SPECBRIDGE: FUTURE-GUIDED SPECTRAL ALIGNMENT FOR POWER OUTAGE FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Reliable power outage forecasting is critical for grid resilience, emergency response, and resource allocation, particularly as outage events can impose substantial operational and economic consequences. However, outage dynamics vary substantially across time and geographic regions, making forecasting highly challenging. Existing time-series forecasting methods primarily learn a direct mapping from historical observations to future values, but often overlook how structural relationships encoded in the past evolve toward future temporal patterns, potentially compromising forecasting accuracy as outage patterns change. We introduce SpecBridge, a future-guided spectral alignment framework for power outage forecasting. SpecBridge constructs latent graphs from numerical and textual representations of historical observations, together with a future-time graph that serves as a reference during training. It aligns the source spectral bases to this reference and adaptively aggregates them using data-driven weights that account for both source-target structural discrepancy and graph density. The aggregated historical subspace and the future-time subspace jointly supervise the forecasting representation, complementing prediction and reconstruction objectives. Empirical evaluations across 11 state-level US power outage datasets demonstrate that our SpecBridge outperforms state-of-the-art baselines.

1 INTRODUCTION

Power outages impose substantial economic and human costs. The August 2003 Northeast blackout left approximately 50 million people without electricity and caused an estimated $5–12 billion in lost economic activity. An epidemiological analysis estimated approximately 90 excess deaths in New York City during the blackout. More recently, Winter Storm Uri in February 2021 left over 4.5 million people in Texas without power, with some outages lasting up to four days. Forecasting how outage levels evolve can help utilities anticipate restoration needs and allocate resources before disruptions intensify (Arab et al., 2015; Gao et al., 2017; Hou et al., 2022). Historical outage records provide a basis for this prediction, but their temporal patterns vary across events and regions. A period of low outage activity may precede a sudden disruption, while similar outage levels can be followed by different recovery trajectories. Learning from these records therefore requires identifying which aspects of the observed history remain informative about the future.

Statistical models such as the autoregressive integrated moving average (ARIMA) capture linear temporal dependencies but suffer from many limitations, e.g., some strong assumptions with respect to the noise terms and loss functions. Deep learning methods, including recurrent neural networks (RNNs) (Medsker & Jain, 1999) and Transformers (Li et al., 2019; Nie et al., 2023; He et al., 2024), learn richer representations of outage histories. Through mechanisms such as temporal patching, autocorrelation, and frequency decomposition, these models capture local fluctuations and dependencies over longer periods. However, they primarily model temporal dependencies or align latent features, leaving graph-based relationships among temporal instances underexplored as a source of supervision. For outage forecasting, these relationships describe which periods exhibit similar fluctuations or recovery patterns, and how such similarities differ between the observed and forecast windows. Using this structure raises two related questions.

How can future relational structure guide learning from outage histories? Outage forecasting requires anticipating how a disruption develops and how recovery unfolds. Similar fluctuations in

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

the observed history may precede different future trajectories, so relationships observed in the past need not remain informative throughout the forecast window. A loss on predicted outage values supervises numerical accuracy but does not explicitly supervise the relationships among temporal representations. We therefore consider an additional source of guidance, i.e., the similarity structure among temporal instances in the future window. Representing these similarities as graphs provides a basis for comparing historical and future relational patterns, while spectral subspaces summarize their connectivity for alignment.

Which historical view should the alignment trust? Different views of an outage history can emphasize competing patterns. A numerical representation may distinguish brief fluctuations that a textual summary compresses into an overall trend; conversely, the summary may emphasize a sustained disruption that is less apparent among local variations. Neither view is necessarily more informative across all forecasting examples. Combining them with fixed weights can dilute useful information when their relationships disagree. Selecting the view closest to the future graph also requires care, because this comparison relies on estimated spectral subspaces whose accuracy can depend on graph density. The resulting challenge is selective transfer: determining how much each historical view should contribute by considering both its agreement with the future target and the uncertainty in its spectral estimate.

Inspired by these questions, we introduce SpecBridge, a graph spectral alignment framework for power outage forecasting. Our central idea is to use future relational structure as supervision for learning which relationships in historical observations to retain. We construct two source graphs from numerical and textual representations of the observed history, together with a target graph from the future window during training. Each graph expresses similarities among corresponding temporal instances, and its spectral subspace provides a low-dimensional representation of these relationships. This construction turns past–future alignment into a comparison of relational structures across views.

Our contributions are summarized as follows: (1) **Graph-based structural supervision for outage forecasting**. We introduce SpecBridge, which constructs latent graphs from numerical and textual views of historical observations and uses future graph structure to supervise the forecasting representation; (2) **Adaptive spectral transfer with theoretical analysis**. We develop a source aggregation and alignment mechanism that accounts for structural discrepancy and graph sparsity. Under stated assumptions, we further characterize the aggregation error and the trade-off between target estimation error and source heterogeneity in the spectral alignment module; (3) **Extensive experiments validate this framework's alignment with state-of-the-art models**. We evaluate power outage forecasting across regions and prediction horizons which covers 11 US state-level outage datasets, 3 forecasting horizons, and comparisons with 11 baselines.

2 Related Work

Power outage prediction. Research on power outage prediction has garnered significant attention in recent years (Guikema et al., 2014; He et al., 2017; Casey et al., 2020; Taylor et al., 2023). Jaech et al. (2018) develops neural predictors of unplanned power outage duration by using environmental factors for initial estimates and natural language processing to update predictions as field reports arrive. Chen et al. (2023) proposes a deep neural architecture that leverages both the utility of a simplicial convolutional layer and a trainable topological layer for detecting outages in distribution networks. Wang et al. (2026) combines cascading failure modeling with a probability-constrained neural network to predict spatio-temporal outage risks. Transformer-based methods provide another approach to modeling temporal dependencies in power systems (L'Heureux et al., 2022; Pentsos et al., 2025). For example, Informer (Zhou et al., 2021) introduces sparse attention for efficient long-sequence modeling. FEDformer (Zhou et al., 2022) incorporates frequency-domain representations into a decomposition-based architecture. Furthermore, PatchTST (Nie et al., 2023) groups observations into patches to capture local temporal patterns. iTransformer (Liu et al., 2024) treats individual variables as tokens to model multivariate dependencies. Time-VLM (Zhong et al., 2025) integrates visual and contextual text representations of time series with numerical representations learned by a Transformer. TimeAlign (Hu et al., 2026) aligns predictive representations with those learned by an auxiliary future-reconstruction branch through global and local objectives. However, these approaches primarily model temporal dependencies or align latent features, leaving the relational

structure captured by graph spectral subspaces underexplored for aligning historical observations with future targets.

Graph spectral learning. Spectral methods represent graph structure through low-dimensional eigenspaces, providing a foundation for clustering and relational representation learning (Smola & Kondor, 2003; Von Luxburg, 2007; Liang et al., 2020; Berahmand et al., 2025). The importance of the graph Laplacian has been well recognized as a tool for embedding, manifold learning, spectral sparsification, clustering and semi-supervised learning (Belkin et al., 2006; Belkin & Niyogi, 2001; Spielman & Teng, 2011). Classical spectral clustering uses low-dimensional eigenspaces to reveal groups of closely connected nodes (Ng et al., 2001), while spectral graph neural networks (GNNs) use graph filters to learn representations from signals defined on nodes. These filters broadly follow two designs. Fixed filters prescribe how information propagates over the graph; for example, APPNP (Gasteiger et al., 2019) uses Personalized PageRank to combine information from different propagation depths, and learnable filters adapt their spectral responses to the task. GPR-GNN learns coefficients that weight successive propagation steps (Chien et al., 2021), and SGL (Kumar et al., 2019) introduces a unified graph learning framework, lying at the integration of the spectral properties of the Laplacian matrix with Gaussian graphical modeling. Ouyang et al. (2026) proposes a spectral aggregation strategy can control global features through frequency components, making the exploration-exploitation control both interpretable and effective. More recently, Guo et al. (2026) studies spectral transfer for community detection under local differential privacy.

3 METHODOLOGY

Notation and Problem Statement. Let $\boldsymbol{x}_\ell \in \mathbb{R}^C$ denote the power outage observations across C channels at time step ℓ. Given a look-back window of length L, we write the historical input and its corresponding future target as $\boldsymbol{X}_p = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L] \in \mathbb{R}^{C\times L}$, $\boldsymbol{X}_f = [\boldsymbol{x}_{L+1}, \ldots, \boldsymbol{x}_{L+T}] \in \mathbb{R}^{C\times T}$, where T is the forecasting horizon. For notational convenience, we write $\boldsymbol{X} = \boldsymbol{X}_p$ and $\boldsymbol{Y} = \boldsymbol{X}_f$ in the following. Power outage forecasting aims to learn a mapping $f_\theta : \mathbb{R}^{C\times L} \to \mathbb{R}^{C\times T}$ that predicts the next T observations from the available history, i.e., $\widehat{\boldsymbol{Y}}_{\text{pred}} = f_\theta(\boldsymbol{X}_p)$, where θ denotes the learnable model parameters. The overall framework is illustrated in Figure 1.

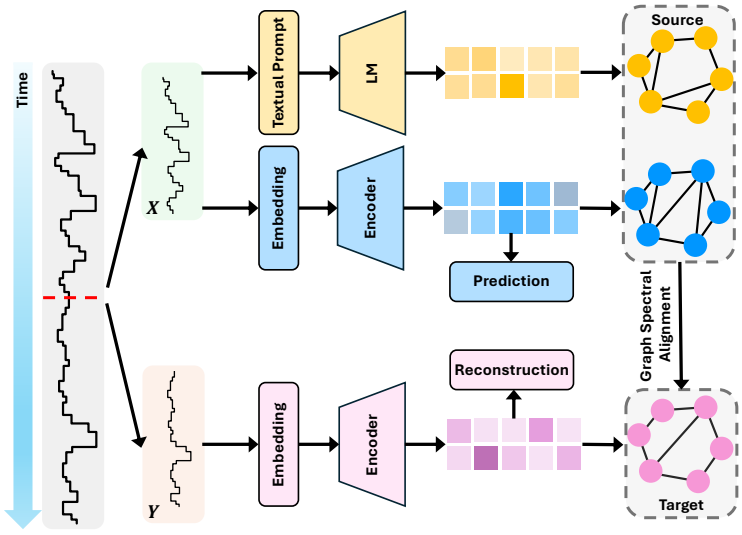

Figure 1: Overview of the SpecBridge framework for power outage prediction.

3.1 PAST AND FUTURE INPUT REPRESENTATION LEARNING

Historical outage records trace the onset, persistence, and recovery of disruptions. Forecasting requires translating this evolving record into representations that connect past observations with future trajectories. We use a lightweight residual attention encoder for this purpose, keeping the backbone

deliberately simple to examine the contribution of structural alignment without introducing substantial architectural complexity. The alignment mechanism acts on latent representations, allowing other forecasting backbones to be used when they provide compatible features. Given a multivariate history $\boldsymbol{X}_p \in \mathbb{R}^{C\times L}$, where C is the number of channels and L is the look-back length, we divide each channel into N temporal patches. Each patch becomes a local fragment of the observed history, projected into a D-dimensional feature space:

$$\boldsymbol{Z}_p = \text{Linear}\left(\text{Patch}\left(\text{LayerNorm}(\boldsymbol{X})\right)\right), \tag{1}$$

where $\text{LayerNorm}(\cdot)$ normalizes the input and $\text{Linear}(\cdot)$ is a learnable projection. The resulting $\boldsymbol{Z}_p \in \mathbb{R}^{(C\cdot N)\times D}$ collects the patch embeddings across all channels. We refine these embeddings through B residual encoder blocks, each combining multi-head self-attention (MSA) with a position-wise feed-forward network. Self-attention allows each patch to draw context from the other patches, connecting local observations through learned dependencies. For head m in block l, we compute:

$$\boldsymbol{Q}_m^{(l)} = \boldsymbol{H}_x^{(l-1)}\boldsymbol{W}_m^{Q,(l)},\ \boldsymbol{K}_m^{(l)} = \boldsymbol{H}_x^{(l-1)}\boldsymbol{W}_m^{K,(l)},\ \boldsymbol{V}_m^{(l)} = \boldsymbol{H}_x^{(l-1)}\boldsymbol{W}_m^{V,(l)}, \tag{2}$$

where $\boldsymbol{H}_x^{(0)} = \boldsymbol{Z}_p$ and the three $\boldsymbol{W}_m$ matrices are learnable projections. The attention output is:

$$\text{head}_m^{(l)} = \text{Softmax}\left(\boldsymbol{Q}_m^{(l)}\left(\boldsymbol{K}_m^{(l)}\right)^\top / \sqrt{d_h}\right)\boldsymbol{V}_m^{(l)}, \tag{3}$$

where d_h is the dimension of each attention head. The M heads provide distinct attention patterns, which are concatenated and projected:

$$\text{MSA}^{(l)}(\boldsymbol{H}_x^{(l-1)}) = \oplus\left(\text{head}_1^{(l)}, \ldots, \text{head}_M^{(l)}\right)\boldsymbol{W}^{O,(l)}, \tag{4}$$

where $\oplus$ denotes concatenation along the feature dimension and $\boldsymbol{W}^{O,(l)}$ is a learnable output projection. After the final block, we regroup the contextualized patch embeddings by channel. To predict the next T time steps, we reshape $\boldsymbol{H}_x^{(B)}$ into C channel-wise representations, each formed by concatenating its N patch embeddings, i.e., $\widetilde{\boldsymbol{H}}_x^{(B)} = \text{Reshape}(\boldsymbol{H}_x^{(B)})$. Then a linear readout maps each representation directly to the forecast horizon:

$$\widehat{\boldsymbol{Y}}_{\text{pred}} = \text{Linear}(\widetilde{\boldsymbol{H}}_x^{(B)}) \in \mathbb{R}^{C\times T}. \tag{5}$$

