ST-GNN-IOT: Spatio-Temporal Graph Neural Networks with Sensor-Proxy-Driven Dynamic Edge-Weight Modulation for Large-Scale Renewable Energy Forecasting

Rudra Yadav[1], Rameshwari Pramod Bhurle[2], Rinku Chakradhar Bawankule[3], Rahul Hemraj Shrirame[4], Roshan Sudam Wani[5]

Govt. College of Engineering, Nagpur, India

rudrayadav080@gmail.com[1], rameshwaribhurle41@gmail.com[2], bawankulerinku@gmail.com[3], hemrajshrirame46@gmail.com[4], roshanwani9th@gmail.com[5]

How to Cite this Article:

Yadav, R., Wani, R. S., Shrirame, R. H., Bawankule, R. C. & Bhurle, R. P. (2026). ST-GNN-IOT: Spatio-Temporal Graph Neural Networks with Sensor-Proxy-Driven Dynamic Edge-Weight Modulation for Large-Scale Renewable Energy Forecasting. International Journal of Creative and Open Research in Engineering and Management, <i>02</i>(03). https://doi.org/10.55041/ijcope.v2i3.117

https://doi.org/10.55041/ijcope.v2i3.117

Abstract

Accurate multi-node renewable energy forecasting is increasingly critical as national grids approach high renewable penetration, yet the spatial correlation structure that governs forecast accuracy changes continuously with meteorological conditions—a non-stationarity that deployed models systematically ignore. The consequence is that existing spatio-temporal graph neural network (ST-GNN) architectures fail most severely during precisely those high-variability events that are most consequential for spinning-reserve procurement and grid stability. This paper addresses this structural limitation analytically rather than empirically, advancing three formally proved contributions: (i) a model-class impossibility result establishing that any node-local estimator incurs irreducible spatial bias growing with the spatial scale of the driving meteorological event, providing a theoretical rather than empirical basis for graph-based forecasting; (ii) a theorem establishing the precise condition under which dynamic adjacency strictly dominates static adjacency in expected forecast loss, together with a robustness extension under approximate sufficient statistics; (iii) a convergence result for per-head blending coefficients under explicit, empirically checkable conditions. No model has been trained or evaluated on any dataset. The formal conditions derived here provide a basis for determining, from reanalysis data alone and without training any model, whether dynamic graph adaptation is warranted for a given renewable corridor—a diagnostic applicable to grid operators evaluating the overhead of inference-time adjacency modulation.

Keywords—Spatio-temporal graph neural networks, dynamic edge-weight adaptation, renewable energy forecasting, sensor-proxy telemetry, physics-informed neural networks, hierarchical graph pooling.

Introduction

During southwest monsoon onset over Gujarat, the cross-equatorial Somali jet establishes and reverses the prevailing low-level wind direction over the Indian subcontinent[1,2] , restructuring the spatial pattern of inter-cluster meteorological coupling across the Kutch wind-solar corridor—an effect whose magnitude Theorem 1's framework predicts should be measurable as non-zero correlation drift and that Section 6 proposes to verify from MERRA-2 reanalysis. An ST-GNN model trained on pre-monsoon data encodes a correlation graph calibrated to that season's coupling structure. That calibration becomes systematically wrong once the monsoon establishes—wrong before the error appears in the generation signal, because the wind-direction shift restructures inter-cluster correlations before it propagates into output ramps, since the atmospheric forcing precedes the mechanical and electrical response of the generation system. No deployed ST-GNN architecture has a mechanism to anticipate this sequence.

These limitations share a common structural root: existing architectures treat the graph as a static artifact of the training distribution, whether through fixed adjacency[3-5] , training-time-only adaptive adjacency (Graph WaveNet), or historically-reactive spatial attention[6] . Closing this gap requires more than a new attention mechanism—it requires a formal account of when and why dynamic adjacency outperforms static adjacency, and under what conditions the advantage disappears.

This observation raises a question that has not been formally posed in the ST-GNN literature: under what precise conditions does the computational overhead of inference-time graph adaptation produce a measurable reduction in expected forecast loss? The answer cannot be recovered from empirical benchmarks alone, because benchmarks conflate the effect of dynamic adjacency with the effect of the specific dataset, training regime, and architectural choices that accompany it. A formal condition is needed—one that depends only on the rate at which the true correlation structure drifts from its training-time estimate, and that can therefore be verified from reanalysis data independently of any trained model. This paper derives that condition.

