Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Forecasting precipitable water vapor at unmonitored coastal sites using a sparse GNSS network and NWP forecasts

Peng Min
Chengdu Jincheng College
Jie Zhao
Chengdu Jincheng College
Linrui Liu
Chengdu Jincheng College
You Yin
Chengdu Jincheng College
Ganzhi Luo
Chengdu Jincheng College
Han Su
jkxy_sh@sicnu.edu.cn

Sichuan Normal University

Research Article

Keywords:

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11322350/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Forecasting precipitable water vapor at unmonitored coastal sites using a sparse GNSS network and NWP forecasts

Peng Min[1,2], Jie Zhao[1], Linrui Liu[1], You Yin[1], Ganzhi Luo[1], Han Su[2,*]

[1] *Department of Electronic Information Engineering, Chengdu Jincheng College, Chengdu, Sichuan 611731, China*
[2] *School of Computer Science, Sichuan Normal University, Chengdu, Sichuan 610101, China*
[*] *Corresponding Author: Han Su (Email: jkxy_sh@sicnu.edu.cn)*

Abstract

Ground-based Global Navigation Satellite System (GNSS) precipitable water vapor (PWV) provides essential observational constraints for coastal convective forecasting. However, existing statistical and deep-learning models rely on multi-year single-station archives, restricting utility to permanently instrumented locations and leaving unmonitored coastal zones dependent on uncorrected numerical weather prediction (NWP). Here, we establish a sparse-GNSS-backbone regional forecasting framework that projects observation-informed calibration beyond sensor footprints by coupling GNSS anchors with continuous NWP tendencies through geographically biased graph attention. Trained with stochastic whole-station history dropout, the architecture operates across a two-tier service paradigm: Level 1 regional coverage projecting calibration across unmonitored footprints without local GNSS hardware or archives, and Level 2 local enhancement providing precision upgrades where receivers exist. In China, Level 1 reduces raw NWP 1–24 h mean absolute error across 19 unmonitored targets by 12.8% (7.0% over calibrated NWP), where on-site GNSS records yield a 9.7% precision upgrade; across the instrumented source backbone, Level 2 reduces error by 7.5% over local fusion. An independent 138-station benchmark along the US coast confirms cross-continental portability: a 30-station backbone reliably projects calibration across 108 verification targets (reducing error by 6.1% over calibrated NWP), with skill sustained across 200–300 km. Under severe sensing disruptions, the architecture maintains a 9.9% error reduction during tropical cyclones, and its advantage increases to 19.9% (China) and 30.7% (US) during extended telemetry outages. This framework demonstrates observational leverage, transforming sparse GNSS arrays into regional predictive assets.

Keywords: *GNSS meteorology; precipitable water vapor; sparse GNSS network; numerical weather prediction; regional forecast post-processing; coastal atmosphere*

1. Introduction

Water vapor supplies the latent heat driving atmospheric convection and tropical cyclogenesis. Along marine–terrestrial boundaries, steep moisture gradients and rapid variations demand accurate short-term (1–24 h) forecasts of precipitable water vapor (PWV) to track convective precursors and frontal convergence (Bevis et al. 1992; Rocken et al. 1997; Guerova et al. 2016; He et al. 2020). Ground-based Global Navigation Satellite System (GNSS) networks continuously infer zenith tropospheric delays in all weather conditions, inverted into PWV with millimeter-level precision (Bevis et al. 1994; Wang et al. 2007). Extending these continuous geodetic observations from passive real-time monitoring to regional forecast post-processing represents a vital frontier for GNSS atmospheric applications (Shoji 2013; Hu et al. 2026).

Despite advances in statistical post-processing and deep learning, operational deployment of GNSS-informed forecasting remains constrained by a single-station design paradigm that calibrates separate functions at individual stations using multi-year collocated archives. This localized fitting philosophy exposes three structural vulnerabilities:

1. **Unmonitored regional footprints:** Expansive coastal stretches, maritime estuaries, and offshore islands lack permanent GNSS receivers, leaving them reliant on uncorrected numerical weather prediction (NWP) outputs.
2. **Cold-start barrier for newly deployed assets:** Newly commissioned stations stream real-time signals but lack the multi-year archives required to train site-specific models.
3. **Vulnerability to telemetry interruptions:** During severe weather or communication cuts, local telemetry is interrupted for hours; models reliant strictly on on-site histories degrade sharply once inputs become stale.

Numerical weather prediction (NWP) models, such as the Global Ensemble Forecast System (GEFS; Zhou et al. 2022; Guan et al. 2022), provide spatially continuous moisture prognostic fields. Nevertheless, grid-scale parameterizations and smoothed coastlines introduce systematic, lead-dependent PWV biases. Direct GNSS data assimilation mitigates errors (Gendt et al. 2004; Shoji et al. 2011) but entails high computational cost and latency. Statistical Model Output Statistics (MOS; Glahn and Lowry 1972) and machine-learning post-processing (Rasp and Lerch 2018; Vannitsem et al. 2018) correct local errors efficiently. Crucially, because synoptic advection and frontal systems organize coastal moisture errors into coherent mesoscale patterns, errors at unmonitored sites are correlated with those at neighboring GNSS stations, allowing sparse backbone corrections to inform surrounding unmonitored

locations (Chen et al. 2018); GNSS PWV has also been used to evaluate and correct NWP water-vapor products (Liu et al. 2024).

Recent studies have explored spatial GNSS tropospheric sensing via tomographic reconstruction (Zhao et al. 2020), regional PWV mapping (Chen et al. 2018), PWV-based rainfall forecasting (Benevides et al. 2015, 2019), multi-station regression (Ghaffari-Razin et al. 2023), deep-learning PWV gap filling (Li et al. 2026), and multivariate forecasting of tropospheric parameters (Bezcioglu 2026). GNSS has also been fused with NWP forecasts (Lu et al. 2024) or surface meteorological observations (Yang et al. 2026) to map tropospheric delay in sparse networks, but these efforts reconstruct current-state fields for positioning rather than multi-hour PWV forecasts at sites without receivers. Graph neural networks also show promise for precipitation nowcasting (Han et al. 2026) and ensemble post-processing (Feik et al. 2024). However, predicting at unmonitored coordinates absent from training remains an open challenge in spatiotemporal learning (Wu et al. 2021; Zheng et al. 2023; Baran and Lakatos 2024; Yang et al. 2025), as learned node-identity embeddings (Shao et al. 2022) are tied to the training locations and cannot be defined for unseen sites. Furthermore, existing models assume uninterrupted telemetry and degrade when streams drop out.

To resolve these challenges, we formulate Numerical weather prediction Exchange Post-processing (NEP). Rather than requiring separate models for different observational scenarios, we introduce a single-network architecture trained with whole-station history dropout ($p = 0.2$), unifying two operational deployment modes: Level 1 (Regional coverage mode, 'NEP without local history') projects calibration from a sparse GNSS backbone across unmonitored regional footprints without requiring local GNSS instrumentation, historical archives, or on-site fitting; Level 2 (Local instrumentation enhancement, 'NEP with local history') incorporates on-site GNSS streams as an optional precision upgrade for high-priority observatories, while ensuring graceful degradation under telemetry outages.

We systematically evaluate this framework across two continental domains: a primary development network of 30 backbone stations and 19 independent external targets along China's coast (Wu et al. 2025), and an independent validation benchmark comprising 30 backbone stations and 108 regional targets along the US Gulf and Atlantic coast processed by the Nevada Geodetic Laboratory (Blewitt et al. 2018).

The primary contributions of this study are fourfold: (1) Inductive regional post-processing framework: Operates interchangeably with or without local GNSS records, projecting calibration across unmonitored coordinates absent

from training while treating physical stations strictly as out-of-sample ground-truth verification probes; (2) Demonstration of regional observational leverage: A 30-station backbone is validated across 108 independent regional targets (3.6 times the number of backbone stations); dedicated network scaling shows that 10 stations capture roughly half the maximum gain, with diminishing returns near 30–45 stations; (3) Operational resilience under sensing stress: Maintains a 9.9% error reduction during tropical cyclones and degrades gracefully under telemetry outages, expanding its advantage over single-station fusion to 19.9% (China) and 30.7% (US) when observations are interrupted for >6 h; and (4) Mechanisms of spatial extrapolation: While point-wise correction transfer decays beyond 100 km, graph attention anchored by geographic bias and NWP tendencies preserves robust gains (4.8–7.9%) out to 200–300 km.

2. Data and evaluation settings

2.1. Coastal GNSS observations and numerical guidance

We utilize quality-controlled hourly GNSS PWV time series from the China coastal GNSS archive documented by Wu et al. (2025). Thirty stations with multi-year continuous observations form the primary source backbone network (Fig. 1a). Nineteen independent coastal stations from the same archive serve as external evaluation targets; these stations are strictly reserved for out-of-sample evaluation and never participate in model parameter optimization or baseline fitting. Zenith wet delay was converted to PWV (Bevis et al. 1994) using weighted mean temperature and surface pressure derived from ERA5 reanalysis (Hersbach et al. 2020); ERA5 fields were not provided as predictive inputs. In real-time operational forecasting, weighted mean temperature can be sourced from empirical mapping functions such as GPT3 (Landskron and Böhm 2018) or operational NWP analyses.

Numerical weather prediction guidance is extracted from the Global Ensemble Forecast System version 12 (GEFS v12; Zhou et al. 2022) reforecast archive (Guan et al. 2022). Three-hourly precipitable water vapor fields from the control run (archived on a 0.25° grid) were bilinearly interpolated to station coordinates and linearly interpolated in time to produce hourly prognostic trajectories spanning lead times of 1–24 h. To reflect real-time operational dissemination constraints, every forecast origin accesses the most recent daily GEFS cycle that is at least 6 h old. This operational latency rule is applied consistently to all evaluated models.

The temporal record is partitioned chronologically into four non-overlapping periods: training from 16 November 2010 to 23 April 2016; validation (hyperparameter tuning and early stopping) from 24 April 2016 to 1 February

2017; a development period to 22 August 2018; and an independent evaluation period from 23 August 2018 to 31 December 2019. The independent evaluation period contains 1,942 six-hourly forecast origins with complete 168 h input windows and corresponding GEFS cycles. Missing observational inputs are handled strictly causally: the most recent valid observation is carried forward for up to 6 h; for longer gaps, the training climatological mean for the matching calendar month and solar hour is substituted. Normalization statistics are computed solely from training-period observations. Only genuine observations are scored. Hyperparameters were fixed during development, and all evaluation-period results reflect strictly causal evaluations under frozen model weights, avoiding temporal information leakage (Kapoor and Narayanan 2023).

