applied sciences

MDPI

Article

Bounded Spatial Residual Adaptation for Multi-Station GNSS Displacement-Residual Forecasting

Shanshan Li [1,2], Zequn Wang [1], Qingjie Liu [1,2,*] and Guan Li [3]

[1] School of Computer Science and Information Security, University of Emergency Management, Beijing 101601, China; liss@cidp.edu.cn (S.L.); 25661614@st.cidp.edu.cn (Z.W.)
[2] Hebei Province University Smart Emergency Application Technology Research and Development Center, Beijing 101601, China
[3] Planning and Development Department, Big Data Center of Ministry of Emergency Management, Beijing 100054, China; lg1268@126.com
* Correspondence: liuqingjie@cidp.edu.cn

Abstract

Multi-station global navigation satellite system (GNSS) displacement-residual forecasting may benefit from neighbouring stations, yet unrestricted spatial aggregation can obscure station-local temporal dynamics and the independent value of spatial information. Whether neighbouring stations provide additional predictive information after a strong station-local forecast has been established remains unclear. Here, we propose the Multi-Scale Graph Residual Adaptation Network (MS-GRAN), a local-first framework that learns bounded spatial residual corrections rather than an unrestricted graph forecast. MS-GRAN first establishes a station-local NLinear forecast and then learns only an incremental residual correction through a training-only correlation–distance graph. The spatial branch integrates causal multi-scale high-frequency residual summaries, sparse non-negative adjacency constraints, neighbour-minus-own residual contrast, and a zero-output-initialized residual head with a gate initialized near zero to ensure a controlled spatial contribution. Experiments on a 90-station Cascadia GNSS benchmark with a fixed 2010–2019/2020–2021/2022–2024 chronological split showed that MS-GRAN achieved a 2.9158 $\pm$ 0.0033 mm one-day frozen-test RMSE, compared with 2.9579 $\pm$ 0.0006 mm for the matched frozen NLinear backbone. The resulting 1.422% reduction is modest but reproducible: 82-84 out of 90 stations improved across three random seeds, paired station-bootstrap confidence intervals were positive, upper-tail errors decreased more than mean RMSE, and gains remained positive across forecast horizons, network densities, seasons, and missingness strata. Additional controls showed gains of 0.635% for self-only adaptation, 1.265% for rewired placebo graphs, -0.343% for a validation-selected causal pre-local common-mode baseline, and 1.335% for an alternative frozen DLinear backbone. These results support conservative spatial residual adaptation in sparse GNSS networks while showing that residual-adaptation capacity contributes substantially, that simple causal common-mode preprocessing does not reproduce the gain, and that training-only graph topology provides a smaller additional benefit. They do not establish universal spatial transfer, physical event attribution, or earthquake-precursor detection.

Keywords: GNSS; displacement-residual forecasting; bounded spatial residual adaptation; local-first forecasting; leakage-controlled graph learning; deformation monitoring

Check for updates

Academic Editor: Christos Bouras

Received: 11 August 2026
Revised: 7 September 2026
Accepted: 10 September 2026
Published: 13 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

1.1. Multi-Station GNSS Displacement-Residual Forecasting

Continuous GNSS networks provide daily east, north, and up displacement records for regional deformation monitoring and reference-frame maintenance [1,2]. Before such records are used as forecast targets, long-term trends, seasonal signals, offsets, coloured noise, and missing observations must be treated explicitly in a manner consistent with the forecasting objective [3–8]. These issues are not only preprocessing details: they determine whether a forecast residual represents a meaningful deviation or an artefact of the coordinate time-series model.

Regional GNSS residuals can also contain spatially coherent components. Hydrological and other environmental loading can produce shared displacement signals [9,10], while common-mode-error studies show that reference-frame effects, processing artefacts, and network-wide noise can remain after standard trajectory modelling [11,12]. At the same time, each station retains local temporal behaviour related to equipment history, observation gaps, local noise, and site-specific environmental responses [8,13,14]. Spatial information is therefore forecast-useful only if it contributes incremental information beyond the target station's own history, rather than merely smoothing neighbouring records.

This creates a specific forecasting problem for sparse GNSS networks. Treating neighbouring stations as unrestricted predictors can dilute station-local temporal structure, whereas ignoring the network may miss coherent residual behaviour that is not captured by a local model. The methodological question is thus not whether a graph can be added, but whether spatial information contributes measurable residual information after a strong station-local temporal forecast has already been learned and fixed.

In this study, we focus on daily displacement-residual forecasting rather than event-label prediction. This scope is important because improved residual forecasts can support downstream deformation monitoring, anomaly screening, and quantitative deviation analysis; however, they do not by themselves establish physical precursors, assign event labels, or demonstrate earthquake predictability.

1.2. Temporal and Spatiotemporal Modelling: Progress and Remaining Gap

Classical GNSS trajectory and residual modelling define the local problem that any learning-based forecast must respect. Parametric trajectory models separate secular motion and deterministic terms [4], seasonal deformation studies show why periodic signals cannot be treated as noise [3,5], and offset-detection work addresses discontinuities that can mimic deformation [8,14]. Coloured-noise and missing-data-aware analyses further show that residual uncertainty depends on both stochastic structure and observation gaps [6,7,13]. Their limitation for the present question is not a lack of geodetic rigour, but scope: they establish station-level residual baselines without directly testing whether neighbouring stations retain marginal predictive information after a strong local forecaster has been fixed. The appropriate treatment of deterministic and seasonal components depends on the analysis objective; the present study therefore adopts a forecasting-oriented residual definition rather than a fully seasonally corrected geophysical residual.

GNSS machine-learning studies have recently expanded from single-model demonstrations to hybrid and workflow-oriented coordinate prediction. Recurrent and sliding-window models established the feasibility of station-wise GNSS denoising and prediction [15–17], while machine-learning studies of discontinuities and interpolation addressed change points and missing values rather than direct residual forecasting [18,19]. Decomposition-assisted and deformation-oriented studies further improved temporal prediction under nonstationary or hazard-specific settings [20–22]. More recent deep-learning work has tested broader GNSS station-position modelling [23], noise-resilient AVMD-

sLSTM-Transformer prediction [24], and a dedicated GNSS prediction toolbox combining signal decomposition with multiple learning models [25]. These studies demonstrate that GNSS time-series prediction is an active and maturing topic, but they generally optimize the temporal or decomposition pipeline rather than isolate the marginal value of neighbouring stations through a frozen-backbone spatial ablation.

General time-series forecasting further motivates the use of a strong local temporal backbone. Recurrent and convolutional models capture sequential dependence and multi-scale temporal context [26–29], while Transformer-family models address long-horizon structure through attention or decomposition mechanisms [30–33]. Patch-based Transformer and linear-decomposition studies are especially relevant here because they show that simple temporal baselines can be highly competitive [34,35]. For the present GNSS residual problem, this literature is useful mainly because it raises the comparison standard: a spatial model should be tested after a credible local forecaster has already been learned, not against a weak temporal baseline.

Spatial coherence motivates a graph pathway, but it also requires caution. GNSS common-mode-error and spatiotemporal filtering studies show that regional residuals may share coherent components, especially when missingness, loading, reference-frame effects, and processing artefacts interact [11,12]. Graph forecasting methods in traffic, sensor-network, and multivariate time-series settings show how edges can encode diffusion over known topology [36,37], adaptive or attention-based dependence [38–40], and learned cross-series structure [41,42]. Daily GNSS networks differ, however, because station spacing is irregular, geographic proximity is not equivalent to forecast-useful coupling, and station-local histories remain highly informative; studies of local effects in graph forecasting therefore motivate controlled spatial correction rather than unrestricted propagation [43]. This motivates training-only graph construction, sparse adjacency, neighbour-minus-own residual contrast, and gated fusion.

Missing-data, imputation, and anomaly-detection studies reinforce the need to separate forecast baselines from event interpretation. Missing-data models and graph imputation methods can exploit cross-series dependence under incomplete observations [44,45], and graph-based or deep anomaly-detection methods can use residual deviations for multivariate screening [46–48]. However, a lower forecast error only improves the expected-residual baseline used for screening; it does not establish event attribution, earthquake predictability, or precursor detection. The remaining gap is therefore not the lack of a more complex spatiotemporal architecture, but the lack of a controlled framework for isolating the marginal predictive value of spatial information after a strong station-local GNSS forecaster has been fixed. Table 1 summarizes this gap-mapping logic.

The broader deformation and forecasting literatures provide context for the modelling choices but do not by themselves resolve the controlled GNSS question. GNSS processing, plate-motion, coordinate-noise, and residual-modelling studies define the geodetic setting considered here [49–52]. InSAR deformation studies provide complementary examples of dense surface-deformation mapping and monitoring, but they address a different observational regime from sparse station-wise GNSS forecasting [53–58]. Modern temporal forecasting models, including probabilistic recurrent, basis-expansion, efficient-attention, and temporal-variation architectures, illustrate the range of sequence predictors against which a local temporal baseline should be considered [59–62]. Graph convolutional and spatiotemporal forecasting models motivate spatial aggregation mechanisms [63–65], while missing-data and anomaly-detection studies highlight why mask handling and residual reliability must be treated carefully in noisy geodetic time series [66–70]. Accordingly, these literatures motivate the present controlled design, but the central question remains

narrower: whether neighbouring GNSS stations add marginal predictive information after a strong station-local forecast has already been fixed.

Table 1. Representative studies informing leakage-controlled GNSS residual forecasting and the remaining need to isolate marginal spatial contribution after local temporal modelling.

Study	Data or Task	Method	Relevance and Remaining Limitations
Bos et al. (2013) [6]	Incomplete GNSS coordinate series	Missing-data-aware coloured-noise analysis	Shows that data gaps affect uncertainty estimation; not a forecasting model.
Bevis and Brown (2014) [4]	Crustal-motion trajectories	Parametric trajectory model	Separates deterministic motion, offsets, and seasonal terms; no learned spatial residual correction.
Gazeaux et al. (2013) [8]	GNSS offsets and discontinuities	Multi-method offset detection	Clarifies that discontinuities can mimic deformation; not an expected-residual forecasting benchmark.
Wei et al. (2025) [10]	GNSS and GRACE loading displacement	3D-CNN loading assessment	Shows that environmental loading can shape GNSS displacement residuals; external drivers are not used by MS-GRAN.
Xiao et al. (2024); Liu et al. (2026) [11,12]	Regional GNSS common-mode errors	PCA/ICA toolbox and MSSA-PCA filtering	Supports spatially coherent residual components; CME extraction is not the same as held-out residual forecasting.
Gao et al. (2022); Chen et al. (2023) [17,20]	GNSS time-series prediction	GBDT/LSTM/SVM and VMD-LSTM	Demonstrates station-wise and decomposition-assisted prediction; marginal neighbour value is not isolated.
Simsek et al. (2025) [23]	GNSS station position time series	Deep temporal models	Shows recent deep-learning feasibility for GNSS position prediction; mainly temporal rather than frozen-backbone spatial ablation.
Jiao et al. (2025) [24]	GNSS coordinate time-series prediction	AVMD–sLSTM–Transformer hybrid model	Addresses noisy and nonstationary coordinate prediction; complex temporal modelling does not isolate neighbour contribution.
He et al. (2026) [25]	GNSS prediction workflow/toolbox	Signal decomposition plus ML/deep-learning toolbox	Shows maturation of GNSS forecasting workflows; not designed to test bounded spatial residual adaptation after a frozen local model.
Nie et al. (2023); Zeng et al. (2023) [34,35]	General time-series forecasting	Linear and patch-based Transformer baselines	Motivates strong local temporal baselines and matched ablations before claiming spatial gain.

Table 1. *Cont.*

Study	Data or Task	Method	Relevance and Remaining Limitations
Li et al. (2018); Wu et al. (2019); Cini et al. (2023) [36,38,43]	Graph spatiotemporal forecasting	Diffusion/adaptive graph models and local-effect analysis	Graphs can exploit multivariate dependence, but local effects and overmixing require controlled spatial correction.
Cini et al. (2022); Deng and Hooi (2021); Ruff et al. (2021) [45,47,48]	Missing data and anomaly scoring	Graph imputation and anomaly-detection frameworks	Residual scores can support screening, but forecast improvement does not establish event attribution.

1.3. Objective and Contributions

The literature leaves a focused gap: for sparse daily GNSS networks, it remains unclear whether neighbouring stations provide reproducible predictive information after a strong station-local temporal forecast has already been learned under a leakage-controlled protocol. This gap matters because even small changes in expected residual error can affect station-level monitoring thresholds, while an unconstrained graph may appear useful simply by absorbing local temporal structure or test-period information.

We therefore formulate a focused and testable research question: whether training-only spatial relationships can improve frozen station-local GNSS displacement-residual forecasts when spatial information is restricted to a bounded residual correction pathway.

(1) We formulate multi-station GNSS displacement-residual forecasting as a local-first residual adaptation problem, in which spatial information is allowed to correct only the component not captured by a frozen station-local predictor.

(2) We develop MS-GRAN, a bounded graph residual adapter that combines causal multi-scale high-frequency residual summaries, training-only correlation–distance graph construction, neighbour-minus-own residual contrast, and a zero-output-initialized residual head and near-zero initial gate to prevent uncontrolled graph propagation.

(3) We evaluate the marginal spatial contribution using matched frozen-backbone ablation, self-only and rewired-placebo graph controls, a causal common-mode baseline, an alternative DLinear frozen backbone, baseline fairness auditing, HAC-robust predictive-ability testing, station-level uncertainty analysis, tail-error diagnostics, and forecast-horizon, network-density, seasonal, and missingness-stratified diagnostics.

2. Materials and Methods

2.1. Study Region, GNSS Data, and Preprocessing

2.1.1. Data Source and Cascadia Station Network

We used the Nevada Geodetic Laboratory (NGL) IGS20 24 h final tenv3 daily GNSS position series in the Cascadia region [1,2]. A fixed archive of 200 candidate stations was screened over the common interval from 1 January 2010 through 31 December 2024. A station was retained when at least 90% of the 5479 calendar days were observed and its longest continuous gap did not exceed 30 days. This deterministic quality-control rule retained 90 stations spanning 40.119–50.544° N and 121.076–126.843° W, thereby avoiding post hoc station selection based on forecast performance.

For each tenv3 record, integer and fractional coordinate fields were combined to recover complete local east, north, and up (E/N/U) coordinates in metres. Station identifiers, coordinates, modified Julian dates, and formal E/N/U uncertainties were retained as metadata. The formal uncertainties were used only for descriptive station-condition

stratification and were not provided to the forecasting model. No synthetic anomalies were injected in the numerical forecasting experiments.