Similarly, during training, we apply the same patch embedding, encoding, and readout procedure in Eqs. 1–5 to the future sequence $\boldsymbol{X}_f$ to obtain its reconstruction $\widehat{\boldsymbol{Y}}_{\text{recon}}$. We supervise this branch using the mean absolute reconstruction error, $\mathcal{L}_{\text{recon}} = \frac{1}{CT}\sum_{c=1}^{C}\sum_{t=1}^{T}|[\hat{\boldsymbol{Y}}_{\text{recon}}]_{c,t} - [\boldsymbol{Y}]_{c,t}|$. This objective encourages the future-window representation to retain the observed outage dynamics, providing a reference for subsequent graph spectral alignment.

3.2 Latent Graph Construction

Power outage forecasting requires recognizing both how individual outage patterns evolve and how relationships among these patterns change over time. Historical periods may exhibit similar fluctuations or recovery profiles, yet these similarities need not persist into the future. Learned embeddings can encode such dependencies, but leave them implicit in feature vectors, making it difficult to directly compare the relational structure of past and future observations. We therefore construct latent graphs that express embedding similarities as explicit connections. These graphs expose which instances form neighborhoods in each representation space, providing a structural basis for aligning historical patterns with future outage behavior. Specifically, we construct two complementary graphs from the historical input, i.e., the first uses learned time-series embeddings to represent similarities in numerical outage patterns; the second uses language-model representations of statistical summaries and domain context, providing a textual perspective on the same observations. During training, we also construct a graph from the learned embeddings of the future target.

Numerical representations. For each training pair $(\boldsymbol{X}_p, \boldsymbol{X}_f)$, we encode these two windows as $\boldsymbol{H}_p$ and $\boldsymbol{H}_f$, and construct a latent graph from the similarities among node embeddings within each window. The past-time graph describes relationships in the observed history; the future-time graph reveals their counterpart in the forecast window. During training, the future graph serves as a reference for learning which historical relationships to emphasize in structural alignment. At inference, this reference is no longer available: the model forecasts from the observed history alone.

Textual representations. Following Zhong et al. (2025), we construct a textual view of the historical outage sequence by summarizing its statistical properties and domain context. We first generate a description $\mathcal{T}_p$ and then encode it with a pretrained language model:

$$\boldsymbol{H}_t = \mathrm{MLP}\left(\mathrm{LM}(\mathcal{T}_p)\right), \tag{6}$$

where $\mathcal{T}_p$ specifies the attributes of $\boldsymbol{X}_p$ to summarize, including the minimum, maximum, median, overall trend, available periodicity information, input window length, and relevant domain context, $\mathrm{LM}(\cdot)$ denotes the pretrained language encoder (here we use the frozen text encoder of a pretrained CLIP model (i.e., ViT-B/32) (Radford et al., 2021)), and $\mathrm{MLP}(\cdot)$ is a lightweight neural network. The resulting $\boldsymbol{H}_t$ is used to construct the historical text graph.

We represent each view over the same N temporal instances with consistent node ordering. For each source view $r \in \mathcal{S} = \{p, t\}$, let $\boldsymbol{H}_r = [\boldsymbol{h}_1^r, \ldots, \boldsymbol{h}_N^r]^\top$. Its graph $\mathcal{G}_r = (\boldsymbol{A}_r, \boldsymbol{H}_r)$ connects nodes according to their similarity in the corresponding latent space. Specifically, for nonzero node embeddings, we compute:

$$s_{uv}^r = (\boldsymbol{h}_u^r)^\top \boldsymbol{h}_v^r \big/ \left(\|\boldsymbol{h}_u^r\|_2 \|\boldsymbol{h}_v^r\|_2\right). \tag{7}$$

Two distinct nodes are connected when their similarity exceeds a view-specific threshold τ_r:

$$[\boldsymbol{A}_r]_{uv} = \mathbb{I}[u \neq v]\, \mathbb{I}[s_{uv}^r > \tau_r], \tag{8}$$

where $\mathbb{I}[\cdot]$ denotes the indicator function. This produces an undirected graph without self-loops. We also apply the same construction to $\boldsymbol{H}_f$, using threshold τ_f, to obtain the future-time graph $\mathcal{G}_f = (\boldsymbol{A}_f, \boldsymbol{H}_f)$. The three graphs share node correspondence but retain distinct connectivity: each expresses which temporal instances are similar under its own representation.

3.3 Graph Spectral Alignment

Past observations and their textual descriptions capture different relationships among the same nodes. Their contributions to forecasting should therefore depend on how well these relationships reflect future temporal structure. We use the future-time graph as a reference to adaptively weight the two source graphs and combine their spectral representations for transfer alignment. Let $\boldsymbol{A}_p$ and $\boldsymbol{A}_t$ denote the graphs constructed from past time-series observations and their textual descriptions, respectively, and let $\boldsymbol{A}_f$ denote the future-time graph (i.e., target graph). We define the source index set as $\mathcal{S} = \{p, t\}$. All three graphs share the same N nodes and node ordering. For each graph ($r \in \mathcal{S}$), we extract the K orthonormal eigenvectors selected by a fixed spectral rule (e.g., the K largest algebraic eigenvalues) from $\boldsymbol{A}_r$ to form $\boldsymbol{U}_r \in \mathbb{R}^{N \times K}$, with $\boldsymbol{U}_r^\top \boldsymbol{U}_r = \boldsymbol{I}_K$. Similarly, we obtain $\boldsymbol{U}_f \in \mathbb{R}^{N \times K}$ using the same spectral rule on the target graph $\boldsymbol{A}_f$, with $\boldsymbol{U}_f^\top \boldsymbol{U}_f = \boldsymbol{I}_K$. These spectral subspaces provide a common representation for comparing and combining the structures encoded by the three graphs.

Adaptive weighting. Directly averaging eigenvectors is ambiguous because different orthonormal bases can represent the same subspace. We first align each source basis to the future-time basis:

$$\boldsymbol{Z}_r = \underset{\boldsymbol{Z}^\top \boldsymbol{Z} = \boldsymbol{I}_K}{\arg\min} \|\boldsymbol{U}_r \boldsymbol{Z} - \boldsymbol{U}_f\|_F^2, \qquad r \in \mathcal{S}. \tag{9}$$

This orthogonal Procrustes problem has the solution $\boldsymbol{Z}_r = \boldsymbol{B}_r \boldsymbol{C}_r^\top$, where $\boldsymbol{U}_r^\top \boldsymbol{U}_f = \boldsymbol{B}_r \boldsymbol{\Sigma}_r \boldsymbol{C}_r^\top$ is a singular value decomposition. Here, $\boldsymbol{Z}_r \in \mathbb{R}^{K \times K}$ is an orthogonal transformation that aligns the source basis with the future-time basis while preserving the source subspace.

We measure structural discrepancy using the corresponding projection matrices:

$$\widehat{h}_r = \left\|\boldsymbol{U}_r \boldsymbol{U}_r^\top - \boldsymbol{U}_f \boldsymbol{U}_f^\top\right\|_2, \qquad \widehat{\rho}_r = \frac{\sum_{u \neq v} [\boldsymbol{A}_r]_{uv}}{N(N-1)}, \tag{10}$$

where $\|\cdot\|_2$ is the spectral norm and $\widehat{\rho}_r$ is the graph density. Unlike a distance between eigenvector bases, $\widehat{h}_r$ is invariant to orthogonal transformations.

We assign each source a weight based on its structural agreement with the future-time graph and its graph density. To robustly integrate heterogeneous temporal and textual graph sources, we develop an adaptive weighting mechanism tailored to SpecBridge that jointly accounts for source-target structural discrepancy and graph sparsity, building on the transfer-learning principles of Guo

et al. (2026), we define:

$$a_r = \left(\widehat{h}_r^2 + c \cdot \frac{\varsigma_N}{N \widetilde{\rho}_r} \right)^{-1}, \qquad w_r = \frac{a_r}{\sum_{s \in \mathcal{S}} a_s}, \tag{11}$$

where $\varsigma_N = \max\{1, (\log N)/|\mathcal{S}|\}$, $|\mathcal{S}| = 2$, $\widetilde{\rho}_r = \max\{\widehat{\rho}_r, \epsilon\}$ for a small density floor $\epsilon = \epsilon_N > 0$, and $c > 0$ is a constant. For a fixed graph density, a smaller source–target discrepancy yields a larger weight. The density-dependent term reduces the contribution of sparse graphs and prevents the weight from being determined solely by structural agreement. The normalized weights satisfy $w_r \geq 0$ and $\sum_{r \in \mathcal{S}} w_r = 1$. We then combine the aligned source bases and orthonormalize the result:

$$\widetilde{\boldsymbol{U}} = \sum_{r \in \mathcal{S}} w_r \boldsymbol{U}_r \boldsymbol{Z}_r, \qquad \overline{\boldsymbol{U}} = \mathrm{qf}(\widetilde{\boldsymbol{U}}), \tag{12}$$

where $\mathrm{qf}(\cdot)$ returns the orthonormal factor of a thin QR decomposition. The resulting $\overline{\boldsymbol{U}}$ represents the combined past-time and text structure, with each source contributing according to its target relevance and density.

Transfer alignment. We align the spectral representation of the forecasting features with the future-time subspace and the aggregated source subspace. To maintain correspondence with the graph nodes, we rearrange the final encoder output $\boldsymbol{H}_x^{(B)}$ into $\boldsymbol{F}_x \in \mathbb{R}^{N \times (CD)}$ (whose i-th row concatenates the features of temporal patch i across all channels). Next, we define $\boldsymbol{V} = \mathrm{Eig}_K\left(\boldsymbol{F}_x \boldsymbol{F}_x^\top\right)$ where $\mathrm{Eig}_K(\cdot)$ returns the orthonormal eigenvectors associated with the K largest eigenvalues, so that $\boldsymbol{V}^\top \boldsymbol{V} = \boldsymbol{I}_K$. We then construct the alignment loss as:

$$\mathcal{L}_{\text{spec}} = K - \mathrm{tr}\left(\boldsymbol{V}^\top \boldsymbol{U}_f \boldsymbol{U}_f^\top \boldsymbol{V}\right) + \frac{\lambda}{2} \left\| \boldsymbol{V}\boldsymbol{V}^\top - \overline{\boldsymbol{U}}\,\overline{\boldsymbol{U}}^\top \right\|_F^2, \tag{13}$$

where K is a hyperparameter that specifies the dimensionality of the spectral subspace (i.e., the number of spectral components), $\lambda \geq 0$ controls the contribution of the aggregated source structure. Thus, the loss encourages the learned subspace to agree with future temporal structure while retaining the adaptively weighted information from past observations and text. Finally, the overall loss function is defined as the sum of three components, i.e., $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{pred}} + \eta_{\text{recon}} \mathcal{L}_{\text{recon}} + \eta_{\text{spec}} \mathcal{L}_{\text{spec}}$ where $\mathcal{L}_{\text{pred}}$ is the mean absolute error (MAE), and η_{recon} and η_{spec} are hyperparameters.