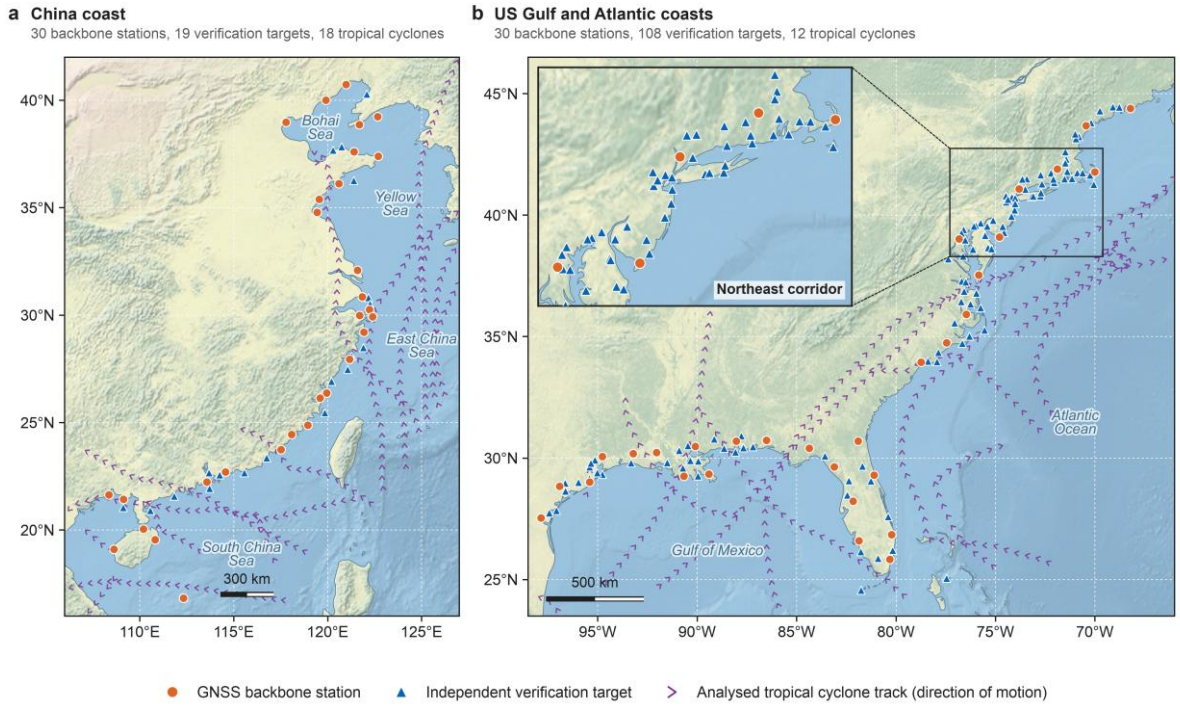


Fig. 1 *GNSS backbones and independent verification targets. (a) China coast: 30 backbone stations (orange circles) and 19 independent verification targets (blue triangles). (b) US Gulf and Atlantic coasts: 30 backbone stations and 108 verification targets; the inset enlarges the Northeast corridor (black box). Purple chevrons trace the analyzed tropical cyclones (18 in China, 12 in the US; IBTrACS v04r01) along their direction of motion, where winds were at least 17.5 m* s^{-1} *(34 kt) within 500 km of a station. Shaded relief: Natural Earth I, 1:50m. Scale bars are exact at the latitude where they are drawn*

2.2. Two-tier evaluation regimes and station partitions

To test observational leverage and spatial projection under operational constraints, we partition monitoring assets into three station tiers in China (Table 1): (1) Source backbone network (30 stations): the core fitting array providing continuous observational anchors (29 stations possess valid target labels during the 2018–2019 evaluation period);

(2) Withheld targets (30 stations, 5-fold cross-validation): the 30 backbone stations are stratified into five geographic clusters of six stations each, where each cluster is completely withheld from model training and validation in turn; and (3) External independent targets (19 stations): an independent coastal network meeting quality criteria (>= 30% coverage and >= 2,000 valid observation hours during evaluation), evaluated under both Level 1 and Level 2 modes.

2.3. Independent US coastal benchmark

To examine geographic portability and algorithmic generalizability across distinct observing systems, we assembled an independent continental benchmark from the Nevada Geodetic Laboratory (NGL; Blewitt et al. 2018) archive. This benchmark operates under independent geodetic processing (GipsyX software with VMF1 tropospheric mapping) and encompasses stations within 50 km of the US Gulf and Atlantic shorelines (24–45°N) from 2011 to 2019. Thirty stations with data availability exceeding 85% and uniform spatial distribution (median nearest-neighbor spacing 148 km) form the US source backbone, while 108 remaining stations serve as external regional targets (Fig. 1b). Numerical GEFS guidance, temporal partitioning, latency constraints, and the 2018–2019 evaluation window strictly mirror the Chinese protocol. The full station partitioning and evaluation settings across both continental domains are summarized in Table 1.

Table 1 Geographic station partitions and operational evaluation settings across China and US coastal domains

Station tier	Stations (total / scored)	In training	Modes	Description
China coast (development domain; PANDA/GAMIT processing)				
Source backbone	30 / 29	Yes	L2	Core fitting network; 1 station lacks evaluation-period targets
Withheld targets	30 / 29	No	L1, L2	5-fold spatial cross-validation; 6 stations withheld per fold
External targets	19 / 19	No	L1, L2	Independent coastal network; ≥ 30% coverage and ≥ 2,000 observed hours
US Gulf and Atlantic coast (portability domain; NGL GipsyX/VMF1 processing)				
Source backbone	30 / 30	Yes	L2	Benchmark network along the Gulf and Atlantic coasts
External targets	108 / 108	No	L1, L2	Independent targets within 50 km of the shoreline

L1, Level 1 (forecasting without target GNSS history); L2, Level 2 (with target GNSS history). China-trained weights are also applied to the US backbone without retraining to test zero-shot transfer (Online Resource 1, Section S9).

3. Methods

3.1. Unified architecture and history-dropout training

The core architecture of NEP couples a station-level temporal encoder with an inter-station graph attention exchange mechanism (Fig. 2). Rather than training separate model checkpoints for observed and unmonitored settings, the

network is trained with whole-station history dropout. During training, the entire 168 h historical observation sequence of each station is replaced with the missing-data token with probability p = 0.2 per sample, while its future NWP inputs and target supervisory labels are preserved. This conditions the network to extract predictive skill simultaneously from local observations (when present) and from the surrounding network state and NWP guidance (when local observations are absent).

At each station, the local temporal branch ingests the 168 h PWV history, observation flags, and age indicators. Following patch-based temporal modeling (Nie et al. 2023), the time series is segmented into 24 h patches with a 6 h stride. Two Transformer encoder layers (Vaswani et al. 2017) map these patches to a 24 h local temporal representation, anchored to the latest available observation via a learned, lead-dependent coefficient. The local temporal forecast is concatenated with the 24 h NWP forecast and the 24 h NWP forecast tendency, projecting each station into a 64-dimensional atmospheric state vector. Two multi-head graph attention layers (Veličković et al. 2018; four heads each) exchange these representations across the network. Attention weights are modulated by a continuous geographic distance prior log(A + ε) (Eq. 1), where A is the normalized geographic adjacency matrix with edge weights exp[-(d_ij / 800 km)²]. The 800 km length scale is intended to represent the synoptic scale of coastal moist air masses and maritime frontal systems; it was treated as a hyperparameter and fixed during the development period. Because both regional networks are dominated by low-elevation coastal stations, horizontal great-circle distance was adopted as the primary geometric prior; elevation-related variations are partly accounted for through the normalization proxy (Eq. 3). To preserve computational efficiency and support linear scalability O(N · k), each station connects to its k = 15 nearest neighbors.

$$\mathrm{Attn}(Q,K,V) = \mathrm{softmax}\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_k}} + \lambda \log(A+\varepsilon)\right)V \tag{1}$$

where Q, K and V are the query, key and value projections of the station state vectors, d_k is the key dimension, λ is a learned scalar for each attention head, and ε is a small constant that keeps the logarithm finite.

Crucially, the inter-station attention layers omit station-identity embeddings. This inductive formulation ensures that spatial information exchange depends exclusively on dynamic atmospheric state vectors and continuous geographic separation, preventing the model from overfitting to memorized station coordinates and enabling direct application to unseen coordinates. The local encoder also keeps a learned offset for each training station; at an unseen site it is

replaced by the mean of the training offsets. The final forecast synthesizes local, NWP tendency, and graph residual terms (Eq. 2), totaling 236,137 trainable parameters.

$$\hat{y}_{i,t,h} = c_{i,t,h} + g_h \Delta n_{i,t,h} + r_{i,t,h} \tag{2}$$

where $c_{i,t,h}$ is the anchored local temporal forecast at station i, origin t and lead h; $\Delta n_{i,t,h} = n_{i,t,h} - n_{i,t,0}$ is the GEFS PWV tendency from the forecast origin; g_h is a learned lead-dependent gain; and $r_{i,t,h}$ is the residual correction produced by the graph exchange layers.

