-range global weather forecasting. *Science* 2023, *382*, 1416-1421.
[26] Zhang, Y.; Long, M.; Chen, K.; Xing, L.; Jin, R.; Jordan, M.I.; Wang, J. Skilful nowcasting of extreme precipitation with NowcastNet. *Nature* 2023, *619*, 526-532.
[27] Gan, W.; Guo, H.; Cao, Y.; Wang, W.; Yao, N.; Wang, Y.; Maeyer, P.D. Spatiotemporal Evolution and Intensification of Extreme Precipitation Events in Mainland China from 1961 to 2022. *Remote Sensing* 2025, *17*, 2037.
[28] Wu, M.; Jin, S.; Li, Z.; Cao, Y.; Ping, F.; Tang, X. High-precision GNSS PWV and its variation characteristics in China based on individual station meteorological data. *Remote Sensing* 2021, *13*, 1296.
[29] Böhm, J.; Niell, A.; Tregoning, P.; Schuh, H. Global Mapping Function (GMF): A new empirical mapping function based on numerical weather model data. *Geophysical research letters* 2006, *33*.
[30] Saastamoinen, J. Atmospheric correction for the troposphere and stratosphere in radio ranging satellites. *The use of artificial satellites for geodesy* 1972, *15*, 247-251.
[31] Askne, J.; Nordius, H. Estimation of tropospheric delay for microwaves from surface weather data. *Radio science* 1987, *22*, 379-386.
[32] Huang, L.; Liu, L.; Chen, H.; Jiang, W. An improved atmospheric weighted mean temperature model and its impact on GNSS precipitable water vapor estimates for China. *GPS solutions* 2019, *23*, 51.
[33] Zhang, J.; Howard, K.; Langston, C.; Vasiloff, S.; Kaney, B.; Arthur, A.; Van Cooten, S.; Kelleher, K.; Kitzmiller, D.; Ding, F. National Mosaic and Multi-Sensor QPE (NMQ) system: Description, results, and future plans. *Bulletin of the American Meteorological Society* 2011, *92*, 1321-1338.
[34] Smith, E.K.; Weintraub, S. The constants in the equation for atmospheric refractive index at radio frequencies. *Proceedings of the IRE* 2007, *41*, 1035-1037.
[35] Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the Proceedings of the European conference on computer vision (ECCV), 2018; pp. 3-19.
[36] Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, Q.; Wang, M. Swin-unet: Unet-like pure transformer for medical image segmentation. In Proceedings of the European conference on computer vision, 2022; pp. 205-218.
[37] Milletari, F.; Navab, N.; Ahmadi, S.-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In Proceedings of the 2016 fourth international conference on 3D vision (3DV), 2016; pp. 565-571.
[38] Shi, X.; Gao, Z.; Lausen, L.; Wang, H.; Yeung, D.-Y.; Wong, W.-k.; Woo, W.-c. Deep learning for precipitation nowcasting: A benchmark and a new model. *Advances in neural information processing systems* 2017, *30*.
[39] Kanopoulos, N.; Vasanthavada, N.; Baker, R.L. Design of an image edge detection filter using the Sobel operator. *IEEE Journal of solid-state circuits* 1988, *23*, 358-367.
[40] Li, J.; Zheng, J.; Li, B.; Min, M.; Liu, Y.; Liu, C.-Y.; Li, Z.; Menzel, W.P.; Schmit, T.J.; Cintineo,

J.L. Quantitative applications of weather satellite data for nowcasting: Progress and challenges. *Journal of Meteorological Research* 2024, *38*, 399-413.

[41] Wang, X.; Chen, F.; Ke, F.; Xu, C. An empirical grid model for precipitable water vapor. *Remote Sensing* 2022, *14*, 6174.

[42] Zhang, W.; Zhang, H.; Liang, H.; Lou, Y.; Cai, Y.; Cao, Y.; Zhou, Y.; Liu, W. On the suitability of ERA5 in hourly GPS precipitable water vapor retrieval over China. Journal of Geodesy 2019, 1-13.