Theoretical guarantees. In this section, we study how source heterogeneity and graph sparsity affect spectral alignment. We first bound the error of weighted source aggregation and show that the adaptive weights attain, up to constants, an error bound governed by informative sources without requiring their identities in advance. We then characterize how regularized alignment balances target estimation error against source heterogeneity. The corresponding assumptions and proofs are provided in Appendix A.

Proposition 1 (Weighted aggregation). Under Assumptions 1– 3 in Appendix A.1, consider any $\boldsymbol{w} \in \Delta_{\mathcal{S}}$. On $\mathcal{E}_N$,

$$\mathrm{dist}\big(\overline{\boldsymbol{U}}(\boldsymbol{w}), \boldsymbol{U}_f^\star\big) \leq C \left\{ \left(\sum_{r \in \mathcal{S}} w_r^2 \frac{\log N}{N \rho_r} \right)^{1/2} + \sum_{r \in \mathcal{S}} w_r h_r \right\}. \tag{14}$$

Furthermore, the weights in Eq. (11) uniquely minimize $\sum_r w_r^2 \widehat{\theta}_r$ over the probability simplex. Consequently, on $\mathcal{E}_N$, they are stabilized plug-in versions of the inverse-risk weights $w_r^\star \propto \theta_r^{-1}$.

The first term of Proposition 1 is the sampling error after source aggregation, while its second term is the weighted population heterogeneity.

Define the baseline scale and informative-source set by

$$\eta_N := \frac{\varsigma_N}{N\rho}, \qquad \mathcal{I} := \left\{ r \in \mathcal{S} : h_r^2 \leq c_I \eta_N \right\}, \qquad m := |\mathcal{I}|, \tag{15}$$

where $c_I > 0$ is fixed. This set is used only for analysis and is not needed by the algorithm. When $\mathcal{I}^c \neq \varnothing$, let

$$G_N := \min_{r \in \mathcal{I}^c} \frac{h_r^2}{\eta_N};$$

set $G_N = +\infty$ when $\mathcal{I}^c = \varnothing$. When $\mathcal{I}^c = \varnothing$, all bad-source sums and all remainders involving $(J-m)/G_N$ or $(J-m)/\sqrt{G_N}$ are defined to be zero. Finally, for $m > 0$, write

$$M_N := \left\{ \frac{1}{m^2} \sum_{r \in \mathcal{I}} \frac{\log N}{N\rho_r} \right\}^{1/2}, \qquad H_N := \frac{1}{m} \sum_{r \in \mathcal{I}} h_r. \tag{16}$$

Theorem 1 (Error-bound-oracle property)**.** *Suppose Assumptions 1–3 in Appendix A.1 hold, $m > 0$, and G_N is bounded below by a sufficiently large constant. Assume further that*

$$\frac{J-m}{m\sqrt{G_N}} \sqrt{\eta_N} = O(M_N + H_N).$$

Then, on $\mathcal{E}_N$, the aggregate $\overline{\boldsymbol{U}}$ constructed using the adaptive weights in Eq. (11) *satisfies*

$$\operatorname{dist}\big(\overline{\boldsymbol{U}}, \boldsymbol{U}_f^\star\big) \lesssim \sqrt{\frac{\log N}{mN\rho}} + \frac{1}{m} \sum_{r \in \mathcal{I}} h_r. \tag{17}$$

Thus, up to constants, the error bound depends only on the informative sources, even though $\mathcal{I}$ is unknown.

For the regularized estimator, let $\widehat{\boldsymbol{U}}(\lambda)$ be a global minimizer of Eq. (13). Since all projectors have rank K, the objective can be written as

$$\mathcal{L}_{\text{spec}} = K(1+\lambda) - \operatorname{tr}\Big[\boldsymbol{V}^\top \Big(\widehat{\boldsymbol{\Pi}}_f + \lambda \overline{\boldsymbol{U}}\overline{\boldsymbol{U}}^\top \Big) \boldsymbol{V} \Big].$$

Hence $\widehat{\boldsymbol{U}}(\lambda)$ spans the leading K-dimensional eigenspace of $\widehat{\boldsymbol{\Pi}}_f + \lambda \overline{\boldsymbol{U}}\overline{\boldsymbol{U}}^\top$. Equivalently, with $\gamma = \lambda/(1+\lambda)$, it uses the convex combination $(1-\gamma)\widehat{\boldsymbol{\Pi}}_f + \gamma \overline{\boldsymbol{U}}\overline{\boldsymbol{U}}^\top$.

Theorem 2 (Regularized transfer)**.** *Suppose the conditions of Theorem 1 hold and $t_f + M_N + H_N = o(1)$. For $H_N > 0$, set*

$$\lambda_{\text{bal}} := \frac{t_f}{H_N}, \qquad \gamma_{\text{bal}} := \frac{\lambda_{\text{bal}}}{1+\lambda_{\text{bal}}} = \frac{t_f}{t_f + H_N};$$

when $H_N = 0$, interpret $\lambda_{\text{bal}} = +\infty$ and $\gamma_{\text{bal}} = 1$, with $\widehat{\boldsymbol{U}}(+\infty)$ defined to span the same subspace as $\overline{\boldsymbol{U}}$. Then, for all sufficiently large N, on $\mathcal{E}_N$,

$$\begin{aligned} \operatorname{dist}\Big(\widehat{\boldsymbol{U}}(\lambda_{\text{bal}}), \boldsymbol{U}_f^\star\Big) &\le C\left\{\gamma_{\text{bal}} M_N + (t_f \wedge H_N)\right\} \\ &\le C\left\{M_N + (t_f \wedge H_N)\right\}. \end{aligned} \tag{18}$$

In particular, under sharp target concentration,

$$\operatorname{dist}\Big(\widehat{\boldsymbol{U}}(\lambda_{\text{bal}}), \boldsymbol{U}_f^\star\Big) \lesssim \sqrt{\frac{\log N}{mN\rho}} + \left\{ \frac{1}{\sqrt{N\rho_f}} \wedge \frac{1}{m} \sum_{r \in \mathcal{I}} h_r \right\}.$$

The source-only endpoint $\lambda = +\infty$ has upper bound $M_N + H_N$, whereas the target-only endpoint $\lambda = 0$ has upper bound t_f. A safeguarded oracle that selects between 0 and λ_{bal} according to the smaller certified bound is therefore no worse in upper-bound order than either endpoint. With the Bernstein choice $t_f \asymp \sqrt{(\log N)/(N\rho_f)}$ and comparable graph densities, $M_N = O(t_f)$, so the unsafeguarded balanced choice is target-rate competitive. Under the sharper $t_f \asymp (N\rho_f)^{-1/2}$ and a fixed number of sources, only the safeguarded choice is guaranteed by these bounds to be target-rate competitive uniformly over H_N.

Remark 1 (Scope of the guarantees). The results above concern only the spectral transfer module and do not by themselves establish end-to-end forecasting generalization. Assumption 1 is a stylized sufficient model: thresholded graphs built from jointly learned embeddings may have dependent edges and dependent layers. For the actual construction, Eq. (21) must instead be verified through stability, sample splitting, or an appropriate dependence argument. Theorem 2 applies to an exact global subspace minimizer of Eq. (13); it does not cover optimization error in the learned forecasting network.

4 EXPERIMENTS

Datasets. We use the EAGLE-I records[1] of customer-level power outages for the 2025 calendar year, covering the 11 US states. Each series is sampled every 15 minutes from 2025-01-01 00:00 to 2025-12-31 23:45, giving $35{,}040$ time steps per state with no missing values. We split each chronologically into training, validation, and test portions in a $7:1:2$ ratio (i.e., $24{,}528$ / $3{,}504$ / $7{,}008$ steps). The look-back window is $L = 720$ steps (i.e., 7.5 days) and the horizons are $T \in \{96, 192, 336\}$ steps (i.e., 1, 2 and 3.5 days). Outage counts are heavy-tailed rather than approximately Gaussian, with a pooled mean of $4{,}696$ customers against a maximum of $3{,}494{,}570$. We therefore apply $\log(1+x)$ followed by a RobustScaler whose statistics are estimated on the training portion only.

Baselines and evaluation metrics. We evaluate our method against 11 state-of-the-art baselines, including: (1) PatchTST (Nie et al., 2023); (2) iTransformer (Liu et al., 2024); (3) TimesNet (Wu et al., 2023); (4) Autoformer (Wu et al., 2021); (5) FEDformer (Zhou et al., 2022); (6) Non-Stationary Transformer (Liu et al., 2022); (7) Informer (Zhou et al., 2021); (8) Reformer (Kitaev et al., 2020); (9) TimeMixer (Wang et al., 2024); (10) Time-VLM (Zhong et al., 2025); and (11) TimeAlign (Hu et al., 2026). To guarantee the robustness of the evaluation, we run each method three times with different random seeds for every prediction horizon. Performance is measured using Mean Absolute Error (MAE) and Mean Squared Error (MSE). **Bold** and underline denote the best and second metrics on each dataset. All baselines use their official implementations and the settings their authors recommend for long-term forecasting, without additional per-dataset tuning. For our model, the search is conducted over the patch count $N \in \{8, 12\}$ with spectral dimension $K < N$ (here K is set to 4), the feature width $D \in \{128, 256, 512, 768\}$ and depth $B \in \{1, \ldots, 4\}$, the learning rate in $\{5{\times}10^{-5}, 10^{-4}, 3{\times}10^{-4}, 5{\times}10^{-4}\}$, dropout in $\{0.1, \ldots, 0.8\}$, $\lambda \in \{0, 0.03, 0.1, 0.3, 1, 3\}$, the loss weights $\eta_{\text{recon}}, \eta_{\text{spec}} \in \{0, 0.1, 0.3, 1, 3, 10\}$, and the graph thresholds $\tau_p \in \{0, 0.25, 0.5, 0.7, 0.9\}$, $\tau_t \in \{0.5, 0.7, 0.9, 0.95, 0.99\}$, and $\tau_f \in \{0.4, 0.6, 0.7, 0.8, 0.9\}$. The look-back length, batch size, and number of attention heads are set to 720, 16, and 8 respectively. Training uses AdamW for at most 10 epochs with early stopping (i.e., patience 3). Codes are available at `https://anonymous.4open.science/r/specbridge-10DF`.

Performance comparison. As presented in Table 1, our SpecBridge achieves competitive performance compared to state-of-the-art baselines. SpecBridge achieves the best horizon-averaged MSE and MAE on all 11 datasets. Across individual horizons, it ranks first in 29 of 33 settings for MSE and all 33 settings for MAE, showing consistent performance across both regions and forecasting horizons. In particular, the improvements are pronounced on TX and MI, where SpecBridge outperforms the runner-up, TimeMixer, by 9.88-11.34% in MSE and 7.66-7.96% in MAE. On PA and WA, it surpasses Time-VLM by 9.04-9.43% in MSE and 4.70-5.87% in MAE, indicating that the benefits extend beyond incorporating textual information alone. SpecBridge also improves upon iTransformer on LA by 6.75% in MSE and 5.21% in MAE, and exceeds Reformer on GA by 8.19% and 4.35%, respectively. On CA, it achieves gains of 5.36% in MSE and 3.15% in MAE over TimeMixer. In addition, our SpecBridge remains competitive as the forecasting horizon increases. For instance, at $h = 336$, it achieves the lowest MSE on 8 out of 11 datasets and the lowest MAE on all 11 datasets. On GA, it improves upon iTransformer by 8.27% in MSE and 5.07% in MAE; on MI, it outperforms TimesNet by 6.71% in MSE and Time-VLM by 6.86% in MAE. The improvements are not uniform: SpecBridge slightly trails iTransformer in MSE on NC at $h = 192$ and $h = 336$, and Reformer and Informer on NY and OH, respectively, at $h = 336$. Nevertheless, it maintains the lowest MAE in each of these settings and the best horizon-averaged performance on these 3 datasets (i.e., NC, NY, and OH). We also report the standard deviation of SpecBridge performance under 3 runs with different random seeds in Table 2, which exhibits that the performance of SpecBridge is stable. In addition, we provide computational efficiency comparison in Table 5 and hyperparameter sensitivity analysis in Table 4, Appendix B (for more details).