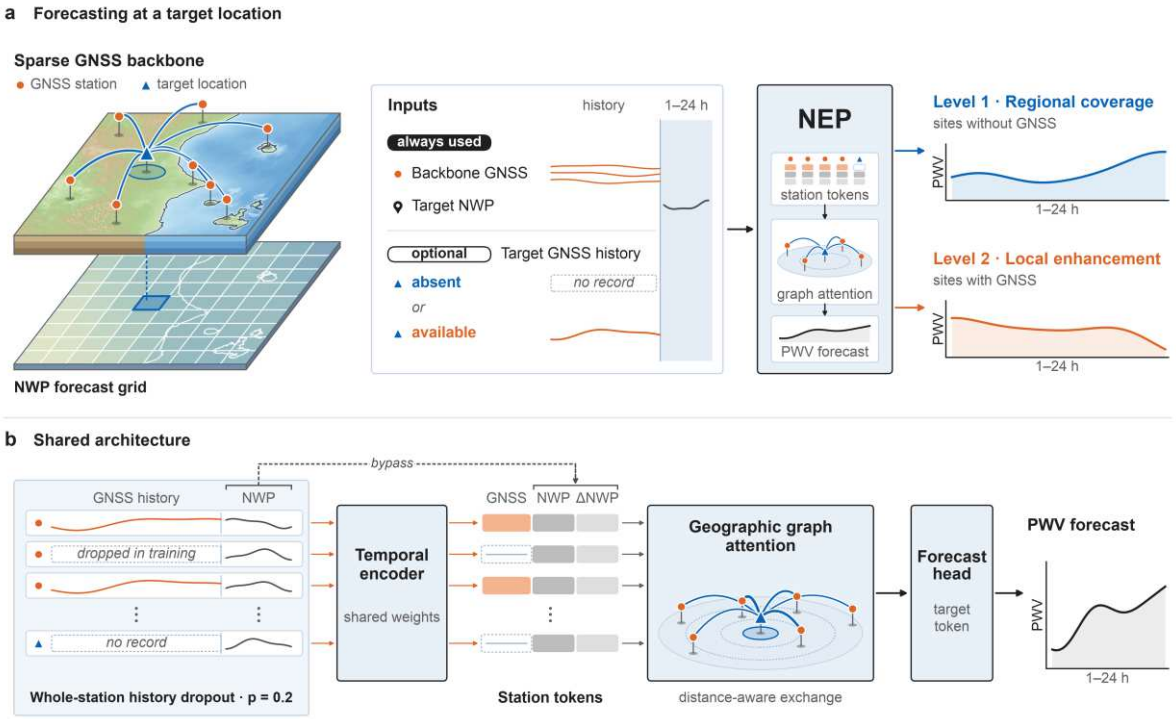


Fig. 2 *NEP with the target's own GNSS history as an optional input. (a) A sparse GNSS backbone and an NWP forecast grid supply the inputs for a target location. The backbone GNSS histories and the target's NWP forecast are always used; the target's own GNSS history is optional and may be absent or available. NEP provides Level 1 regional coverage at sites without GNSS when the history is absent and Level 2 local enhancement when it is available. Arc widths indicate that nearer stations contribute more. (b) Shared architecture. A temporal encoder with weights shared across stations encodes each station's GNSS history; the station's NWP forecast and NWP tendency bypass the encoder (dashed arrow), giving one GNSS–NWP feature token per station. Geographic graph attention passes information from the backbone stations to the target, with weights modulated by the distance between stations, and a forecast head maps the target token to the 1–24 h PWV forecast. During training, whole-station history dropout (p = 0.2) makes the local history an optional input. The figure is a schematic: terrain, station layout, NWP field and curves are illustrative*

3.2. Two-tier service paradigm: regional coverage and local enhancement

The trained network operates across a two-tier service hierarchy: (1) Level 1: Regional coverage mode ('NEP without local history') is deployed to unmonitored geographic target locations (lon, lat, alt) across the regional

domain where no GNSS receiver or local observation archive exists. Because on-site observational records are absent, the target's local temporal input branch is assigned a null unmonitored token (matching the whole-station history dropout state during training). Its feature normalization parameters (mean and standard deviation) are predicted non-locally through a physical proxy fitted strictly on source backbone stations using coordinate elevation and NWP climatological moments (Eq. 3), thereby avoiding any requirement for local historical archives (detailed in Online Resource 1, Section S2); (2) Level 2: Local instrumentation enhancement ('NEP with local history') is deployed at high-priority sites equipped with a physical GNSS receiver, supplying the station's recent 168 h observation stream to the temporal encoder to deliver an operational precision upgrade while naturally degrading gracefully to Level 1 during telemetry dropouts.

$$\hat{\mu}_x = a_0 + a_1\mu_x^{\mathrm{GEFS}} + a_2 z_x, \quad \hat{\sigma}_x = b_0 + b_1\sigma_x^{\mathrm{GEFS}} + b_2\mu_x^{\mathrm{GEFS}} \tag{3}$$

where μ_x^{GEFS} and σ_x^{GEFS} are the training-period mean and standard deviation of GEFS PWV at location x, z_x is its elevation, and the coefficients a and b are fitted by linear regression on the source backbone stations only.

Importantly, because Level 1 requires no local sensor hardware and no station-specific parameters (an unseen site uses the mean of the training offsets), the architecture is capable in principle of evaluating at arbitrary target coordinates (lon, lat, alt) where numerical weather guidance is available. The 19 Chinese and 108 US external stations employed in this study serve strictly as independent ground-truth validation probes to quantify predictive accuracy against geodetic observations, rather than structural boundary constraints of the forecasting framework itself.

3.3. Comparators and training

We benchmark NEP against operational and machine-learning baselines: (1) Calibrated NWP (MOS), a lead-dependent linear model output statistics baseline (Wilks 2011; Eq. 4); (2) Local Fusion, a linear post-processing baseline integrating local observations and NWP analysis at forecast origin (Eq. 5); (3) Matched No-Exchange Control, an identical patch Transformer architecture trained under the same history-dropout regime (p=0.2) but replacing inter-station graph attention with station-wise feedforward layers; (4) Correction Transfer, a training-free spatial interpolation comparator transferring residual corrections from the three nearest backbone stations to target coordinates via inverse-distance weighting (Eq. 6); and (5) Representative pure-observation deep-learning models (TimesNet, Wu et al. 2023; PatchTST, Nie et al. 2023; STGCN, Yu et al. 2018; Crossformer, Zhang and Yan 2023;

Online Resource 1, Table S1). Benchmarking against pure-observation deep learning architectures serves specifically to evaluate whether coastal moisture can be reliably forecasted purely from autoregressive temporal sequences; lacking NWP dynamical constraints, such models severely degrade to 3.77–4.33 mm MAE, establishing that numerical weather prediction provides an essential physical foundation for 1–24 h forecasting. To strictly isolate the value of inter-station spatial graph exchange without confounding temporal model capacity, our primary controlled baseline is the matched No-Exchange control, which receives identical NWP forecasts and tendencies. Neural networks were trained with Adam, cosine annealing, early stopping (patience 15), gradient clipping at 1.0, and exponential moving average weight smoothing (decay 0.999). Scores report means over five random seeds (42–46).

$$\hat{y}_{i,t,h}^{\mathrm{MOS}} = \alpha_{i,h} + \beta_{i,h} n_{i,t,h} \tag{4}$$

$$\hat{y}_{i,t,h}^{\mathrm{LF}} = \alpha_{i,h} + \beta_{i,h} n_{i,t,h} + \gamma_{i,h} y_{i,t} + \delta_{i,h} n_{i,t,0} \tag{5}$$

$$\hat{y}_{x,t,h}^{\mathrm{CT}} = n_{x,t,h} + \sum_{s\in\mathcal{N}_3(x)} w_s\left(\hat{y}_{s,t,h} - n_{s,t,h}\right), \quad w_s = \frac{d_{sx}^{-2}}{\sum_{k\in\mathcal{N}_3(x)} d_{kx}^{-2}} \tag{6}$$

where $n_{i,t,h}$ is the GEFS PWV forecast, $n_{i,t,0}$ is the GEFS PWV valid at the forecast origin, $y_{i,t}$ is the latest GNSS PWV observation, $\mathcal{N}_3(x)$ is the set of the three backbone stations nearest to x, $\hat{y}_{s,t,h}$ is the NEP forecast at backbone station s, and d_{sx} is the great-circle distance (floored at 1 km). MOS and Local Fusion coefficients are fitted for each lead time and source station; at external targets they are pooled across source stations without access to target observations.

3.4. Forecast assessment and stress testing

Forecast accuracy is evaluated using mean absolute error (MAE) and root-mean-square error (RMSE) in millimeters across 1–24 h leads. Statistical uncertainty is estimated via moving-block bootstrap (Künsch 1989) using 2,000 resamples of 28 consecutive six-hourly forecast cycles (7-day blocks). Operational stress testing incorporates two challenging regimes: (1) Controlled and natural telemetry outages, where the last k hours of target observations (k in {0, 1, 3, 6, 12, 24, 48, 72} h) are withheld prior to issuance across 20,004 (China) and 184,392 (US) eligible station–origin pairs, and natural outages where the latest observation is >6 h old; and (2) Tropical cyclone environments,

evaluated across 798 station-origin cases from 18 storms in China and 1,934 cases from 12 storms in the US (IBTrACS; Knapp et al. 2010), resampling whole storms.

4. Results

4.1. Regional coverage at unmonitored targets (Level 1)

We first evaluate the primary operational objective of the framework: can a sparse GNSS backbone calibrate numerical forecasts across unmonitored coastal areas where no receiver exists (Level 1 service mode)? In this regime, the model synthesizes predictions exclusively from surrounding backbone observations and the target's local NWP guidance, while on-site GNSS records serve strictly as independent ground-truth verification probes.

Across the 19 independent external targets along China's coast, Level 1 regional coverage achieves an MAE of 2.680 ± 0.007 mm and an RMSE of 3.910 mm (Table 2). This represents a 12.8% reduction in forecast error relative to raw GEFS (3.074 mm) and a 7.0% reduction over calibrated NWP (MOS) [95% CI: 5.8, 8.1]%, outperforming MOS at all 19 target locations. This capability reproduces consistently across the 29 withheld cross-validation targets (Online Resource 1, Table S2), where nearest-backbone separation is wider (median 96 vs. 66 km), yielding a 13.6% error reduction over raw GEFS and outperforming training-free correction transfer by 5.0% [95% CI: 4.1, 5.9]%.

Table 2 Forecast accuracy without target GNSS history at the 19 external targets during the independent evaluation period. Neural results are means over five random seeds (± seed standard deviation for NEP)

Model	MAE (mm)	RMSE (mm)	MAE reduction (%)	
			vs. raw GEFS	vs. MOS
Raw GEFS	3.074	4.391	–	–
Calibrated NWP (MOS)	2.881	4.185	+6.3	–
No-exchange control	2.834	4.083	+7.8	+1.6
Correction transfer	2.715	3.942	+11.7	+5.8 [4.4, 7.0]
NEP without target history (Level 1)	**2.680 ± 0.007**	**3.910**	**+12.8**	**+7.0 [5.8, 8.1]**

Brackets give 95% confidence intervals.

