Spatio-Temporal Graph Neural Networks for Multi-Horizon River-Stage Forecasting in the Amazon: A Case Study on the Manaus–Santarém Corridor

Isabela A. Aguiar[1], Gabrielly F. Rodrigues[1], Raphael I. Winovski Vilhena[1], Antonio Fontenelle[1], Wington L. Vital[1], Italo P. Caliari[1], Hendrio Bragança[1]

[1]Núcleo de Capacitação em Inteligência Artificial (NCIA), Fundação Paulo Feitoza - FPFtech, Av. Danilo de Matos Areosa, 1170, Distrito Industrial, Manaus, Amazonas, 69075-351, Brazil.

{isabela.252312, gabrielly.252311, raphael.252325, antonio.fontenele, wington.vital, italo.caliari, hendrio.braganca}@fpf.br

Abstract. *Inland waterway transport in the Brazilian Amazon depends critically on river navigability, which is strongly governed by hydroclimatic variability and increasingly severe hydrological extremes. This paper proposes a multi-model framework for daily multi-horizon river-stage forecasting in the Manaus–Santarém corridor, combining official hydrometric observations from ANA with a spatio-temporal graph neural network (STGNN) that encodes upstream–downstream connectivity and travel-time-informed edge weights. The STGNN is evaluated against temporal benchmarks, recurrent baselines, a statistical model, and operational baselines across six forecast horizons (1–90 days) under a unified protocol with a chronological hold-out test set covering 2024. The STGNN achieves the best MAE, RMSE, and NSE at every horizon, maintaining NSE $>$ 0.99 through 30 days, a threshold no competing model reaches, with RMSE gains of 29–35% over the strongest baseline at 14–30 days, the window governed by upstream signal propagation according to the physical travel times encoded in the graph. Results demonstrate that explicitly representing the river network as a directed graph provides consistent and physically interpretable predictive advantages over purely temporal architectures for multi-horizon Amazonian hydrological forecasting.*

1. Introduction

In the Brazilian Amazon, rivers are the primary infrastructure sustaining mobility, trade, and territorial integration across vast areas with limited road connectivity. Inland waterway transport is therefore central to the circulation of food, fuel, and consumer goods, and to access to essential services for riverine populations. The operational reliability of this system, however, is strongly conditioned by hydroclimatic variability: seasonal oscillations in river stage, variable discharge, sediment redistribution, and prolonged drought events interact to create a dynamic navigational environment that directly affects safety, route feasibility, and logistics planning [Fassoni-Andrade et al. 2021, de Lima et al. 2024, Espinoza et al. 2022, Chagas et al. 2022].

These pressures have intensified recently, with the Amazon experiencing increasingly severe extremes, the 2021 flood exceeded historical benchmarks in Manaus while the 2023 drought produced exceptionally low water levels across large portions of the basin [Espinoza et al. 2022, Espinoza et al. 2024]. Severe droughts have been shown to reduce river navigability, isolate communities, and disrupt freight transport, reflecting a broader pattern of increasing hydrological variability driven by ocean–atmosphere interactions, land-use change, and global warming [de Lima et al. 2024, Marengo et al. 2024].

Despite the availability of long hydrometric records through official Brazilian monitoring infrastructure, sush as particularly ANA's HidroWeb and the SNIRH network [Agência Nacional de Águas e Saneamento Básico (ANA) 2025a, Agência Nacional de Águas e Saneamento Básico (ANA) 2025b], operational decision-making in Amazonian river logistics still relies heavily on fragmented information and the tacit knowledge of experienced pilots.

A growing body of literature has advanced the understanding of Amazon flood pulses, inundation dynamics, and the climatic controls of extreme river stages [Fassoni-Andrade et al. 2021, Chagas et al. 2022, Marengo et al. 2024], while machine learning and deep learning methods have demonstrated strong results for nonlinear river-stage prediction and hydrological time-series modeling [Zhao et al. 2024, Li and Jun 2024, Domenighini 2024]. Research on navigational risk in inland waterways has likewise progressed, proposing data-driven models based on vessel-maneuvering behavior and sequence models such as LSTM variants; however, these are largely calibrated for densely regulated channels where traffic density, rather than hydroclimatic forcing, is the dominant risk driver.