Ablation study. Table 3 evaluates the contributions of key components of SpecBridge, including reconstruction objective and graph spectral alignments. Specifically, we compare SpecBridge with two variants: (1) *w/o graph spectral*: removes graph spectral alignment; and (2) *w/o reconstruction*: removes the reconstruction objective. We can observe that all ablated variants underperform the full model, highlighting the importance of each component. In particular, removing graph spectral alignment increases MSE and MAE by an average of 4.59% and 2.69%, respectively, with increases

[1] https://www.osti.gov/biblio/3012826

Methods		SpecBridge (**Ours**)		TimeAlign		TimeMixer		iTransformer		Time-VLM		Reformer		Informer		Autoformer		FEDformer		TimesNet		PatchTST		Stationary	
Metric		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
TX	96	**0.3703**	**0.4538**	0.5175	0.5536	0.4272	0.5006	0.4340	0.5093	0.4180	0.4910	0.4945	0.5275	0.5151	0.5443	0.6045	0.6097	0.5308	0.5608	0.4410	0.5139	0.4188	0.4978	0.5889	0.5728
	192	**0.4534**	**0.5013**	0.6673	0.6321	0.4901	0.5339	0.5290	0.5617	0.5273	0.5504	0.5418	0.5685	0.6366	0.6265	0.6669	0.6347	0.5815	0.5841	0.5205	0.5500	0.6338	0.6207	0.7000	0.6425
	336	**0.5120**	**0.5325**	0.7308	0.6703	0.5504	0.5672	0.5543	0.5722	0.5517	0.5753	0.5474	0.5551	0.8411	0.7301	0.8082	0.7004	0.5979	0.5982	0.5931	0.5940	0.8582	0.7143	0.6867	0.6320
	Avg	**0.4452**	**0.4959**	0.6385	0.6187	0.4892	0.5339	0.5058	0.5478	0.4990	0.5389	0.5279	0.5504	0.6642	0.6336	0.6932	0.6483	0.5701	0.5811	0.5182	0.5526	0.6369	0.6109	0.6585	0.6157
NY	96	**0.6253**	**0.5972**	0.9962	0.7742	0.7624	0.6704	0.7450	0.6706	0.6728	0.6279	0.8901	0.7073	0.8999	0.7202	0.9607	0.7646	0.8423	0.7090	0.7381	0.6663	0.7369	0.6535	0.7693	0.6723
	192	**0.7790**	**0.6697**	1.1439	0.8342	0.9470	0.7463	0.8405	0.7087	0.8747	0.7207	1.0125	0.7664	0.8996	0.7294	1.0303	0.7892	0.9040	0.7304	0.8716	0.7296	0.9771	0.7564	1.0046	0.7712
	336	0.8371	**0.6998**	1.1522	0.8311	1.0650	0.8008	0.8625	0.7222	0.9240	0.7441	**0.8247**	0.7000	0.9769	0.7668	0.9984	0.7737	0.8674	0.7237	0.9435	0.7622	0.9651	0.7564	1.1193	0.8139
	Avg	**0.7471**	**0.6556**	1.0974	0.8132	0.9248	0.7392	0.8160	0.7005	0.8238	0.6976	0.9091	0.7245	0.9255	0.7388	0.9965	0.7758	0.8712	0.7210	0.8511	0.7194	0.8930	0.7221	0.9644	0.7525
FL	96	**0.4291**	**0.5014**	0.5292	0.5645	0.4773	0.5281	0.4813	0.5381	0.4549	0.5191	0.6084	0.6300	0.6075	0.6216	0.5961	0.6001	0.5227	0.5598	0.4649	0.5235	0.4819	0.5377	0.5885	0.6045
	192	**0.4795**	**0.5361**	0.6022	0.5969	0.5024	0.5437	0.5017	0.5520	0.5044	0.5516	0.6538	0.6572	0.7150	0.6851	0.5989	0.6020	0.5378	0.5691	0.5124	0.5526	0.5404	0.5721	0.6852	0.6560
	336	**0.4924**	**0.5414**	0.6706	0.6372	0.5433	0.5625	0.5244	0.5632	0.5409	0.5692	0.5928	0.6197	0.7968	0.7184	0.5613	0.5834	0.5412	0.5696	0.5597	0.5806	0.5929	0.6036	0.6741	0.6462
	Avg	**0.4670**	**0.5263**	0.6007	0.5995	0.5077	0.5448	0.5024	0.5511	0.5001	0.5466	0.6183	0.6356	0.7064	0.6751	0.5854	0.5952	0.5339	0.5661	0.5123	0.5522	0.5384	0.5711	0.6493	0.6356
NC	96	**0.4465**	**0.5150**	0.6095	0.5985	0.5031	0.5445	0.4869	0.5464	0.4982	0.5462	0.4759	0.5355	0.5613	0.5786	0.5636	0.5879	0.5195	0.5665	0.5148	0.5579	0.5778	0.5886	0.5643	0.5825
	192	0.4989	**0.5494**	0.6434	0.6219	0.5931	0.6010	**0.4969**	0.5524	0.5498	0.5747	0.5161	0.5633	0.6160	0.6072	0.5713	0.5959	0.5249	0.5716	0.5285	0.5674	0.5955	0.6011	0.6002	0.6059
	336	0.5134	**0.5573**	0.6575	0.6329	0.6060	0.6078	**0.5043**	0.5583	0.5749	0.5908	0.6501	0.6364	0.5585	0.5821	0.5722	0.5981	0.5363	0.5779	0.5688	0.5917	0.7067	0.6597	0.6805	0.6469
	Avg	**0.4863**	**0.5406**	0.6368	0.6178	0.5674	0.5844	0.4960	0.5524	0.5410	0.5706	0.5474	0.5784	0.5786	0.5893	0.5690	0.5940	0.5269	0.5720	0.5374	0.5723	0.6267	0.6165	0.6150	0.6118
GA	96	**0.4668**	**0.5312**	0.6328	0.6206	0.4989	0.5523	0.5207	0.5689	0.5074	0.5584	0.5154	0.5553	0.5211	0.5728	0.5962	0.6182	0.5611	0.5990	0.5106	0.5608	0.5428	0.5778	0.5560	0.5765
	192	**0.5179**	**0.5703**	0.6983	0.6644	0.5616	0.5904	0.5504	0.5895	0.5474	0.5857	0.5367	0.5821	0.6269	0.6382	0.6272	0.6328	0.5842	0.6110	0.5582	0.5913	0.5853	0.6063	0.8230	0.6954
	336	**0.5103**	**0.5677**	0.7610	0.7013	0.5683	0.6015	0.5525	0.5965	0.5802	0.6115	0.5651	0.6045	0.6892	0.6728	0.6214	0.6357	0.5842	0.6146	0.5719	0.6044	0.6740	0.6614	0.6863	0.6425
	Avg	**0.4983**	**0.5564**	0.6974	0.6621	0.5429	0.5814	0.5412	0.5850	0.5450	0.5852	0.5391	0.5806	0.6124	0.6279	0.6150	0.6289	0.5765	0.6082	0.5469	0.5855	0.6007	0.6152	0.6884	0.6381
LA	96	**0.5084**	**0.5421**	0.7303	0.6609	0.6518	0.6144	0.5653	0.5857	0.5852	0.5843	0.5936	0.6043	0.6315	0.6164	0.5933	0.6091	0.5761	0.5945	0.6176	0.6025	0.6552	0.6274	0.6068	0.5990
	192	**0.5635**	**0.5748**	0.8312	0.7111	0.6815	0.6310	0.6052	0.6046	0.6556	0.6277	0.5890	0.6007	0.7629	0.6914	0.6499	0.6385	0.6093	0.6127	0.6887	0.6346	0.7225	0.6596	0.7615	0.6680
	336	**0.5807**	**0.5869**	0.8459	0.7190	0.6845	0.6392	0.5937	0.6022	0.6840	0.6439	0.6216	0.6218	0.8080	0.7024	0.7356	0.6822	0.5981	0.6071	0.7205	0.6574	0.7658	0.6844	0.7809	0.6800
	Avg	**0.5509**	**0.5679**	0.8025	0.6970	0.6726	0.6282	0.5881	0.5975	0.6416	0.6186	0.6014	0.6090	0.7341	0.6700	0.6596	0.6433	0.5945	0.6047	0.6756	0.6315	0.7145	0.6571	0.7164	0.6490
CA	96	**0.5331**	**0.5362**	0.6476	0.6036	0.5852	0.5622	0.6281	0.5927	0.6083	0.5739	0.6542	0.5838	0.6419	0.5943	0.7144	0.6470	0.6964	0.6255	0.5994	0.5736	0.6388	0.5957	0.6069	0.5791
	192	**0.6376**	**0.5890**	0.7948	0.6628	0.6760	0.6103	0.7081	0.6314	0.7100	0.6234	0.7190	0.6146	0.7495	0.6436	0.7757	0.6735	0.7323	0.6389	0.6636	0.6057	0.8763	0.7001	0.6960	0.6225
	336	**0.7090**	**0.6255**	0.8939	0.7055	0.7195	0.6336	0.7509	0.6498	0.7671	0.6518	0.7778	0.6467	0.8572	0.6967	0.7727	0.6561	0.7581	0.6497	0.7486	0.6526	0.8749	0.7104	0.7942	0.6554
	Avg	**0.6266**	**0.5836**	0.7787	0.6573	0.6602	0.6020	0.6957	0.6247	0.6951	0.6164	0.7170	0.6150	0.7495	0.6449	0.7543	0.6589	0.7289	0.6381	0.6705	0.6106	0.7967	0.6688	0.6990	0.6190
MI	96	**0.3325**	**0.4383**	0.4887	0.5361	0.3802	0.4711	0.4027	0.5032	0.3753	0.4760	0.4590	0.5128	0.4642	0.5372	0.5104	0.5693	0.5054	0.5616	0.3980	0.4892	0.4436	0.5172	0.4794	0.5401
	192	**0.4332**	**0.4986**	0.6608	0.6258	0.4655	0.5291	0.4951	0.5513	0.5209	0.5557	0.6867	0.5978	0.5275	0.5612	0.7342	0.6753	0.5444	0.5734	0.4889	0.5408	0.6553	0.6167	0.5507	0.5771
	336	**0.5170**	**0.5410**	0.7345	0.6632	0.5826	0.5953	0.5658	0.5848	0.5601	0.5781	0.6675	0.6020	0.5550	0.5799	0.7361	0.6764	0.5560	0.5855	0.5517	0.5859	0.6634	0.6283	0.7427	0.6606
	Avg	**0.4276**	**0.4926**	0.6280	0.6084	0.4761	0.5318	0.4879	0.5465	0.4855	0.5366	0.6044	0.5708	0.5156	0.5594	0.6602	0.6403	0.5352	0.5735	0.4795	0.5386	0.5875	0.5874	0.5909	0.5926
OH	96	**0.4853**	**0.5448**	0.6986	0.6516	0.5411	0.5726	0.5621	0.5914	0.5151	0.5631	0.5366	0.5735	0.4952	0.5580	0.5962	0.6139	0.5602	0.5912	0.5280	0.5715	0.5294	0.5700	0.5926	0.6021
	192	**0.5555**	**0.5840**	0.7954	0.7032	0.6134	0.6137	0.5877	0.6034	0.5798	0.5975	0.5695	0.5981	0.5623	0.5991	0.6741	0.6473	0.5797	0.5968	0.6047	0.6077	0.7008	0.6539	0.6983	0.6557
	336	0.5541	**0.5865**	0.7964	0.7058	0.6721	0.6445	0.5934	0.6069	0.6002	0.6130	0.6781	0.6550	**0.5445**	0.5947	0.6085	0.6133	0.5643	0.5966	0.6482	0.6349	0.7206	0.6663	0.6973	0.6544
	Avg	**0.5316**	**0.5718**	0.7635	0.6869	0.6088	0.6102	0.5811	0.6006	0.5650	0.5912	0.5947	0.6088	0.5340	0.5839	0.6263	0.6248	0.5681	0.5948	0.5936	0.6047	0.6503	0.6301	0.6627	0.6374
PA	96	**0.4761**	**0.5113**	0.7479	0.6551	0.5957	0.5782	0.5497	0.5648	0.5108	0.5398	0.6869	0.6238	0.5834	0.5755	0.7819	0.7007	0.6699	0.6446	0.5545	0.5672	0.5392	0.5555	0.5713	0.5710
	192	**0.5990**	**0.5809**	0.9726	0.7582	0.7221	0.6402	0.7024	0.6525	0.6650	0.6224	0.8881	0.7101	0.6791	0.6263	0.8162	0.7178	0.7096	0.6583	0.6836	0.6331	0.7109	0.6426	1.1165	0.7747
	336	**0.6458**	**0.6146**	1.1357	0.8222	0.8136	0.6910	0.7331	0.6683	0.7073	0.6448	0.9562	0.7445	0.7192	0.6569	0.8182	0.7171	0.6912	0.6513	0.8191	0.7021	0.8162	0.6932	1.1061	0.7803
	Avg	**0.5736**	**0.5689**	0.9521	0.7451	0.7105	0.6365	0.6617	0.6285	0.6277	0.6023	0.8437	0.6928	0.6606	0.6196	0.8055	0.7119	0.6902	0.6514	0.6857	0.6341	0.6888	0.6304	0.9313	0.7087
WA	96	**0.6488**	**0.6216**	0.9472	0.7430	0.8019	0.6927	0.8284	0.7215	0.7045	0.6506	1.0089	0.7626	1.0733	0.8063	0.9216	0.7620	0.9581	0.7758	0.7313	0.6676	0.7665	0.6874	0.6942	0.6388
	192	**0.8184**	**0.7048**	1.1319	0.8069	0.9080	0.7378	0.9267	0.7642	0.8929	0.7385	1.0326	0.7772	1.2437	0.8636	1.0102	0.7966	1.0409	0.8121	0.8956	0.7369	0.9360	0.7581	1.0486	0.7818
	336	**0.9655**	**0.7685**	1.3158	0.8894	1.0220	0.7874	1.0647	0.8211	1.0554	0.8041	1.1546	0.8202	1.1950	0.8463	1.1494	0.8517	1.1403	0.8519	1.0926	0.8187	1.0465	0.8061	1.0304	0.7863
	Avg	**0.8109**	**0.6983**	1.1316	0.8131	0.9106	0.7393	0.9399	0.7689	0.8842	0.7311	1.0654	0.7867	1.1707	0.8387	1.0271	0.8034	1.0465	0.8133	0.9065	0.7411	0.9164	0.7505	0.9244	0.7356