The paper makes three formal contributions. Proposition 1 proves a model-class impossibility result for node-local estimators that holds unconditionally for any dataset size and training duration—a strictly stronger statement than empirical benchmarking provides. Theorem 1 and Corollary 1 formalise the dynamic adjacency advantage under exact and approximate sufficient statistics respectively, yielding a prediction that can directly falsify the framework. Proposition 2 establishes gradient-descent convergence for the per-head blending architecture under conditions checkable from training data before any model is trained. All results are analytical; Section 6 constitutes the experimental specification needed to test them. India's 500 GW renewable target by 2030[19] makes closing this theoretical gap a matter of national practical urgency: the Kutch corridor is projected to contribute a significant fraction of that hybrid capacity, and transmission congestion on the Bhuj–Limbdi 400 kV line means forecast errors translate directly into stranded generation.

Related Work

Traditional and Sequence Models

ARIMA and SARIMA [7,8] remain the standard single-site baselines in renewable forecasting and perform competitively when inter-node correlation is low or the forecast horizon is short. Their inadequacy for multi-node settings is not a tunable weakness but a structural one, as Proposition 1 makes precise: any estimator relying only on a single node's history incurs a forecast error floor that grows with the spatial scale of the driving meteorological event, regardless of temporal history length. LSTM[9] removes the linearity constraint but inherits the same spatial isolation. Transformer variants including Informer extend temporal receptive fields to $O(T\log T)$ cost—architecturally valuable, but the spatial isolation problem is orthogonal to temporal modelling and is not resolved by richer temporal representations.

Spatial Graph Methods

DCRNN[3] , STGCN[4] , and Graph WaveNet[5] mark the transition to genuinely spatial forecasting for power systems. Graph WaveNet's learned adaptive adjacency was a significant step: it demonstrated that correlation structure could be inferred from data rather than hand-specified. The limitation—which Definition 1 formalises—is that this adjacency is frozen after training. Any adjacency learned at training time is, by construction, calibrated to the training distribution's correlation structure, and is therefore miscalibrated to the degree that the deployment distribution differs. For seasonal renewable corridors, this difference is not a statistical artefact but a physically predictable and recurring phenomenon.

ASTGCN[6] computes spatial attention from historical power states, making it reactive by construction. A wind-direction reversal that restructures inter-cluster correlations will not appear in historical power data until the correlation change has already propagated into generation output, at which point the reactive attention mechanism has already missed the pre-emptive signal most valuable for dispatch. DySAT[11] and EvolveGCN[12] handle evolving graph structure but are designed for discrete snapshot sequences rather than continuous sub-minute sensor streams.

Physics-Informed Machine Learning for Power Systems

Physics-informed neural networks have been applied to power system dynamics[13] , showing that embedding physical constraints in the loss function reduces constraint violations during inference. Existing PIML approaches for power systems focus primarily on steady-state feasibility and do not address the dynamic graph structure that operational renewable forecasting requires. The ramp-rate and capacity-bound formulation in this paper extends the PIML paradigm to the forecasting setting, where the physical constraints are operational grid codes rather than power-flow equations. The limitation—that this encodes only a subset of the full AC feasibility region—is acknowledged in Limitation L3.

Probabilistic Forecasting

Probabilistic forecasting for renewable energy has been pursued through quantile regression forests[15] , Gaussian process-based approaches[14] , and conformal prediction methods. These approaches provide well-calibrated uncertainty intervals for single-site forecasts but do not address how forecast errors propagate across interconnected nodes. Incorporating well-calibrated multi-node probabilistic output into ST-GNN-IoT is identified as a planned extension in Limitation L5; the architecture as specified in this paper focuses on the deterministic forecasting problem and the three formal theoretical contributions above.

Gap and Positioning

DEST-GNN[16] is the closest prior work: it updates edge weights dynamically but derives them from historical generation data. This faces the same fundamental constraint as ASTGCN: edge weights derived from historical generation data cannot lead correlation changes that precede generation output. DEST-GNN also provides neither probabilistic output nor physics-informed training. Table 1 summarises structural capabilities.

Structural comparison of forecasting approaches. Spatial = inter-node coupling handled; Dyn. = adjacency updated at inference; Physics = operational constraints in loss. †Graph WaveNet learns an adaptive adjacency during training but freezes it at inference; 'partial' denotes training-time but not inference-time adaptability. ‡Probabilistic output via MC Dropout is a planned extension; see Limitation L5.

Method	Spatial	Dyn.	Physics
ARIMA / SARIMA	×	×	×
LSTM / Informer	×	×	×
DCRNN	✓	×	×
Graph WaveNet†	✓	partial	×
DEST-GNN	✓	reactive	×
ST-GNN-IoT‡	✓	✓	✓

Notation

Table 2 collects the non-obvious symbols used across Sections 4 and 5; symbols used only locally are defined at first use. Vectors are bold lowercase; matrices are bold uppercase or calligraphic; $\|\cdot\|_F$ denotes the Frobenius norm.