The waterfall progression in Fig. 3a illustrates that this unmonitored skill is primarily unlocked by inter-station exchange (+5.4%) rather than temporal representation capacity alone (+1.6%). Lead-time analysis shows region-specific behavior (Fig. 3c, d; Online Resource 1, Tables S5 and S7). In China, NEP without local history is 12.6% and 3.6% more accurate than correction transfer at 1 h and 3 h, and the two agree within 1% from 6 h onward. In the US the pattern reverses at short leads: correction transfer is 10.3% and 4.0% more accurate at 1 h and 3 h, and NEP

is 0.5–1.6% more accurate from 6 h onward. The clearer difference is with distance: correction transfer loses skill with separation from the nearest backbone station, whereas NEP retains most of its skill (Section 4.4).

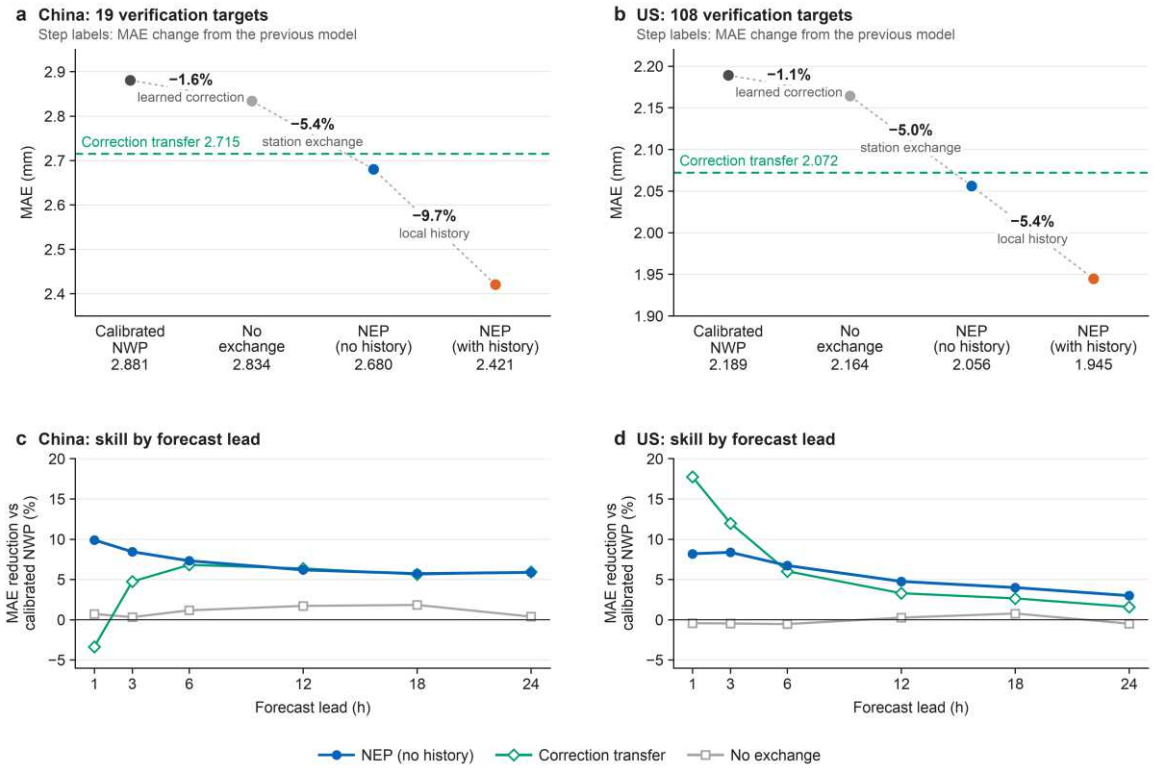


Fig. 3 *Regional forecasting skill and the benefit of local history. (a, b) Mean absolute error (MAE) at the independent verification targets in China (19) and the US (108). Points follow the progression from calibrated NWP to a model without station exchange, NEP without local history and NEP with local history; labels give the MAE change from the previous model, and values below the model names are MAE in mm. The dashed green line marks correction transfer. Both NEP modes use the same trained weights. (c, d) MAE reduction relative to calibrated NWP by forecast lead for NEP without local history, correction transfer and the no-exchange model; positive values indicate lower error*

4.2. Local instrumentation enhancement (Level 2)

Where high-priority sites are equipped with GNSS tracking hardware, supplying the station's own recent 168 h observation history provides an optional precision upgrade (Level 2). This raises a practical network design question: is investing in on-site GNSS hardware still worthwhile given the substantial skill already achieved by regional Level 1 coverage?

The results confirm that Level 2 delivers a consistent, high-value precision upgrade across all observing tiers (Table 3). Most importantly for the two-tier service paradigm, deploying on-site GNSS tracking at the 19 independent external targets reduces MAE from 2.680 mm to 2.421 mm, corresponding to a 9.7% precision upgrade over Level 1 regional coverage [95% CI: 8.8, 10.6]%.

Furthermore, spatial graph exchange continues to provide substantial value even when complete on-site observations are available. Across the source backbone (29 stations), withheld targets (29 stations), and external targets (19 stations), NEP consistently outperforms operational local fusion by 5.7% to 7.5% (all $p < 0.001$; Table 3), with inter-station exchange contributing an incremental 2.8–3.9% gain over the matched no-exchange control. In contrast, pure observation deep-learning baselines lacking NWP dynamical constraints degrade sharply to 3.77–4.33 mm (Online Resource 1, Table S1), demonstrating that local observations serve as an enhancement to, rather than a replacement for, numerical weather prediction. Complete station-wise error metrics are catalogued in Table 3.

Table 3 Precision upgrade for instrumented stations with local observation histories across evaluation tiers. Neural results are means over five random seeds

Evaluation tier	MAE (mm)			NEP MAE reduction (%)	
	Local fusion	No-exchange control	NEP (Level 2)	vs. local fusion	vs. no-exchange control
Source backbone (29 stations)	2.212	2.131	**2.047**	+7.5 [6.8, 8.2]	+3.9 [3.4, 4.5]
Withheld targets (29 stations, 5-fold)	2.391	2.309	**2.235**	+6.5 [5.6, 7.3]	+3.2 [2.5, 3.9]
External targets (19 stations)	2.568	2.492	**2.421**	+5.7 [4.7, 6.7]	+2.8 [2.4, 3.3]

The reduction relative to the no-exchange control is the marginal contribution of inter-station exchange. Brackets give 95% confidence intervals; seed standard deviations at the backbone are listed in Table 6.

4.3. Cross-continental validation and US regional scaling

Deploying the framework across the 138-station US Gulf and Atlantic coastal network examines whether regional calibration transfers across distinct geodetic processing software (NGL GipsyX vs. PANDA/GAMIT) and maritime climate regimes (Table 4).

Across the 108 US external regional targets, Level 1 regional coverage without local history reduces MAE by 6.1% over calibrated NWP (MOS) [95% CI: 5.2, 7.0]%, improving accuracy at 103 of 108 targets (105 of 108 relative to the no-exchange control), with inter-station exchange contributing 5.0% [95% CI: 4.5, 5.5]%. Where local GNSS hardware is available (Level 2), accuracy improves by 5.2% over local fusion across the 30-station backbone and delivers an equivalent 5.4% upgrade at regional targets (Table 4).

Crucially, zero-shot cross-ocean transfer—applying China-trained model weights directly to the US benchmark without fine-tuning—yields an MAE of 2.21 mm. This zero-shot model outperforms locally fitted US MOS (2.39 mm) by 7.6% and matches locally trained US fusion (2.22 mm) within 0.5% (Online Resource 1, Section S9). This demonstrates that the network captures transferable relationships between NWP tendencies and tropospheric moisture errors that generalize across the two coastal domains and independent geodetic processing regimes.

Table 4 Cross-continental summary of forecast improvements across China and US coastal domains (August 2018 to December 2019). Values report relative MAE reductions (%); positive values indicate superior performance

Metric	China coast	US Gulf and Atlantic coast
Level 1: without target GNSS history		
NEP vs. calibrated NWP (MOS)	+7.0 [5.8, 8.1]	+6.1 [5.2, 7.0]
Correction transfer vs. MOS	+5.8 [4.4, 7.0]	+5.4 [4.3, 6.4]
NEP vs. no-exchange control (exchange contribution)	+5.4 [4.5, 6.3]	+5.0 [4.5, 5.5]
Level 2: with target GNSS history		
Added value of the on-site receiver (Level 2 vs. Level 1)	+9.7 [8.8, 10.6]	+5.4 [5.1, 5.8]
NEP vs. local fusion, source backbone	+7.5 [6.8, 8.2]	+5.2 [4.6, 5.8]
NEP vs. local fusion, external targets	+5.7 [4.7, 6.7]	+5.4 [4.9, 5.9]
NEP vs. no-exchange control at the backbone (exchange contribution)	+3.9 [3.4, 4.5]	+1.5 [1.1, 1.8]
Stress tests		
Telemetry outage (> 6 h gap at origin), NEP vs. local fusion	+19.9 [15.0, 24.6]	+30.7 [26.0, 34.7]
Tropical cyclones, NEP vs. local fusion	+9.9 [4.5, 14.4]	+7.8 [3.2, 11.0]
Tropical cyclones, NEP vs. no-exchange control	+4.9 [0.7, 8.9]	+0.6 [−1.7, 2.5]

Brackets give 95% confidence intervals.

4.4. Backbone density scaling and distance decay

To determine how many backbone stations are required to establish an effective regional forecasting service, we executed a designed network scaling experiment across the US coastal domain. Forty fixed regional targets were selected via farthest-point sampling across the 108 external stations. Three nested farthest-point backbone chains (sizes N in {10, 20, 30, 45, 60}) were constructed from candidate stations, spanning the entire coastline (Table 5; Fig. 4a, b).