Possible equipment changes and processing discontinuities are an important limitation of long GNSS records. The forecasting pipeline uses the NGL tenv3 daily position time series rather than reprocessing raw RINEX observations with full equipment metadata, and equipment or processing-step metadata are not used for model training, graph construction, validation selection, or frozen-test evaluation. Such changes can increase local variance, introduce persistent offsets, and perturb training-period inter-station correlations; they could therefore weaken or, in isolated cases, inflate an apparent spatial benefit. To assess whether registered discontinuities drive the reported test gain, we additionally performed a post hoc diagnostic using the public NGL steps.txt database, as discussed in Section 4.2. The main conclusions remain restricted to forecasting performance on the processed displacement-residual benchmark; more complete equipment-history analysis would still be required for geophysical attribution.

2.1.2. Calendar Alignment, Missing-Data Masks, and Chronological Split

The selected station files are mapped to a common daily calendar, producing an array of 5479 days $\times$ 90 stations $\times$ 3 components. This alignment step performs no interpolation, outlier deletion, or missing-value imputation. A missing coordinate remains NaN in the archived array and is represented by a Boolean station-day observation mask. The resulting observed fraction is 0.989102. Preserving the original gaps avoids treating interpolated values as observations in the forecasting benchmark.

The data are divided before model training: 2010–2019 for training, 2020–2021 for validation, and 2022–2024 for held-out testing. All preprocessing parameters and graph relationships are estimated without access to validation or test targets, while model parameters are optimized using the training split with validation-selected checkpoints. Forecast windows are required to lie wholly inside one chronological partition. Missing station-days remain missing; the observation mask is carried through window construction, model input, loss calculation, and evaluation. Figure 1 shows the fixed Cascadia station network, representative E/N/U residual records, and the chronological split used consistently throughout all experiments.

2.1.3. Detrending, Residual Construction, and Robust Normalization

Let $p_{t,i,c}$ denote the raw GNSS coordinate in metres on day t at station i and component $c \in \{E, N, U\}$; let $m_{t,i}$ be one when the station-day is observed and zero otherwise. With $\tau(t)$ denoting the number of days since the first epoch and $\mathcal{T}_{\text{train}}$ the training set, Equation (1) estimates the station- and component-specific intercept and linear trend using observed training epochs only:

$$a_{i,c}, b_{i,c} = \underset{a,b}{\operatorname{argmin}} \sum_{t \in \mathcal{T}_{\text{train}}} m_{t,i}[p_{t,i,c} - a - b\tau(t)]^2 \quad (1)$$

The fitted trend is extrapolated unchanged to the validation and test epochs. Equation (2) defines the detrended displacement residual $r_{t,i,c}$ in millimetres, where the factor 1000 converts metres to millimetres. This detrended residual representation is used throughout the spatial branch:

In the present forecasting benchmark, only the station- and component-specific linear trend is removed. Annual and semiannual harmonics are intentionally retained so that all forecasting methods operate on the same residual signal and can learn predictable seasonal variability from the historical context under identical information. Accordingly, "displacement residual" in this study denotes the linearly detrended and train-only normalized

displacement series rather than a fully seasonally corrected geophysical residual. Seasonal and environmentally driven components that remain in this residual may therefore contribute to the spatial correlations used by the graph model; this contribution is treated as part of forecastable residual structure and is not interpreted as physical attribution. A harmonic baseline is included as a classical comparison, but all learning models are evaluated on the same residual definition.

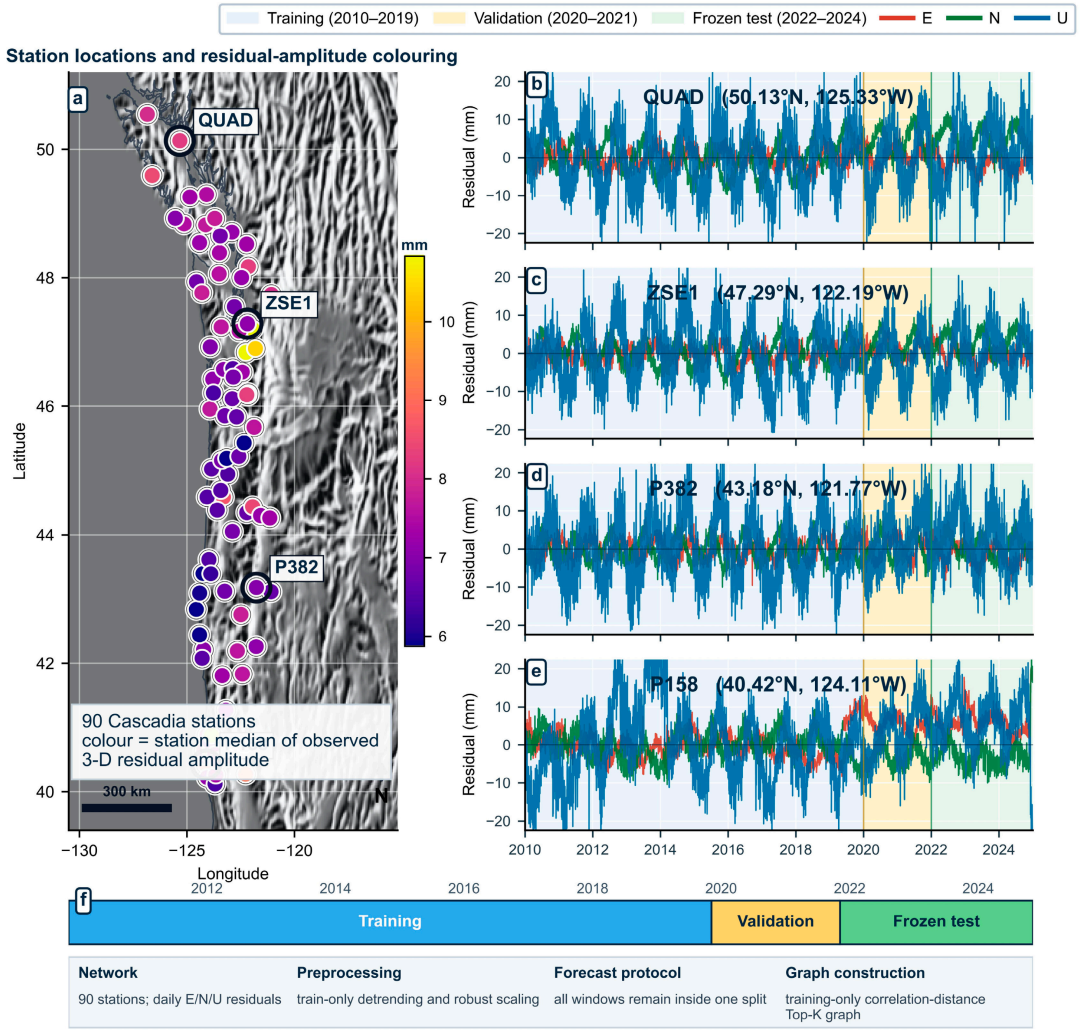


Figure 1. Cascadia GNSS network, representative displacement-residual series, and forecasting protocol. (**a**) Spatial distribution of the selected 90 Cascadia GNSS stations, where marker colours indicate the median observed three-dimensional residual amplitude. (**b**) Representative detrended E/N/U displacement-residual time series at station QUAD (50.13° N, 125.33° W). (**c**) Representative detrended E/N/U displacement-residual time series at station ZSE1 (47.29° N, 122.19° W). (**d**) Representative detrended E/N/U displacement-residual time series at station P382 (43.18° N, 121.77° W). (**e**) Representative detrended E/N/U displacement-residual time series at station P158 (40.42° N, 124.11° W). (**f**) Chronological data partition and forecasting protocol, including training (2010–2019), validation (2020–2021), frozen-test (2022–2024) periods, and the main preprocessing and graph-construction procedures.

$$r_{t,i,c} = 1000[p_{t,i,c} - a_{i,c} - b_{i,c}\tau(t)] \quad (2)$$

Train-only robust centring and scaling are then applied to $r_{t,i,c}$. Equation (3) and (4) define the training-only robust location and scale parameters, where the factor 1.4826 converts the median absolute deviation (MAD) to a robust Gaussian standard-deviation es-

timate, and the 0.1 mm floor prevents numerical instability for nearly constant components. Equation (5) defines the normalized residual $z_{t,i,c}$:

$$\mu_{i,c} = \text{median}_{t \in \mathcal{T}_{\text{train}}, m_{t,i}=1}(r_{t,i,c}) \tag{3}$$

$$\kappa_{i,c} = \max\left(1.4826\text{median}_{t \in \mathcal{T}_{\text{train}}, m_{t,i}=1}\left|r_{t,i,c} - \mu_{i,c}\right|, 0.1\right) \tag{4}$$

$$z_{t,i,c} = \frac{r_{t,i,c} - \mu_{i,c}}{\kappa_{i,c}} \tag{5}$$

The fitted trend and robust location-scale parameters are frozen after training-period estimation and applied unchanged to validation and test observations, ensuring that no future information is used during preprocessing.

2.2. Forecasting Problem Formulation

2.2.1. Input, Target, Forecast Horizon, and Mask

Let B denote the batch size, $N = 90$ the number of GNSS stations, $C = 3$ the displacement components, $L = 60$ the historical context length, and H the direct forecast horizon. For forecast origin t, Equation (6) defines the historical input tensor, future target tensor, context observation mask, target observation mask, and forecasting function. Only samples whose $L + H$ days remain entirely within the same chronological partition are retained, preventing train–validation–test window crossing.

$$\begin{aligned} X_t &\in \mathbb{R}^{B \times L \times N \times C}, \quad Y_t \in \mathbb{R}^{B \times H \times N \times C}, \\ M_t &\in \left\{0,1\right\}^{B \times L \times N}, \quad M_t^+ \in \left\{0,1\right\}^{B \times H \times N}, \\ \hat{Y}_t &= F^{\theta}(X_t, M_t, A), \quad \hat{Y}_t \in \mathbb{R}^{B \times H \times N \times C}. \end{aligned} \tag{6}$$

Here, X_t contains the historical detrended and normalized displacement residuals, while Y_t represents the future residual targets. The context mask M_t indicates observed historical station-days and is provided to the model to distinguish valid observations from missing values. The future target mask M_t records the availability of target observations and is used only for masked loss computation and evaluation, not as an input to the forecasting model.

2.2.2. Leakage-Controlled Learning Protocol

All preprocessing parameters, including trend coefficients and robust location-scale parameters, are estimated exclusively from the 2010–2019 training period. The spatial graph is also constructed using training-period observations only, based on the predefined correlation–distance criterion, and remains fixed throughout validation and testing. No validation or test observations are used to update preprocessing statistics, modify graph relationships, or select model parameters.

During model construction, missing normalized inputs are replaced by zeros immediately before neural operations to avoid undefined numerical operations, while the corresponding observation masks are retained and propagated through the spatial encoder, loss calculation, and evaluation procedures. Therefore, zero replacement serves only as a numerical implementation step and does not introduce artificial observations or remove missing-data information.

The forecasting model follows a strict chronological evaluation protocol. The station-local temporal backbone is trained using the training period and selected according to validation performance before being frozen. The residual adaptation branch is subsequently optimized without modifying the frozen local predictor. Predictions are inverse-

transformed back to millimetres only for metric reporting. This protocol ensures that any observed improvement can be interpreted as arising from the learned residual adaptation pathway rather than from future information access, adaptive preprocessing, or test-period graph reconstruction.

2.3. Local-First MS-GRAN Framework

2.3.1. Design Principle and Overall Architecture

MS-GRAN is designed to quantify incremental residual-adaptation contributions without replacing station-local temporal forecasting. The station-local NLinear backbone first establishes a strong station-local temporal reference and is subsequently frozen. The spatial residual branch receives only causal residual summaries and a training-derived adjacency matrix and influences the final forecast only through a bounded residual correction. This design ensures that improvements over the local baseline arise from the bounded residual-adaptation pathway rather than from a second unrestricted temporal forecaster; the self-only and rewired controls are then used to distinguish target-only adaptation from topology-specific cross-station effects.

The architecture follows a local-first principle. The NLinear backbone receives normalized E/N/U residual histories and produces the primary station-local forecast. The spatial residual branch then addresses a narrower question: whether neighbouring stations, selected only from the training split, can explain part of the remaining residual variability through a conservative correction pathway. Figure 2 summarizes the resulting local-first residual adaptation workflow.

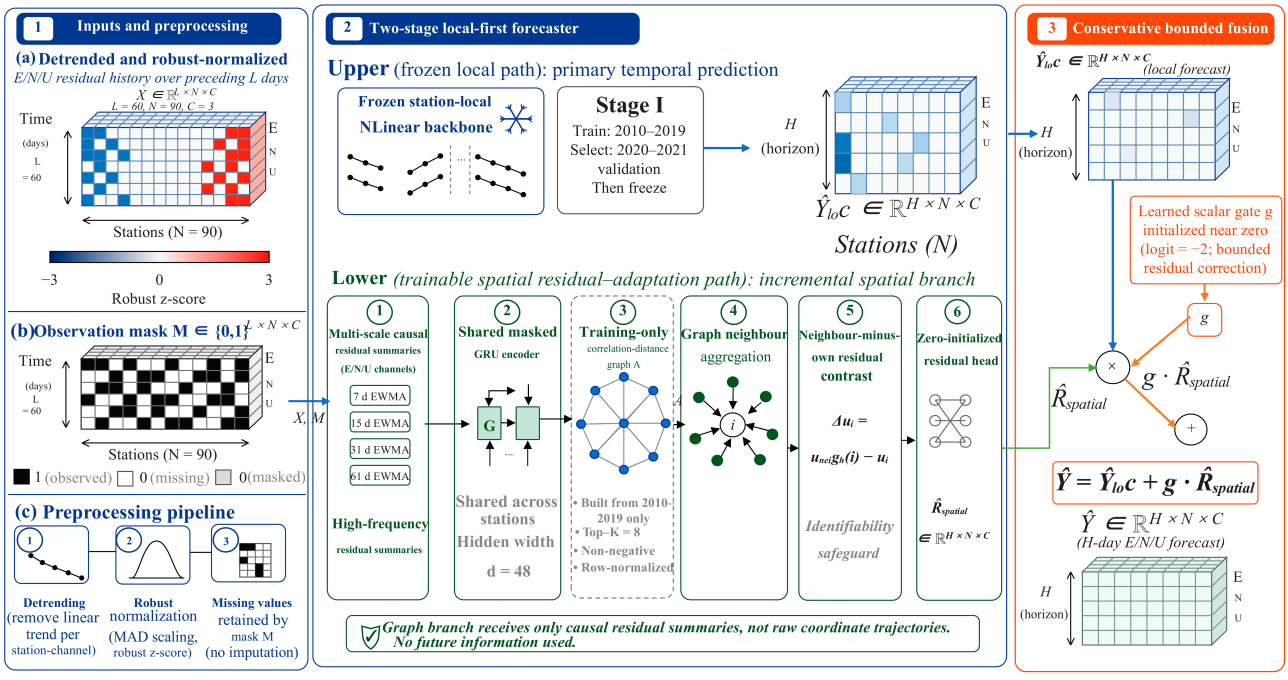


Figure 2. Simplified local-first architecture of MS-GRAN. A frozen station-local NLinear path produces the primary forecast. A separate trainable path encodes causal 7-/15-/31-/61-day residual summaries, aggregates a frozen training-only graph, forms neighbour-minus-own contrast, and outputs a zero-initialized spatial residual correction. A scalar gate initialized near zero controls the additive correction. The enlarged schematic omits implementation-level decoration to remain readable at publication size.