Across all three strands, a persistent gap remains: hydrological studies characterize environmental variability without translating it into route-level risk indicators; machine-learning models stop at environmental forecasting without producing navigation-centered outputs; and inland-waterway risk models do not generalize to the seasonal extremes, long-range hydrological propagation, and sparse infrastructure of the Amazon. Moreover, most approaches treat each gauging station as an isolated time series, ignoring that river-stage propagation is inherently spatial and depends on upstream, downstream connectivity, a limitation that is especially consequential in large river systems such as the Manaus–Santarém corridor.

This study addresses these gaps by proposing a multi-model framework for daily multi-horizon river-stage forecasting in the Manaus–Santarém corridor. Daily multi-station series from ANA are organized into a common temporal grid; the river network is represented as a directed graph based on upstream–downstream connectivity and travel-time-informed links; and a spatio-temporal graph neural network (STGNN) is evaluated against temporal benchmarks under a unified experimental protocol.

The framework contributes: (i) a physically oriented river connectivity representation combined with edge weights derived from lagged statistical associations in the training data; (ii) horizon-specific training with a fixed 365-day input window and upstream lag features across six forecast lead times (1, 7, 14, 30, 60, and 90 days); (iii) assessment under both a main chronological split and a walk-forward validation regime.

By linking long-term hydrometric monitoring to spatially explicit predictive modeling, the framework provides a reproducible basis for anticipating river-stage dynamics in one of the world's most hydrologically variable regions, with direct relevance to river logistics planning in the Amazon.

2. Methodology

This section describes the multi-model framework adopted for daily river-stage forecasting in the Manaus–Santarém fluvial corridor. The central objective is to investigate the role of explicit fluvial network connectivity in multi-horizon water-level prediction.

The pipeline is organized into four sequential modules: Data Layer, Pre-Processing, Modeling & Training, and Selection & Evaluation, as illustrated in Figure 1. Across all experiments, the same supervised dataset, temporal partitions, and evaluation protocol were maintained to ensure full comparability between models.

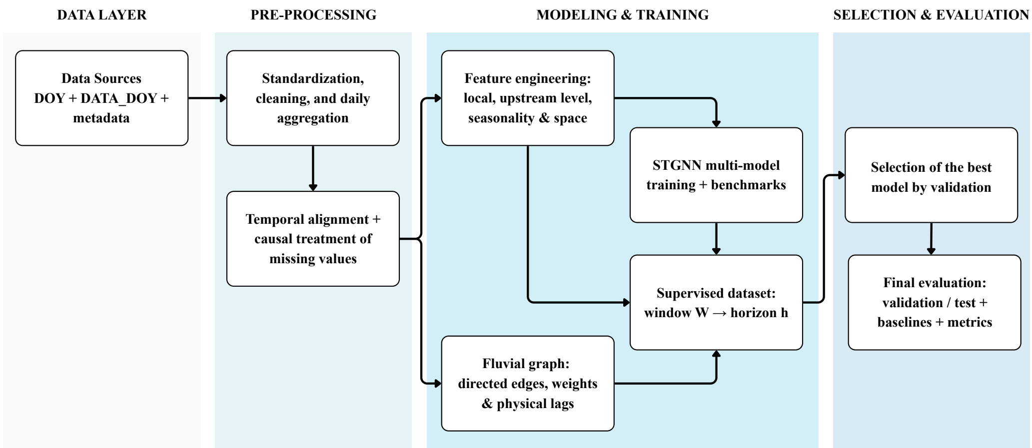


Figure 1. Methodological flow of the proposed framework, organized into four pipeline modules: Data Layer, Pre-Processing, Modeling & Training, and Selection & Evaluation.

2.1. Data Layer

The study uses daily river-stage series spanning 1 January 2016 to 31 December 2024 (3,288 time steps after trimming), acquired from the Agência Nacional de Águas e Saneamento Básico (ANA) through the HidroWeb platform and SNIRH infrastructure [Agência Nacional de Águas e Saneamento Básico (ANA) 2025a, Agência Nacional de Águas e Saneamento Básico (ANA) 2025b]. The monitoring network consists of seven hydrometric stations along the Manaus–Santarém corridor—four telemetric and three conventional—spanning approximately 1,100 km of the main Amazon channel. Telemetric stations were prioritized as the primary source given their greater operational reliability; conventional stations completed spatial and temporal coverage where telemetric records were insufficient.