The scaling trajectory reveals pronounced observational leverage across the coastal network (Table 5; Fig. 4a, b). As few as 10 backbone stations (median target distance 131 km) achieve a 3.6% error reduction over calibrated NWP, capturing roughly half (47%, ratio 95% CI: [0.40, 0.53]) of the maximum benefit achieved by the densest 60-station configuration (7.6%). Regional calibration skill exhibits diminishing marginal returns: the smallest backbone that reaches 90% of the plateau gain (7.4%) is 45 stations in the point estimate (7.1% gain, median distance 47 km) and 30 stations in 36% of bootstrap realizations, so the plateau begins at about 30–45 stations. In ultra-sparse regimes (N=10), designed geometric farthest-point siting outperforms random station placement (3.6% vs. 2.7% gain), though this gap closes once network size reaches 20 stations (4.7% vs. 4.6%; Online Resource 1, Table S9).

Decomposing the sources of skill highlights an evolving trade-off between spatial network exchange and on-site receiver deployment. As the backbone densifies from 10 to 60 stations, the contribution of inter-station exchange

rises monotonically from 1.9% to 5.9%, whereas the incremental precision upgrade from installing an on-site GNSS receiver declines from 7.7% to 5.4% (Table 5; Fig. 4b). At approximately 45 stations, the regional gain supplied by network exchange becomes comparable in magnitude to the benefit of deploying dedicated local hardware. Meanwhile, the matched no-exchange control remains virtually flat at ~1.9% across all network sizes, indicating that the scaling benefit stems primarily from inter-station spatial exchange rather than local model capacity.

Table 5 Backbone density scaling experiment across 40 fixed US regional targets. Median distance reflects target-to-backbone separation. Gains are relative MAE reductions; positive values indicate lower error

Backbone size (stations)	**10**	**20**	**30**	**45**	**60**
Median target-to-backbone distance (km)	131	79	67	47	42
Level 1: without target GNSS history (MAE reduction, %)					
NEP vs. MOS	+3.6	+4.7	+6.5	+7.1	+7.6
No-exchange control vs. MOS	+1.8	+1.9	+1.9	+2.0	+1.9
NEP vs. no-exchange control (exchange contribution)	+1.9	+2.9	+4.7	+5.2	+5.9
Targets where NEP beats MOS (of 40)	37	38	39	39	40
Level 2: with target GNSS history (MAE reduction, %)					
Added value of the on-site receiver (Level 2 vs. Level 1)	+7.7	+7.1	+6.2	+5.6	+5.4

Distance decay analysis explains why NEP outperforms spatial interpolation across regional scales (Fig. 4c, d; Online Resource 1, Table S8). In the US, inverse-distance correction transfer decays sharply with distance: from 7.1% at 40–70 km to 3.7% at 70–100 km, falling to 2.1% [95% CI: −0.2, 3.4]% beyond 150 km. In contrast, NEP without local history sustains a 4.9% error reduction beyond 150 km. In China, correction transfer falls from 6.6% (<40 km) to 2.3% (>100 km), whereas NEP maintains a 7.9% [95% CI: 2.8, 11.8]% reduction beyond 100 km. Graph attention anchored by continuous geographic bias and NWP tendencies preserves synoptic error coherence across 200–300 km regional footprints where point-wise interpolation collapses.

Regional gains also display a pronounced latitudinal gradient, with rank correlations of −0.83 in the US and −0.76 in China against target latitude. In the US, Level 1 gains decrease from 8.7% (<30°N) to 2.4% (>= 40°N); in China, gains decline from 10.1% in the tropical south to −1.7% in the temperate north, consistent with larger benefits in warm, humid coastal regimes characterized by high absolute PWV and active convective processes.

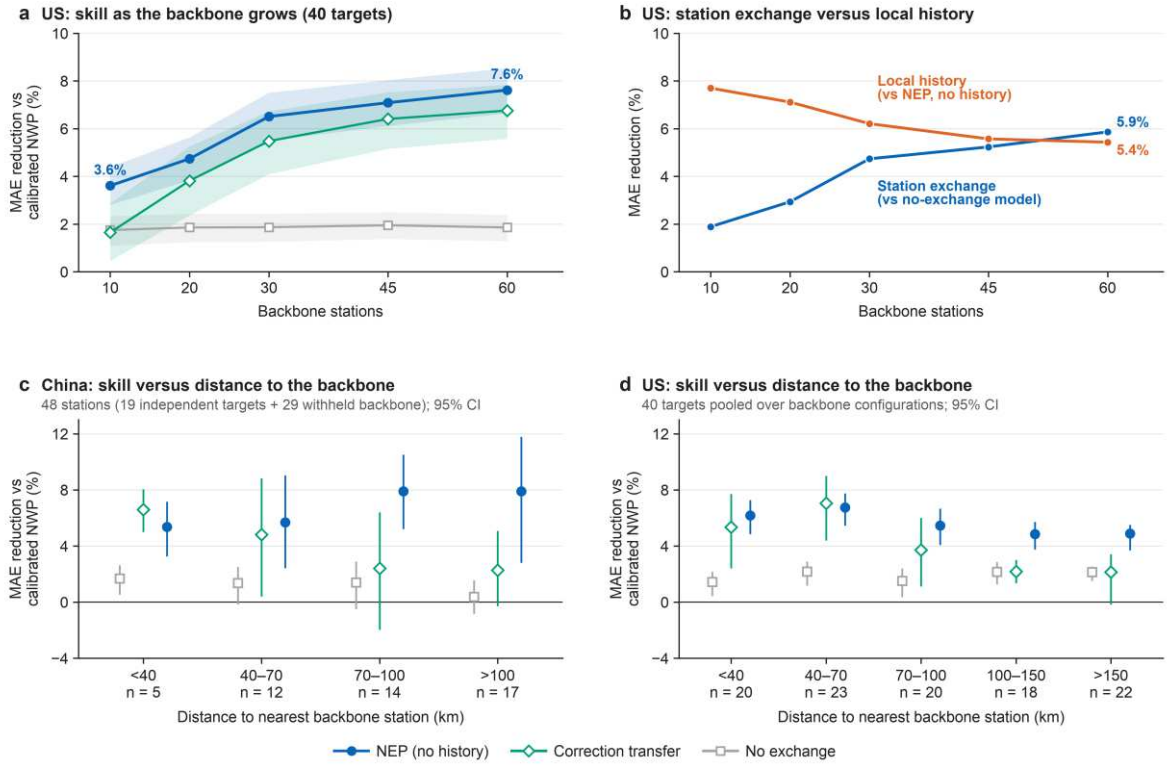


Fig. 4 *Forecast skill as the backbone expands and as distance to the backbone increases. (a) US: MAE reduction relative to calibrated NWP at 40 targets as the backbone grows from 10 to 60 stations; shading shows 95% confidence intervals. (b) US: gain from station exchange (NEP without local history relative to the no-exchange model) and additional gain from local history (NEP with local history relative to NEP without local history); the two curves use different reference models and are not additive. (c, d) MAE reduction relative to calibrated NWP grouped by distance to the nearest backbone station; bars show 95% confidence intervals and n is the number of stations (China) or targets (US) in each class. China: 48 evaluation stations (19 independent targets and 29 withheld backbone stations); US: 40 targets pooled across backbone configurations*

4.5. Operational resilience under sensor outages and tropical cyclones

A major vulnerability of operational post-processing systems is performance degradation when sensor observations are interrupted. Traditional models trained strictly on complete histories diverge sharply under missing data: without history dropout, the architecture deteriorates to +49% (US) and +10% (China) worse than MOS when records are interrupted for >= 12 h, and local fusion to +72% and +60% (Fig. 5a, b). Because NEP is trained with whole-station history dropout (p=0.2), it achieves graceful degradation under both controlled and natural telemetry outages (Fig. 5a, b). When the last k hours of local observations are withheld (k in {0, 1, 3, 6, 12, 24, 48, 72} h), NEP consistently outperforms calibrated NWP: in the US, MAE remains 1.1–11.2% superior to MOS; in China, MAE remains 6.0–17.0% superior. Conversely, the no-exchange control degrades to +8.7% (US) and +5.3% (China) worse than MOS at k=12 h.

The value of inter-station exchange expands dramatically as local observations age: in the US, the exchange advantage rises from 2.3% (k=0) to 9.0% (k=12 h); in China, it surges from 2.7% to 10.7% (Fig. 5c, d). At 1 h forecast lead under extended outages (k >= 12 h), inter-station exchange provides a 19–23% error reduction in the US and 20–26% in China. During natural operational outages where the latest observation is >6 h old, NEP's advantage over local fusion surges from 7.0% (China) and 4.6% (US) under fresh observations to 19.9% in China [95% CI: 15.0, 24.6]% and 30.7% in the US [95% CI: 26.0, 34.7]%. Surrounding backbone stations effectively substitute for the missing on-site data.