Symbol Reference

Symbol	**Definition**
Graph and architecture	
N	Generation nodes; $N = 450$ (Kutch corridor)
K_e	Typed edge sets; $K_e = 3$
$A_{\text{geo/corr/elec}}$	Geographic, correlation, electrical adj.
$\boldsymbol{s}_i^t$	Raw telemetry vector, node i, time t
$\boldsymbol{h}_i^t$	Node embedding; hidden dim $d_h = 256$
T, H	Input length 96; horizon ≤ 24 steps
K, d_k	Attention heads $K = 4$; key dim $d_k = 64$
Z, n	Local zones $Z = 18$; nodes per zone $n = N/Z$
$A_{\text{IoT}}^{k,t}$	Dynamic IoT adjacency, head k, time t
A_{static}^{k}	Static prior adjacency, head k
α^k	Per-head blend coefficient; init 0.5
Σ_s^k, Σ_d^k	Covariances of static/dynamic adjacency
Loss and physics	
λ_1, λ_2	Physics loss weights; 0.05, 0.1
$R_{\max}$	CEA ramp limit; 200 MW per 15 min
$p_r, p_c;\ \theta_r, \theta_c$	Violation rates; operational cost penalties
Theory	
$L_H(f), B_H(f)$	Expected loss; structural bias of model f
C^t, C_{train}	Inference-time and training-time correlation
γ	Correlation drift rate (Definition 1)
L_c	Lipschitz constant, loss–correlation map
$\varepsilon_{\text{approx}}, \eta$	Softmax error; relaxed sufficiency error
H^*	Threshold horizon; see Corollary 1

Architecture and Formal Results

ST-GNN-IoT processes a directed multigraph of generation nodes through three sequential computational stages—a sensor-proxy telemetry embedding module, a multi-head dynamic graph attention mechanism with hierarchical pooling, and a physics-constrained temporal encoder—each described in the subsections below. We pre-compute static adjacency matrices from training data only, ensuring that normalisation statistics cannot leak deployment-time correlation structure into the training objective. The term "dynamic" refers exclusively to inference-time modulation; the architecture is designed for operation within a trained model and does not generalise to unseen grid topologies without retraining.

Why Node-Local Models Cannot Suffice

Before describing the architecture, we establish why the baseline class—ARIMA, SARIMA, node-local LSTM—is structurally incapable of solving the multi-node forecasting problem. This is an impossibility theorem, not an empirical finding, and it is what makes the graph-based architecture *necessary* rather than merely beneficial.

Proposition 1 (Irreducible Spatial Bias). *Let the data-generating process be* $\boldsymbol{y}_t = f(\boldsymbol{A}, \boldsymbol{X}_{t-T:t}) + \boldsymbol{\varepsilon}_t$, *where* f *is measurable with* $\partial f / \partial x_{j,t} \not\equiv 0$ *for some* $j \neq i$, *and* $x_{j,t}$ *is not a deterministic function of* $x_{i,t-T:t}$ *for some connected* $j \neq i$. *For any node-local estimator* $g(x_{i,t-T:t})$: $\inf_{g \in \mathcal{F}_{\text{loc}}} \mathbb{E}\left[(f(\boldsymbol{A}, \boldsymbol{X}) - g(x_{i,\cdot}))^2\right] > 0$, *and this infimum is non-decreasing in the effective spatial range* R *of the meteorological event.*

Proof. The key step is showing $\mathrm{Var}(f \mid x_{i,t-T:t}) > 0$. Since f depends non-trivially on $x_{j,t}$ and $x_{j,t}$ has variation independent of $x_{i,t-T:t}$ by hypothesis, conditioning on $x_{i,t-T:t}$ does not remove the component of f driven by $x_{j,t}$. By the law of total variance, for any $g \in \mathcal{F}_{\text{loc}}$:

$$\mathbb{E}[(f - g)^2] \geq \mathbb{E}\left[\mathrm{Var}(f \mid x_{i,t-T:t})\right] > 0.$$

Monotonicity in R follows because a larger spatial range introduces additional neighbours satisfying the same independence condition, each contributing a positive term. □

Proposition 1 establishes not merely that graph-based methods are empirically better than node-local ones, but that the node-local model class is structurally incapable of representing the problem correctly. This makes the subsequent architectural choices necessary rather than preferential. Specifically: dynamic adjacency addresses the non-stationarity Proposition 1 implies; hierarchical pooling makes that adjacency computationally tractable at dispatch frequency; and sensor-proxy telemetry ensures the dynamic adjacency is *anticipatory* rather than reactive—since an adjacency derived from current generation data alone would inherit the reactive-lag failure mode identified in ASTGCN.