Table 1: Full long-term forecasting results. 'Avg' denotes the average across 3 prediction horizons $\{96, 192, 336\}$.

Dataset	**TX**	**NY**	**FL**	**NC**	**GA**	**LA**	**CA**	**MI**	**OH**	**PA**	**WA**
MSE	0.4452 ± 0.0076	0.7471 ± 0.0016	0.4670 ± 0.0081	0.4863 ± 0.0050	0.4983 ± 0.0065	0.5509 ± 0.0004	0.6266 ± 0.0036	0.4276 ± 0.0050	0.5316 ± 0.0018	0.5736 ± 0.0055	0.8109 ± 0.0047
MAE	0.4959 ± 0.0045	0.6556 ± 0.0011	0.5263 ± 0.0042	0.5406 ± 0.0023	0.5564 ± 0.0031	0.5679 ± 0.0005	0.5836 ± 0.0009	0.4926 ± 0.0020	0.5718 ± 0.0010	0.5689 ± 0.0013	0.6983 ± 0.0016

Table 2: Robustness of SpecBridge performance. The results are obtained from 3 random seeds, averaged over 3 prediction lengths $\{96, 192, 336\}$.

of 8.73% and 4.58% on WA. These results demonstrate that graph spectral alignment provides effective structural guidance for forecasting by transferring informative relationships from source graphs to the target representation. Removing the reconstruction objective also consistently degrades performance, leading to average increases of 1.29% in MSE and 0.65% in MAE.

Model	**TX**		**FL**		**NC**		**GA**		**CA**		**MI**		**OH**		**PA**		**WA**	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
SpecBridge	**0.4452**	**0.4959**	**0.4670**	**0.5263**	**0.4863**	**0.5406**	**0.4983**	**0.5564**	**0.6266**	**0.5836**	**0.4276**	**0.4926**	**0.5316**	**0.5718**	**0.5736**	**0.5689**	**0.8109**	**0.6983**
W/o graph spectral	0.4528	0.5067	0.4863	0.5372	0.5095	0.5539	0.5229	0.5721	0.6623	0.5990	0.4461	0.5090	0.5425	0.5783	0.6019	0.5859	0.8817	0.7303
W/o reconstruction	0.4624	0.5047	0.4697	0.5273	0.4883	0.5418	0.5068	0.5600	0.6334	0.5877	0.4282	0.4946	0.5393	0.5762	0.5797	0.5713	0.8218	0.7033

Table 3: Ablation studies, averaged over prediction lengths {96, 192, 336}.

5 Conclusion & Limitations

We presented SpecBridge, a framework that uses graph spectral alignment to transfer relational information from outage histories to future prediction. SpecBridge combines numerical and textual views through adaptive weights that account for their spectral agreement with the future target and graph density. Future observations provide structural supervision during training, while inference uses only historical inputs. Extensive experiments on 11 US state-level datasets with different forecasting horizons demonstrate the effectiveness of SpecBridge. **Limitations and Future Works.** The performance of SpecBridge depends on the quality of the time-series data and the representations derived from it. Current implementation constructs multiple latent graphs and computes their spectral representations for alignment, which can increase resource consumption during training and hyperparameter tuning. For future exploration, we will investigate transfer to unseen regions and incorporate weather and infrastructure information to better characterize outage development and recovery.

REFERENCES

Ali Arab, Amin Khodaei, Suresh K Khator, Kevin Ding, Valentine A Emesih, and Zhu Han. Stochastic pre-hurricane restoration planning for electric power systems infrastructure. *IEEE Transactions on Smart Grid*, 6(2):1046–1054, 2015.

Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. *Advances in neural information processing systems*, 14, 2001.

Mikhail Belkin, Partha Niyogi, and Vikas Sindhwani. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *Journal of machine learning research*, 7 (11), 2006.

Kamal Berahmand, Farid Saberi-Movahed, Razieh Sheikhpour, Yuefeng Li, and Mahdi Jalili. A comprehensive survey on spectral clustering with graph structure learning. *arXiv preprint arXiv:2501.13597*, 2025.

Joan A Casey, Mihoka Fukurai, Diana Hernández, Satchit Balsari, and Mathew V Kiang. Power outages and community health: a narrative review. *Current environmental health reports*, 7(4): 371–383, 2020.

Yuzhou Chen, Roshni Anna Jacob, Yulia R Gel, Jie Zhang, and H Vincent Poor. Learning power grid outages with higher-order topological neural networks. *IEEE Transactions on Power Systems*, 39 (1):720–732, 2023.

Eli Chien, Jianhao Peng, Pan Li, and Olgica Milenkovic. Adaptive universal generalized pagerank graph neural network. In *International Conference on Learning Representations*, 2021.

Haixiang Gao, Ying Chen, Shengwei Mei, Shaowei Huang, and Yin Xu. Resilience-oriented pre-hurricane resource allocation in distribution systems considering electric buses. *Proceedings of the IEEE*, 105(7):1214–1233, 2017.

Johannes Gasteiger, Aleksandar Bojchevski, and Stephan Günnemann. Combining neural networks with personalized pagerank for classification on graphs. In *International Conference on Learning Representations*, 2019.

Seth David Guikema, Roshanak Nateghi, Steven M Quiring, Andrea Staid, Allison C Reilly, and Michael Gao. Predicting hurricane power outages to support storm response planning. *Ieee Access*, 2:1364–1373, 2014.

Xiao Guo, Xuming He, Xiangyu Chang, and Shujie Ma. Privacy-preserving transfer learning for community detection using locally distributed multiple networks. *Journal of the American Statistical Association*, pp. 1–67, 2026. URL https://arxiv.org/abs/2504.00890.

Jichao He, David W Wanik, Brian M Hartman, Emmanouil N Anagnostou, Marina Astitha, and Maria EB Frediani. Nonparametric tree-based predictive modeling of storm outages on an electric distribution network. *Risk Analysis*, 37(3):441–458, 2017.

Jindong He, Shanshan Lei, and Lijin Wu. Multi-source data fusion for power outage warning based on transformer model. In *Journal of Physics: Conference Series*, volume 2816, pp. 012030, 2024.

Hui Hou, Zhiwei Zhang, Ruizeng Wei, Yong Huang, Yongchao Liang, and Xianqiang Li. Review of failure risk and outage prediction in power system under wind hazards. *Electric Power Systems Research*, 210:108098, 2022.

Yifan Hu, Jie Yang, Tian Zhou, Peiyuan Liu, Yujin Tang, Rong Jin, and Liang Sun. Bridging past and future: Distribution-aware alignment for time series forecasting. In *The Fourteenth International Conference on Learning Representations*, 2026.

Aaron Jaech, Baosen Zhang, Mari Ostendorf, and Daniel S Kirschen. Real-time prediction of the duration of distribution system outages. *IEEE Transactions on Power Systems*, 34(1):773–781, 2018.

Nikita Kitaev, Lukasz Kaiser, and Anselm Levskaya. Reformer: The efficient transformer. In *ICLR*, 2020.

Sandeep Kumar, Jiaxi Ying, José Vinícius de Miranda Cardoso, and Daniel Palomar. Structured graph learning via laplacian spectral constraints. *Advances in neural information processing systems*, 32, 2019.

Shiyang Li, Xiaoyong Jin, Yao Xuan, Xiyou Zhou, Wenhu Chen, Yu-Xiang Wang, and Xifeng Yan. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. *Advances in neural information processing systems*, 32, 2019.

Weixuan Liang, Sihang Zhou, Jian Xiong, Xinwang Liu, Siwei Wang, En Zhu, Zhiping Cai, and Xin Xu. Multi-view spectral clustering with high-order optimal neighborhood laplacian matrix. *IEEE Transactions on Knowledge and Data Engineering*, 34(7):3418–3430, 2020.

Yong Liu, Haixu Wu, Jianmin Wang, and Mingsheng Long. Non-stationary transformers: Exploring the stationarity in time series forecasting. In *NeurIPS*, volume 35, pp. 9881–9893, 2022.

Yong Liu, Tengge Hu, Haoran Zhang, Haixu Wu, Shiyu Wang, Lintao Ma, and Mingsheng Long. itransformer: Inverted transformers are effective for time series forecasting. In *The Twelfth International Conference on Learning Representations*, 2024.

Alexandra L'Heureux, Katarina Grolinger, and Miriam AM Capretz. Transformer-based model for electrical load forecasting. *Energies*, 15(14):4993, 2022.

Larry Medsker and Lakhmi C Jain. *Recurrent neural networks: design and applications*. CRC press, 1999.

Andrew Ng, Michael Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 14, 2001.

Yuqi Nie, Nam H Nguyen, Phanwadee Sinthong, and Jayant Kalagnanam. A time series is worth 64 words: Long-term forecasting with transformers. In *The Eleventh International Conference on Learning Representations*, 2023.

Kaichen Ouyang, Zong Ke, Shengwei Fu, Lingjie Liu, Puning Zhao, and Dayu Hu. Learn from global correlations: Enhancing evolutionary algorithm via spectral gnn. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 24665–24673, 2026.

Vasileios Pentsos, Spyros Tragoudas, Jason Wibbenmeyer, and Nasser Khdeer. A hybrid lstm-transformer model for power load forecasting. *IEEE Transactions on Smart Grid*, 16(3):2624–2634, 2025.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, pp. 8748–8763, 2021.

Alexander J Smola and Risi Kondor. Kernels and regularization on graphs. In *Learning theory and kernel machines: 16th annual conference on learning theory and 7th kernel workshop, COLT/kernel 2003, Washington, DC, USA, august 24-27, 2003. Proceedings*, pp. 144–158. Springer, 2003.

Daniel A Spielman and Shang-Hua Teng. Spectral sparsification of graphs. *SIAM Journal on Computing*, 40(4):981–1025, 2011.

William O Taylor, Diego Cerrai, David Wanik, Marika Koukoula, and Emmanouil N Anagnostou. Community power outage prediction modeling for the eastern united states. *Energy Reports*, 10: 4148–4169, 2023.

Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.

Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma, James Y. Zhang, and Jun Zhou. Timemixer: Decomposable multiscale mixing for time series forecasting. In *ICLR*, 2024.

Yuxuan Wang, Bin Zhou, Yan Xu, Chi Yung Chung, Yun Yang, Zhao Yuan, and Chunchao Hu. Spatio-temporal power outage risk prediction for interdependent urban electricity and drainage networks under rainstorm disasters. *IEEE Transactions on Smart Grid*, 2026.