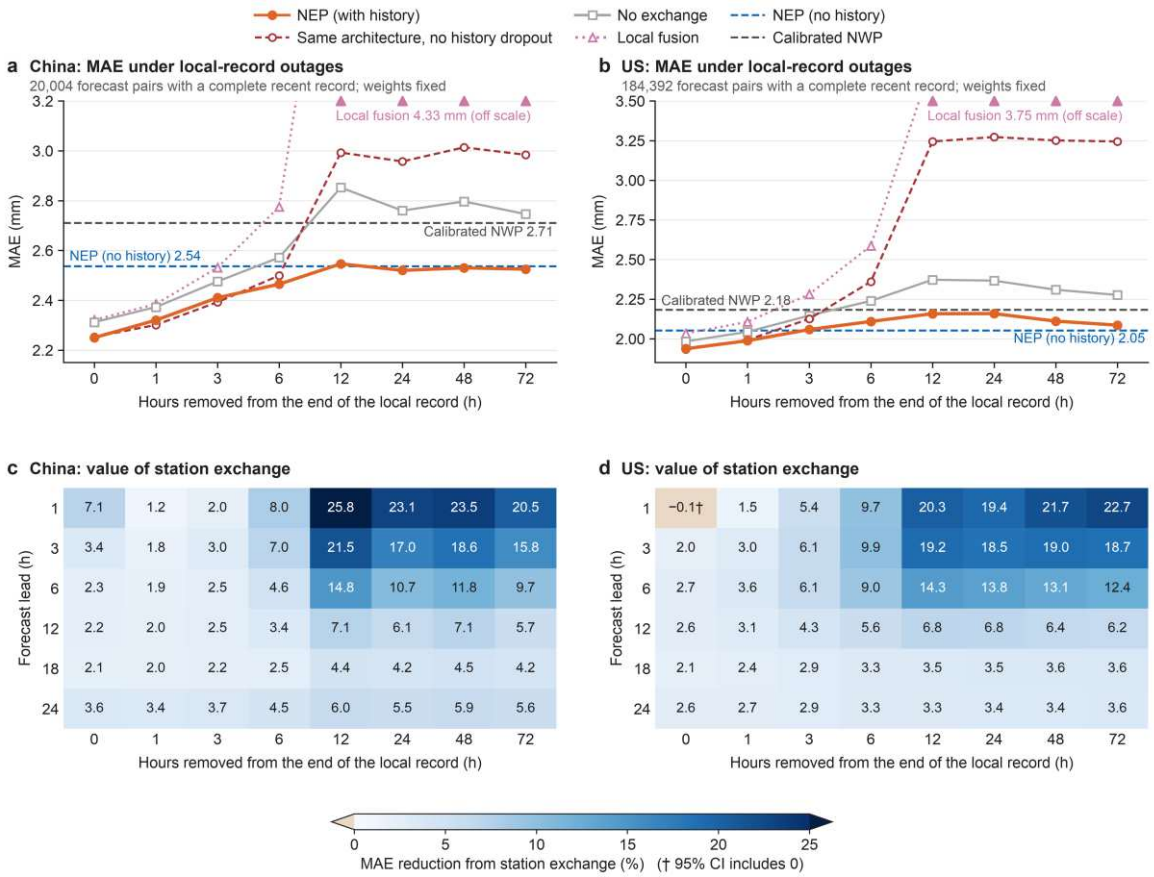


Fig. 5 *Forecast behavior when the local GNSS record is interrupted. (a, b) MAE when the last k hours of the target's GNSS record are removed before each forecast (k = 0–72 h), with model weights fixed. Only forecast windows with a complete recent record are used (China: 20,004 forecast pairs; US: 184,392), so values differ from Fig. 3. The red dashed line with open circles shows the same architecture trained without history dropout. Horizontal dashed lines give calibrated NWP and NEP without local history; local-fusion values beyond the axis are marked at the top edge. (c, d) MAE reduction from station exchange, that is, NEP with local history relative to the matched no-exchange model under the same interruption, by forecast lead; † marks the only cell whose 95% confidence interval includes zero. Both heat maps share one color scale*

During severe tropical cyclones, where rapid moisture advection challenges empirical post-processing (He et al. 2020; Zhu et al. 2020), NEP sustains strong predictive skill across 798 storm cases in China (18 cyclones; Fig. 6, Fig. S1) and 1,934 cases in the US (12 cyclones; Table 4). Across China storms, NEP with local history achieves an MAE of 2.911 mm, outperforming local fusion (3.232 mm) by 9.9% [storm-clustered 95% CI: 4.5, 14.4]%. Inter-station exchange contributes a 4.9% error reduction [95% CI: 0.7, 8.9]% over the no-exchange control (3.061 mm). In Level 1 leave-one-out testing without local history, NEP achieves an 8.7% gain over MOS [95% CI: 4.8, 12.9]% during China cyclones (Fig. S1). During Super Typhoon Lekima (August 2019; Fig. 6), station DWZH experienced a close passage (42 km; Fig. 6a); NEP without local history captured the moisture increase following upstream backbone anomalies (Fig. 6b, c), outperforming calibrated NWP in 26 of 36 forecasts issued during the event (Fig. 6d) and yielding a 4.9% error reduction across all 20 Lekima-affected stations.

In the US, NEP achieves a 7.8% gain over local fusion [95% CI: 3.2, 11.0]%. However, the marginal contribution of inter-station exchange during tropical cyclones is statistically neutral (+0.6%, [95% CI: −1.7, 2.5]%; Table 4), indicating that cyclone skill derives primarily from single-station nonlinear GNSS–NWP integration rather than multi-station advective exchange. In addition, analysis of peak moisture amplitudes indicates that all empirical models under-forecast peak cyclone PWV relative to calibrated NWP (median error 3.21 vs. 3.08 mm in China; superior in 56% of 32 events).

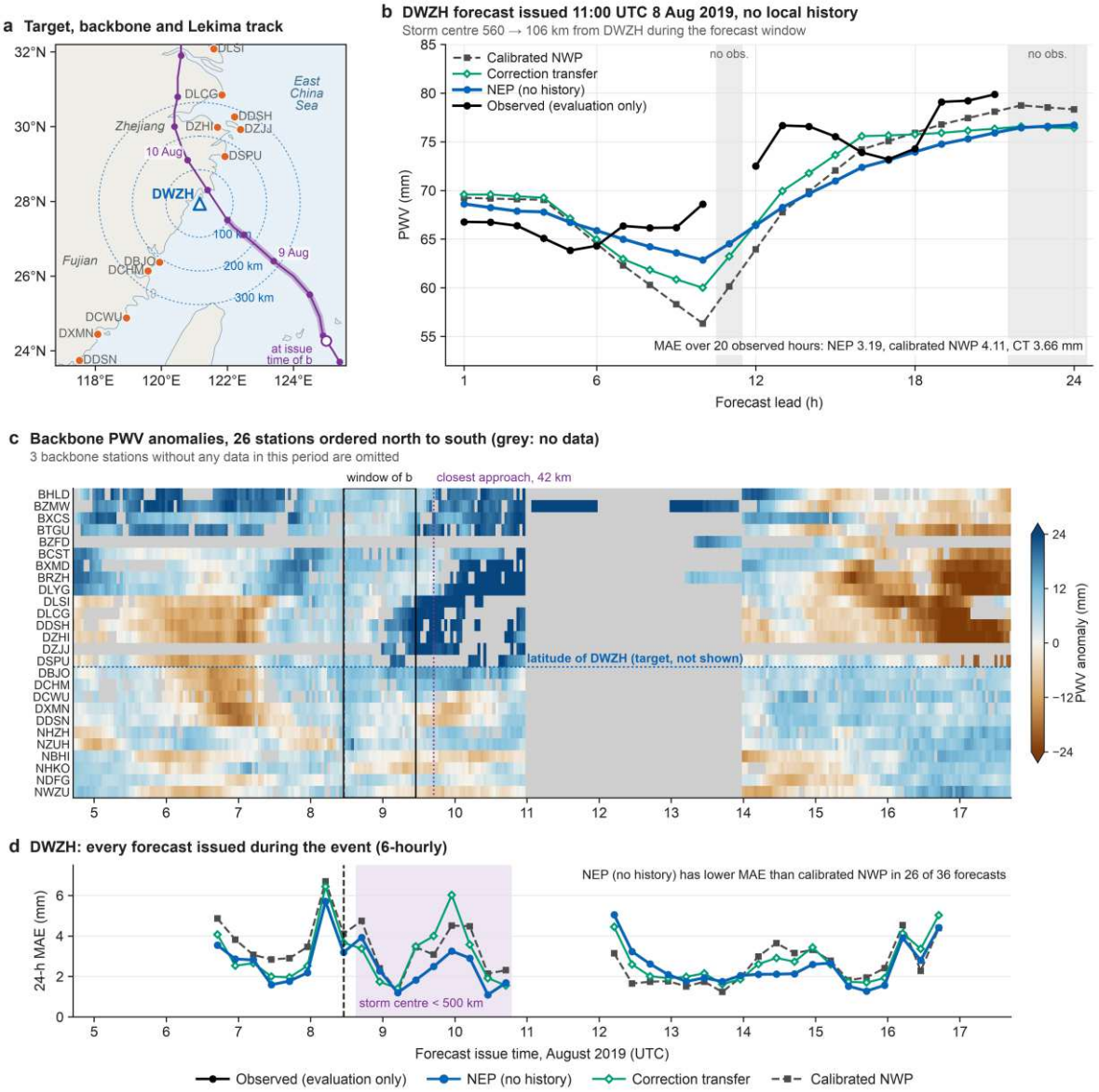


Fig. 6 *A no-history forecast at DWZH as Super Typhoon Lekima approached, August 2019. (a) DWZH (target; its observations are used for evaluation only) and the nearby backbone stations; dotted circles mark 100, 200 and 300 km from DWZH. The purple line is the best track of Lekima (IBTrACS v04r01, 2019214N15134; Knapp et al. 2010) with 6-hourly positions; the thick segment marks the 24 h window of (b) and the open circle the position at issue time. (b) The 24 h forecast issued at 11:00 UTC on 8 August 2019 without local history, with calibrated NWP, correction transfer and observations at DWZH; gray shading marks hours without observations. (c) PWV anomalies at 26 backbone stations ordered from north to south; gray cells are missing data, and three stations without data in this period are omitted. The box marks the forecast window of (b) and the dotted line the closest approach of the storm center (42 km, 17:00 UTC 9 August). (d) 24 h MAE at DWZH for every forecast issued at 6 h intervals during the event, plotted at issue time; shading marks the period when the storm center was within 500 km of DWZH. Across the 20 stations affected by Lekima the gain of NEP without local history over calibrated NWP was 4.9%, and across all tropical-cyclone forecasts in China it was 8.7% [4.8, 12.9] (Fig. S1). All times are UTC*

4.6. Component ablation in the history-dropout regime

To determine the architectural mechanisms enabling spatial transfer, we conducted systematic ablations in the history-dropout regime across both the source backbone network and external targets (Table 6).

At source backbone stations, most skill gains are already captured by the single-station neural core and unconstrained spatial exchange (reducing MAE from 2.212 to 2.052 mm), while removing the continuous geographic distance bias changes MAE by only 0.005 mm (0.25%; Table 6).

However, evaluating the identical model weights across the 19 unseen external targets reveals the decisive role of the physical distance prior. Without local history (Level 1), removing the geographic distance bias inflates MAE from 2.680 mm to 2.845 mm—a 5.8% performance loss [95% CI: 4.9, 6.7]%, causing spatial exchange to fall behind the matched no-exchange control (2.845 vs. 2.834 mm, −0.4%). With local history (Level 2), removing geographic bias incurs an analogous 4.6% loss (2.421 to 2.539 mm [95% CI: 4.0, 5.3]%), again dropping below the no-exchange control (2.492 mm, −1.9%). Removing NWP tendencies results in smaller penalties (0.2–0.5%). This contrast demonstrates that while unconstrained graph attention can fit training coordinates through arbitrary learned weights, the continuous geographic distance prior $\log(A + \varepsilon)$ serves as a critical geographic regularizer preventing spatial overfitting and enabling reliable spatial transfer across unmonitored targets.