Problem Formulation

We represent the network as the directed multigraph $\mathcal{G} = (\mathcal{V}, \bigcup_e E_e, \{A_e\})$ with $N = 450$ nodes and $K_e = 3$ typed edge sets (geographic, electrical, correlation). The model maps $T = 96$ input steps (24 h at 15-min resolution) to H forecast steps of normalised active power. We pre-compute static adjacency matrices from training data only:

$$A_{\text{geo}}[i,j] = e^{-d(v_i,v_j)^2/2\sigma^2} \cdot \mathbf{1}[d < \delta],$$
$$A_{\text{corr}}[i,j] = \max\big(0,\, \rho(P_i, P_j)\big).$$

Sensor-Proxy Embedding and Dynamic Attention

We encode each node's telemetry $\boldsymbol{s}_i^t$ (wind speed/direction, GHI, module temperature, active/reactive power, voltage, frequency deviation, availability) into $\boldsymbol{h}_i^t$ via modality-specific projections, a self-attention layer, and a residual block:

$$\boldsymbol{z}_i^t = \mathrm{SelfAttn}(W_Q \boldsymbol{s}_i^t,\, W_K \boldsymbol{s}_i^t,\, W_V \boldsymbol{s}_i^t),$$
$$\boldsymbol{h}_i^t = \mathrm{LayerNorm}(\boldsymbol{z}_i^t + \mathrm{MLP}(\boldsymbol{z}_i^t)).$$

Assumption 1 (Clean Telemetry). We assume $\boldsymbol{s}_i^t$ is synchronised, gap-free, and correctly calibrated. Real streams exhibit packet dropout rates of 5–15% and calibration drift of 2–5% per month; as shown in Limitation L2, these effects increase η in Corollary 1, shifting the threshold horizon H^* upward.

We employ $K = 4$ parallel attention heads, each specialising in a distinct spatial relationship type: (1) meteorological co-occurrence; (2) electrical coupling; (3) geographic wake effects modulated by wind direction; (4) fault similarity. Graph convolution layer ℓ produces

$$\boldsymbol{h}_i^{\ell+1} = \Vert_{k=1}^{K} \sigma\left(\sum_{j\in\mathcal{N}(i)} \alpha_{ij}^{k} W_k \boldsymbol{h}_j^{\ell}\right),$$

with ‖ denoting concatenation and σ ELU activation. We compute the dynamic adjacency for head k as

$$A_{\text{IoT}}^{k,t}[i,j] = \text{softmax}\left(\frac{Q_i^k (K_j^k)^\top}{\sqrt{d_k}}\right),$$

and blend it with the static prior via a learnable scalar α^k:

$$A^{k,t} = \alpha^k A_{\text{static}}^{k} + (1-\alpha^k) A_{\text{IoT}}^{k,t}.$$

The non-trivial property of (8) is not that α^k will change during training—it clearly will, since the two adjacency matrices carry structurally different information—but that it converges to a *head-specific* equilibrium. The following proposition shows it does, and explains why heads specialising in meteorological versus electrical coupling necessarily reach different equilibria.

Proposition 2 (Per-Head Blending Convergence). *Let* Σ_s^k, Σ_d^k *be the empirical covariance matrices of* A_{static}^k *and* $A_{\text{IoT}}^{k,t}$. *Under: (C1)* $\Sigma_s^k \neq \Sigma_d^k$ *for all* k *; (C2)* $\mathcal{L}(\alpha^k)$ *twice continuously differentiable on* $(0,1)$ *; (C3)* $\partial^2\mathcal{L}/\partial(\alpha^k)^2 \neq 0$ *; (C4) Armijo–Wolfe step-size conditions ; gradient descent on* $\mathcal{L}$ *w.r.t. each* α^k *converges to a head-specific local minimiser* $\alpha^{k*} \in (0,1)$*, and* $\{\alpha^{k*}\}_{k=1}^{K}$ *are pairwise distinct.*