Haixu Wu, Jiehui Xu, Jianmin Wang, and Mingsheng Long. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. In *NeurIPS*, volume 34, pp. 22419–22430, 2021.

Haixu Wu, Tengge Hu, Yong Liu, Hang Zhou, Jianmin Wang, and Mingsheng Long. Timesnet: Temporal 2d-variation modeling for general time series analysis. In *ICLR*, 2023.

Siru Zhong, Weilin Ruan, Ming Jin, Huan Li, Qingsong Wen, and Yuxuan Liang. Time-VLM: Exploring multimodal vision-language models for augmented time series forecasting. In *Forty-second International Conference on Machine Learning*, 2025.

Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 11106–11115, 2021.

Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun, and Rong Jin. FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *ICML*, 2022.

A Theoretical Results

A.1 Assumptions

For $j \in \mathcal{S} \cup \{f\}$, let $\boldsymbol{P}_j = \mathbb{E}(\boldsymbol{A}_j)$ be the population adjacency operator and define

$$\rho_j = \frac{1}{N(N-1)} \sum_{u \neq v} [\boldsymbol{P}_j]_{uv}.$$

Define $\widehat{\rho}_j$ analogously from $\boldsymbol{A}_j$; for $j \in \mathcal{S}$, this agrees with Eq. (10). Let $\boldsymbol{U}_j^\star \in \mathbb{R}^{N \times K}$ span the population invariant subspace corresponding to the same K eigenvalues used to construct $\widehat{\boldsymbol{U}}_j$, and write

$$\boldsymbol{\Pi}_j^\star = \boldsymbol{U}_j^\star (\boldsymbol{U}_j^\star)^\top, \qquad \widehat{\boldsymbol{\Pi}}_j = \widehat{\boldsymbol{U}}_j \widehat{\boldsymbol{U}}_j^\top.$$

For two orthonormal bases, define the projection distance

$$\operatorname{dist}(\boldsymbol{U}, \boldsymbol{V}) := \|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{V}\boldsymbol{V}^\top\|_2, \qquad h_r := \operatorname{dist}(\boldsymbol{U}_r^\star, \boldsymbol{U}_f^\star).$$

Thus, h_r is the population source–target heterogeneity and $\widehat{h}_r$ in Eq. (10) is its plug-in counterpart. Let $J = |\mathcal{S}|$ and define

$$\theta_r := h_r^2 + c\frac{\varsigma_N}{N\rho_r}, \qquad r \in \mathcal{S}. \tag{19}$$

Let $\Delta_{\mathcal{S}} := \{\boldsymbol{w} \in [0,1]^J : \sum_{r \in \mathcal{S}} w_r = 1\}$ denote the source-weight simplex. The empirical score in Eq. (11) is

$$\widehat{\theta}_r := a_r^{-1} = \widehat{h}_r^2 + c\frac{\varsigma_N}{N\widehat{\rho}_r}, \qquad r \in \mathcal{S}.$$

Assumption 1 (Population graph model)**.** The matrices $\boldsymbol{A}_j$ are symmetric with zero diagonal. Conditional on $\boldsymbol{P}_j$, their upper-triangular entries are independent Bernoulli variables with means $[\boldsymbol{P}_j]_{uv}$, and the graph layers are independent. Moreover, $\max_u \sum_v [\boldsymbol{P}_j]_{uv} \leq CN\rho_j$ and all graph densities have a common order: $c\rho \leq \rho_j \leq C\rho$ for $j \in \mathcal{S} \cup \{f\}$. The rank K and the number of sources J are fixed as N grows.

Assumption 2 (Eigengap, sparsity, and nondegeneracy)**.** Let δ_j be the separation between the K population eigenvalues defining $\boldsymbol{U}_j^\star$ and the remaining spectrum of $\boldsymbol{P}_j$. For constants $c_\delta, c_\star > 0$,

$$\delta_j \geq c_\delta N\rho_j, \qquad N \min_j \rho_j \gtrsim \log N, \qquad \frac{N\rho}{\varsigma_N} \longrightarrow \infty.$$

The density floor obeys $0 < \epsilon_N \leq \frac{1}{2}\min_j \rho_j$. In addition, for every source, assume the population Procrustes map is stable: $\sigma_K((\boldsymbol{U}_r^\star)^\top \boldsymbol{U}_f^\star) \geq c_Z > 0$. If

$$\boldsymbol{Z}_r^\star \in \operatorname*{arg\,min}_{\boldsymbol{Z}^\top \boldsymbol{Z} = \boldsymbol{I}_K} \|\boldsymbol{U}_r^\star \boldsymbol{Z} - \boldsymbol{U}_f^\star\|_F,$$

then the population Procrustes average is nondegenerate uniformly over the weight simplex: $\sigma_K(\sum_r w_r \boldsymbol{U}_r^\star \boldsymbol{Z}_r^\star) \geq c_\star$ for every $\boldsymbol{w} \in \Delta_{\mathcal{S}}$.

For a weight vector $\boldsymbol{w}$, define the sample and population aggregates

$$\overline{\boldsymbol{U}}(\boldsymbol{w}) := \mathrm{qf}\left(\sum_{r\in\mathcal{S}} w_r \boldsymbol{U}_r \boldsymbol{Z}_r\right), \qquad \overline{\boldsymbol{U}}^\star(\boldsymbol{w}) := \mathrm{qf}\left(\sum_{r\in\mathcal{S}} w_r \boldsymbol{U}_r^\star \boldsymbol{Z}_r^\star\right).$$

Assumption 3 (Joint spectral and score concentration)**.** Let $\mathcal{J} := \mathcal{S} \cup \{f\}$. There exists an event $\mathcal{E}_N$, with $\mathbb{P}(\mathcal{E}_N) \geq 1 - CJN^{-c}$, on which

$$\max\left\{\max_{j\in\mathcal{J}} \frac{\|\boldsymbol{A}_j - \boldsymbol{P}_j\|_2}{\sqrt{N\rho_j \log N}}, \max_{j\in\mathcal{J}} \frac{N|\widehat{\rho}_j - \rho_j|}{\sqrt{\rho_j \log N}}, \frac{\mathrm{dist}(\boldsymbol{U}_f, \boldsymbol{U}_f^\star)}{t_f}\right\} \leq C, \tag{20}$$

and, uniformly over $\boldsymbol{w} \in \Delta_{\mathcal{S}}$,

$$\mathrm{dist}\Big(\overline{\boldsymbol{U}}(\boldsymbol{w}), \overline{\boldsymbol{U}}^\star(\boldsymbol{w})\Big) \leq C\left(\sum_{r\in\mathcal{S}} w_r^2 \frac{\log N}{N\rho_r}\right)^{1/2}, \qquad c_0\theta_r \leq \widehat{\theta}_r \leq C_0\theta_r \quad (r \in \mathcal{S}). \tag{21}$$

Here $t_f \asymp (N\rho_f)^{-1/2}$ under sharp independent-edge concentration, or $t_f \asymp \sqrt{\log N/(N\rho_f)}$ under matrix Bernstein.

Assumption 3 isolates the only probabilistic ingredient required below. For independent random graphs, it can be established from eigenspace perturbation expansions and matrix concentration when combined with the Procrustes-stability condition in Assumption 2 and a uniform first-order concentration argument. Sample splitting is a sufficient alternative to proving uniformity for weights computed from the same graphs.

A.2 Proofs of the theoretical results in 'Theoretical guarantees' of Section 3.3

Throughout this section, $C, c > 0$ denote constants that may change from line to line. They may depend on the fixed rank K, the fixed number of sources J, and the constants in the assumptions and theorem conditions, but not on N. All stochastic statements are made on the common event $\mathcal{E}_N$ from Assumption 3; consequently, they hold with probability at least $1 - CJN^{-c}$. The assertions below are understood for all sufficiently large N.

A.2.1 A deterministic QR bound

We first record a fact used repeatedly below. Let $\boldsymbol{U}_0 \in \mathbb{R}^{N\times K}$ have orthonormal columns, write $\boldsymbol{\Pi}_0 = \boldsymbol{U}_0\boldsymbol{U}_0^\top$, and let $\boldsymbol{X} \in \mathbb{R}^{N\times K}$ have full column rank. If $\boldsymbol{Q} = \mathrm{qf}(\boldsymbol{X})$, then

$$\mathrm{dist}(\boldsymbol{Q}, \boldsymbol{U}_0) \leq \frac{\|(\boldsymbol{I}_N - \boldsymbol{\Pi}_0)\boldsymbol{X}\|_2}{\sigma_K(\boldsymbol{X})}. \tag{22}$$

Indeed, write the thin QR decomposition as $\boldsymbol{X} = \boldsymbol{Q}\boldsymbol{R}$. Then $\boldsymbol{Q} = \boldsymbol{X}\boldsymbol{R}^{-1}$ and $\|\boldsymbol{R}^{-1}\|_2 = \sigma_K(\boldsymbol{X})^{-1}$. Since both subspaces have dimension K, the principal-angle identity gives

$$\mathrm{dist}(\boldsymbol{Q}, \boldsymbol{U}_0) = \|(\boldsymbol{I}_N - \boldsymbol{\Pi}_0)\boldsymbol{Q}\|_2,$$

which proves Eq. (22).

The population averages have full column rank by Assumption 2. The sample averages are also well-defined uniformly over $\Delta_{\mathcal{S}}$ on $\mathcal{E}_N$. To see this, the eigengap and spectral concentration conditions, together with the Davis–Kahan perturbation bound, give

$$\max_{j\in\mathcal{S}\cup\{f\}} \operatorname{dist}(\boldsymbol{U}_j, \boldsymbol{U}_j^\star) \le C\sqrt{\frac{\log N}{N\rho}} = o(1).$$

Here we used fixed J and $N\rho/\varsigma_N \to \infty$, which imply $(\log N)/(N\rho) \to 0$. After aligning each sample basis with its population counterpart, singular-value perturbation and Assumption 2 imply

$$\sigma_K(\boldsymbol{U}_r^\top \boldsymbol{U}_f) \ge c_Z/2 \qquad (r \in \mathcal{S}).$$

If $\boldsymbol{U}_r^\top \boldsymbol{U}_f = \boldsymbol{B}_r \boldsymbol{\Sigma}_r \boldsymbol{C}_r^\top$, the Procrustes solution satisfies

$$\boldsymbol{U}_f^\top \boldsymbol{U}_r \boldsymbol{Z}_r = \boldsymbol{C}_r \boldsymbol{\Sigma}_r \boldsymbol{C}_r^\top \succeq (c_Z/2)\boldsymbol{I}_K.$$

Consequently, for every $\boldsymbol{w} \in \Delta_{\mathcal{S}}$,

$$\sigma_K\left(\sum_{r\in\mathcal{S}} w_r \boldsymbol{U}_r \boldsymbol{Z}_r\right) \ge \sigma_K\left(\boldsymbol{U}_f^\top \sum_{r\in\mathcal{S}} w_r \boldsymbol{U}_r \boldsymbol{Z}_r\right) \ge c_Z/2.$$

Apply Eq. (22) to

$$\boldsymbol{X}^\star(\boldsymbol{w}) := \sum_{r\in\mathcal{S}} w_r \boldsymbol{U}_r^\star \boldsymbol{Z}_r^\star, \qquad \boldsymbol{U}_0 = \boldsymbol{U}_f^\star.$$

For equal-dimensional subspaces,

$$\|(\boldsymbol{I}_N - \boldsymbol{\Pi}_f^\star)\boldsymbol{U}_r^\star \boldsymbol{Z}_r^\star\|_2 = h_r.$$

The nondegeneracy condition therefore yields, uniformly over $\boldsymbol{w} \in \Delta_{\mathcal{S}}$,

$$\operatorname{dist}\Big(\overline{\boldsymbol{U}}^\star(\boldsymbol{w}), \boldsymbol{U}_f^\star\Big) \le c_\star^{-1} \sum_{r\in\mathcal{S}} w_r h_r. \tag{23}$$

A.2.2 Proof of Proposition 1

Proof. The triangle inequality, Assumption 3, and Eq. (23) give, uniformly over the simplex,

$$\begin{aligned}\operatorname{dist}\big(\overline{\boldsymbol{U}}(\boldsymbol{w}), \boldsymbol{U}_f^\star\big) &\le \operatorname{dist}\Big(\overline{\boldsymbol{U}}(\boldsymbol{w}), \overline{\boldsymbol{U}}^\star(\boldsymbol{w})\Big) + \operatorname{dist}\Big(\overline{\boldsymbol{U}}^\star(\boldsymbol{w}), \boldsymbol{U}_f^\star\Big) \\ &\le C\left(\sum_{r\in\mathcal{S}} w_r^2 \frac{\log N}{N\rho_r}\right)^{1/2} + c_\star^{-1}\sum_{r\in\mathcal{S}} w_r h_r.\end{aligned}$$

This proves Eq. (14). Uniformity makes the result applicable to the data-dependent weights in Eq. (11).