2.3.2. Station-Local NLinear Temporal Backbone

The first computational component is a station-local temporal backbone that provides the controlled temporal reference for evaluating marginal spatial contributions. For each

station and displacement component, the normalized residual history is processed independently without information exchange between stations. In the NLinear formulation adopted in this study, the last entry of the zero-filled normalized input window is used as the local-level anchor, and Equation (7) maps the normalized residual context to the direct forecast horizon through a shared linear projection:

$$y_{i,c}^{\text{local}} = W_N[x_{i,c} - \mathbf{1}_L x_{L,i,c}] + \mathbf{1}_H x_{L,i,c} W_N, \quad W_N \in \mathbb{R}^{H \times L} \tag{7}$$

where $x_{i,c}$ denotes the normalized residual history of station i and component c, $x_{L,i,c}$ represents the last-value anchor repeated across the historical window and removed before projection, $\mathbf{1}_L$ and $\mathbf{1}_H$ are all-one vectors of lengths L and H, respectively, and $x_{L,i,c}$ is the final normalized input entry after missing inputs have been set to zero, consistent with the adopted NLinear formulation used throughout model training and evaluation. Thus, the repeated level term in Equation (7) is the final zero-filled input entry rather than the temporal mean of the input window or a backward-searched last valid observation.

The same NLinear projection is shared across all stations and components, so the number of backbone parameters remains independent of the station count. This design provides a strong and scalable station-local temporal reference while avoiding implicit cross-station information exchange. The backbone is trained first using the training period, and the validation-selected checkpoint is frozen before any spatial residual-adaptation parameters are optimized.

Freezing the local backbone transforms the experiment from a direct model competition into a controlled evaluation of residual adaptation beyond the target station's own frozen forecast. Therefore, any subsequent improvement obtained by MS-GRAN reflects the learned residual-adaptation pathway rather than additional retraining of the local temporal backbone; the self-only and rewired controls then distinguish target-only adaptation from topology-specific cross-station effects.

2.3.3. Multi-Scale High-Frequency Residual Encoder

The spatial branch does not receive raw coordinate histories directly. Instead, it is restricted to causal high-frequency summaries of the detrended residual $\mathrm{r}_{\mathrm{t,i,c}}$, and these summaries are normalized only after their residual-scale parameters have been estimated from training data. Equation (8) defines the causal EWMA residual state used to construct these summaries. This design prevents the spatial module from accessing the full local trajectory representation and focuses spatial aggregation on short- to intermediate-scale residual deviations shared across stations.

$$\ell_{\mathrm{t,i,c}}^{(\mathrm{s})} = \mathrm{r}_{\mathrm{t,i,c}}$$

for the first observed epoch, and

$$\ell_{\mathrm{t,i,c}}^{(\mathrm{s})} = (1 - \beta_{\mathrm{s}})\ell_{\mathrm{t-1,i,c}}^{(\mathrm{s})} + \beta_{\mathrm{s}}\mathrm{r}_{\mathrm{t,i,c}}, \quad \mathrm{m}_{\mathrm{t,i}} = 1 \tag{8}$$

For each EWMA span s ∈ {7, 15, 31, 61}, the causal residual state l(t, i, c, s) is updated only from previously observed detrended residuals. The corresponding high-frequency residual summary is the difference between the current residual and the EWMA state when observations are available, while missing values are masked. Equation (9) normalizes, clips, and concatenates the four multi-scale residual summaries with the contemporaneous observation mask before neural encoding. The ordering is scale-major and component-major within each scale: 7-day E/N/U, 15-day E/N/U, 31-day E/N/U, and 61-day E/N/U residual summaries, followed by the E/N/U mask indicators. Thus, entries 1–12 are the

four residual summaries for each of the three components and entries 13–15 are the three binary mask indicators.

$$h^{(s)}_{t,i,c} = \frac{r_{t,i,c} - \ell^{(s)}_{t,i,c}}{\kappa_{i,c}}, \quad a_{t,i} = \left[\mathrm{clip}(h^{7}), \mathrm{clip}(h^{15}), \mathrm{clip}(h^{31}), \mathrm{clip}(h^{61}), m_{t,i}\right] \in \mathbb{R}^{15} \tag{9}$$

A single GRU encoder with hidden width $d = 48$ is shared across all stations. Equations (10) and (11) define the GRU update and the resulting residual representation u_i from the residual sequence $\{a_{1,i}, \ldots, a_{L,i}\}$.

$$\gamma_t = \sigma\left(W_{i\gamma} a_t + b_{i\gamma} + W_{h\gamma} h_{t-1} + b_{h\gamma}\right),$$

$$z_t = \sigma(W_{iz} a_t + b_{iz} + W_{hz} h_{t-1} + b_{hz}) \tag{10}$$

$$\tilde{h}_t = \tanh(W_{in} a_t + b_{in} + \gamma_t \odot (W_{hn} h_{t-1} + b_{hn})),$$

$$h_t = (1 - z_t) \odot \tilde{h}_t + z_t \odot h_{t-1}, \quad u_i = h_L \tag{11}$$

This encoder is not a second temporal forecaster. It exists only to summarize short- and intermediate-scale residual states before cross-station comparison. Equation (12) collects the encoded station residual states used by the graph module. Using multiple EWMA spans allows the adapter to respond to residual variations at several time scales while preserving the residual-only role of the graph pathway. Each station representation is $u_i \in \mathbb{R}^d$, where $d = 48$ is the GRU hidden width.

$$U_t = [u_1, u_2, \ldots, u_N]^T \in \mathbb{R}^{N \times d} \tag{12}$$

Only the encoded residual representation, observation-mask information, and frozen adjacency matrix are passed to the spatial module. Raw coordinates, validation-test statistics, and future targets are excluded from graph construction and spatial aggregation.

The role of this module is deliberately restricted: it provides station residual states for testing marginal cross-station information rather than generating an independent forecast. Neighbour stations are allowed to contribute only through the downstream graph aggregation and gated residual-fusion pathway. This separation keeps local temporal prediction, residual encoding, graph aggregation, and forecast fusion independently interpretable.

2.3.4. Training-Only Correlation–Distance Graph

The graph is constructed once from the 2010–2019 training partition and then frozen. Equation (13) defines component-wise Pearson correlation from mutually observed 31-day high-frequency residuals, with at least 60 common station-days required for a valid pair. Geographic distance is used as a physical support constraint rather than as proof of predictive dependence. For station coordinates, Equations (14) and (15) define the great-circle distance and the Gaussian distance kernel: φ_i and φ_j are the station latitudes and λ_i and λ_j are the station longitudes, all expressed in radians, and $R = 6371$ km is the Earth's radius. For each station, Equation (16) retains the K highest-scoring candidate neighbours after combining absolute correlation and geographic support. The mixture coefficient $\alpha = 0.75$ was fixed before frozen-test evaluation so that training-period residual correlation provides the primary evidence (75%) while geographic distance acts as a secondary physical-support constraint (25%); it was not selected using 2022–2024 test performance and is not claimed to be universally optimal. $K = 8$ was likewise fixed before frozen-test evaluation as a prespecified sparse operating point, rather than selected from frozen-test performance. The extended $K = \{4, 8, 12, 16, 24, 32\}$ analysis is a robustness diagnostic: denser graphs yield

only marginal additional frozen-test improvement, while increasing K from 8 to 24–32 increases the retained neighbourhood size by 3–4 times, raises graph density, and increases sparse aggregation and graph-storage cost. Although K = 24/32 achieve slightly lower frozen-test RMSE, these post hoc values were not used to replace the primary K = 8 configuration. K = 8 is therefore retained as a compact operating point that balances predictive performance, graph sparsity, and computational cost, rather than being claimed as universally optimal. Self-loops are then imposed, the directed candidate matrix is symmetrized by the element-wise maximum, and Equation (17) row-normalizes the resulting non-negative matrix. Here, $\varepsilon = 10^{-8}$ is a fixed numerical-stability constant used only in denominator terms to avoid division by zero; it is not a tuneable hyperparameter. The same ε is used in Equations (17), (23) and (24). Validation data are used only for checkpoint selection; neither validation nor frozen-test residuals revise graph edges or weights.

$$\rho_{ij} = \mathrm{median}_{c\in\{E,N,U\}}\mathrm{Corr}\left(r_i^c, r_j^c \middle| m_i = m_j = 1, \mathcal{T}_{\mathrm{train}}\right), \quad N_i = \mathrm{TopK}\left(\left|\rho_{ij}\right|\right), K = 8 \tag{13}$$

$$d_{ij} = 2R\cdot\arcsin\left(\sqrt{\left[\sin^2\left(\frac{\varphi_i - \varphi_j}{2}\right) + \cos(\varphi_i)\cos(\varphi_j)\sin^2\left(\frac{\lambda_i - \lambda_j}{2}\right)\right]}\right), R = 6371\ \mathrm{km} \tag{14}$$

$$D_{ij} = \exp\left[-0.5\left(\frac{d_{ij}}{100\ \mathrm{km}}\right)^2\right] \text{for geographic nearest neighbours, and} D_{ij} = 0 \text{ otherwise} \tag{15}$$

The unnormalized edge magnitude combines absolute training correlation with geographic support. All non-selected off-diagonal pairs are set to zero. The mixing coefficient is fixed as $\alpha = 0.75$, and the diagonal self-reference term is set to $S_{ii} = 1$.

$$S_{ij} = \begin{cases} 1, & i = j, \\ \alpha\left|\rho_{ij}\right| + (1-\alpha)D_{ij}, & i, j \in N_i, \\ 0, & otherwise. \end{cases} \tag{16}$$

$$\widetilde{S} = \max\left(S, S^T\right), A_{ij} = \frac{\widetilde{S}_{ij}}{\sum_k \widetilde{S}_{ik} + \varepsilon} \tag{17}$$

The element-wise maximum symmetrizes the selected support, after which row normalization produces a single non-negative adjacency matrix **A**. Absolute training correlation determines neighbour relevance; the sign is not transmitted as a separate channel in the final adopted model, keeping the spatial pathway sparse, stable, and auditable.

2.3.5. Neighbour-Minus-Own Residual Contrast

Given station states and the frozen adjacency matrix, Equation (18) computes the graph-supported neighbourhood state. MS-GRAN then subtracts the target station state, so the correction head receives only the information that neighbouring stations add relative to the station's own residual state. This operation suppresses the station-common component and retains only neighbour-relative residual information.

$$v_i = \sum_j A_{ij}u_j, \quad \Delta u_i = v_i - u_i \tag{18}$$

The neighbour-minus-own contrast in Equation (19) is an identifiability safeguard: an identity graph gives identical neighbour and own states, and therefore a zero contrast. The graph branch cannot become an unconstrained duplicate local predictor, and any

non-zero correction must originate from differences between the station and its permitted training-derived neighbours.

$$\Delta\hat{y}_i^{graph} = W_2 GELU(W_1 \Delta u_i), W_1 \in R^{\mathrm{d} \times \mathrm{d}}, W_2 \in R^{\mathrm{HC} \times \mathrm{d}} \tag{19}$$

The correction head is initialized exactly to zero and has no bias. Stage II therefore begins with a complete forecast identical to the frozen NLinear backbone. A non-zero spatial correction is learned only when it reduces validation loss, making the graph contribution auditable and conservative.

2.3.6. Zero-Initialized Gated Residual Correction

The graph residual correction is not added freely. Equation (20) defines the zero-initialized residual correction head, and Equation (21) applies the scalar gate that controls its global amplitude. Together, these equations limit the spatial branch to an incremental correction rather than an unrestricted graph forecast.

$$g = \sigma(\gamma), \gamma_0 = -2, g_0 \approx 0.119 \tag{20}$$

$$\hat{Y}_t = \hat{Y}_t^{local} + g\Delta\hat{Y}_t^{graph} \tag{21}$$

For H > 1, both the NLinear projection and graph correction head emit all H future days directly; recursive feedback is not used. Across the three final one-day MS-GRAN runs, the scalar gate converged to 0.162–0.164, confirming that the spatial pathway remained a small residual adjustment rather than a dominant replacement of the station-local forecast.

2.3.7. Masked Huber Objective and Two-Stage Training

The temporal backbone and graph adapter are trained with the same masked Huber objective. Equation (22) defines the pointwise Huber loss, and Equation (23) averages it over observed targets only using the numerical-stability constant ε defined above. Equation (24) defines RMSE and MAE over observed targets using the same denominator safeguard, Equation (25) defines upper-tail absolute-error percentiles, and Equation (26) defines the station-level relative RMSE reduction.

$$Huber_\delta(e) = \begin{cases} \frac{1}{2}e^2, |e| \le \delta, \\ \delta\left(|e| - \frac{1}{2}\delta\right), |e| > \delta. \end{cases} \tag{22}$$

$$\mathcal{L}(\theta) = \frac{\sum\limits_{q\in\Omega} m_q^+ \,\mathrm{Huber}_\delta\left(\hat{y}_q - y_q\right)}{\sum\limits_{q\in\Omega} m_q^+ + \varepsilon} \tag{23}$$

$$\mathrm{RMSE} = \left(\frac{\sum_{q\in Q} m_q^+ \left(\hat{y}_q - y_q\right)^2}{\sum_{q\in Q} m_q^+ + \epsilon}\right)^{1/2}, \mathrm{MAE} = \frac{\sum_{q\in Q} m_q^+ \left|\hat{y}_q - y_q\right|}{\sum_{q\in Q} m_q^+ + \epsilon} \tag{24}$$

$$\mathrm{P95} = \mathrm{percentile}_{95}\left\{\left|\hat{y}_q - y_q\right| : m_q^+ = 1\right\}, \mathrm{P99} = \mathrm{percentile}_{99}\left\{\left|\hat{y}_q - y_q\right| : m_q^+ = 1\right\} \tag{25}$$

$$G_i = 100 \times \left[1 - \frac{\mathrm{RMSE}_\mathrm{i}^{\mathrm{MS-GRAN}}}{\mathrm{RMSE}_\mathrm{i}^{\mathrm{NLinear}}}\right] \tag{26}$$

Optimization follows a two-stage protocol. Stage I trains the NLinear backbone using the masked Huber objective and freezes the validation-selected checkpoint. Stage II constructs the adjacency matrix A from training data only and optimizes the shared multi-

scale GRU encoder, graph residual head, and scalar gate, while the NLinear parameters remain fixed.