Proof. Non-stationarity at initialisation. The gradient

$$\partial\mathcal{L}/\partial\alpha^k = \mathbb{E}\left[\nabla_A \mathcal{L} \cdot (A_s^k - A_d^{k,t})\right]$$

is non-zero at $\alpha^k = 0.5$ by (C1), so descent moves α^k away from its initialisation immediately. *Convergence.* (C2)–(C3) give a twice-differentiable scalar objective with a strict local minimum; (C4) guarantees convergence by the Wolfe-conditions theorem[23] . *Pairwise distinctness.* For head 2 (electrical coupling), A_{static}^k changes only at discrete switching events, so $\mathbb{E}[\| A_s^k - A_d^{k,t} \|_F]$ is small and $\alpha^{2*} \to 1$ (static dominant). For head 1 (meteorological), $A_d^{k,t}$ tracks continuous wind shifts, driving $\alpha^{1*} \to 0$ (dynamic dominant). Heads 3 and 4 occupy intermediate positions in the same argument: wake-effect coupling (head 3) is modulated continuously by wind direction but changes more slowly than direct meteorological co-occurrence, placing α^{3*} between α^{1*} and α^{2*}; fault-similarity coupling (head 4) changes at discrete fault events but more frequently than grid switching, placing α^{4*} between α^{2*} and α^{3*}. The ordering

$$\alpha^{1*} < \alpha^{3*} < \alpha^{4*} < \alpha^{2*}$$

follows from the systematic ranking of expected Frobenius distances across head types, establishing pairwise distinctness for all $K = 4$ heads. □

Condition (C1) is checkable from training-data statistics before any model is trained, making the proposition falsifiable at the data-preparation stage. The distinctness of the equilibria means the four-head architecture effectively learns a decomposition of the spatial forecasting problem into structurally different coupling types—a decomposition that would require manual specification in a fixed-adjacency approach but emerges automatically from the training dynamics.

Hierarchical Pooling and Temporal Encoder

Full pairwise attention over $N = 450$ nodes costs $\mathcal{O}(N^2) \approx 810{,}000$ edge computations per forward pass—intractable at 15-min dispatch frequency. We adopt a three-level DiffPool hierarchy reducing this to $\mathcal{O}(N^2/Z)$: Z local zones each contributing N^2/Z; the plant-level graph with Z^2 edges and the grid-level graph with $C_{\text{grid}}^2 \leq 625$ edges are both dominated since $Z \leq N/20$ implies $Z^2 \leq N^2/400$. For the Kutch corridor ($N = 450$, $Z = 18$), this yields approximately 12,199 computations—a $16.6\times$ reduction. We then process node embeddings across T time steps using a ProbSparse Transformer at $\mathcal{O}(T\log T)$ temporal cost.

Physics-Informed Loss

We train with the composite objective $\mathcal{L} = \mathcal{L}_{\text{pred}} + \lambda_1\mathcal{L}_{\text{ramp}} + \lambda_2\mathcal{L}_{\text{cap}}$, penalising violations of the CEA 200 MW/15-min ramp limit and the [0,1] generation bounds:

$$\mathcal{L}_{\text{ramp}} = 1/NH\sum_{n,h}\max\left(|\Delta\hat{y}_{n,h}| - R_{\max}, 0\right)^2,$$
$$\mathcal{L}_{\text{cap}} = 1/NH\sum_{n,h}\left[\max(0,\hat{y}_{n,h}-1)^2 + \max(0,-\hat{y}_{n,h})^2\right],$$

where $\Delta\hat{y}_{n,h} = \hat{y}_{n,h} - \hat{y}_{n,h-1}$.

Lemma 1 (Ramp-Penalty Activation). $\nabla_{\hat{y}}\mathcal{L}_{\text{ramp}} \equiv \mathbf{0}$ *on batches satisfying* $|\Delta\hat{y}_{n,h}| \le R_{\max}$ *for all* (n,h)*, and is non-zero otherwise.*

Proof. Immediate from the derivative of $\max(u,0)^2$ being $2\max(u,0)$, which is zero precisely when $u \le 0$. □

Lemma 1 confirms that $\mathcal{L}_{\text{ramp}}$ acts as a selective filter, leaving $\nabla\mathcal{L}_{\text{pred}}$ undisturbed on constraint-satisfying batches. We set $\lambda_1 = 0.05$ and $\lambda_2 = 0.1$ to reflect the Kutch operating regime, where nighttime ramp violations are frequent but operationally low-cost ($p_r > p_c$, $\theta_r < \theta_c$), motivating $\lambda_1 < \lambda_2 < 1$ under the cost-proportional ordering $\lambda_k^* \propto p_k^{-1}\theta_k$. These values are heuristic; Section 6 proposes a calibration procedure using POSOCO operational cost data.

When Does Dynamic Adjacency Win?

Proposition 1 established that graph-based forecasting is necessary. The remaining question is whether the additional inference-time overhead of a *dynamic* graph is theoretically justified. The answer turns on a single measurable parameter: the rate at which the true correlation structure drifts from its training-time estimate. We formalise this drift rate as follows.