The adaptive weighting score is also consistent with the scale of this bound. By Cauchy–Schwarz and $\log N \le J\varsigma_N$,

$$\begin{aligned}\left\{\left(\sum_r w_r^2 \frac{\log N}{N\rho_r}\right)^{1/2} + \sum_r w_r h_r\right\}^2 &\le 2\sum_r w_r^2 \frac{\log N}{N\rho_r} + 2J\sum_r w_r^2 h_r^2 \\ &\le 2J\sum_r w_r^2 \theta_r.\end{aligned}$$

Because $\widetilde{\rho}_r \ge \epsilon_N > 0$, every empirical score $\widehat{\theta}_r$ is strictly positive. Hence $F(\boldsymbol{w}) = \sum_r \widehat{\theta}_r w_r^2$ is strictly convex. The stationarity equations for the equality-constrained problem are $2\widehat{\theta}_r w_r = \mu$. Normalization gives

$$w_r = \frac{\widehat{\theta}_r^{-1}}{\sum_{s\in\mathcal{S}} \widehat{\theta}_s^{-1}}.$$

Every coordinate is positive, so the nonnegativity constraints are inactive and this is the unique global minimizer. It is exactly Eq. (11).

Define $w_r^\star = \theta_r^{-1} / \sum_s \theta_s^{-1}$. The score comparison in Assumption 3 implies

$$\frac{c_0}{C_0} w_r^\star \le w_r \le \frac{C_0}{c_0} w_r^\star, \qquad r \in \mathcal{S}.$$

This proves the final assertion. □

A.2.3 Proof of Theorem 1

Proof. We work on the event $\mathcal{E}_N$. Let $d = J - m$ and $\widehat{D} = \sum_{s\in\mathcal{S}} \widehat{\theta}_s^{-1}$. All sums and remainders over $\mathcal{I}^c$ are zero when $d = 0$. By the definition of η_N and density comparability,

$$\frac{\varsigma_N/(N\rho_r)}{\eta_N} = \frac{\rho}{\rho_r} \asymp 1, \qquad r \in \mathcal{S}. \tag{24}$$

For $r \in \mathcal{I}$, the definition $h_r^2 \le c_I \eta_N$, Eq. (24), and score comparability give

$$\theta_r \asymp \eta_N, \qquad \widehat{\theta}_r \asymp \eta_N. \tag{25}$$

For $r \in \mathcal{I}^c$, we have

$$\theta_r \ge h_r^2 \ge G_N \eta_N, \qquad \widehat{\theta}_r \ge c_0 G_N \eta_N. \tag{26}$$

Consequently,

$$c\frac{m}{\eta_N} \le \widehat{D} \le C\frac{m}{\eta_N} + C\frac{d}{G_N\eta_N} \le C'\frac{m}{\eta_N},$$

where the last inequality uses fixed J, $m \ge 1$, and the constant lower bound on G_N. Thus $\widehat{D} \asymp m/\eta_N$. It follows that

$$w_r \asymp m^{-1} \quad (r \in \mathcal{I}), \qquad w_r \le C\frac{\eta_N}{m\theta_r} \quad (r \in \mathcal{S}).$$

For $r \in \mathcal{I}^c$, this gives $w_r \le C/(mG_N)$, and hence

$$\sum_{r\in\mathcal{I}^c} w_r \le C\frac{d}{mG_N}.$$

We next substitute these weights into Proposition 1. Since

$$\frac{\log N/(N\rho_r)}{\eta_N} = \frac{\log N}{\varsigma_N}\frac{\rho}{\rho_r} \le CJ, \tag{27}$$

and J is fixed, splitting the stochastic term gives

$$\begin{aligned}
\left(\sum_{r\in\mathcal{S}} w_r^2 \frac{\log N}{N\rho_r}\right)^{1/2} &\le \left(\sum_{r\in\mathcal{I}} w_r^2 \frac{\log N}{N\rho_r}\right)^{1/2} + \left(\sum_{r\in\mathcal{I}^c} w_r^2 \frac{\log N}{N\rho_r}\right)^{1/2} \\
&\le CM_N + C\frac{\sqrt{d}}{mG_N}\sqrt{\eta_N} \\
&\le C\left\{M_N + \frac{d}{m\sqrt{G_N}}\sqrt{\eta_N}\right\}.
\end{aligned}$$

The last step uses $G_N \ge 1$ and, when $d > 0$, $\sqrt{d} \le d$.

For the heterogeneity term,

$$\sum_{r\in\mathcal{I}} w_r h_r \le CH_N.$$

For each $r \in \mathcal{I}^c$, $h_r^2 \le \theta_r$ and Eq. (26) yield

$$w_r h_r \le C\frac{\eta_N h_r}{m\theta_r} \le C\frac{\eta_N}{m\sqrt{\theta_r}} \le C\frac{\sqrt{\eta_N}}{m\sqrt{G_N}}.$$

Summing over $\mathcal{I}^c$ gives

$$\sum_{r\in\mathcal{I}^c} w_r h_r \le C\frac{d}{m\sqrt{G_N}}\sqrt{\eta_N}.$$

Combining these bounds with Proposition 1, we obtain

$$\operatorname{dist}\big(\overline{\boldsymbol{U}}, \boldsymbol{U}_f^\star\big) \le C\left\{M_N + H_N + \frac{J-m}{m\sqrt{G_N}}\sqrt{\eta_N}\right\}. \tag{28}$$

By the condition

$$\frac{J-m}{m\sqrt{G_N}}\sqrt{\eta_N} = O(M_N + H_N)$$

in the theorem, it follows that

$$\operatorname{dist}\big(\overline{\boldsymbol{U}}, \boldsymbol{U}_f^\star\big) \le C(M_N + H_N).$$

Finally, density comparability implies

$$M_N = \left\{\frac{1}{m^2}\sum_{r\in\mathcal{I}}\frac{\log N}{N\rho_r}\right\}^{1/2} \asymp \sqrt{\frac{\log N}{mN\rho}}.$$

Together with $H_N = m^{-1}\sum_{r\in\mathcal{I}} h_r$, this proves Eq. (17). □

A.3 Proof of Theorem 2

Proof. For an orthonormal matrix $\boldsymbol{V}$, write $\boldsymbol{P}_V = \boldsymbol{V}\boldsymbol{V}^\top$ and $\boldsymbol{Q} = \overline{\boldsymbol{U}}\,\overline{\boldsymbol{U}}^\top$. Both are rank-$K$ orthogonal projectors, so

$$\frac{\lambda}{2}\|\boldsymbol{P}_V - \boldsymbol{Q}\|_F^2 = \lambda K - \lambda\,\mathrm{tr}(\boldsymbol{V}^\top\boldsymbol{Q}\boldsymbol{V}).$$

Thus

$$\mathcal{L}_{\text{spec}}(\boldsymbol{V}) = (1+\lambda)K - \mathrm{tr}\Big[\boldsymbol{V}^\top(\widehat{\boldsymbol{\Pi}}_f + \lambda\boldsymbol{Q})\boldsymbol{V}\Big].$$

For finite λ, set $\gamma = \lambda/(1+\lambda)$ and

$$\boldsymbol{A}_\gamma := (1-\gamma)\widehat{\boldsymbol{\Pi}}_f + \gamma\boldsymbol{Q}.$$

The variational characterization of the leading eigenspace shows that a global minimizer spans its leading K-dimensional subspace. At $\gamma = 1$, the endpoint convention identifies this subspace with $\overline{\boldsymbol{U}}$.

Let $\boldsymbol{E}_\gamma = \boldsymbol{A}_\gamma - \boldsymbol{\Pi}_f^\star$ and $e_\gamma = \|\boldsymbol{E}_\gamma\|_2$. Assumption 3, Theorem 1, and the remainder condition give, uniformly for $0 \le \gamma \le 1$,

$$e_\gamma \le C\{(1-\gamma)t_f + \gamma(M_N + H_N)\}. \tag{29}$$

The Kth and $(K+1)$st eigenvalues of $\boldsymbol{\Pi}_f^\star$ are one and zero. Since $t_f + M_N + H_N = o(1)$, we have $e_\gamma \le 1/4$ uniformly for large N. Weyl's inequality gives

$$\lambda_K(\boldsymbol{A}_\gamma) \ge 1 - e_\gamma, \qquad \lambda_{K+1}(\boldsymbol{A}_\gamma) \le e_\gamma.$$

The leading subspace is therefore unique, with gap at least $1/2$.

Let $\widehat{\boldsymbol{U}}_\gamma$ contain its orthonormal eigenvectors, with eigenvalue matrix $\widehat{\boldsymbol{\Lambda}}_\gamma$, and let $\boldsymbol{U}_{f,\perp}^\star$ be an orthonormal complement of $\boldsymbol{U}_f^\star$. Premultiplying the eigenvalue equation by $(\boldsymbol{U}_{f,\perp}^\star)^\top$ gives

$$(\boldsymbol{U}_{f,\perp}^\star)^\top\widehat{\boldsymbol{U}}_\gamma\widehat{\boldsymbol{\Lambda}}_\gamma = (\boldsymbol{U}_{f,\perp}^\star)^\top\boldsymbol{E}_\gamma\widehat{\boldsymbol{U}}_\gamma.$$

Consequently,

$$\begin{aligned}\operatorname{dist}(\widehat{\boldsymbol{U}}_\gamma, \boldsymbol{U}_f^\star) &= \|(\boldsymbol{U}_{f,\perp}^\star)^\top\widehat{\boldsymbol{U}}_\gamma\|_2 \\ &\le \frac{e_\gamma}{1-e_\gamma} \le \frac{4}{3}e_\gamma \\ &\le C\{(1-\gamma)t_f + \gamma(M_N + H_N)\}.\end{aligned} \tag{30}$$

If $H_N > 0$, substituting $\gamma_{\mathrm{bal}} = t_f/(t_f + H_N)$ gives

$$(1-\gamma_{\mathrm{bal}})t_f + \gamma_{\mathrm{bal}}(M_N + H_N) = \gamma_{\mathrm{bal}}M_N + \frac{2t_f H_N}{t_f + H_N} \leq \gamma_{\mathrm{bal}}M_N + 2(t_f \wedge H_N).$$

If $H_N = 0$, the convention $\gamma_{\mathrm{bal}} = 1$ gives the bound CM_N. Since $\gamma_{\mathrm{bal}} \leq 1$, these observations prove both inequalities in Eq. (18). Substituting $M_N \asymp \sqrt{\log N/(mN\rho)}$ and $t_f \asymp (N\rho_f)^{-1/2}$ proves the displayed sharp-target form.

For the endpoint comparison, $\lambda = 0$ returns the target subspace, with error at most Ct_f, and $\lambda = +\infty$ returns the source aggregate, with error at most $C(M_N + H_N)$. Using a common upper-bound constant, the safeguarded oracle that compares t_f with $\gamma_{\mathrm{bal}}M_N + (t_f \wedge H_N)$ satisfies

$$\mathrm{dist}(\widehat{\boldsymbol{U}}_{\mathrm{sg}}, \boldsymbol{U}_f^\star) \leq C\min\{t_f, \gamma_{\mathrm{bal}}M_N + (t_f \wedge H_N)\} \leq C\min\{t_f, M_N + H_N\}.$$

This is an oracle comparison; it does not assert an implementable selection rule based on unknown population errors. For the Bernstein target scale, density comparability gives $M_N = O(t_f)$, so the balanced choice itself is target-rate competitive. For the sharp target scale with fixed m, the ratio M_N/t_f can be of order $\sqrt{\log N/m}$; the same conclusion does not follow uniformly without the safeguard. □

B Experimental Details

Hyperparameter sensitivity analysis. Table 4 evaluates SpecBridge under different loss weights and graph construction thresholds. Across the 11 datasets, the model exhibits stable forecasting performance over a broad range of η_{recon}, η_{spec}, τ_p, τ_t, and τ_f, with no single hyperparameter choice being uniformly optimal across all regions. This behavior is desirable, i.e., it suggests that the gains of SpecBridge do not arise from a particular carefully tuned graph or loss configuration. Notably, performance remains consistently strong when varying the 3 graph thresholds, even though these thresholds directly control the connectivity of the past-time, textual, and future-time graphs. This suggests that the spectral alignment operates on structural information that persists across reasonable variations in graph construction, rather than exploiting a specific adjacency pattern. Similarly, the reconstruction and alignment objectives remain effective across a range of weighting coefficients, while excessively emphasizing either auxiliary objective can lead to moderate degradation. Overall, SpecBridge is robust to its key design hyperparameters, supporting the generality of future-guided spectral alignment across heterogeneous outage dynamics.