The five target nodes (Manaus, Itacoatiara, Parintins, Óbidos, and Santarém) define the stations whose future water levels constitute the prediction objective. São Gabriel da Cachoeira and Barcelos serve as upstream context nodes, contributing to the spatio-temporal network representation but excluded from the loss function. Input sources include conventional DOY spreadsheets, telemetric `DATA_DOY` spreadsheets, station-to-file and station-to-hydrometric-code mapping files, main-channel connectivity, and static spatial attributes per station.

Table 1. Stations used in the experiment.

Station	Type	Role
São Gabriel da Cachoeira	Conventional	Spatial and temporal context
Barcelos	Conventional	Spatial and temporal context
Manaus	Telemetric	Target node
Itacoatiara	Telemetric	Target node
Parintins	Conventional	Target node
Óbidos	Telemetric	Target node
Santarém	Telemetric	Target node

2.2. Pre-Processing

Raw records were transformed through six sequential steps: conversion to long format, metadata standardization, daily aggregation, temporal alignment, causal imputation, and annual representativeness filtering.

All stations were synchronized into a *single temporal matrix* in which each row represents the simultaneous state of the full network, enabling models to learn spatial correlations across nodes. A minimum annual coverage threshold of 60% was applied per station to balance historical span with information density.

Gap imputation followed a strictly causal protocol to prevent data leakage: windows of up to seven days were filled by forward fill; longer gaps used the cumulative historical mean of the series, computed exclusively from past observations. The historical series begins at the first date of complete multi-station operation, ensuring coherence between conventional and telemetric records across all partitions.

2.3. Modeling and Training

2.3.1. Fluvial Graph: Directed Edges, Weights, and Physical Lags

The river network is modeled as a directed graph $G = (V, E)$, where nodes V are monitoring stations and edges E encode physical upstream–downstream connections, introducing a hydrological inductive bias that propagates information along the true river topology rather than Euclidean distances. The graph contains six directed edges: São Gabriel da Cachoeira $\rightarrow$ Barcelos $\rightarrow$ Manaus $\rightarrow$ Itacoatiara $\rightarrow$ Parintins $\rightarrow$ Óbidos $\rightarrow$ Santarém.

Each edge carries a physical lag (Table 2), rounded to the nearest integer day from the `travel_time_days` attribute, with a minimum floor of one day. Edge weights were estimated on the training split only. For a directed edge $i \rightarrow j$ with lag l_{ij}, the weight is defined as Equation 1.

$$w_{ij} = \max\bigl(0.05,\ \bigl|\mathrm{corr}\bigl(x_{1:T-l_{ij}},\ z_{1+l_{ij}:T}\bigr)\bigr|\bigr)\,, \tag{1}$$

where $x_t = y_t^{(i)}$ and $z_t = y_t^{(j)}$. Edge direction is physically fixed; intensity is calibrated by the lag-shifted Pearson correlation observed in the training period, with a safety floor of 0.05.

Table 2. Physical lags per graph edge.

Edge	Lag (days)
São Gabriel da Cachoeira $\rightarrow$ Barcelos	5
Barcelos $\rightarrow$ Manaus	4
Manaus $\rightarrow$ Itacoatiara	2
Itacoatiara $\rightarrow$ Parintins	2
Parintins $\rightarrow$ Óbidos	2
Óbidos $\rightarrow$ Santarém	2

2.3.2. Feature Engineering

For each time step t and station n, the input tensor $X \in \mathbb{R}^{T \times N \times F}$ combines four feature groups: (i) local water level, (ii) upstream lag attributes at fixed lags $l \in \{3, 7, 14\}$ days, (iii) seasonal encoding (sine and cosine of day-of-year), and (iv) static spatial attributes (latitude, longitude, relative upstream distance, and main-channel position score).

Fixed upstream lags complement the physical graph lags by capturing short- to medium-range upstream influences that may not be fully described by a single travel-time delay. All features were standardized with statistics estimated on the training set; the target was normalized per station and converted back to meters for evaluation.