2.3.8. Computational Complexity and Algorithmic Scope

The code-equivalent execution order is summarized below to connect the mathematical formulation to the reproducible training and evaluation workflow.

Step 1. Estimate detrending coefficients and robust-normalization statistics from observed 2010–2019 training samples only and apply the frozen transformations to all partitions.

Step 2. Construct within-partition $L + H$ windows and retain both input masks and future target masks.

Step 3. Train the station-local NLinear backbone, select by 2020–2021 validation loss, and freeze θ_N.

Step 4. Compute causal multi-scale EWMA residuals and build the static training-only correlation–distance adjacency A.

Step 5. Evaluate frozen NLinear, encode masked multi-scale residual histories, aggregate graph neighbours, and form the neighbour-minus-own residual contrast.

Step 6. Optimize only MS-GRAN adapter parameters with the masked Huber objective; retain the minimum-validation-loss adapter checkpoint.

Step 7. Freeze all parameters and evaluate each selected seed and horizon once on 2022–2024; convert predictions to millimetres before metric calculation.

For one-day forecasting, the main computational costs are the shared residual encoder, sparse graph aggregation, and Top-K graph storage. The GRU encoder scales approximately linearly with the number of stations and sequence length; sparse graph aggregation scales with the number of retained edges; and graph storage scales with the Top-K neighbourhood size.

At inference, MS-GRAN uses only the preceding L days, their masks, and the frozen training-derived adjacency. It is a numerical displacement-residual forecaster: it does not update the graph from future observations and does not assign earthquake, precursor, or anomaly labels.

2.4. Experimental Protocol

2.4.1. Comparison Baselines

The baselines provide controlled context across classical GNSS trajectory models, local temporal neural models, and a representative spatiotemporal graph model. Persistence, seasonal naive, trend plus annual/semiannual harmonics, harmonic plus AR residual, Ridge AR, DLinear, NLinear, LSTM, TCN, PatchTST, and STGCN are evaluated under the same chronological split, masking rule, and frozen-test protocol. Neural baselines use masked Huber loss [71], AdamW optimization [72], validation-selected checkpoints, and at most 20 training epochs. The main controlled comparison remains the matched frozen NLinear ablation, because it isolates the marginal contribution of the spatial residual pathway after the station-local forecast has already been learned.

To address baseline fairness without overloading Section 2, Table 2 summarizes the common experimental protocol and the core MS-GRAN settings, while Supplementary Table S1 reports the complete baseline-specific architectures, fixed values, and validation-selection status. All trainable neural candidates used the same processed inputs and masks, masked Huber objective, AdamW learning rate of 10^{-3}, weight decay of 10^{-4}, gradient clipping at 1.0, 20-epoch maximum, and minimum-validation-loss checkpoint rule. Persistence, seasonal naive, trend-harmonic, Harmonic + AR, and Ridge AR were fitted deterministically or in closed form; DLinear, NLinear, TCN, and PatchTST used

one prespecified architecture each. For LSTM and STGCN, five explicit candidates were evaluated using 2020–2021 validation loss only; the selected settings were then fixed and rerun for three seeds. The complete candidate sets and selected configurations are listed in Supplementary Table S1. No baseline architecture, hyperparameter, or checkpoint was selected using the 2022–2024 frozen-test period, and below-persistence comparator results are reported only as outcomes under the documented finite tuning budget.

Table 2. Core dataset, experimental protocol, and MS-GRAN configuration.

Item	Adopted Value	Selection Basis/Role
Network	90 Cascadia GNSS stations	Fixed benchmark
Observations	Daily E/N/U residuals, 2010–2024	Identical input for all methods
Chronological split	2010–2019/2020–2021/2022–2024	Train–validation–frozen test
Context	60 days	Prespecified compact local history
Forecast horizon	1 day; robustness at 3 and 7 days	Direct multi-horizon audit
Graph	Training-only Top-K; main K = 8	Fixed before frozen-test evaluation; K = 4–32 is a robustness analysis only
Graph mixture	alpha = 0.75	Fixed before frozen-test evaluation; 75% correlation and 25% distance support
Residual scales	EWMA 7, 15, 31, 61 days	Weekly-to-bimonthly causal summaries
Residual adapter	Shared GRU, d = 48	Compact station-shared encoder
Initialization	Residual head = 0; gate logit = -2	Exactly zero initial correction; g $\approx$ 0.119
Optimizer	AdamW; lr = 1×10^{-3}; weight decay = 1×10^{-4}	Common neural-training protocol
Loss/clipping	Masked Huber, delta = 1.5; norm = 1.0	Robust observed-target objective
Training budget	Batch 32; 20 epochs	Same maximum budget
Uncertainty	3 seeds; 10,000 station bootstraps	Seed variability; descriptive station robustness
Checkpoint rule	Minimum validation loss only	No frozen-test model selection
Hardware	NVIDIA RTX 4090-class GPU	Neural training and timing

2.4.2. Main Comparison and Ablation Design

The primary experiment compares all methods on the identical one-day, 90-station frozen-test set. The primary matched ablation isolates the full residual adapter from the frozen NLinear backbone and evaluates whether the adapter gain can be explained by added target-station capacity, placebo graph connectivity, simple regional common-mode processing, or dependence on the NLinear backbone. The controls include a self-only residual adapter with no neighbour aggregation, a rewired-placebo graph preserving sparse degree and weight distributions, two causal regional common-mode controls, and a DLinear frozen-backbone residual-adaptation test.

The first common-mode control adds a training-fitted regional-median correction after the frozen NLinear forecast and is retained as a simple forecast-time stack. To follow the conventional preprocessing order explicitly, the stricter pre-local control constructs a daily regional-median E/N/U stack, estimates station-component loadings from 2010 to 2019 only, removes the loaded common mode before robust normalization and NLinear training, and reconstructs the next-day raw residual by adding a causal common-mode forecast. Persistence and a train-only AR(1) with intercept were compared on 2020–2021 validation RMSE; AR(1) was selected before frozen-test evaluation. No validation or test target enters the loading estimation, local-model training, or forecast-time common-mode estimate.

The reporting boundary is fixed by experimental role rather than outcome. Every confirmatory control requested to distinguish target-only capacity, arbitrary connectivity, regional common mode, and the adopted graph is included in the corresponding experimental analysis. The complete prespecified Top-K grid K = {4, 8, 12, 16, 24, 32} includes settings that outperform K = 8 on the frozen test; K = 8 was retained because it was the

pre-specified main configuration. Earlier architecture screens, signed-channel diagnostics, and spatial hold-out prototypes were exploratory diagnostics, were not used to select the reported primary model, and are not presented as confirmatory evidence. All model, hyperparameter, and checkpoint selection used the 2020–2021 validation period only; the 2022–2024 frozen-test period was not used to revise the adopted main configuration. The additional controls and robustness grids are reported as fixed diagnostic evaluations rather than as test-driven selection procedures, and the corresponding validation and frozen-test results are presented side by side.

2.4.3. Forecast-Horizon and Network-Density Tests

Operating-boundary experiments cross direct 1-, 3-, and 7-day forecast horizons with fixed nested networks of 90, 70, 50, and 30 stations. Every cell is trained with three random seeds under the same train–validation–test chronology. These tests ask whether the spatial residual pathway remains useful when prediction becomes harder and station density is reduced. The 70-, 50-, and 30-station subsets were selected deterministically before model evaluation by farthest-point geographic sampling from station coordinates to preserve broad spatial coverage; no station was selected or removed based on validation or test forecast performance.

2.4.4. Metrics and Statistical Analysis

RMSE and MAE are evaluated over observed station-component targets in the frozen-test period. P95 and P99 are defined in Equation (25), and station-wise graph gain is defined in Equation (26). Ten thousand paired station-wise bootstrap resamples are used to quantify uncertainty in station-level gain estimates [73]. Because station residuals are spatially coherent, the bootstrap intervals are interpreted as descriptive station-level robustness evidence rather than as an independence-based proof of predictive ability.

Forecast-comparison inference first follows the autocorrelation-robust framework of Newey and West [74] and the paired predictive-accuracy logic of Diebold and Mariano [75]. Because MS-GRAN nests the frozen NLinear forecast when the residual correction is zero, the formal test uses the Clark–West adjusted loss differential [76]. For each frozen-test day, squared errors are averaged over observed station-component targets and then across the three independently trained seeds. The adjusted differential is $d_{CW,t} = e_{N,t}^2 - [e_{M,t}^2 - (f_{N,t} - f_{M,t})^2]$, where positive values favour MS-GRAN. Its mean is tested against zero with a one-sided normal approximation and a Newey–West HAC variance; seven days is the primary lag and 30 days is a sensitivity check. The unadjusted paired loss difference is retained only descriptively, while the station bootstrap assesses how gains are distributed rather than providing an independence-based test.

2.4.5. Implementation and Efficiency Evaluation

All trainable baselines use the same processed inputs, chronological partitions, observed-target masks, masked Huber objective, optimizer settings, 20-epoch maximum, and validation-only checkpoint rule. All experiments were implemented using Python 3.12 under Ubuntu 22.04, with PyTorch 2.5.1 and CUDA 12.4. The models were trained on a single NVIDIA RTX 4090D GPU with 24 GB memory, equipped with an Intel Xeon Platinum 8352V CPU (16 vCPUs) and 62 GB system memory. Table 2 summarizes the common protocol and core MS-GRAN settings; complete comparator architectures, fixed values, and the five-candidate LSTM/STGCN validation audits are provided in Supplementary Table S1. No model architecture, hyperparameter, or checkpoint was selected using the 2022–2024 frozen-test period. Parameter counts and latency use the same evaluation script and hardware, and efficiency is not interpreted as a hardware-independent ranking.

For reproducibility, the adopted MS-GRAN setting uses a 60-day context, direct one-day primary horizon, EWMA spans of 7, 15, 31, and 61 days, shared-GRU hidden width d = 48, pre-specified Top-K = 8 graph, and correlation–distance coefficient alpha = 0.75. Training uses batch size 32, AdamW learning rate 1×10^{-3}, weight decay 1×10^{-4}, Huber delta 1.5, gradient-norm clipping at 1.0, and a 20-epoch budget. The residual head's final layer is initialized to zero, while the scalar-gate logit is initialized at -2 ($g \approx 0.119$), so the spatial output begins at exactly zero without incorrectly describing the sigmoid gate itself as zero. Checkpoints are selected only by 2020–2021 validation loss on an NVIDIA RTX 4090-class GPU; the 2022–2024 test is frozen until final reporting. These values were treated as prespecified engineering settings rather than the result of a frozen-test hyperparameter search. The 7/15/31/61-day EWMA spans provide multiple causal residual time scales, while the remaining compact architecture and training settings were fixed before frozen-test evaluation.

3. Results

3.1. Overall Forecasting Accuracy and Computational Efficiency

The primary benchmark asks whether bounded spatial residual adaptation improves a strong station-local predictor under the same data, split, masks, and evaluation metric. MS-GRAN achieved the lowest one-day frozen-test RMSE among the evaluated methods, but the main result is not the absolute ranking alone. Table 3 reports the full numerical comparison, and Figure 3 visualizes the same result as an accuracy ranking, multi-metric heatmap, and accuracy-efficiency frontier. The critical comparison is against the matched frozen NLinear backbone: MS-GRAN achieved 2.9158 ± 0.0033 mm, compared with 2.9579 ± 0.0006 mm for NLinear, corresponding to a modest but reproducible 1.422% RMSE reduction.

The validation-selected tuning audit did not alter the benchmark ordering. With the selected settings fixed and rerun across three seeds, LSTM achieved 3.8914 ± 0.0729 mm frozen-test RMSE and STGCN achieved 4.5744 ± 0.1804 mm. Both remained below persistence (3.4640 mm); LSTM's RMSE was about 12.3% higher, and the gap for STGCN was larger. The one-day forecasting task contains a strong station-local temporal structure, so a higher-capacity temporal model such as LSTM does not necessarily translate its greater flexibility into lower test error under the same finite training and validation-search budget. For STGCN, this challenge is compounded by the relatively sparse GNSS network and the modest additional exploitable cross-station signal, because the model must jointly estimate temporal representations and graph-based spatial propagation. We therefore report these values as outcomes under the declared data conditions, training budget, and validation-only search range rather than as a general ranking of the model classes.

Table 3. Three-seed frozen one-day test performance.

Method	RMSE (mm)	MAE (mm)	Skill vs. Persistence (%)	Parameters	ms/Window
Persistence	3.4640	2.0209	0.00	0	-
Seasonal naive	5.4722	3.4375	−57.97	0	-
Trend + annual/ semiannual harmonics	8.5440	4.8580	−146.65	6 per component	-
Harmonic + AR residual	3.0419	1.8094	12.19	Station-wise	-
DLinear	2.9662 ± 0.0015	1.7499 ± 0.0009	14.37	122	0.0056
LSTM	3.8914 ± 0.0729	2.1192 ± 0.0182	−12.34	51,907	0.0700 ± 0.0010

Table 3. *Cont.*

Method	RMSE (mm)	MAE (mm)	Skill vs. Persistence (%)	Parameters	ms/Window
TCN	3.2840 ± 0.0577	1.9225 ± 0.0234	5.20	56,931	0.1367
PatchTST	3.3292 ± 0.0742	1.9140 ± 0.0282	3.89	102,849	0.1919
STGCN	4.5744 ± 0.1804	2.6343 ± 0.0630	−32.06	10,563	0.0455
NLinear	2.9579 ± 0.0006	1.7455 ± 0.0007	14.61	61	0.0050
Ridge AR	2.9968	1.7740	13.49	16,470	0.0034
MS-GRAN	2.9158 ± 0.0033	1.7110 ± 0.0022	15.83 ± 0.10	11,870	0.101 ± 0.047

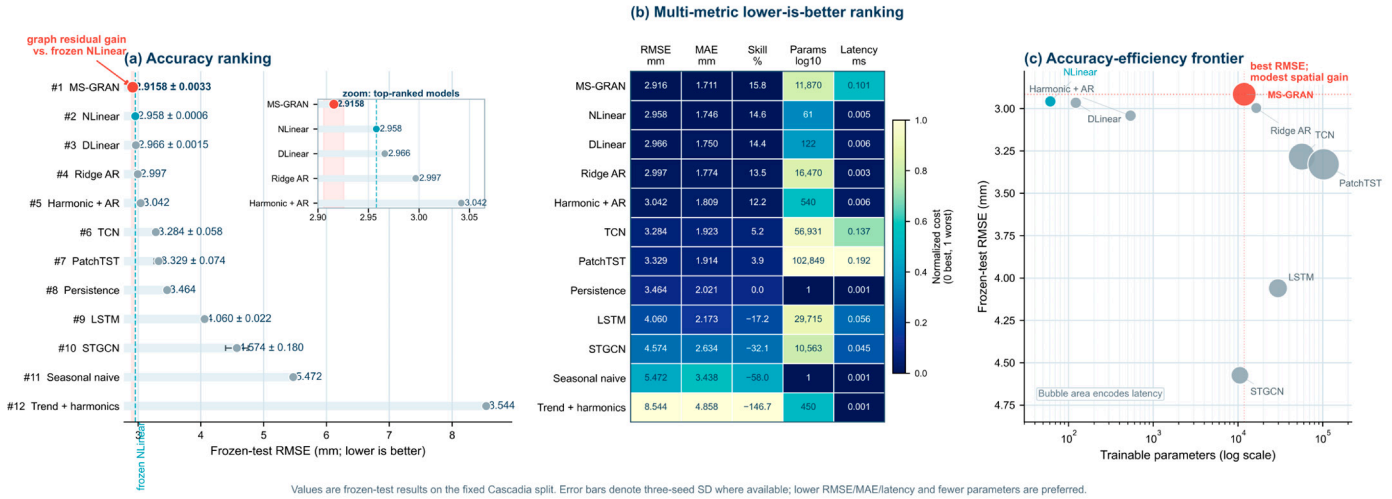


Figure 3. Three-seed benchmark accuracy, multi-metric ranking, and efficiency frontier. (**a**) Frozen-test RMSE ranking with one-standard-deviation intervals. (**b**) Normalized accuracy and efficiency metrics. (**c**) Accuracy–parameter–latency frontier. MS-GRAN uses 11,870 parameters and retains a compact operating point while achieving the lowest RMSE.