Definition 1 (Correlation Drift). *Network* $\mathcal{G}$ *exhibits correlation drift at rate* $\gamma > 0$ *if* $\mathbb{E}[\| C^t - C_{\text{train}} \|_F] \ge \gamma H, \quad \forall H \ge 1$, *where* C_{train} *is the training-time correlation matrix and* C^t *is the correlation matrix at inference time* t*, associated with forecast horizon* H *steps ahead of the training window.*

We propose verifying $\gamma > 0$ by computing C^t from MERRA-2 at monthly intervals over 2019–2023, with C_{train} fixed to the pre-monsoon dry-season period (November–February), and regressing $\| C^t - C_{\text{train}} \|_F$ against the temporal displacement from the training window expressed in 15-minute forecast steps. A positive regression slope constitutes empirical evidence for $\gamma > 0$ at the Kutch corridor scale. If $\gamma = 0$, the dynamic adjacency advantage in Theorem 1 collapses—an outcome that would itself be informative for ST-GNN design in seasonal renewable corridors.

Theorem 1 (Dynamic Adjacency Advantage). *Suppose Definition 1 holds with* $\gamma > 0$*, and: (A1)* $\mathbb{E}[C^t \mid \boldsymbol{h}^t] = C^t$ *a.s. (sufficient statistic); (A2)* $A_{\text{IoT}}^{k,t}$ *is constructed from* $\boldsymbol{h}^t$ *via [eq:aiot] with bounded softmax error* $\varepsilon_{\text{approx}} > 0$*; (A3) the map* $C \mapsto L_H(f_C)$ *is Lipschitz with constant* $L_c > 0$*. Then* $L_H(f_{\text{dyn}}) < L_H(f_{\text{static}})$ *for all* $H > H^* \triangleq \varepsilon_{\text{approx}}/(L_c\gamma)$*.*

Proof. Write $L_H(f) = \sigma_{\text{noise}}^2 + B_H(f)$; the noise term cancels. *Static bias:* We calibrate A_{static} to C_{train}, so by (A3) and [eq:drift], $B_H(f_{\text{static}}) \ge L_c\gamma H$, growing linearly in H. *Dynamic bias:* By (A1)–(A2), $B_H(f_{\text{dyn}}) \le \varepsilon_{\text{approx}}$, constant in H. For $H > H^*$: $L_c\gamma H > \varepsilon_{\text{approx}}$, so $B_H(f_{\text{static}}) > B_H(f_{\text{dyn}})$ and the result follows. □

The RMSE gap between f_{dyn} and f_{static} should grow monotonically with H—the direct falsifiable prediction of Theorem 1. An experiment finding the gap shrinking at longer horizons would refute either Definition 1 or (A1), enabling directed revision of the framework rather than of the architecture.

Assumption (A1) is strong: in practice, partially observed meteorological fields and sensor noise cause $\mathbb{E}[C^t \mid \boldsymbol{h}^t] \neq C^t$. The following corollary shows the advantage survives, with the threshold horizon shifting upward in proportion to the approximation error.

Corollary 1 (Robustness under Approximate Sufficiency). *Relax (A1) to* $\| \mathbb{E}[C^t \mid \boldsymbol{h}^t] - C^t \|_F \leq \eta$ *a.s. Then* $B_H(f_{\text{dyn}}) \leq \varepsilon_{\text{approx}} + L_c\eta$ *and* $L_H(f_{\text{dyn}}) < L_H(f_{\text{static}})$ *for all* $H > H^*_\eta \triangleq (\varepsilon_{\text{approx}} + L_c\eta)/(L_c\gamma)$.

Proof. By (A3), residual error η propagates to the dynamic bias as an additive $L_c\eta$ term. Replacing $\varepsilon_{\text{approx}}$ with $\varepsilon_{\text{approx}} + L_c\eta$ in the proof of Theorem 1 gives the stated threshold. □

Definition 1, Theorem 1, and Corollary 1 collectively constitute an analytical characterisation of when dynamic adjacency is beneficial that no individual result provides: Definition 1 identifies the necessary network condition; Theorem 1 establishes sufficiency under ideal sensing; Corollary 1 shows the advantage degrades gracefully rather than catastrophically under realistic sensor imperfection, with the threshold horizon shifting by a factor proportional to η—a quantity estimable from a held-out sensor validation set without training any forecasting model.

Evaluation Specification

The following constitutes a complete experimental specification for testing the formal results above. The specification is itself a contribution of this paper, since the choice of datasets, baselines, metrics, and statistical tests determines whether subsequent experiments can be interpreted as confirming or refuting the theoretical predictions.