2.3.3. Supervised Formulation

The problem is formulated as sliding-window regression with input window $W = 365$ days and forecast horizons $H = \{1, 7, 14, 30, 60, 90\}$ days. The prediction target is the future level change as showed in Equation 2, reducing prediction amplitude and facilitating learning at longer horizons. Models were trained independently per horizon in a direct single-horizon regime; each sample spans the full network rather than isolated per-station observations.

$$\Delta y_{t+h} = y_{t+h} - y_t, \tag{2}$$

2.4. Selection and Evaluation

Model selection was performed by minimum validation loss on the 2023 hold-out year, using early stopping ($\texttt{patience} = 12$, minimum improvement 10^{-5}) decoupled from best-checkpoint saving to avoid sensitivity to short-term oscillations. Final evaluation was conducted on the 2024 test set. A walk-forward scheme with three annual folds (test years 2022, 2023, 2024) was used during development to verify that rankings were stable across different hydrological regimes.

Performance is assessed with three complementary metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Nash–Sutcliffe Efficiency (NSE), aggregated across the five target nodes. RMSE is the primary benchmark for cross-model comparison given its sensitivity to larger deviations; NSE measures skill relative to the observed series mean and is treated as an independent hydrological indicator, not a surrogate for R^2.

3. Experimental Setup

3.1. Evaluation Scenarios

Two evaluation regimes were employed: a primary hold-out split for final performance reporting, and a walk-forward scheme used to support model selection and verify robustness during development.

Primary evaluation — Main chronological split. A single hold-out partition was defined a priori, respecting strict temporal causality with no shuffling and no target overlap between splits. All quantitative results reported in Section 4 correspond to this scenario. To ensure sufficient historical context for the first samples in each split, a left-context window of length $W = 365$ days was prepended. This context was used exclusively as model input; prediction targets remained restricted to the corresponding split interval.

Table 3. Main temporal partition.

Split	Start	End	Days	Purpose
Train	2016-01-01	2022-12-31	2,557	Parameter fitting
Validation	2023-01-01	2023-12-31	365	Model and hyperparameter selection
Test	2024-01-01	2024-12-31	366	Final evaluation

Robustness check — Walk-forward validation. To guard against conclusions dependent on a single train–test cut and to support model selection across diverse hydrological years, a walk-forward scheme was applied with three contiguous annual folds during development. In each fold, one year is withheld for testing, the year immediately preceding it serves as validation, and all earlier data form the training set; the test block advances one year per fold.

Table 4. Walk-forward validation folds.

Fold	Train	Validation	Test
1	2016-01-01 – 2020-12-31	2021-01-01 – 2021-12-31	2022-01-01 – 2022-12-31
2	2016-01-01 – 2021-12-31	2022-01-01 – 2022-12-31	2023-01-01 – 2023-12-31
3	2016-01-01 – 2022-12-31	2023-01-01 – 2023-12-31	2024-01-01 – 2024-12-31

Fold 3 is structurally equivalent to the main split, providing a direct consistency check between the two regimes. Walk-forward fold results confirmed the ranking observed on the main test set and are used here to support the robustness interpretation in Section 5; per-fold quantitative breakdowns are left for future extended reporting.

3.2. Dataset

All experiments use daily river-stage observations acquired from the Agência Nacional de Águas e Saneamento Básico (ANA) through the HidroWeb platform and SNIRH infrastructure [Agência Nacional de Águas e Saneamento Básico (ANA) 2025a, Agência Nacional de Águas e Saneamento Básico (ANA) 2025b]. The dataset covers the period from 1 January 2016 to 31 December 2024, yielding 3,288 time steps after temporal alignment, causal imputation, and annual representativeness filtering. Seven hydrometric stations along the Manaus–Santarém corridor were selected, spanning approximately 1,100 km of the main Amazon channel. Of these, five serve as target nodes (Manaus, Itacoatiara, Parintins, Óbidos, and Santarém) and two as upstream context nodes (São Gabriel da Cachoeira and Barcelos). Table 5 summarizes the key properties of the experimental dataset.

Table 5. Summary of the experimental dataset.