		TX		NY		FL		NC		GA		LA		CA		MI		OH		PA		WA	
Hyperparameter	**Value**	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
η_{recon}	0.1	0.4635	0.5052	0.7552	0.6591	0.4689	0.5276	0.4934	0.5451	0.5033	0.5587	0.5639	0.5744	0.6347	0.5873	0.4288	0.4956	0.5419	0.5777	0.5890	0.5753	0.8128	0.7004
	0.3	0.4659	0.5064	0.7665	0.6652	0.4694	0.5280	0.4965	0.5463	0.5052	0.5603	0.5658	0.5754	0.6298	0.5851	0.4245	0.4908	0.5365	0.5749	0.5828	0.5731	0.8121	0.6998
	1	0.4509	0.4992	0.7563	0.6600	0.4689	0.5274	0.4964	0.5463	0.4999	0.5578	0.5681	0.5771	0.6266	0.5836	0.4276	0.4926	0.5351	0.5739	0.5742	0.5687	0.8132	0.6999
	3	0.4574	0.5030	0.7515	0.6568	0.4684	0.5273	0.4935	0.5448	0.4999	0.5573	0.5625	0.5733	0.6354	0.5874	0.4296	0.4944	0.5342	0.5727	0.5801	0.5725	0.8187	0.7017
	10	0.4650	0.5075	0.7584	0.6617	0.4700	0.5285	0.4993	0.5480	0.5009	0.5578	0.5694	0.5773	0.6305	0.5853	0.4287	0.4954	0.5363	0.5743	0.5831	0.5737	0.8083	0.6977
η_{spec}	0.1	0.4538	0.5007	0.7502	0.6586	0.4789	0.5324	0.5101	0.5523	0.5120	0.5648	0.5716	0.5803	0.6534	0.5948	0.4440	0.5060	0.5421	0.5773	0.5981	0.5801	0.8817	0.7289
	0.3	0.4507	0.4993	0.7546	0.6597	0.4771	0.5311	0.5128	0.5533	0.5095	0.5641	0.5698	0.5789	0.6479	0.5922	0.4379	0.5001	0.5418	0.5773	0.5948	0.5754	0.8854	0.7254
	1	0.4555	0.5011	0.7656	0.6628	0.4739	0.5293	0.5080	0.5517	0.5092	0.5618	0.5718	0.5779	0.6296	0.5842	0.4359	0.5008	0.5440	0.5782	0.5834	0.5678	0.8501	0.7126
	3	0.4509	0.4992	0.7563	0.6600	0.4689	0.5274	0.4964	0.5463	0.4999	0.5578	0.5681	0.5771	0.6266	0.5836	0.4276	0.4926	0.5351	0.5739	0.5742	0.5687	0.8132	0.6999
	10	0.4613	0.5061	0.7692	0.6681	0.4636	0.5250	0.4843	0.5403	0.5209	0.5694	0.5738	0.5813	0.6305	0.5868	0.4379	0.4990	0.5411	0.5780	0.6059	0.5876	0.8157	0.7024
τ_p	0	0.4507	0.4968	0.7557	0.6589	0.4656	0.5247	0.4973	0.5462	0.5141	0.5653	0.5766	0.5848	0.6434	0.5909	0.4370	0.4995	0.5257	0.5681	0.5915	0.5752	0.8157	0.7017
	0.25	0.4477	0.4961	0.7500	0.6564	0.4666	0.5262	0.4970	0.5460	0.5083	0.5617	0.5904	0.5911	0.6328	0.5857	0.4361	0.4996	0.5309	0.5714	0.5993	0.5815	0.8325	0.7084
	0.5	0.4509	0.4992	0.7563	0.6600	0.4689	0.5274	0.4964	0.5463	0.4999	0.5578	0.5681	0.5771	0.6266	0.5836	0.4276	0.4926	0.5351	0.5739	0.5742	0.5687	0.8132	0.6999
	0.7	0.4526	0.5008	0.7558	0.6598	0.4677	0.5265	0.4909	0.5436	0.5063	0.5608	0.5664	0.5768	0.6331	0.5871	0.4219	0.4922	0.5309	0.5713	0.6005	0.5813	0.8195	0.7028
	0.9	0.4584	0.5027	0.7444	0.6531	0.4740	0.5297	0.4891	0.5417	0.5019	0.5586	0.5654	0.5752	0.6301	0.5856	0.4317	0.4969	0.5347	0.5734	0.5805	0.5694	0.8227	0.7036
τ_t	0.5	0.4458	0.4935	0.7589	0.6617	0.4630	0.5232	0.4950	0.5441	0.5176	0.5663	0.5707	0.5784	0.6385	0.5884	0.4277	0.4962	0.5367	0.5750	0.5925	0.5749	0.8307	0.7076
	0.7	0.4549	0.5000	0.7789	0.6674	0.4673	0.5259	0.4908	0.5428	0.5082	0.5599	0.5738	0.5799	0.6348	0.5884	0.4300	0.4971	0.5334	0.5722	0.5887	0.5756	0.8259	0.7075
	0.9	0.4509	0.4992	0.7563	0.6600	0.4689	0.5274	0.4964	0.5463	0.4999	0.5578	0.5681	0.5771	0.6266	0.5836	0.4276	0.4926	0.5351	0.5739	0.5742	0.5687	0.8132	0.6999
	0.95	0.4598	0.5051	0.7605	0.6626	0.4675	0.5255	0.4944	0.5461	0.5044	0.5595	0.5633	0.5743	0.6352	0.5880	0.4307	0.4963	0.5360	0.5743	0.5937	0.5796	0.8174	0.7020
	0.99	0.4556	0.5021	0.7621	0.6624	0.4633	0.5233	0.4919	0.5444	0.5133	0.5645	0.5636	0.5745	0.6347	0.5881	0.4291	0.4949	0.5376	0.5752	0.5789	0.5725	0.8156	0.7012
τ_f	0.4	0.4533	0.5017	0.7435	0.6536	0.4652	0.5251	0.4908	0.5443	0.5080	0.5631	0.5609	0.5742	0.6431	0.5906	0.4223	0.4915	0.5307	0.5713	0.5762	0.5718	0.7994	0.6947
	0.6	0.4450	0.4970	0.7480	0.6570	0.4666	0.5253	0.4942	0.5460	0.5136	0.5660	0.5556	0.5705	0.6389	0.5879	0.4363	0.5001	0.5300	0.5708	0.5741	0.5695	0.8066	0.6970
	0.7	0.4613	0.5045	0.7420	0.6530	0.4659	0.5252	0.4960	0.5467	0.5170	0.5673	0.5489	0.5669	0.6359	0.5866	0.4218	0.4916	0.5279	0.5695	0.5910	0.5768	0.8079	0.6972
	0.8	0.4509	0.4992	0.7563	0.6600	0.4689	0.5274	0.4964	0.5463	0.4999	0.5578	0.5681	0.5771	0.6266	0.5836	0.4276	0.4926	0.5351	0.5739	0.5742	0.5687	0.8132	0.6999
	0.9	0.4601	0.5049	0.7580	0.6597	0.4833	0.5369	0.5118	0.5547	0.5227	0.5682	0.6013	0.5966	0.6425	0.5898	0.4353	0.4980	0.5510	0.5834	0.5982	0.5773	0.8399	0.7134

Table 4: Hyperparameter sensitivity analysis, averaged over 3 prediction lengths $\{96, 192, 336\}$.

Computational efficiency. Table 5 compares SpecBridge with 5 representative baselines (i.e., TimeAlign, PatchTST, TimesNet, Informer, and Time-VLM) in terms of parameter count, memory usage, and training time per epoch (in seconds). We can observe that SpecBridge has the lowest memory usage among the compared methods at 198 MiB and the second-smallest parameter count at 7.93M. Compared with Time-VLM, it reduces parameter count by 93.1% and memory usage

by 66.9%, while training approximately 9.1 times faster based on average epoch time. Relative to PatchTST, SpecBridge uses 26.2% fewer parameters and 38.3% less memory, with comparable average epoch times of 13.15 and 13.49 seconds, respectively. Although TimeAlign trains faster and TimesNet uses fewer parameters, SpecBridge requires less memory than either method. Together with its lowest horizon-averaged MSE and MAE on all 11 datasets, these results demonstrate a favorable balance between forecasting accuracy and computational cost.

Method	Metric	TX	NY	FL	NC	GA	LA	CA	MI	OH	PA	WA
SpecBridge (ours)	Param. (M)	7.93	7.93	7.93	7.93	7.93	7.93	7.93	7.93	7.93	7.93	7.93
	Mem. (MiB)	198	198	198	198	198	198	198	198	198	198	198
	Time (s/epoch)	12.7	13.0	13.1	13.7	13.3	12.7	12.8	12.7	12.8	14.1	13.8
TimeAlign	Param. (M)	9.36	9.36	9.36	9.36	9.36	9.36	9.36	9.36	9.36	9.36	9.36
	Mem. (MiB)	235	235	235	235	235	235	235	235	235	235	235
	Time (s/epoch)	3.6	3.4	3.3	3.4	3.4	3.7	3.6	3.5	3.4	3.5	3.5
PatchTST	Param. (M)	10.74	10.74	10.74	10.74	10.74	10.74	10.74	10.74	10.74	10.74	10.74
	Mem. (MiB)	321	321	321	321	321	321	321	321	321	321	321
	Time (s/epoch)	13.1	13.2	13.6	13.5	13.6	13.5	13.6	13.5	13.7	13.6	13.5
TimesNet	Param. (M)	1.17	1.17	1.17	1.17	1.17	1.17	1.17	1.17	1.17	1.17	1.17
	Mem. (MiB)	483	435	479	476	478	475	478	479	475	444	480
	Time (s/epoch)	63.2	67.6	67.8	67.1	63.1	66.9	65.7	63.0	68.1	61.0	70.0
Informer	Param. (M)	11.31	11.31	11.31	11.31	11.31	11.31	11.31	11.31	11.31	11.31	11.31
	Mem. (MiB)	1130	1130	1130	1130	1130	1130	1130	1130	1130	1130	1130
	Time (s/epoch)	76.8	77.8	77.3	77.6	77.6	77.3	77.0	77.7	77.8	77.5	76.7
Time-VLM	Param. (M)	114.67	114.67	114.67	114.67	114.67	114.67	114.67	114.67	114.67	114.67	114.67
	Mem. (MiB)	599	599	599	599	599	599	599	599	599	599	599
	Time (s/epoch)	119.2	119.3	118.9	118.9	118.8	119.8	118.5	119.4	119.4	119.2	119.5

Table 5: Computational efficiency comparison.

AI USE STATEMENT

We used large language models to improve the grammar, clarity, conciseness, and presentation of author-written text, and OpenAI Codex to assist with code development. We reviewed and verified all AI-assisted contributions and retain full responsibility for the scientific ideas, problem formulation, theoretical results and proofs, algorithms, experimental design, implementation, data processing, numerical results, analyses, and conclusions.

ETHICS STATEMENT

This work advances power outage forecasting by transferring relational structure from future observations to historical representations during training. More accurate forecasts can help utilities anticipate outage dynamics, allocate restoration resources, and support emergency response, with potential benefits for grid resilience and affected communities. Because forecasting errors may influence consequential decisions, model predictions should complement operational evidence and expert judgment, with careful consideration of uncertainty and performance differences across regions.

REPRODUCIBILITY STATEMENT

Section 3 presents the representation learning, latent graph construction, and spectral alignment components, together with the training objectives. Section 4 describes the datasets, preprocessing, chronological splits, baselines, hyperparameter search, and evaluation protocol. Appendix B reports hyperparameter sensitivity analyses and computational efficiency comparisons. The paper provides open access to an anonymized code repository, enabling reproducibility of the main experimental results.