3.2. Independent Contributions of the MS-GRAN Components

Table 4 separates the one-day residual-adaptation gain into confirmatory controls, includes the alternative DLinear-backbone check, and places validation RMSE beside frozen-test metrics for all listed controls. The adopted real graph reduces the test RMSE from 2.9579 ± 0.0006 to 2.9158 ± 0.0033 mm. Self-only adaptation improves it by 0.635%, showing that trainable target-station residual capacity explains part of the gain. Rewired graphs retain a 1.265% gain, so topology alone cannot explain the full result; nevertheless, the real training-only graph remains better than both controls. The post-forecast regional-median correction yields only a 0.082% gain. In the stricter pre-local experiment, validation selected the train-only AR(1) common-mode reconstruction over persistence (2.5839 ± 0.0008 versus 2.6006 ± 0.0008 mm validation RMSE), but its frozen-test RMSE was 2.9680 ± 0.0025 mm, corresponding to a −0.343 ± 0.084% gain relative to frozen NLinear. Accordingly, the MS-GRAN improvement exceeds the pre-local CME baseline by 1.765 percentage points and the post-forecast regional-median correction by 1.340 percentage points. In a separate backbone check, the frozen DLinear backbone has a validation RMSE of 2.5875 ± 0.0017 mm and a frozen-test RMSE of 2.9662 ± 0.0015 mm, while the DLinear adapter reaches 2.5446 ± 0.0014 mm on validation and 2.9266 ± 0.0007 mm on the frozen test, corresponding to a 1.335 ± 0.040% backbone-matched test gain, supporting a positive residual-adaptation effect beyond NLinear without claiming universal backbone invariance.

Table 4. Confirmatory residual-adaptation controls, including the alternative DLinear-backbone check, for one-day 90-station forecasting.

Variant	Gate/Control	Validation RMSE	Test RMSE	Test MAE	Backbone-Matched Gain (%)	Improved
Frozen NLinear	None	2.5878 ± 0.0012	2.9579 ± 0.0006	1.7455 ± 0.0007	0.000%	N/A
MS-GRAN	Real train-only graph	2.5438 ± 0.0013	2.9158 ± 0.0033	1.7110 ± 0.0022	1.422 ± 0.122%	82–84/90
Self-only adapter	No neighbours	2.5733 ± 0.0013	2.9391 ± 0.0014	1.7301 ± 0.0007	0.635 ± 0.031%	Control
Rewired graph	3 rewires × 3 seeds	2.5535 ± 0.0012	2.9205 ± 0.0024	1.7146	1.265 ± 0.092%	Control
Regional-stack correction	Post-forecast median	2.5874 ± 0.0013	2.9555 ± 0.0013	1.7443 ± 0.0003	0.082 ± 0.022%	Control
Pre-local CME + NLinear	Train-only median loadings; AR(1)	2.5839 ± 0.0008	2.9680 ± 0.0025	1.7664 ± 0.0010	−0.343 ± 0.084%	Control
Frozen DLinear	None	2.5875 ± 0.0017	2.9662 ± 0.0015	1.7499 ± 0.0009	0.000%	N/A
DLinear + adapter	Real train-only graph	2.5446 ± 0.0014	2.9266 ± 0.0007	1.7173 ± 0.0010	1.335 ± 0.040%	Backbone control

Note: Gain is relative to the matched frozen local backbone unless otherwise stated. Rewired placebo values summarize nine runs (three random initializations times three rewiring draws). The DLinear control reports gain relative to the matched frozen DLinear backbone. Controls are used for mechanism interpretation and are not retrospectively selected as the main model. N/A indicates not applicable because no corresponding gain is calculated for the reference baseline.

Table 5 reports the Top-K sensitivity analysis, now extended from K = 4–16 to K = 24 and K = 32 and paired with validation metrics. Gains remain positive across all tested sparse graph densities. Larger K values slightly improve the frozen-test RMSE, but the gain largely saturates between K = 24 and K = 32. Relative to K = 8, increasing K to 24 or 32 enlarges the retained neighbourhood by 3–4 times, increases graph density, and correspondingly raises sparse aggregation and graph-storage cost for only a modest additional predictive improvement. K = 8 is therefore retained as the pre-specified compact operating point that balances predictive performance, graph sparsity, and computational cost, rather than being retrospectively selected from frozen-test performance; the slightly better K = 24/32 test values are reported as sensitivity evidence, not as a basis for changing the main configuration.

Table 5. Sensitivity of MS-GRAN to the number of graph neighbours on the one-day 90-station frozen test.

K	Validation RMSE (mm)	Validation Gain (%)	Test RMSE (mm)	Test Gain (%)	Improved Stations
4	2.5485 ± 0.0016	1.519 ± 0.096	2.9206 ± 0.0021	1.262 ± 0.082	79–83/90
8	2.5438 ± 0.0013	1.699 ± 0.093	2.9158 ± 0.0033	1.422 ± 0.122	82–84/90
12	2.5416 ± 0.0014	1.782 ± 0.092	2.9151 ± 0.0024	1.446 ± 0.096	83–84/90
16	2.5410 ± 0.0023	1.809 ± 0.130	2.9141 ± 0.0023	1.482 ± 0.094	83–84/90
24	2.5397 ± 0.0021	1.857 ± 0.109	2.9124 ± 0.0019	1.539 ± 0.079	81–84/90
32	2.5403 ± 0.0014	1.834 ± 0.092	2.9122 ± 0.0021	1.545 ± 0.088	82–84/90

3.3. Component- and Station-Level Robustness Analysis

Because the aggregate one-day RMSE combines east, north, and up displacement residuals, we additionally evaluated the frozen-test performance by coordinate component. This analysis tests whether the marginal graph benefit is broadly distributed across E/N/U or is dominated by a single component. Table 6 reports the component-wise metrics. Component-wise evaluation showed positive RMSE reductions across E, N, and U, indicating that the aggregate improvement was not produced by a single coordinate

component. The larger improvement in the up component suggests that the marginal contribution of spatial residual information may vary across displacement components.

Table 6. Component-wise one-day frozen-test performance of NLinear and MS-GRAN on the 90-station Cascadia benchmark.

Component	Valid Targets	NLinear RMSE (mm)	MS-GRAN RMSE (mm)	RMSE Gain (%)	NLinear MAE (mm)	MS-GRAN MAE (mm)	MAE Gain (%)
East (E)	90,427	1.3798 ± 0.0024	1.3769 ± 0.0025	0.2123 ± 0.0426	0.9762 ± 0.0018	0.9667 ± 0.0021	0.973
North (N)	90,427	1.4043 ± 0.0007	1.3940 ± 0.0010	0.7352 ± 0.0990	1.0130 ± 0.0009	1.0014 ± 0.0011	1.145
Up (U)	90,427	4.7298 ± 0.0021	4.6548 ± 0.0057	1.5868 ± 0.1335	3.2473 ± 0.0007	3.1649 ± 0.0033	2.537
All E/N/U	271,281	2.9579 ± 0.0006	2.9158 ± 0.0033	1.4223 ± 0.1220	1.7455 ± 0.0007	1.7110 ± 0.0022	1.977

The station-level analysis further shows that the spatial residual contribution is broadly distributed across the network while retaining visible station-to-station heterogeneity. Table 7 summarizes the station-level uncertainty and tail-error evidence: across the three one-day seeds, MS-GRAN improves 83/90, 82/90, and 84/90 stations, with mean station-wise RMSE reductions of 1.875%, 1.825%, and 1.996%. Every paired station-bootstrap confidence interval remains positive. Tail metrics also improve: P95 and P99 errors decrease more than the mean RMSE, indicating that the residual graph pathway mainly reduces moderate-to-large forecast deviations rather than uniformly shifting every prediction.

Table 7. Station-level uncertainty and frozen-test tail-error reductions across three one-day seeds.

Seed	Improved	Mean Gain (%)	95% CI (%)	MAE Red. (%)	P95 Red. (%)	P99 Red. (%)
1	83/90	1.875	1.403–2.064	1.961	2.323	1.650
2	82/90	1.825	1.259–1.988	1.893	2.350	1.662
3	84/90	1.996	1.460–2.198	2.081	2.543	1.764

The station-wise gain and pooled RMSE gain are intentionally reported as complementary summaries because they answer different questions. The pooled RMSE reduction is the primary overall benchmark metric: it aggregates all observed station-component targets and is therefore more strongly influenced by components with larger absolute errors, particularly the vertical component. In contrast, the mean station-wise gain first computes a relative RMSE reduction within each station and then assigns equal weight to every station, making it useful for assessing how broadly the improvement is distributed across the network. This difference in weighting explains why the mean station-wise gains in Table 7 are larger than the pooled 1.422% RMSE reduction reported in Tables 3, 4 and 6. Throughout the manuscript, 1.422% is retained as the headline overall performance effect, whereas the station-wise values are used only to characterize station-level distribution and consistency.

3.4. Spatiotemporal Mechanism and Representative Forecasts

3.4.1. Spatiotemporal Diagnostics and Representative Forecasts

Figure 4 examines whether the aggregate accuracy improvement of MS-GRAN is accompanied by spatially coherent reductions in forecast error. For four representative frozen-test dates, the diagnostic maps compare observed residual amplitude, NLinear absolute error, MS-GRAN absolute error, and the NLinear-minus-MS-GRAN error reduction. Positive error reductions are observed at multiple station locations, while the spatial patterns indicate that the learned residual pathway contributes beyond a uniform averaging effect. These patterns are consistent with the intended role of MS-GRAN as a selective and controlled spatial residual adaptation module, where neighbouring stations provide additional residual information rather than replacing the local temporal forecast. Together with

the component-wise and station-level analyses, Figure 4 provides spatially interpretable evidence that the model improves localized forecast accuracy without implying continuous deformation-field reconstruction.

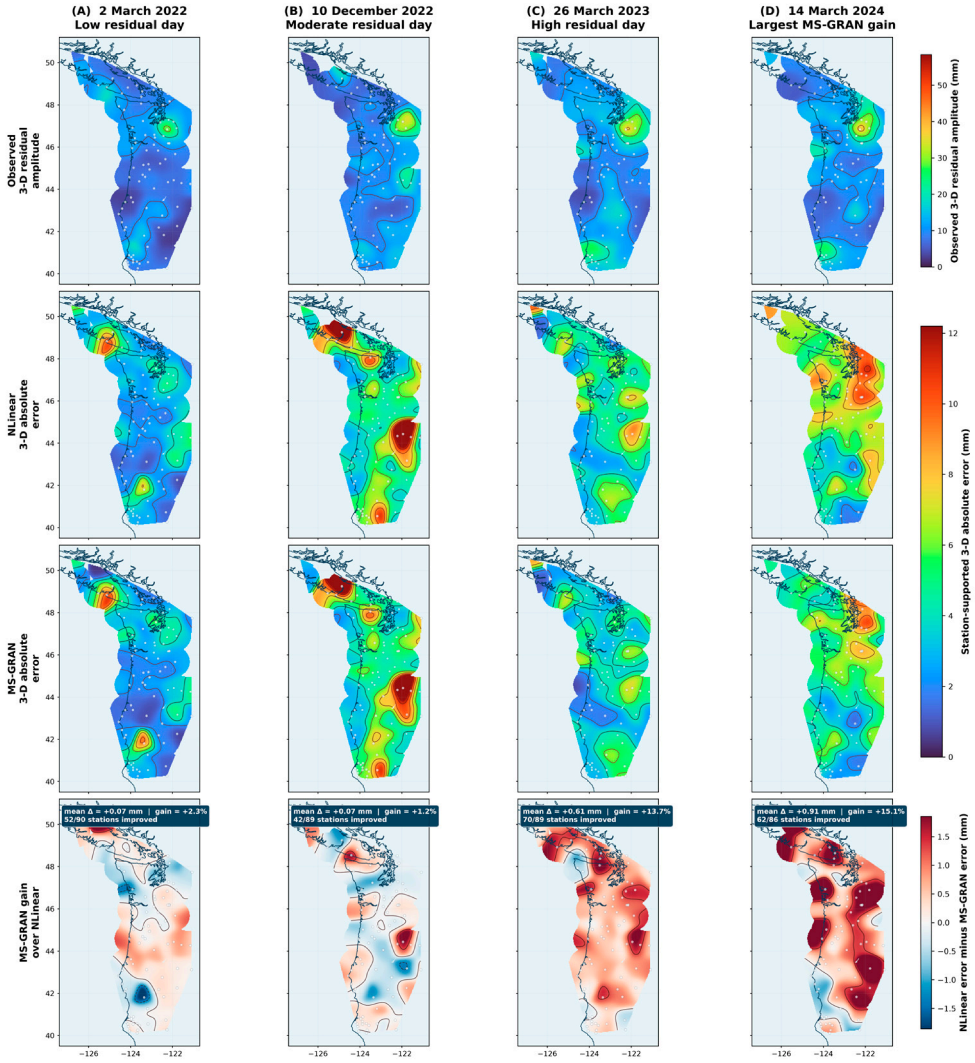


Figure 4. Station-supported diagnostic atlas of GNSS residual amplitude and one-day forecast error reduction. Rows show observed 3D residual amplitude, NLinear absolute error, MS-GRAN absolute error, and NLinear-minus-MS-GRAN absolute error reduction for representative frozen-test dates. Positive values indicate localized forecast-error reduction. Interpolated backgrounds are used only as visual summaries over the observed-station footprint and should not be interpreted as continuous deformation fields. White dots indicate GNSS stations with valid observations on the corresponding date, and the overlaid lines are contour lines of the station-supported Gaussian-kernel summaries.