Property	Value
Source	ANA / HidroWeb (SNIRH)
Variable	Daily river stage (m)
Period	2016-01-01 – 2024-12-31
Total time steps	3,288 days
Number of stations	7 (4 telemetric, 3 conventional)
Target nodes	5 (Manaus, Itacoatiara, Parintins, Óbidos, Santarém)
Context nodes	2 (São Gabriel da Cachoeira, Barcelos)
Annual coverage threshold	$\geq$ 60% per station-year
Gap imputation	Forward fill ($\leq$7 days); causal historical mean (¿7 days)

The nine-year span covers a wide range of hydrological regimes, including the record high-water event of 2021 in Manaus and the severe drought of 2023, which produced exceptionally low river levels across the basin. This variability makes the dataset representative of the hydroclimatic extremes that the forecasting framework must handle in practice.

3.3. Supervised Formulation

The prediction problem was formulated as sliding-window regression with an input window fixed at $W = 365$ days. The set of forecast horizons evaluated is presented in Equation 3. Each supervised sample consists of: the W-day historical window preceding the reference time step, the observed level at the reference time step, the future target at horizon h, and supervision restricted to target nodes. Crucially, each sample contains a complete historical snapshot of the full network, not isolated per-station observations. The prediction target is defined as the future level change relative to the reference instant as presented in Equation 4. Formulating the target as a delta reduces the amplitude of the prediction problem and facilitates learning of future variations, particularly at longer horizons where absolute-level prediction is substantially harder.

$$H = \{1,\ 7,\ 14,\ 30,\ 60,\ 90\}\ \text{days}. \tag{3}$$

$$\Delta y_{t+h} = y_{t+h} - y_t. \tag{4}$$

3.4. Training, Optimization, and Model Selection

Training was standardized across all experiments to preserve comparability. All models were trained with the AdamW optimizer and a `ReduceLROnPlateau` scheduler (`patience` $= 3$, `factor` $= 0.5$, minimum learning rate 1×10^{-5}). Central hyperparameters are listed in Table 6.

The loss function used in all reported experiments was the Huber loss with $\delta = 0.7$, which balances sensitivity to small errors with robustness to outliers. Because

Table 6. Central training hyperparameters.

Parameter	Value
Initial learning rate	3×10^{-4}
Batch size	32
Maximum epochs	80
Weight decay	1×10^{-5}
Hidden dimension (backbone)	128
Dropout	0.20
Gradient clipping	max norm 1.0

the protocol trains one model per horizon, the effective loss in each run is single-horizon; horizon weighting therefore does not affect the final reported checkpoints. For autoregressive models, the teacher-forcing ratio decays linearly from 1.0 to 0.1 over 40 epochs, progressively reducing dependence on ground-truth inputs and better approximating inference conditions. Model selection was performed by minimum validation loss. Early stopping used $\texttt{patience} = 12$ with a minimum improvement threshold of 1×10^{-5}. Decoupling best-checkpoint saving from early stopping reduces sensitivity to short-term oscillations without sacrificing the best generalization point observed on validation.

3.5. Models Evaluated

3.5.1. STGNN

The spatio-temporal graph neural network (STGNN) is implemented as a hybrid STGCN–DCRNN architecture for multi-horizon river-stage forecasting. The model is composed of: (i) a gated temporal convolutional block (kernel size = 3, dilation = 2) for extraction of local temporal patterns; (ii) diffusion-style recurrent graph cells derived from the DCRNN framework for cross-node propagation; (iii) a combination of physical adjacency, adaptive adjacency, and optional dynamic adjacency; and (iv) an autoregressive decoder with scheduled teacher forcing at the target forecast horizon. The physical graph provides a hydrological inductive bias by propagating information along the true upstream–downstream topology of the river rather than treating each station as an isolated series.