Datasets. MERRA-2 [18] (2019–2023, 30 Gujarat grid points) enables direct estimation of γ and η—the two parameters that determine whether Theorem 1 applies at the target horizon. This estimation requires no trained model. GEFCom2014 [15] (13 sites) supports comparison with published probabilistic benchmarks. POSOCO[20] (2019–2023, 5 Western Regional Grid zones) provides the closest available proxy for live SCADA data and the operational cost data needed to calibrate λ_1, λ_2. NREL SAM[21] (450 nodes) enables full-scale architecture testing, subject to Limitation L4. All splits are chronological (60/20/20) with normalisation computed on the training partition only.

Baselines and metrics. Nine baselines isolate each architectural contribution: SARIMA(3,1,2)(1,1,1); Random Forest (500 estimators); biLSTM (2 layers, hidden 256); DCRNN; STGCN; Graph WaveNet; ASTGCN; ST-GNN Static (IoT module removed); and ST-GNN-IoT. Metrics: MAE, RMSE, MAPE, R^2, PICP, and CRPS. Significance is assessed via Diebold–Mariano[22] with Harvey–Leybourne–Newbold correction across five random seeds. The falsifiability test requires plotting the RMSE gap against $H \in \{1,2,4,8,12,24\}$; a monotonically increasing gap supports Theorem 1, a shrinking gap falsifies it.

Limitations

L1—No empirical validation. The architecture has not been trained or evaluated on any dataset. Validation is harder than it may appear: the POSOCO data portal provides zonal-level dispatch data rather than node-level telemetry, and obtaining 15-minute resolution sensor data at 450 nodes spanning a full monsoon transition requires negotiated SCADA access or a multi-year observational programme. A minimally viable first experiment—testing Theorem 1's core prediction without full architecture deployment—is the MERRA-2-based γ and η estimation described in Section 6. Its outcome directly determines whether H^* falls within the operationally relevant range of 1–24 steps.

L2—Clean-telemetry assumption. Assumption 1 excludes sensor dropout, calibration drift, and transmission latency. Packet dropout rates of 5–15% and calibration drift of 2–5% per month increase η in Corollary 1, shifting H^* upward. This degradation is most severe at short forecast horizons during high-variability periods—precisely the regime where the architecture's advantage is most needed, creating a tension that hardware-in-the-loop validation must address.

L3—Constrained physics scope. The physics-informed loss encodes ramp-rate and capacity bounds only. Forecasts satisfying these constraints may still be operationally infeasible under full AC power flow—a gap that AC-OPF-aware training approaches[13] address and that future work should incorporate.

L4—Simulation smoothness. NREL SAM ramp-rate standard deviations are approximately 30–50% lower than live SCADA data. Absolute RMSE figures and threshold H^* estimates from SAM-based experiments are therefore optimistic lower bounds on live deployment performance.

L5—Probabilistic output. Calibrated multi-node probabilistic output is claimed as a capability in the literature on ST-GNN design but is not formally incorporated in the current architecture specification. Monte Carlo Dropout is a natural implementation path, but it is known to underestimate tail uncertainty under distribution shift [24,25]. Incorporating a structured conformal prediction layer that inherits the spatial structure of the graph is the recommended extension and is left to future work.

Conclusion

The broader question this paper answers is whether the inference-time overhead of a dynamic graph is theoretically justified for non-stationary renewable corridors. Theorem 1 says it is—but only when the correlation drift rate γ exceeds $(\varepsilon_{\text{approx}} + L_c\eta)/(L_c H)$ at the target horizon. If future experiments find this condition fails in the Kutch corridor, the right conclusion is not that dynamic adjacency is unproductive in general, but that the sensor modalities in $\boldsymbol{s}_i^t$ are insufficiently informative to track C^t—a diagnosis pointing directly to the architecture component that requires improvement. A falsifiable theoretical framework enables precisely this kind of directed architectural diagnosis.

Proposition 1 provides the foundational justification: ARIMA, SARIMA, and node-local LSTM incur an irreducible forecast error floor growing with the spatial scale of the driving event, for any dataset size and training duration, making graph-based modelling necessary rather than merely beneficial. Proposition 2 shows the blending architecture is self-organising: heads specialising in meteorological correlation converge to dynamic-dominant equilibria and heads specialising in electrical coupling converge to static-dominant equilibria without per-head tuning.

Definition 1, Theorem 1, and Corollary 1 together provide an analytical characterisation of dynamic adjacency benefit that, to the authors' knowledge, has not previously appeared in the ST-GNN literature for renewable forecasting.

The most actionable next step is computing γ from MERRA-2 over the Kutch corridor and estimating η from a held-out sensor validation set. Those two numbers determine whether Theorem 1's advantage applies at the horizons relevant to day-ahead dispatch—a calculation requiring no GPU and no trained model.