Authors

Jing Sun is pursuing the Ph.D degree in Space Science and Technology at the School of Space and Earth Sciences, Beihang University, Beijing, China. Her research interests include GNSS water vapor retrieval, artificial intelligence, image processing, and multimodal data fusion.

Jiale Wang received the Ph.D. degree in Communication and Information Systems from Beihang University, Beijing, China, in 2024, and the B.E. degree in Communication Engineering from Beijing Jiaotong University, Beijing, China, in 2019. He is an Assistant Professor at Beihang University, Beijing, China, and also an Assistant Researcher at The Hong Kong Polytechnic University, Hong Kong, China. His current research interests include satellite navigation, multi-sensor fusion, the Artificial Internet of Things (A-IoT), and wearable sensors.

Chuang Shi (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees from the Wuhan Technical University of Surveying and Mapping (now Wuhan University), Wuhan, China, in 1991, 1995, and 1998, respectively. He is currently an Academician of the Chinese Academy of Sciences (CAS) and a Professor with the School of Space and Earth Sciences, Beihang University, Beijing, China. He also serves as the Director of the Key Laboratory of Satellite Navigation and Mobile Communication Fusion Technology, Ministry of Industry and Information Technology. He is a Distinguished Professor of the Chang Jiang Scholars Program and a recipient of the National Science Fund for Distinguished Young Scholars. He has led the development of China's autonomous software systems for precision orbit determination and positioning of navigation satellites, as well as the wide-area real-time precise positioning system and the critical technology underpinning the BeiDou ground-based augmentation system. These contributions have been successfully integrated into national key initiatives, such as the BeiDou Navigation Major Project, the XiHe System, and the National BeiDou Ground-Based Augmentation System. He has authored or coauthored more than 300 academic papers and holds 50 patents and five software copyrights. Prof. Shi's exemplary work has been recognized with two first-place National Science and Technology Progress Prizes and more than ten awards at the provincial or ministry level. His research interests center on the advanced theoretical innovation, methodological development and practical application of high-precision Positioning, Navigation, Timing, Remote Sensing, and Communication (PNTRC) technologies, with a specific focus on their integration and optimization for Global Navigation Satellite System (GNSS) and low Earth orbit (LEO) satellite constellations.

Yunchang Cao is currently a Senior Research Scientist with the Meteorological Observation Center, China Meteorological Administration, Beijing, China. He has been involved in the study of the global navigation satellite system (GNSS) meteorological applications for decades. For the study of the precipitable water vapor (PWV) estimation, the tomography, and its assimilation technique, he and his colleagues designed and operated the China GNSS PWV monitoring network. As the Project Scientist, he is in charge of the GNSS meteorological demonstration project, the BeiDou-based meteorological application in the atmospheric, marine, and space weather. He and his colleagues have implemented the demonstration of the BeiDou radiosonde system, the BeiDou PWV and TEC monitoring systems, and the BeiDou GNSS-R systems. Currently, his group is working on the operation and the study of the BeiDou-based GNSS/MET application in China.

Hong Liang received the Ph.D. degree in meteorology from the University of Chinese Academy of Sciences, Beijing, China, in 2012. He was a Visiting Scholar at the Scripps Institu-tion of Oceanography, University of California San Diego, San Diego, CA, USA, from 2017 to 2018. He is currently a Researcher with the Meteorological Observation Center, China Meteorological Adminis-tration (CMA), Beijing. His main research interests are GNSS meteorology, GNSS-refractometry, and GNSS-reflectometry.

Wu Chen is currently a Professor with the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University (HK PolyU), Hong Kong. He has been actively working on global navigation satellite system (GNSS)-related research for over 30 years.

His main research interests are geodesy and geodynamics, seamless positioning technologies, GNSS positioning and applications, system integration, GNSS performance evaluation, regional global positioning system (GPS) networks, and simultaneous localization and mapping (SLAM).