3.5.2. Baselines

Two simple baselines were used as lower-bound references. The *persistence baseline* projects the last available observation forward to all horizons (Equation 5). The *seasonal baseline* uses the observation from 365 days prior (Equation 6). When that value is unavailable, the seasonal baseline falls back to persistence.

$$\hat{y}^{\text{persist}}_{t+h|t} = y_t. \tag{5}$$

$$\hat{y}^{\text{seasonal}}_{t+h|t} = y_{t+h-365}. \tag{6}$$

Prophet was included as a univariate statistical baseline, following recent evidence of its applicability to short-range hydrological level forecasting [Elneel et al. 2024]. A separate model was fitted for each hydrometric station from its local historical series, with explicit annual seasonality modeled by Fourier series with a period of 365 days, consistent with the input window adopted in the main model. Predictions were generated directly for multiple target horizons from the test-set reference dates.

Long Short-Term Memory (LSTM) networks were adopted as a recurrent deep-learning baseline given their wide application in hydrological level forecasting [Luo et al. 2025]. Gated Recurrent Units (GRU) were included as an additional baseline owing to their efficiency in modeling temporal dependencies and lower computational cost relative to LSTM. Both were configured in a multi-step, multi-output regime: a separate model was trained per forecast horizon, and each model predicts simultaneously over all target stations. Experimental settings were kept consistent with the STGNN to ensure fair comparison.

Three additional temporal benchmarks were evaluated: the Temporal Fusion Transformer (TFT), PatchTST, and N-BEATS. These models serve as strong temporal references and were trained under the same experimental protocol, without explicit use of the fluvial graph within their architectures. All three process the multivariate input tensor but model inter-station relationships implicitly through attention or basis-decomposition mechanisms rather than through physically structured graph connectivity.

3.6. Evaluation Metrics

Results are reported both aggregated by model and horizon, and disaggregated by target station. The primary metrics are Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Nash–Sutcliffe Efficiency (NSE). NSE is treated as an independent hydrological skill metric and not as a surrogate for R^2.

Table 7. Definition of primary evaluation metrics.

Metric	Abbrev.	Formula	Interpretation
Mean Absolute Error	MAE	$\frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$	Average error magnitude in meters.
Root Mean Squared Error	RMSE	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$	Penalizes larger errors more heavily than MAE.
Nash–Sutcliffe Efficiency	NSE	$1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$	Model skill relative to the observed series mean.

4. Results

Table 8 reports MAE, RMSE, and NSE for all models on the held-out test set (January–December 2024) across six forecast horizons. The STGNN achieves the best score on all three metrics at every horizon. Figure 2 complements the table by showing the full RMSE density distributions, including PatchTST, N-BEATS, and Persistence, which are discussed qualitatively below.

Short horizons (1–7 days). At 1 day, STGNN (RMSE 0.044 m, NSE 0.99996) and TFT (RMSE 0.047 m, NSE 0.99996) are nearly equivalent, reflecting the dominance of local temporal autocorrelation at short lead times. GRU lags behind (RMSE 0.137 m), and Prophet is unusable across all horizons (RMSE >2.0 m, NSE <0.65). At 7 days, GRU becomes the second-best model (RMSE 0.282 m), narrowly ahead of TFT (0.307 m) and LSTM (0.329 m), while the STGNN maintains a clear lead (RMSE 0.225 m, NSE 0.99914).

Table 8. Performance Comparison of Models Across Forecast Horizons

Forecast Horizon (days)	Model	MAE (m)	RMSE (m)	NSE
1	STGNN	**0.0301**	**0.0439**	**0.99996**
	TFT	0.0323	0.0467	0.99996
	LSTM	0.0561	0.0792	0.99990
	GRU	0.0864	0.1373	0.99970
	Prophet	1.5294	2.0194	0.64722
7	STGNN	**0.1607**	**0.2254**	**0.99914**
	GRU	0.1937	0.2822	0.99876
	TFT	0.2118	0.3073	0.99840
	LSTM	0.2243	0.3291	0.99832
	Prophet	1.5691	2.0835	0.62149
14	STGNN	**0.2745**	**0.3977**	**0.99733**
	TFT	0.3799	0.5599	0.99472
	LSTM	0.3883	0.6016	0.99440
	GRU	0.4027	0.5898	0.99461
	Prophet	1.6168	2.1609	0.58992
30	STGNN	**0.5272**	**0.7622**	**0.99031**
	GRU	0.7927	1.1750	0.97858
	LSTM	0.8042	1.1746	0.97860
	TFT	0.8552	1.2493	0.97398
60	STGNN	**0.9025**	**1.2610**	**0.97396**
	LSTM	1.1136	1.5370	0.96325
	GRU	1.2010	1.7017	0.95495
	TFT	1.2974	1.5781	0.95922
	Prophet	1.9146	2.5455	0.42426
90	STGNN	**1.3253**	**1.6259**	**0.95735**
	LSTM	1.3660	1.9132	0.94296
	GRU	1.3767	1.8967	0.94394
	TFT	1.8277	2.0857	0.92982
	Prophet	2.1106	2.7341	0.33149