3.4.2. Representative Station-Time Forecasts

Figure 5 further examines whether the spatial residual correction produces consistent station-time error reductions rather than isolated spatial snapshots. Using a common 90-day frozen-test window at eight north-to-south stations, the upper arrays compare matched NLinear and MS-GRAN absolute vertical errors over the same station-time grid, while the lower panels show daily vertical-error reduction at representative stations. Positive reduction episodes indicate station-days on which MS-GRAN lowers the absolute error relative to the frozen NLinear backbone. The figure shows that the graph residual adapter contributes through repeated small error reductions across selected station-time regions, rather than through a single visually dominant event or subset-specific retraining.

This supports the interpretation that the MS-GRAN gain comes from persistent residual correction in the held-out period. The vertical component is shown for compact visualization because it has the largest absolute residual scale; Table 6 confirms that positive RMSE reductions are also present across all E/N/U components.

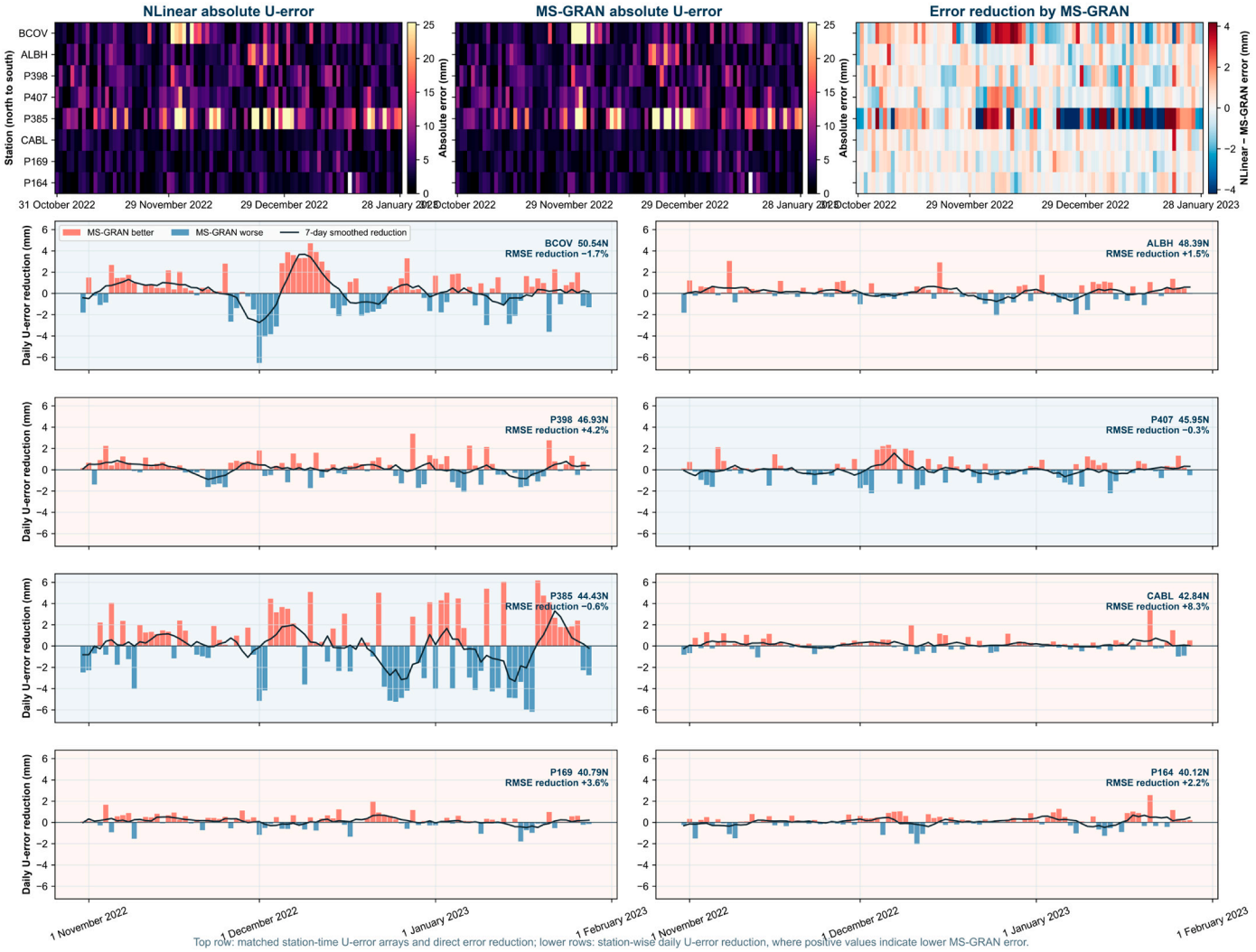


Figure 5. Common-window station-time forecast-error diagnostic using the final MS-GRAN three-seed ensemble. The upper arrays compare NLinear and MS-GRAN absolute vertical errors and their direct difference across the same eight stations and 90 held-out days. Lower panels show station-wise daily vertical-error reduction, defined as NLinear absolute error minus MS-GRAN absolute error; positive values indicate lower MS-GRAN error. No subset-specific retraining or interpolation is used.

3.5. Operating Boundaries Across Forecast Horizons and Network Densities

All 12 horizon-density cells produced positive aggregate RMSE reductions with the adopted MS-GRAN configuration. Table 8 gives the numerical 1-/3-/7-day by 90-/70-/50-/30-station robustness matrix, while Figure 6 maps the station-level evidence for the same design on a fixed Cascadia basemap. Mean gains range from 0.525% to 1.422%. The largest gain occurs for one-day forecasting at 90 stations, while positive gains persist at 30 stations and at seven-day horizons. This pattern supports a bounded interpretation: graph residual correction is most effective when the network is denser and the forecast horizon is shorter, while retaining positive aggregate gains under moderate sparsification and longer direct horizons. These 12 horizon-density cells were evaluated as a fixed operating-boundary matrix and were not used to select the one-day 90-station main model; their paired validation

and frozen-test metrics are reported together in Table 8 for transparency about the full set of evaluations.

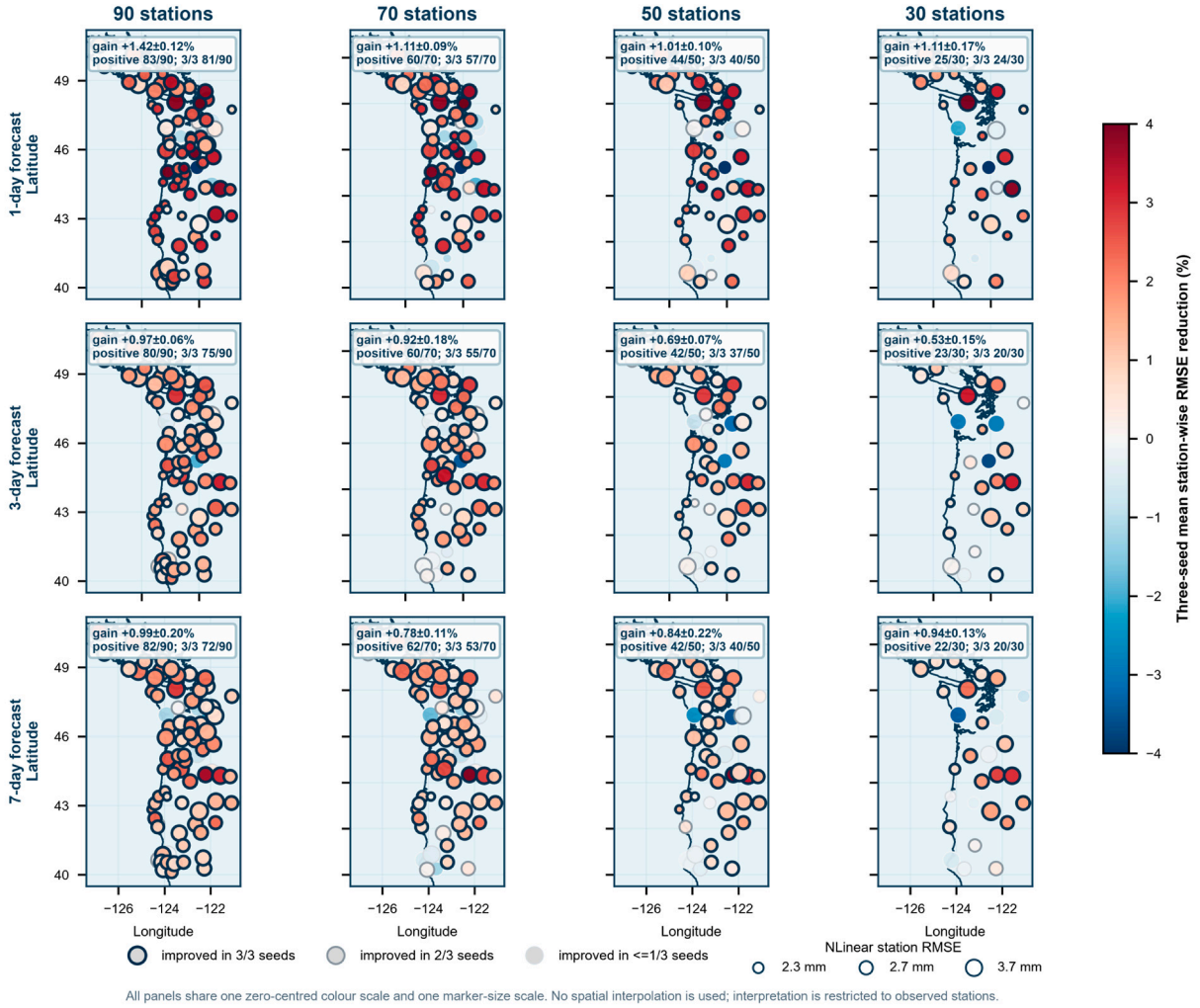


Figure 6. Operating boundaries of graph residual correction across forecast horizons and GNSS network densities. Each panel uses the same Cascadia station map; point colour gives three-seed mean station-wise RMSE reduction, point size gives local NLinear difficulty, and point outline indicates seed consistency. The figure tests whether the marginal spatial residual contribution remains positive across 1-, 3-, and 7-day horizons and nested 90-, 70-, 50-, and 30-station networks.

Table 8. Three-seed robustness across forecast horizons and nested network densities. Validation and frozen-test metrics are shown for every cell; the matrix is a robustness analysis and is not used for model selection.

H	N	Val. NLinear	Val. MS-GRAN	Val. Gain (%)	Test NLinear	Test MS-GRAN	Test Gain (%)
1	90	2.5878 ± 0.0012	2.5438 ± 0.0013	1.699 ± 0.093	2.9579 ± 0.0006	2.9158 ± 0.0033	1.422 ± 0.122
1	70	2.6364 ± 0.0012	2.5903 ± 0.0018	1.748 ± 0.091	3.0712 ± 0.0010	3.0371 ± 0.0025	1.109 ± 0.089
1	50	2.6929 ± 0.0015	2.6448 ± 0.0011	1.786 ± 0.092	3.1786 ± 0.0005	3.1464 ± 0.0031	1.011 ± 0.102
1	30	2.5995 ± 0.0010	2.5503 ± 0.0046	1.893 ± 0.208	3.1572 ± 0.0014	3.1222 ± 0.0060	1.109 ± 0.170
3	90	2.7363 ± 0.0004	2.7068 ± 0.0021	1.078 ± 0.088	3.1724 ± 0.0020	3.1415 ± 0.0013	0.971 ± 0.057
3	70	2.7827 ± 0.0008	2.7512 ± 0.0035	1.133 ± 0.124	3.2938 ± 0.0032	3.2634 ± 0.0049	0.923 ± 0.177
3	50	2.8370 ± 0.0004	2.8072 ± 0.0056	1.049 ± 0.207	3.4128 ± 0.0022	3.3895 ± 0.0025	0.685 ± 0.075
3	30	2.7554 ± 0.0010	2.7333 ± 0.0047	0.800 ± 0.205	3.3872 ± 0.0019	3.3695 ± 0.0057	0.525 ± 0.148
7	90	2.8637 ± 0.0007	2.8336 ± 0.0053	1.051 ± 0.209	3.3703 ± 0.0007	3.3368 ± 0.0064	0.995 ± 0.204

Table 8. *Cont.*

H	N	Val. NLinear	Val. MS-GRAN	Val. Gain (%)	Test NLinear	Test MS-GRAN	Test Gain (%)
7	70	2.9034 ± 0.0023	2.8757 ± 0.0009	0.955 ± 0.048	3.4948 ± 0.0046	3.4674 ± 0.0027	0.782 ± 0.114
7	50	2.9624 ± 0.0022	2.9289 ± 0.0034	1.128 ± 0.152	3.6299 ± 0.0039	3.5993 ± 0.0068	0.843 ± 0.217
7	30	2.8901 ± 0.0005	2.8653 ± 0.0008	0.860 ± 0.027	3.6007 ± 0.0023	3.5669 ± 0.0064	0.939 ± 0.134

3.6. Graph Structure Analysis and Statistical Reliability

3.6.1. Training-Only Graph Structure and Station-Level Mechanism

Training-only grouping indicates that larger station-level gains are more clearly associated with stronger correlation support. Figure 7a compares station-wise graph gain with three station graph descriptors: mean absolute top-8 training correlation, mean correlated-neighbour distance, and formal coordinate noise. Across stations, gain shows a moderate positive association with the mean absolute correlation of the top-eight training neighbours, while associations with neighbour distance and formal coordinate noise are weaker. This pattern is consistent with the model design: distance constrains plausible support, but empirical residual coherence provides stronger evidence of neighbour usefulness.