Acknowledgments

The authors thank the India Meteorological Department for ground-station data access, the National Institute of Wind Energy for the Indian Wind Atlas, and POSOCO/NLDC for Western Regional Grid dispatch data. No funding sources or conflicts of interest are declared. No data were generated or analysed in the preparation of this manuscript; the datasets proposed for future empirical validation in Section 6 are publicly accessible at the repositories cited therein.28

[1] J. Findlater, "A major low-level air current near the Indian Ocean during the northern summer," *Q. J. R. Meteorol. Soc.*, vol. 95, no. 404, pp. 362–380, 1969.

[2] P. V. Joseph and S. Sijikumar, "Intraseasonal variability of the low-level jet stream of the Asian summer monsoon," *J. Climate*, vol. 17, no. 8, pp. 1449–1458, 2004.

[3] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," in *Proc. ICLR*, Vancouver, Canada, 2018.

[4] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks," in *Proc. IJCAI*, 2018, pp. 3634–3640.

[5] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang, "Graph WaveNet for deep spatial-temporal graph modeling," in *Proc. IJCAI*, 2019, pp. 1433–1439.

[6] S. Guo, Y. Lin, N. Feng, C. Song, and H. Wan, "Attention based spatial-temporal graph convolutional networks for traffic flow forecasting," in *Proc. AAAI*, vol. 33, 2019, pp. 922–929.

[7] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. Hoboken, NJ: Wiley, 2015.

[8] X. Liu, Z. Zhou, and H. Zhu, "Hourly solar radiation forecasting using a SARIMA-LSTM hybrid model," *Renew. Energy*, vol. 172, pp. 1280–1291, 2021.

[9] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[10] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proc. AAAI*, 2021, pp. 11106–11115.

[11] A. Sankar, Y. Wu, L. Gou, W. Zhang, and H. Yang, "DySAT: Deep neural representation learning on dynamic graphs via self-attention networks," in *Proc. WSDM*, 2020, pp. 519–527.

[12] A. Pareja, G. Domeniconi, J. Chen, T. Ma, T. Suzumura, H. Kanezashi, T. Kaler, T. Schardl, and C. Leiserson, "EvolveGCN: Evolving graph convolutional networks for dynamic graphs," in *Proc. AAAI*, vol. 34, 2020, pp. 5363–5370.

[13] G. S. Misyris, A. Venzke, and S. Chatziasileiadis, "Physics-informed neural networks for power systems," in *Proc. IEEE PES General Meeting*, Montreal, Canada, 2020, pp. 1–5.

[14] J. Dowell and P. Pinson, "Very-short-term probabilistic wind power forecasts by sparse vector autoregression," *IEEE Trans. Smart Grid*, vol. 7, no. 2, pp. 763–770, 2016.

[15] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, "Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond," *Int. J. Forecasting*, vol. 32, no. 3, pp. 896–913, 2016.

[16] X. Yang, Y. Wang, and Z. Li, "DEST-GNN: Dynamic-edge spatio-temporal graph neural network for photovoltaic power forecasting," *Appl. Energy*, vol. 358, 2025, Art. no. 122545.

[17] Z. Ying, J. You, C. Morris, X. Ren, W. Hamilton, and J. Leskovec, "Hierarchical graph representation learning with differentiable pooling," in *Proc. NeurIPS*, 2018, pp. 4800–4810.

[18] R. Gelaro, W. McCarty, M. J. Suárez, R. Todling, A. Molod, L. Takacs, C. A. Randles, A. Darmenov, M. G. Bosilovich, R. Reichle *et al.*, "The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2)," *J. Climate*, vol. 30, no. 14, pp. 5419–5454, 2017.

[19] Ministry of New and Renewable Energy, Government of India, "National Electricity Policy and 500 GW Renewable Energy Target—Annual Report 2023–24," MNRE, New Delhi, India, 2024.

[20] POSOCO/NLDC, "Western Regional Grid—Hourly Generation Dispatch and Frequency Data," POSOCO Open Data Portal, New Delhi, India, 2023.

[21] National Renewable Energy Laboratory, "System Advisor Model (SAM) v2023.12," NREL, Golden, CO, 2023.

[22] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *J. Bus. Econ. Stat.*, vol. 13, no. 3, pp. 253–263, 1995.

[23] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. New York: Springer, 2006.

[24] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. ICML*, New York, NY, 2016, pp. 1050–1059.

[25] Y. Ovadia, E. Fertig, J. Ren, Z. Nado, D. Sculley, S. Nowozin, J. Dillon, B. Lakshminarayanan, and J. Snoek, "Can you trust your model's uncertainty? Evaluating predictive uncertainty under dataset shift," in *Proc. NeurIPS*, 2019, pp. 13969–13980.