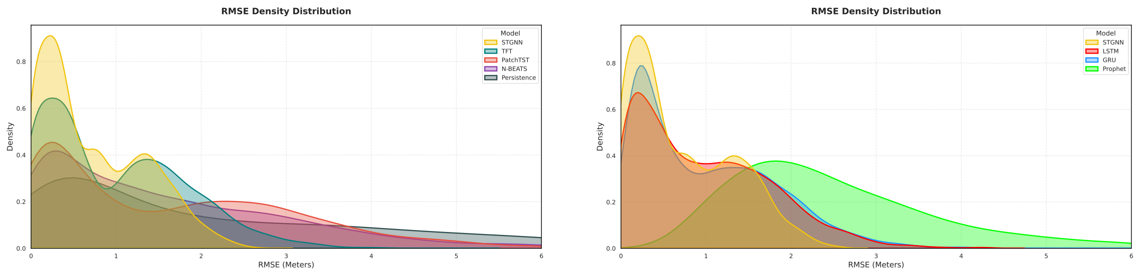
Medium horizons (14–30 days). This is where the STGNN advantage is most pronounced. At 14 days, it outperforms the next best model (TFT, RMSE 0.560 m) by 28.9% in RMSE and holds NSE 0.997 versus 0.995 for TFT. At 30 days, gains reach 35.1% over both LSTM and GRU (RMSE 1.175 m), with STGNN NSE still above 0.99—a threshold no competitor reaches beyond the 7-day horizon. This window corresponds to the cumulative travel times encoded in the graph (2–17 days), making upstream propagation the dominant predictive signal.

Long horizons (60–90 days). Error growth is substantial for all models, but the STGNN remains the most robust. At 60 days, LSTM is the closest competitor (RMSE 1.537 m vs. 1.261 m, NSE 0.963 vs. 0.974). At 90 days, the STGNN (RMSE 1.626 m, NSE 0.957) outperforms GRU (RMSE 1.897 m), LSTM (1.913 m), and TFT (2.086 m), while Prophet collapses to NSE 0.331, confirming its inadequacy for operational multi-horizon forecasting in this corridor.

5. Discussion

The STGNN's consistent superiority across all horizons confirms that explicitly encoding upstream–downstream connectivity and travel-time lags provides predictive advantages beyond what the same input data can yield through purely temporal architectures trained under identical conditions.

The horizon-dependent gain profile is physically coherent: advantages are mod-



(a) Baseline models (STGNN, TFT, PatchTST, N-BEATS, Persistence).

(b) Recurrent and statistical models (STGNN, LSTM, GRU, Prophet).

Figure 2. RMSE density distributions on the test set. The STGNN distribution is shifted toward lower errors and more tightly concentrated across both model groups.

est at 1 day, where local temporal autocorrelation dominates, grow substantially at 14–30 days, where cumulative travel times from upstream context nodes (2–17 days) make spatial propagation most relevant, and narrow beyond 60 days as large-scale hydroclimatic uncertainty increasingly dominates the signal. From an operational standpoint, STGNN errors of 0.225 m at 7 days and 0.762 m at 30 days are practically useful for short-term scheduling and anticipating flood-pulse transitions, respectively; predictions at 60–90 days are better interpreted as broad seasonal regime indicators than precise navigability assessments.

The main limitations of this work are the absence of per-station error disaggregation, the use of static rather than regime-dependent travel-time lags, and the exclusion of meteorological covariates and climate indices that could extend predictive skill beyond 30 days. These constitute direct extensions of the present framework.