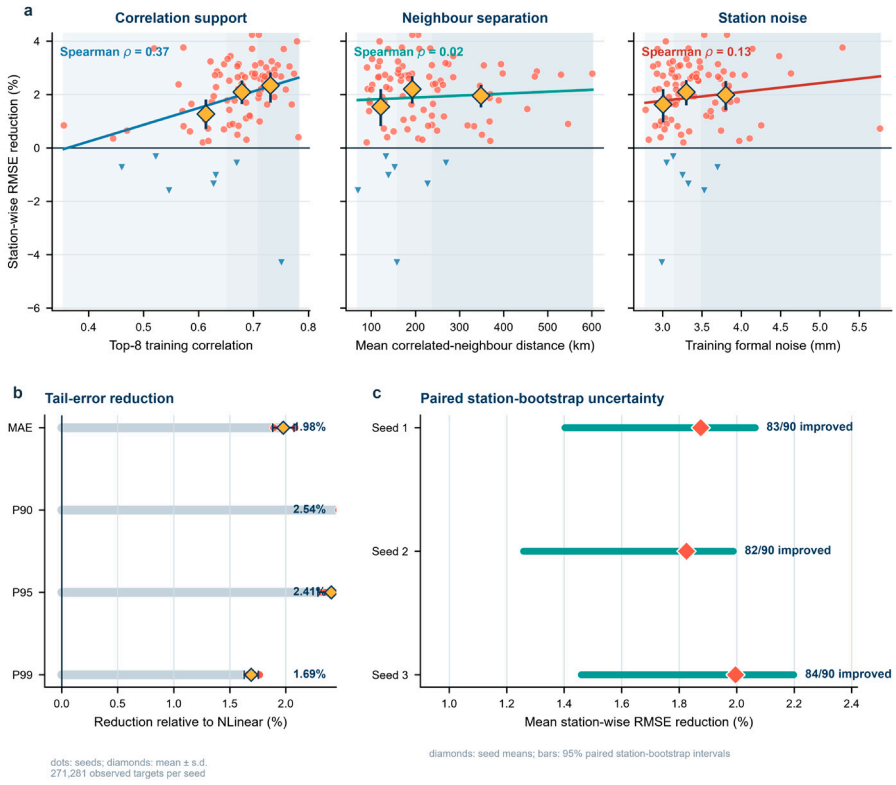


Figure 7. Graph-structure and statistical evidence for MS-GRAN. (**a**) Station-wise RMSE reduction versus training-only top-8 correlation strength, mean correlated-neighbour distance, and formal coordinate noise; diamonds show tertile means with 95% bootstrap intervals. (**b**) Three-seed reductions in MAE and P90/P95/P99 absolute errors. (**c**) Seed-specific mean station-wise gains with paired station-bootstrap 95% intervals and counts of improved stations.

A graph audit found that all 720 directed candidate edges selected in the final 90-station Top-K graph had positive training correlations; symmetrization produced 623 positive station pairs and no selected negative edge. Accordingly, the final manuscript reports the non-negative MS-GRAN graph as the adopted model and treats the signed-channel

implementation only as a diagnostic variant, not as evidence for a signed-neighbour mechanism in this benchmark.

3.6.2. Tail-Error and Bootstrap Reliability

The graph residual pathway improves the error distribution beyond mean RMSE. Figure 7b,c show that, across the three one-day seeds, MAE decreases by 1.89–2.08%, P95 by 2.32–2.54%, and P99 by 1.65–1.76%, with positive paired station-bootstrap intervals for every seed. The larger reductions at P95 than at the mean-error level indicate that MS-GRAN is particularly useful for reducing moderate-to-large forecast deviations, which are important for residual-baseline monitoring. The seed-specific bootstrap results descriptively show that station-wise gains remain positive under resampling and are not concentrated in one favourable seed, one dominant station, or an isolated subset of easy predictions. Because stations are spatially coherent, these intervals are not treated as independence-based inferential proof. Formal evidence for the nested forecast comparison is provided by the Clark–West test with HAC variance in the following paragraph.

The nested-forecast test led to the same conclusion. The three-seed mean Clark–West adjusted daily loss differential was 0.7114 mm^2. With Newey–West variance, the 7-day specification gave z = 17.03 and one-sided $p = 2.62 \times 10^{-65}$; the 30-day sensitivity specification gave z = 18.35 and $p = 1.72 \times 10^{-75}$. The corresponding unadjusted mean loss difference was 0.2487 mm^2. Seasonal gains remained positive at 0.794%, 1.893%, 1.569%, and 1.592% for DJF, MAM, JJA, and SON. To test whether stations with less complete local histories obtain greater predictive benefit, we computed each station's frozen-test gap fraction and stratified the 90 stations using tertile cut points of 0.00965 and 0.01737; ties produced 39 low-, 24 medium-, and 27 high-gap stations. Their mean station-level RMSE gains were 1.183%, 1.230%, and 0.996%, respectively. All three groups therefore retained positive gains, but the high-gap group did not show the largest improvement, providing no evidence that station-level gain systematically increases as local-history completeness decreases. An internal year-wise diagnostic was also positive in 2022, 2023, and 2024 (1.649%, 1.268%, and 1.398%), so no single calendar year dominates the aggregate result.

4. Discussion

4.1. Principal Findings

The central finding is more specific than a general claim that graph topology alone drives the improvement. Under the frozen-backbone comparison, MS-GRAN improves one-day RMSE by 1.422% relative to NLinear and improves most stations across three seeds. The self-only control shows that residual-adaptation capacity contributes part of the gain, and rewired placebo graphs retain a substantial share. Accordingly, the result is best interpreted as modest but reproducible evidence for conservative residual adaptation, with the real training-only graph providing an additional but smaller topology-specific advantage over self-only and placebo connectivity.

4.2. Why the Modest Gain Matters

The modest effect size is expected rather than problematic. Daily GNSS displacement residuals contain strong station-specific temporal structure, coloured noise, missingness, equipment effects, and local environmental responses; after a competitive local model has been trained and frozen, little predictable variance may remain. In this setting, a 1.422% gain is meaningful only because it is reproducible under strict controls: three-seed averages remained positive, paired station-bootstrap intervals were positive, upper-tail errors decreased more than mean RMSE, seasonal and missingness-stratified diagnostics remained positive, and the Clark–West test with HAC variance strongly favoured MS-GRAN under

both lag choices. The post-forecast regional-median control produced only a 0.082% gain, while the validation-selected pre-local CME baseline yielded −0.343%. Relative to the tested simple common-mode procedures, MS-GRAN retains a larger predictive improvement: the advantage is 1.765 percentage points over the validation-selected pre-local CME baseline and 1.340 percentage points over the post-forecast regional-median correction. A post hoc audit against the public NGL steps.txt database further showed that registered Code 1/3 equipment, software, or product step neighbourhoods were sparse in the frozen-test predictions: the ±15-day neighbourhood contained 305 observed station-days, representing 0.337% of the observed frozen-test station-days. Excluding these windows preserved the MS-GRAN gain over NLinear (1.482%), whereas the step-neighbourhood subset itself was negative (−8.455%). The conclusion was unchanged when the exclusion window was varied from ±7 to ±30 days, with positive exclusion-window gains of 1.463–1.484%, indicating that these discontinuity windows are difficult cases rather than the source of the aggregate improvement. These seasonal results support the practical relevance of the modest forecasting gain at the forecasting level, while its downstream impact on operational monitoring, such as anomaly screening or false-alarm reduction, remains to be evaluated in application-specific settings. Because annual and semiannual components are intentionally retained in the forecasting target, seasonal or environmentally driven coherence may contribute to the graph signal; the reported benefit is therefore interpreted as forecasting value under the stated residual definition rather than as evidence of a specific physical coupling mechanism.

4.3. Why Conservative Residual Correction Is Preferable

For sparse GNSS networks, unrestricted graph propagation can be misleading because geographic proximity is not equivalent to predictive dependence, and strong station-local histories can be diluted by neighbours with different noise, equipment, or environmental responses. MS-GRAN therefore introduces spatial information only as a conservative residual correction after the station-local forecast has been fixed. The neighbour-minus-own contrast discourages duplication of the local predictor, and the zero-output-initialized residual head and near-zero initial gate require the spatial term to earn its contribution during validation. The station-level evidence indicates that the correction is broadly distributed yet locally selective, which is consistent with the goal of obtaining a transparent and validation-controlled spatial contribution rather than an opaque increase in model complexity. The positive 1.335% gain obtained with the frozen DLinear backbone further supports a residual-adaptation benefit beyond NLinear in this benchmark, although one alternative backbone is insufficient to establish universal backbone independence.

4.4. Operational Relevance and Network Transfer

The practical value of MS-GRAN lies in producing a leakage-controlled expected residual baseline and auditable station-level forecast-deviation indicators for routine deformation monitoring. The horizon-density experiment shows that this residual-adaptation mechanism retains positive aggregate gains across several realistic operating conditions, including reduced station densities and longer direct forecast horizons. Such forecasts may support downstream anomaly screening, but they do not assign physical event labels, and the present regional evidence does not establish transfer to another tectonic setting, a newly deployed network, or spatially held-out stations without independent recalibration and validation.

4.5. Limitations and Future Work

Three limitations define the scope of the findings. First, the benchmark is regional; transfer to a Chinese GNSS network or another tectonic setting requires train-only recali-

bration and independent held-out validation. Second, the gain beyond NLinear is modest, although it is consistent across seeds, stations, horizons, and network densities. Third, MS-GRAN is evaluated as a numerical residual forecaster, not as an earthquake-precursor detector or event classifier.

These limits are also design choices. The compact experiment isolates a transparent question about spatial information before adding external covariates, adaptive edges, multimodal geophysical observations, or event-label objectives.

5. Conclusions

This study tested whether neighbouring GNSS stations provide measurable residual information beyond a strong station-local temporal forecast. MS-GRAN addresses this question through bounded spatial residual adaptation: a station-local NLinear backbone is first trained and frozen, and a training-only graph branch then learns an incremental residual correction. On the frozen Cascadia benchmark, MS-GRAN achieved a 2.9158 ± 0.0033 mm one-day RMSE, improved most stations across three seeds, reduced upper-tail errors, and retained positive aggregate gains across 1-, 3-, and 7-day horizons and nested 90-, 70-, 50-, and 30-station networks. The controls narrow the interpretation of this result. Self-only adaptation retained a 0.635% gain and rewired placebo graphs retained 1.265%, compared with 1.422% for the real graph. The evidence therefore supports a modest but reproducible residual-adaptation effect, while the real training-only topology provides a smaller additional advantage and cannot be identified as the primary source of the improvement. The resulting forecasts can serve as auditable residual baselines for deformation-monitoring workflows, but the present study does not establish universal spatial transfer, physical-event attribution, or earthquake-precursor detection.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/app16189088/s1: Supplementary Table S1 provides the complete baseline-specific configurations, tuning candidates, and validation-selection status used in the benchmark.

Author Contributions: Conceptualization, S.L. and Q.L.; methodology, S.L. and Z.W.; software, S.L.; validation, S.L. and Z.W.; formal analysis, S.L.; investigation, S.L. and Z.W.; resources, Q.L. and G.L.; data curation, S.L. and Z.W.; writing—original draft preparation, S.L.; writing—review and editing, Q.L., G.L. and Z.W.; visualization, S.L.; supervision, Q.L.; project administration, Q.L. and G.L.; funding acquisition, Q.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China, grant number 2024YFC3908000.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The source daily GNSS position products are publicly available from the Nevada Geodetic Laboratory (https://geodesy.unr.edu/, accessed on 3 August 2026). The reproducibility materials for this study are publicly available at https://github.com/lishanshan2026/MSGRAN (accessed on 9 September 2026). The repository includes the 90 station identifiers and coordinates, the fixed train–validation–test split, the training-only adjacency matrix, and the analysis and training code required to reproduce the reported experiments and independently inspect the leakage-control protocol. It also provides the supporting reproducibility files used for the reported analyses. The repository contains no personally identifiable or proprietary data.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Blewitt, G.; Kreemer, C.; Hammond, W.C.; Gazeaux, J. Harnessing the GPS Data Explosion for Interdisciplinary Science. *Eos* **2018**, *99*, e2020943118. [CrossRef]
2. Nevada Geodetic Laboratory. GNSS Time Series and Products. Available online: https://geodesy.unr.edu/ (accessed on 3 August 2026).
3. Dong, D.; Fang, P.; Bock, Y.; Cheng, M.K.; Miyazaki, S. Anatomy of Apparent Seasonal Variations from GPS-Derived Site Position Time Series. *J. Geophys. Res. Solid Earth* **2002**, *107*, ETG 9-1–ETG 9-16. [CrossRef]
4. Bevis, M.; Brown, A. Trajectory Models and Reference Frames for Crustal Motion Geodesy. *J. Geod.* **2014**, *88*, 283–311. [CrossRef]
5. Blewitt, G.; Lavallee, D.; Clarke, P.; Nurutdinov, K. A New Global Mode of Earth Deformation: Seasonal Cycle Detected. *Science* **2001**, *294*, 2342–2345. [CrossRef] [PubMed]
6. Bos, M.S.; Fernandes, R.M.S.; Williams, S.D.P.; Bastos, L. Fast Error Analysis of Continuous GNSS Observations with Missing Data. *J. Geod.* **2013**, *87*, 351–360. [CrossRef]
7. Williams, S.D.P. The Effect of Coloured Noise on the Uncertainties of Rates Estimated from Geodetic Time Series. *J. Geod.* **2003**, *76*, 483–494. [CrossRef]
8. Gazeaux, J.; Williams, S.; King, M.; Bos, M.; Dach, R.; Deo, M.; Moore, A.W.; Ostini, L.; Petrie, E.; Roggero, M.; et al. Detecting Offsets in GPS Time Series: First Results from the Detection of Offsets in GPS Experiment. *J. Geophys. Res. Solid Earth* **2013**, *118*, 2397–2407. [CrossRef]
9. Borsa, A.A.; Agnew, D.C.; Cayan, D.R. Ongoing Drought-Induced Uplift in the Western United States. *Science* **2014**, *345*, 1587–1590. [CrossRef] [PubMed]
10. Wei, C.; Zhou, M.; Du, Z.; Han, L.; Gao, H. Assessment of Hydrological Loading Displacement from GNSS and GRACE Data Using Deep Learning Algorithms. *Sci. Rep.* **2025**, *15*, 6070. [CrossRef] [PubMed]
11. Xiao, Z.; Liu, B.; Dai, W.; Ma, X.; Xing, X.; Luo, Y. GTS_CME: An Open-Source MATLAB-Based Software for the Analysis of Common Mode Errors in GNSS Coordinate Time Series. *GPS Solut.* **2024**, *28*, 131. [CrossRef]
12. Liu, S.; Guo, J.; Zheng, F.; Song, S.; Liu, X. Extraction of Common Mode Errors from GNSS Coordinate Time Series Based on the Integration of MSSA and PCA. *Adv. Space Res.* **2026**, *78*, 2257–2273. [CrossRef]
13. Langbein, J. Noise in GPS Displacement Measurements from Southern California and Southern Nevada. *J. Geophys. Res. Solid Earth* **2008**, *113*, B05405. [CrossRef]
14. Blewitt, G.; Kreemer, C.; Hammond, W.C.; Gazeaux, J. MIDAS Robust Trend Estimator for Accurate GPS Station Velocities without Step Detection. *J. Geophys. Res. Solid Earth* **2016**, *121*, 2054–2068. [CrossRef] [PubMed]
15. Loli Piccolomini, E.; Gandolfi, S.; Poluzzi, L.; Tavasci, L.; Cascarano, P.; Pascucci, A. Recurrent Neural Networks Applied to GNSS Time Series for Denoising and Prediction. In Proceedings of the 26th International Symposium on Temporal Representation and Reasoning, Malaga, Spain, 16–18 October 2019. [CrossRef]
16. Wang, J.; Jiang, W.; Li, Z.; Lu, Y. A New Multi-Scale Sliding Window LSTM Framework (MSSW-LSTM): A Case Study for GNSS Time-Series Prediction. *Remote Sens.* **2021**, *13*, 3328. [CrossRef]
17. Gao, W.; Li, Z.; Chen, Q.; Jiang, W.; Feng, Y. Modelling and Prediction of GNSS Time Series Using GBDT, LSTM and SVM Machine Learning Approaches. *J. Geod.* **2022**, *96*, 71. [CrossRef]
18. Crocetti, L.; Schartner, M.; Soja, B. Discontinuity Detection in GNSS Station Coordinate Time Series Using Machine Learning. *Remote Sens.* **2021**, *13*, 3906. [CrossRef]
19. Li, Z.; Lu, T.; Yu, K.; Wang, J. Interpolation of GNSS Position Time Series Using GBDT, XGBoost, and RF Machine Learning Algorithms and Models Error Analysis. *Remote Sens.* **2023**, *15*, 4374. [CrossRef]
20. Chen, H.; Lu, T.; Huang, J.; He, X.; Yu, K.; Sun, X.; Ma, X.; Huang, Z. An Improved VMD-LSTM Model for Time-Varying GNSS Time Series Prediction with Temporally Correlated Noise. *Remote Sens.* **2023**, *15*, 3694. [CrossRef]
21. Xie, Y.; Wang, J.; Li, H.; Dong, A.; Kang, Y.; Zhu, J.; Wang, Y.; Yang, Y. Deep Learning CNN-GRU Method for GNSS Deformation Monitoring Prediction. *Appl. Sci.* **2024**, *14*, 4004. [CrossRef]
22. Wang, J.; Nie, G.; Gao, S.; Wu, S.; Li, H.; Ren, X. Landslide Deformation Prediction Based on a GNSS Time Series Analysis and Recurrent Neural Network Model. *Remote Sens.* **2021**, *13*, 1055. [CrossRef]
23. Simsek, M.; Taskiran, M.; Dogan, U. Modelling of GNSS Station Position Time Series Using Deep Learning Approaches. *Earth Sci. Inform.* **2025**, *18*, 96. [CrossRef]
24. Jiao, J.; Wang, H.; Dang, Y.; Ren, Y.; Yue, C.; Wu, X.; Cui, H.; Wang, X. Noise-Resilient GNSS Coordinate Time Series Prediction Using AVMD-sLSTM-Transformer Hybrid Model. *Adv. Space Res.* **2025**, *76*, 6863–6881. [CrossRef]
25. He, X.; Zhou, Y.; Li, J.; Kermarrec, G.; Fernandes, R.; Montillet, J.-P.; Zhang, S.; Hu, S. TS_Predictor: A Deep Learning Toolbox for GNSS Time Series Prediction with Signal Decomposition and Nonlinear Modelling. *Adv. Space Res.* **2026**, *77*, 11566–11575. [CrossRef]
26. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]