Table 6 Component ablation in the history-dropout regime across the source backbone network and the 19 external targets. Values are MAE (mm); neural results are means (± seed standard deviation) over five random seeds

Configuration	Exchange	Geographic bias	NWP tendency	Source backbone	External targets	
				Level 2	Level 1	Level 2
Calibrated NWP (MOS)				–	2.881	–
Local fusion				2.212	–	2.568
No-exchange control	–	–	✓	2.131 ± 0.002	2.834	2.492
Exchange without geographic bias	✓	–	✓	2.052 ± 0.004	2.845	2.539
Exchange without NWP tendency	✓	✓	–	2.050 ± 0.003	2.693	2.426
Full network (NEP)	**✓**	**✓**	**✓**	**2.047 ± 0.003**	**2.680**	**2.421**

Level 1 is scored against MOS and Level 2 against local fusion; – marks a setting that is not applicable or a component that is removed.

5. Discussion

5.1. Observational leverage and two-tier operational deployment

The core scientific finding of this work is that a sparse GNSS backbone can project predictive skill far beyond its physical sensor footprint. In traditional GNSS meteorology, geodetic receivers are treated as localized point sensors,

requiring dense physical deployment to achieve regional coverage. Here, we demonstrate observational leverage: a sparse array of 30 stations was validated across 108 additional regional targets—3.6 times the number of backbone stations. Furthermore, our network scaling experiments demonstrate that as few as 10 stations capture roughly half (47%) of the gain achieved by a dense 60-station configuration, with gains plateauing near 30–45 stations.

This establishes a pragmatic two-tier operational service paradigm: (1) Tier 1 (Regional coverage): National meteorological services can deploy NEP without local history to deliver immediate, continuous post-processing corrections across unmonitored coastal infrastructure, offshore islands, and agricultural districts using existing backbone networks, without requiring the deployment of additional GNSS hardware at every forecast location; and (2) Tier 2 (Targeted precision upgrade): Where critical facilities (e.g., coastal megacities, major ports, launch sites) require enhanced precision, installing a local GNSS receiver delivers a 9.7% accuracy upgrade while benefiting from graceful degradation during sensor outages.

5.2. Climatological moisture coherence and spatial extrapolation mechanisms

Why does NEP maintain predictive skill across 200–300 km footprints while inverse-distance correction transfer decays sharply beyond 100 km? Correction transfer relies strictly on point-wise interpolation of surface residuals, implicitly assuming that NWP errors are stationary in space. However, coastal atmospheric moisture errors are dynamically organized by advective displacement of fronts, marine boundary-layer depth variations, and coastal convergence lines. NEP couples GNSS observation departures with continuous NWP tendencies through geographically biased graph attention. Rather than interpolating static errors, this formulation is designed to capture patterns consistent with large-scale dynamical modulation of regional error structures. The geographic regularizer $\log(A + \varepsilon)$ anchors the attention mechanism, enabling the model to exploit synoptic-scale error coherence across hundreds of kilometers without overfitting to training coordinates.

Regional variability in the exchange contribution likely reflects underlying climatological differences in moisture coherence. In China, station exchange contributes 3.9% across the backbone, ranging from 1.3–1.9% in the temperate Bohai–Yellow Sea domain to 10.7% [95% CI: 9.3, 12.0]% in the South China Sea (Online Resource 1, Table S6); in the US, exchange contributes 1.5%. Tropical and subtropical coastlines characterized by sustained monsoonal moisture channels and convective inflow exhibit stronger spatial autocorrelation, which is consistent with upstream GNSS observations providing transferable predictive guidance. In contrast, mid-latitude regimes

governed by fast-moving baroclinic waves exhibit more fragmented, transient error structures where spatial exchange yields smaller marginal benefits.

5.3. Operational hazards, severe weather, and disclosures

Operationally, while empirical verification is anchored to discrete geodetic sites with ground-truth PWV, the architecture interfaces directly with continuous NWP prognostic grids (0.25° × 0.25°). By querying elevation models and NWP profiles, the trained backbone can dynamically project corrections across maritime bays, offshore shipping channels, and unmonitored coastal plains, extending isolated GNSS networks into gridded regional guidance. While Level 1 can evaluate at arbitrary regional coordinates with numerical guidance, empirical validation here is strictly anchored to independent coastal stations. Future field campaigns with maritime GNSS buoys, shipborne receivers, or airborne dropsonde profiles will extend ground-truth validation across open oceanic expanses.

Extending GNSS post-processing to convective hazards provides valuable early warning utility (Benevides et al. 2019). As detailed in Online Resource 1, Section S5, coupling NEP PWV forecasts with GEFS precipitation yields an ROC AUC of 0.948 for heavy-rain discrimination (increasing detection probability to 0.603) and improves extreme water vapor event identification (90th percentile) by reducing false alarms.

In adherence to rigorous scientific standards, several limitations must be explicitly disclosed: (1) Tropical cyclone peak amplitudes: While NEP reduces aggregate MAE during cyclones by 9.9% in China and 7.8% in the US, it does not improve peak moisture amplitude forecasting relative to calibrated NWP (better in 56% of China events and 42% of US events). This under-forecasting stems from conditional-mean training objectives (MAE and MSE), which exhibit regression-to-the-mean tendencies during rare, sharp moisture spikes; (2) US tropical cyclone exchange neutrality: In the US, the marginal exchange contribution during cyclones is statistically neutral (+0.6%, 95% CI [−1.7, 2.5]%), indicating skill is driven by local GNSS–NWP nonlinear fusion rather than multi-station advection. This reflects storm-track geometry: while typhoons landfall perpendicular to China's coastal backbone providing upstream precursors, US hurricanes frequently track parallel to the coast or recurve from open Gulf waters lacking upstream sensors; (3) Latitudinal sensitivity: Gains concentrate in warm, humid coastal zones (rank correlation −0.83 in US, −0.76 in China) and diminish at higher latitudes (>= 40°N); and (4) Topographic

complexity: The continuous geographic prior was validated along coastal plains and rolling hills; regions with extreme vertical relief will require incorporating explicit topographic lapse rates into the graph formulation.

6. Conclusions

This study introduces a sparse-GNSS-backbone regional forecasting framework addressing a fundamental operational bottleneck in GNSS meteorology: dependence on dense, collocated sensor archives. By coupling sparse geodetic anchors with continuous NWP prognostic tendencies through geographically regularized graph attention, NEP shifts the operational paradigm from isolated single-station fitting to a coordinated network service. Trained with whole-station history dropout, a single architecture unifies a two-tier operational hierarchy without requiring dedicated models for different observation environments.

The empirical findings establish three core advances: (1) Regional projection beyond sensor footprints: In unmonitored coastal areas lacking GNSS receivers, Level 1 regional coverage projects calibration from surrounding backbone stations, reducing raw numerical forecast error by 12.8% (7.0% over calibrated NWP) and sustaining skill across 200–300 km regional footprints where conventional interpolation degrades; (2) Modular precision enhancement via local instrumentation: Where physical GNSS receivers exist, on-site records provide an optional precision upgrade (yielding a 9.7% error reduction over regional coverage), enabling agencies to prioritize dense instrumentation only at critical infrastructure; and (3) Observational leverage and operational resilience: As few as 10 backbone stations capture roughly half of the maximum regional benefit, with marginal gains tapering near 30–45 stations. Under telemetry interruptions and severe tropical cyclones, inter-station exchange acts as a dynamic buffer, expanding predictive advantages over local baselines when on-site measurements fail. By projecting high-precision corrections across unmonitored regional domains from a sparse geodetic backbone, this framework provides operational agencies with a scalable, resource-efficient strategy for coastal convective hazard forecasting.

Statements and declarations

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Competing interests

The authors declare no competing financial or non-financial interests directly related to the work described in this paper.

Author contributions

All authors contributed to the study conception and design. Peng Min: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft. Jie Zhao: Project administration, Supervision, Methodology, Writing – review & editing. Linrui Liu: Methodology, Data curation, Visualization. You Yin: Investigation, Validation. Ganzhi Luo: Resources, Supervision. Han Su: Conceptualization, Supervision, Writing – review & editing. All authors read and approved the final manuscript.

Data availability

The China coastal GNSS precipitable water vapor dataset is accessible via the National Earth System Science Data Center (Wu et al. 2025). Nevada Geodetic Laboratory tropospheric products are publicly available from the University of Nevada, Reno (https://geodesy.unr.edu; Blewitt et al. 2018). GEFS v12 reforecast guidance is available from NOAA (Guan et al. 2022). Tropical cyclone tracks are curated from the IBTrACS archive (Knapp et al. 2010).

Code availability

The complete Python analysis scripts, benchmark configuration files, and evaluation pipelines will be made publicly available in a persistent public repository upon publication (and are available to reviewers upon request during peer review).

Generative AI disclosure

During the preparation of this manuscript, the authors utilized ChatGPT (OpenAI) for English text polishing, style formatting, and automation scripting. Generative AI tools were not used to generate or interpret scientific results, perform statistical inference, or determine the conclusions of the study. All outputs were reviewed and verified by the authors, who take full responsibility for the contents of this publication.