6. Conclusion

This paper proposed a spatio-temporal graph neural network framework for daily multi-horizon river-stage forecasting in the Manaus–Santarém corridor, using directed fluvial graph structure derived from official ANA hydrometric data. Evaluated against eight competing models across six horizons (1–90 days), the STGNN achieved the lowest RMSE at every lead time, with gains of 29–35% over the best temporal baseline at 14–30 days, the window governed by physical upstream–downstream propagation.

The results demonstrate that representing the river as a directed graph, rather than a collection of independent series, yields consistent and physically interpretable improvements for Amazonian hydrological forecasting, with direct applicability to logistics planning in a region facing intensifying hydroclimatic extremes.

Future work will extend reporting to per-station breakdowns, controlled ablation studies, dynamic lags, meteorological covariates, and probabilistic forecast formulations.

Acknowledgments

This research was conducted with the support of Fundação Amazônica de Amparo à Pesquisa e Desenvolvimento Tecnológico Desembargador Paulo dos Anjos Feitoza (FPFtech) and Foxconn Brasil.

References

Agência Nacional de Águas e Saneamento Básico (ANA) (2025a). Hidroweb: Sistema nacional de informações sobre recursos hídricos (snirh). `https://www.snirh.gov.br/hidroweb/`. Hydrometeorological data platform. Accessed: 10 Mar 2026.

Agência Nacional de Águas e Saneamento Básico (ANA) (2025b). Portal de dados abertos da ana. `https://dadosabertos.ana.gov.br`. Acesso em: 10 mar. 2026.

Chagas, V. B., Chaffe, P. L., and Blöschl, G. (2022). Process controls on flood seasonality in brazil. *Geophysical Research Letters*, 49.

de Lima, L. S., e. Silva, F. E. O., Anastácio, P. R. D., de Paula Kolanski, M. M., Pereira, A. C. P., Menezes, M. S. R., Cunha, E. L. T. P., and Macedo, M. N. (2024). Severe droughts reduce river navigability and isolate communities in the brazilian amazon. *Communications Earth and Environment*, 5.

Domenighini, C. (2024). Autonomous inland navigation: a literature review and extra-contractual liability issues. *Journal of Shipping and Trade*, 9.

Elneel, L., Zitouni, M. S., Mukhtar, H., and Al-Ahmad, H. (2024). Examining sea levels forecasting using autoregressive and prophet models. *Scientific Reports*, 14(1):14337.

Espinoza, J. C., Jimenez, J. C., Marengo, J. A., Schongart, J., Ronchail, J., Lavado-Casimiro, W., and Ribeiro, J. V. M. (2024). The new record of drought and warmth in the amazon in 2023 related to regional and global climatic features. *Scientific Reports*, 14.

Espinoza, J. C., Marengo, J. A., Schongart, J., and Jimenez, J. C. (2022). The new historical flood of 2021 in the amazon river compared to major floods of the 21st century: Atmospheric features in the context of the intensification of floods. *Weather and Climate Extremes*, 35.

Fassoni-Andrade, A. C., Fleischmann, A. S., Papa, F., de Paiva, R. C. D., Wongchuig, S., Melack, J. M., Moreira, A. A., Paris, A., Ruhoff, A., Barbosa, C., Maciel, D. A., Novo, E., Durand, F., Frappart, F., Aires, F., Abrahão, G. M., Ferreira-Ferreira, J., Espinoza, J. C., Laipelt, L., Costa, M. H., Espinoza-Villar, R., Calmant, S., and Pellet, V. (2021). Amazon hydrology from space: Scientific advances and future challenges.

Li, L. and Jun, K. S. (2024). Review of machine learning methods for river flood routing.

Luo, J., Zhu, D., and Li, D. (2025). Classification-enhanced lstm model for predicting river water levels. *Journal of Hydrology*, 650:132535.

Marengo, J. A., Espinoza, J. C., Fu, R., Muñoz, J. C. J., Alves, L. M., ROCHA, H. R. D., and Schöngart, J. (2024). Long-term variability, extremes and changes in temperature and hydrometeorology in the amazon region: A review. *Acta Amazonica*, 54.

Zhao, X., Wang, H., Bai, M., Xu, Y., Dong, S., Rao, H., and Ming, W. (2024). A comprehensive review of methods for hydrological forecasting based on deep learning.