27. Cho, K.; van Merrienboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 25–29 October 2014; pp. 1724–1734. [CrossRef]
28. Bai, S.; Kolter, J.Z.; Koltun, V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *arXiv* **2018**, arXiv:1803.01271.
29. Lai, G.; Chang, W.C.; Yang, Y.; Liu, H. Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks. In Proceedings of the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval, Ann Arbor, MI, USA, 8–12 July 2018; pp. 95–104. [CrossRef]
30. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. In *Advances in Neural Information Processing Systems 30*; Curran Associates: Red Hook, NY, USA, 2017; pp. 5998–6008.
31. Lim, B.; Arik, S.O.; Loeff, N.; Pfister, T. Temporal Fusion Transformers for Interpretable Multi-Horizon Time Series Forecasting. *Int. J. Forecast.* **2021**, *37*, 1748–1764. [CrossRef]
32. Wu, H.; Xu, J.; Wang, J.; Long, M. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. In *Advances in Neural Information Processing Systems 34*; Curran Associates: Red Hook, NY, USA, 2021; pp. 22419–22430.
33. Zhou, T.; Ma, Z.; Wen, Q.; Wang, X.; Sun, L.; Jin, R. FEDformer: Frequency Enhanced Decomposed Transformer for Long-Term Series Forecasting. In Proceedings of the 39th International Conference on Machine Learning, Baltimore, MD, USA, 17–23 July 2022; pp. 27268–27286.
34. Nie, Y.; Nguyen, N.H.; Sinthong, P.; Kalagnanam, J. A Time Series Is Worth 64 Words: Long-Term Forecasting with Transformers. In Proceedings of the International Conference on Learning Representations, Kigali, Rwanda, 1–5 May 2023.
35. Zeng, A.; Chen, M.; Zhang, L.; Xu, Q. Are Transformers Effective for Time Series Forecasting? In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; Volume 37, pp. 11121–11128. [CrossRef]
36. Li, Y.; Yu, R.; Shahabi, C.; Liu, Y. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018.
37. Yu, B.; Yin, H.; Zhu, Z. Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting. In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; pp. 3634–3640. [CrossRef] [PubMed]
38. Wu, Z.; Pan, S.; Long, G.; Jiang, J.; Zhang, C. Graph WaveNet for Deep Spatial-Temporal Graph Modeling. In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, Macao, China, 10–16 August 2019; pp. 1907–1913. [CrossRef] [PubMed]
39. Bai, L.; Yao, L.; Kanhere, S.S.; Wang, X.; Sheng, Q.Z. Adaptive Graph Convolutional Recurrent Network for Traffic Forecasting. In *Advances in Neural Information Processing Systems 33*; Curran Associates: Red Hook, NY, USA, 2020; pp. 17804–17815.
40. Zheng, C.; Fan, X.; Wang, C.; Qi, J. GMAN: A Graph Multi-Attention Network for Traffic Prediction. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 1234–1241. [CrossRef]
41. Wu, Z.; Pan, S.; Long, G.; Jiang, J.; Chang, X.; Zhang, C. Connecting the Dots: Multivariate Time Series Forecasting with Graph Neural Networks. In Proceedings of the 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, 23–27 August 2020; pp. 753–763. [CrossRef]
42. Shang, C.; Chen, J.; Bi, J. Discrete Graph Structure Learning for Forecasting Multiple Time Series. In Proceedings of the International Conference on Learning Representations, Virtual Event, 3–7 May 2021.
43. Cini, A.; Marisca, I.; Zambon, D.; Alippi, C. Taming Local Effects in Graph-Based Spatiotemporal Forecasting. In *Advances in Neural Information Processing Systems 36*; Curran Associates: Red Hook, NY, USA, 2023; pp. 31290–31314. [CrossRef]
44. Franceschi, L.; Niepert, M.; Pontil, M.; He, X. Learning Discrete Structures for Graph Neural Networks. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 1972–1982.
45. Cini, A.; Marisca, I.; Alippi, C. Filling the G_ap_s: Multivariate Time Series Imputation by Graph Neural Networks. In Proceedings of the International Conference on Learning Representations, Virtual Event, 25–29 April 2022.
46. Fortuin, V.; Baranchuk, D.; Raetsch, G.; Mandt, S. GP-VAE: Deep Probabilistic Time Series Imputation. In Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics, Palermo, Italy, 26–28 August 2020; pp. 1651–1661.
47. Deng, A.; Hooi, B. Graph Neural Network-Based Anomaly Detection in Multivariate Time Series. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual Event, 2–9 February 2021; Volume 35, pp. 4027–4035. [CrossRef]
48. Ruff, L.; Kauffmann, J.R.; Vandermeulen, R.A.; Montavon, G.; Samek, W.; Kloft, M.; Dietterich, T.G.; Muller, K.R. A Unifying Review of Deep and Shallow Anomaly Detection. *Proc. IEEE* **2021**, *109*, 756–795. [CrossRef]
49. Herring, T.A.; King, R.W.; Floyd, M.A.; McClusky, S.C. *Introduction to GAMIT/GLOBK, Release 10.7*; Massachusetts Institute of Technology: Cambridge, MA, USA, 2018.

50. Kreemer, C.; Blewitt, G.; Hammond, W.C.; Gazeaux, J. A Geodetic Plate Motion and Global Strain Rate Model. *Geochem. Geophys. Geosyst.* **2014**, *15*, 3849–3889. [CrossRef]
51. Mao, A.; Harrison, C.G.A.; Dixon, T.H. Noise in GPS Coordinate Time Series. *J. Geophys. Res. Solid Earth* **1999**, *104*, 2797–2816. [CrossRef]
52. Ruttner, P.; Hohensinn, R.; D'Aronco, S.; Wegner, J.D.; Soja, B. Modeling of Residual GNSS Station Motions through Meteorological Data in a Machine Learning Approach. *Remote Sens.* **2022**, *14*, 17. [CrossRef]
53. Burgmann, R.; Rosen, P.A.; Fielding, E.J. Synthetic Aperture Radar Interferometry to Measure Earth's Surface Topography and Its Deformation. *Annu. Rev. Earth Planet. Sci.* **2000**, *28*, 169–209. [CrossRef]
54. Ferretti, A.; Prati, C.; Rocca, F. Permanent Scatterers in SAR Interferometry. *IEEE Trans. Geosci. Remote Sens.* **2001**, *39*, 8–20. [CrossRef]
55. Berardino, P.; Fornaro, G.; Lanari, R.; Sansosti, E. A New Algorithm for Surface Deformation Monitoring Based on Small Baseline Differential SAR Interferograms. *IEEE Trans. Geosci. Remote Sens.* **2002**, *40*, 2375–2383. [CrossRef]
56. Hooper, A.; Bekaert, D.; Spaans, K.; Arikan, M. Recent Advances in SAR Interferometry Time Series Analysis for Measuring Crustal Deformation. *Tectonophysics* **2012**, *514–517*, 1–13. [CrossRef]
57. Biggs, J.; Wright, T.J. How Satellite InSAR Has Grown from Opportunistic Science to Routine Monitoring over the Last Decade. *Nat. Commun.* **2020**, *11*, 3863. [CrossRef] [PubMed]
58. Rosen, P.A.; Hensley, S.; Joughin, I.R.; Li, F.K.; Madsen, S.N.; Rodriguez, E.; Goldstein, R.M. Synthetic Aperture Radar Interferometry. *Proc. IEEE* **2000**, *88*, 333–382. [CrossRef]
59. Salinas, D.; Flunkert, V.; Gasthaus, J.; Januschowski, T. DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks. *Int. J. Forecast.* **2020**, *36*, 1181–1191. [CrossRef]
60. Oreshkin, B.N.; Carpov, D.; Chapados, N.; Bengio, Y. N-BEATS: Neural Basis Expansion Analysis for Interpretable Time Series Forecasting. In Proceedings of the International Conference on Learning Representations, Addis Ababa, Ethiopia, 26–30 April 2020.
61. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual Event, 2–9 February 2021; Volume 35, pp. 11106–11115. [CrossRef]
62. Wu, H.; Hu, T.; Liu, Y.; Zhou, H.; Wang, J.; Long, M. TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis. In Proceedings of the International Conference on Learning Representations, Kigali, Rwanda, 1–5 May 2023.
63. Kipf, T.N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. In Proceedings of the International Conference on Learning Representations, Toulon, France, 24–26 April 2017.
64. Guo, S.; Lin, Y.; Feng, N.; Song, C.; Wan, H. Attention Based Spatial-Temporal Graph Convolutional Networks for Traffic Flow Forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 922–929. [CrossRef]
65. Cao, D.; Wang, Y.; Duan, J.; Zhang, C.; Zhu, X.; Huang, C.; Tong, Y.; Xu, B.; Bai, J.; Tong, J.; et al. Spectral Temporal Graph Neural Network for Multivariate Time-Series Forecasting. In *Advances in Neural Information Processing Systems 33*; Curran Associates: Red Hook, NY, USA, 2020; pp. 17766–17778.
66. Yoon, J.; Jordon, J.; van der Schaar, M. GAIN: Missing Data Imputation Using Generative Adversarial Nets. In Proceedings of the 35th International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 5689–5698.
67. Cao, W.; Wang, D.; Li, J.; Zhou, H.; Li, L.; Li, Y. BRITS: Bidirectional Recurrent Imputation for Time Series. In *Advances in Neural Information Processing Systems 31*; Curran Associates: Red Hook, NY, USA, 2018; pp. 6775–6785.
68. Su, Y.; Zhao, Y.; Niu, C.; Liu, R.; Sun, W.; Pei, D. Robust Anomaly Detection for Multivariate Time Series through Stochastic Recurrent Neural Network. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 2828–2837. [CrossRef]
69. Audibert, J.; Michiardi, P.; Guyard, F.; Marti, S.; Zuluaga, M.A. USAD: UnSupervised Anomaly Detection on Multivariate Time Series. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining Workshops, Virtual Event, 24 August 2020.
70. Chandola, V.; Banerjee, A.; Kumar, V. Anomaly Detection: A Survey. *ACM Comput. Surv.* **2009**, *41*, 15. [CrossRef]
71. Huber, P.J. Robust Estimation of a Location Parameter. *Ann. Math. Stat.* **1964**, *35*, 73–101. [CrossRef]
72. Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization. In Proceedings of the International Conference on Learning Representations, New Orleans, LA, USA, 6–9 May 2019.
73. Efron, B.; Tibshirani, R.J. *An Introduction to the Bootstrap*; Chapman and Hall/CRC: Boca Raton, FL, USA, 1994.
74. Newey, W.K.; West, K.D. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* **1987**, *55*, 703–708. [CrossRef]

75. Diebold, F.X.; Mariano, R.S. Comparing Predictive Accuracy. *J. Bus. Econ. Stat.* **1995**, *13*, 253–263. [CrossRef]
76. Clark, T.E.; West, K.D. Approximately normal tests for equal predictive accuracy in nested models. *J. Econom.* **2007**, *138*, 291–311. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.