References

Baran S, Lakatos M (2024) Clustering-based spatial interpolation of parametric postprocessing models. Weather Forecast 39:1591–1604. https://doi.org/10.1175/WAF-D-24-0016.1

Benevides P, Catalao J, Miranda PMA (2015) On the inclusion of GPS precipitable water vapour in the nowcasting of rainfall. Nat Hazards Earth Syst Sci 15:2605–2616. https://doi.org/10.5194/nhess-15-2605-2015

Benevides P, Catalao J, Nico G (2019) Neural network approach to forecast hourly intense rainfall using GNSS precipitable water vapor and meteorological sensors. Remote Sens 11:966. https://doi.org/10.3390/rs11080966

Bevis M, Businger S, Chiswell S, Herring TA, Anthes RA, Rocken C, Ware RH (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

Bevis M, Businger S, Herring TA, Rocken C, Anthes RA, Ware RH (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

Bezcioglu M (2026) Long-sequence deep learning frameworks for multivariate forecasting of tropospheric parameters. GPS Solut 30:90. https://doi.org/10.1007/s10291-026-02054-4

Blewitt G, Hammond WC, Kreemer C (2018) Harnessing the GPS data explosion for interdisciplinary science. Eos 99. https://doi.org/10.1029/2018EO104623

Chen B, Dai W, Liu Z, Wu L, Kuang C, Ao M (2018) Constructing a precipitable water vapor map from regional GNSS network observations without collocated meteorological data for weather forecasting. Atmos Meas Tech 11:5153–5166. https://doi.org/10.5194/amt-11-5153-2018

Feik M, Lerch S, Stühmer J (2024) Graph neural networks and spatial information learning for post-processing ensemble weather forecasts. In: ICML Workshop on Machine Learning for Earth System Modeling. https://arxiv.org/abs/2407.11050

Gendt G, Dick G, Reigber C, Tomassini M, Liu R, Ramatschi M (2004) Demonstrating near real-time GPS water vapor ground reception for numerical weather prediction. J Meteorol Soc Jpn 82:361–370. https://doi.org/10.2151/jmsj.2004.361

Ghaffari-Razin SR, Davari Majd R, Hooshangi N (2023) Regional modeling and forecasting of precipitable water vapor using least square support vector regression. Adv Space Res 71:4725–4738. https://doi.org/10.1016/j.asr.2023.01.030

Glahn HR, Lowry DA (1972) The use of Model Output Statistics (MOS) in objective weather forecasting. J Appl Meteorol 11:1203–1211. https://doi.org/10.1175/1520-0450(1972)011<1203:TUOMOS>2.0.CO;2

Guan H, Zhu Y, Sinsky E, Fu B, Li W, Zhou X, Xue X, Hou D, Peng J, Nageswararao MM, Tallapragada V, Hamill TM, Whitaker JS, Bates G, Pegion P, Frederick S, Rosencrans M, Kumar A (2022) GEFSv12 reforecast dataset for supporting subseasonal and hydrometeorological applications. Mon Weather Rev 150:647–665. https://doi.org/10.1175/MWR-D-21-0245.1

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

Han X, Wang X, Wu J, Wang Y, Awange J, Zhang B, Chan TO (2026) A spatiotemporal framework for GNSS-based rainfall nowcasting with FluxFormer and attention-based graph neural networks. GPS Solut 30:160. https://doi.org/10.1007/s10291-026-02121-w

He Q, Zhang K, Wu S, Zhao Q, Wang X, Shen Z, Li L, Wan M, Liu X (2020) Real-time GNSS-derived PWV for typhoon characterizations: a case study for Super Typhoon Mangkhut in Hong Kong. Remote Sens 12:104. https://doi.org/10.3390/rs12010104

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 RJ, Hólm E, Janisková M, Keeley S, Laloyaux P, Lopez P, Lupu C, Radnoti G, de Rosnay P, Rozum I, Vamborg F, Villaume S, Thépaut JN (2020) The ERA5 global reanalysis. Q J R Meteorol Soc 146:1999–2049. https://doi.org/10.1002/qj.3803

Hu F, Wei P, Sha Z, Ye S, Xia P, Zhao L, Luo J (2026) A fusion ZTD grid model utilizing ERA5 and GNSS raw observations for enhancing PPP and optimizing PWV forecasts under extreme weather conditions. GPS Solut 30:65. https://doi.org/10.1007/s10291-026-02028-6

Kapoor S, Narayanan A (2023) Leakage and the reproducibility crisis in machine-learning-based science. Patterns 4:100804. https://doi.org/10.1016/j.patter.2023.100804

Knapp KR, Kruk MC, Levinson DH, Diamond HJ, Neumann CJ (2010) The International Best Track Archive for Climate Stewardship (IBTrACS). Bull Am Meteorol Soc 91:363–376. https://doi.org/10.1175/2009BAMS2755.1

Künsch HR (1989) The jackknife and the bootstrap for general stationary observations. Ann Stat 17:1217–1241. https://doi.org/10.1214/aos/1176347265

Landskron D, Böhm J (2018) 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, Zhao Q, Guo H, Li Z, Ma Y, Yao Y, Yin J, Zhai Y, Liang H, Xiong Z (2026) A GNSS PWV filling and short-term forecasting framework fused hybrid neural network. Atmos Res 329:108508. https://doi.org/10.1016/j.atmosres.2025.108508

Liu HL, Zhou XQ, Zhu YY, Duan MZ, Chen B, Zhang SL (2024) Evaluation and correction of GFS water vapor products over United States using GPS data. Remote Sens 16:3043. https://doi.org/10.3390/rs16163043

Lu C, Zhang X, Zheng Y, Liu C, He B (2024) Real-time high-resolution tropospheric delay mapping based on GFS forecasts and GNSS. GPS Solut 28:175. https://doi.org/10.1007/s10291-024-01722-7

Nie Y, Nguyen NH, Sinthong P, Kalagnanam J (2023) A time series is worth 64 words: Long-term forecasting with Transformers. In: International Conference on Learning Representations (ICLR). https://arxiv.org/abs/2211.14730

Rasp S, Lerch S (2018) Neural networks for postprocessing ensemble weather forecasts. Mon Weather Rev 146:3885–3900. https://doi.org/10.1175/MWR-D-18-0187.1

Rocken C, Van Hove T, Ware R (1997) Near real-time GPS sensing of atmospheric water vapor. Geophys Res Lett 24:3221–3224. https://doi.org/10.1029/97GL03312

Shao Z, Zhang Z, Wang F, Wei W, Xu Y (2022) Spatial-temporal identity: a simple yet effective baseline for multivariate time series forecasting. In: Proceedings of the 31st ACM International Conference on Information & Knowledge Management (CIKM), pp 4454–4458. https://doi.org/10.1145/3511808.3557702

Shoji Y (2013) Retrieval of water vapor inhomogeneity using the Japanese nationwide GPS array and its potential for prediction of convective precipitation. J Meteorol Soc Jpn 91:43–62. https://doi.org/10.2151/jmsj.2013-103

Shoji Y, Kunii M, Saito K (2011) Mesoscale data assimilation of Myanmar Cyclone Nargis Part II: assimilation of GPS-derived precipitable water vapor. J Meteorol Soc Jpn 89:67–88. https://doi.org/10.2151/jmsj.2011-105

Vannitsem S, Wilks DS, Messner JW (eds) (2018) Statistical postprocessing of ensemble forecasts. Elsevier, Amsterdam. https://doi.org/10.1016/C2016-0-03244-8

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. In: Advances in neural information processing systems 30, pp 5998–6008

Veličković P, Cucurull G, Casanova A, Romero A, Liò P, Bengio Y (2018) Graph attention networks. In: International Conference on Learning Representations (ICLR). https://arxiv.org/abs/1710.10903

Wang J, Zhang L, Dai A, Van Hove T, Van Baelen J (2007) A near-global, 2-hourly data set of atmospheric precipitable water from ground-based GPS measurements. J Geophys Res Atmos 112:D11107. https://doi.org/10.1029/2006JD007529

Wilks DS (2011) Statistical methods in the atmospheric sciences, 3rd edn. Academic Press, Oxford

Wu H, Hu T, Liu Y, Zhou H, Wang J, Long M (2023) TimesNet: Temporal 2D-variation modeling for general time series analysis. In: International Conference on Learning Representations (ICLR). https://openreview.net/forum?id=ju_Uqw384Oq

Wu Y, Zhuang D, Labbe A, Sun L (2021) Inductive graph neural networks for spatiotemporal kriging. In: Proceedings of the AAAI Conference on Artificial Intelligence 35, pp 4478–4485. https://doi.org/10.1609/aaai.v35i5.16575

Wu Z, Li B, Liu Q, Liu Y, Zhang H, Zhou D, Liu Y (2025) China coastal GNSS network: advancing precipitable water vapor monitoring and applications in climate analysis. Earth Syst Sci Data 17:5841–5858. https://doi.org/10.5194/essd-17-5841-2025

Yang Q, Giezendanner J, Civitarese DS, Jakubik J, Schmitt E, Chandra A, Vila J, Hohl D, Hill C, Watson C, Wang S (2025) Local off-grid weather forecasting with multi-modal Earth observation data. J Adv Model Earth Syst 17:e2025MS005207. https://doi.org/10.1029/2025MS005207

Yang L, Wang Y, Guo B, Gao J, Chen D, Miao L (2026) A new ZTD modeling method for areas with sparse GNSS stations by fusing surface meteorological observations. GPS Solut 30:36. https://doi.org/10.1007/s10291-025-01999-2

Yu B, Yin H, Zhu Z (2018) Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI), pp 3634–3640. https://doi.org/10.24963/ijcai.2018/505

Zhang Y, Yan J (2023) Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting. In: International Conference on Learning Representations (ICLR). https://openreview.net/forum?id=vSVLM2j9eie

Zhao Q, Yao W, Yao Y, Li X (2020) An improved GNSS tropospheric tomography method with the GPT2w model. GPS Solut 24:60. https://doi.org/10.1007/s10291-020-0974-4

Zheng C, Fan X, Wang C, Qi J, Chen C, Chen L (2023) INCREASE: inductive graph representation learning for spatio-temporal kriging. In: Proceedings of the ACM Web Conference 2023, pp 673–683. https://doi.org/10.1145/3543507.3583525

Zhou X, Zhu Y, Hou D, Fu B, Li W, Guan H, Sinsky E, Kolczynski W, Xue X, Luo Y, Peng J, Yang B, Tallapragada V, Pegion P (2022) The development of the NCEP Global Ensemble Forecast System version 12. Weather Forecast 37:1069–1084. https://doi.org/10.1175/WAF-D-21-0112.1

Zhu M, Liu Z, Hu W (2020) Observing water vapor variability during three super typhoon events in Hong Kong based on GPS water vapor tomographic modeling technique. J Geophys Res Atmos 125:e2019JD032318. https://doi.org/10.1029/2019JD032318

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ESM1.pdf